



**Universidade Federal de Uberlândia
Instituto de Matemática e Estatística**

Bacharelado em Estatística

**INFERÊNCIA BAYESIANA SOBRE
PARÂMETROS GENÉTICOS DE
BOVINOS DA RAÇA GIROLANDO**

Pedro Henrique Galvão dos Santos

**Uberlândia-MG
2025**

Pedro Henrique Galvão dos Santos

**INFERÊNCIA BAYESIANA SOBRE
PARÂMETROS GENÉTICOS DE
BOVINOS DA RAÇA GIROLANDO**

Trabalho de conclusão de curso apresentado à Coordenação do Curso de Bacharelado em Estatística como requisito parcial para obtenção do grau de Bacharel em Estatística.

Orientador: Prof. Dr. José Waldemar da Silva

Uberlândia-MG
2025



**Universidade Federal de Uberlândia
Instituto de Matemática e Estatística**

Coordenação do Curso de Bacharelado em Estatística

A banca examinadora, conforme abaixo assinado, certifica a adequação deste trabalho de conclusão de curso para obtenção do grau de Bacharel em Estatística.

Uberlândia, ____ de _____ de 20____

BANCA EXAMINADORA

Prof. Dr. José Waldemar da Silva

Prof. Dr. Pedro Franklin

Profa. Dra. Maria Imaculada ou Marcelo

**Uberlândia-MG
2025**

AGRADECIMENTOS

Em primeiro lugar, agradeço à minha esposa, pelo amor, paciência e incentivo incondicionais ao longo de toda essa jornada. Sua compreensão e apoio foram fundamentais para que eu pudesse seguir em frente nos momentos mais desafiadores.

Aos meus familiares, que sempre acreditaram no meu potencial e me apoiaram com palavras de motivação, carinho e conselhos valiosos. Vocês são a base de tudo o que conquistei até aqui.

Ao meu sogro, pelo fornecimento da base de dados e pelas valiosas contribuições que tornaram possível a realização deste trabalho.

Ao meu orientador, cuja orientação, conhecimento e paciência foram essenciais para o desenvolvimento deste trabalho. Sou imensamente grato pela sua dedicação e por todas as contribuições valiosas que me guiaram na conclusão deste projeto.

A todos vocês, meu muito obrigado!

RESUMO

Este estudo teve como objetivo estimar parâmetros genéticos associados à produção de leite em gado Girolando, utilizando um modelo bayesiano com o amostrador de Monte Carlo Hamiltoniano (HMC), implementado por meio do pacote `rstan` através do software R. A base de dados utilizada compreende 295 animais únicos, totalizando 485 observações. Os resultados mostraram que a herdabilidade estimada foi de 0,235, com intervalo de máxima densidade à posteriori, em inglês HPD (Highest Posterior Density), de 0,09 a 0,38, sugerindo que cerca de 23,5% da variabilidade na produção de leite é explicada por fatores genéticos, embora fatores ambientais e de manejo também sejam importantes. A análise dos efeitos fixos indicou que os graus de sangue $\frac{5}{8}H$, $\frac{3}{4}H$ e $\frac{1}{2}H$ apresentaram os maiores impactos na produção, sugerindo que esses graus genéticos têm maior potencial para melhorar a produção de leite. As frações acima indicam grau de sangue da raça Holandesa (H) em cruzamento com a raça Gir. Os objetivos do estudo foram atingidos com sucesso, pois foi possível estimar a herdabilidade da produção de leite e identificar os graus de sangue com maior impacto produtivo, fornecendo subsídios para estratégias de seleção genética no rebanho. Como perspectiva futura, recomenda-se incluir outras variáveis de efeito fixo e mais informações sobre o parentesco dos animais para melhorar a análise do desempenho produtivo, considerando tanto os efeitos genéticos quanto os não genéticos.

Palavras-chave: Gado girolando; herdabilidade; HMC; melhoramento genético; modelo Bayesiano; produção de leite; rstan; valor genético aditivo.

ABSTRACT

This study aimed to estimate genetic parameters associated with milk production in Girolando cattle, using a Bayesian model with the Hamiltonian Monte Carlo (HMC) sampler, implemented through the `rstan` package in the R software. The dataset used comprises 295 unique animals, totaling 485 observations. The results showed that the estimated heritability was 0.235, with a HPD interval (Highest Posterior Density) from 0.09 to 0.38, suggesting that approximately 23.5% of the variability in milk production is explained by genetic factors, although environmental and management factors also play an important role. The analysis of fixed effects indicated that the blood compositions $\frac{5}{8}H$, $\frac{3}{4}H$, and $\frac{1}{2}H$ had the greatest impacts on production, suggesting that these genetic compositions have greater potential for improving milk yield. The fractions above indicate the Holstein (H) blood percentage in crosses with the Gir breed. The objectives of the study were successfully achieved, as it was possible to estimate the heritability of milk production and identify the blood compositions with the greatest productive impact, providing support for genetic selection strategies in the herd. As a future perspective, it is recommended to include other fixed-effect variables and more detailed pedigree information to improve the analysis of productive performance, considering both genetic and non-genetic effects.

Keywords: Additive genetic value, Bayesian model, genetic improvement, Girolando cattle, heritability, HMC, milk production, RStan.

SUMÁRIO

Lista de Figuras	I
Lista de Tabelas	II
1 Introdução	1
1.1 Objetivos	3
1.1.1 Objetivo Geral	3
1.1.2 Objetivos Específicos	3
2 Metodologia	4
2.1 Os dados	15
2.2 Implementação	16
3 Resultados	19
4 Conclusões	41
Referências Bibliográficas	42
Apêndice A Apêndice: Código do Modelo Bayesiano em Stan.	45
Apêndice B Apêndice: Código R.	46

LISTA DE FIGURAS

2.1	Exemplo unidimensional de HMC - movimento de um objeto em uma curva suave e sem atrito. [36]	
(a)	Aplica-se uma força com direção e intensidade aleatórias sobre o objeto.	
(b)	O objeto se move ao longo da curva, com sua energia cinética sendo convertida em energia potencial.	
(c)	No ponto mais baixo da curva, toda a energia está na forma de cinética.	
(d)	A energia cinética é novamente convertida em energia potencial, e o movimento continua em direção oposta.	13
3.1	Gráfico de traço das cadeias MCMC dos parâmetros h_2 e σ_a^2	19
3.2	Gráfico de traço das cadeias MCMC dos parâmetros β_2 , a_{50} , a_{64} e a_{81}	20
3.3	Gráfico de traço das cadeias MCMC dos parâmetros a_{141} , a_{146} , a_{163} e a_{195}	21
3.4	Gráfico de traço das cadeias MCMC dos parâmetros a_{215} , a_{228} , a_{251} e a_{258}	22
3.5	Gráfico de floresta das estimativas dos efeitos fixos β_1 a β_9 genéticos e seus intervalos HPD (95%) obtidos a partir das cadeias MCMC.	27
3.6	Gráfico de floresta das estimativas dos efeitos aditivos genéticos e seus intervalos HPD (95%) obtidos a partir das cadeias MCMC.	31
3.7	Gráfico de floresta das estimativas dos efeitos aditivos genéticos e seus intervalos HPD (95%) obtidos a partir das cadeias MCMC.	32
3.8	Gráfico de floresta das estimativas dos efeitos aditivos genéticos e seus intervalos HPD (95%) obtidos a partir das cadeias MCMC.	33
3.9	Gráfico de floresta das estimativas dos efeitos aditivos genéticos e seus intervalos HPD (95%) obtidos a partir das cadeias MCMC.	34
3.10	Gráfico de floresta das estimativas dos efeitos aditivos genéticos e seus intervalos HPD (95%) obtidos a partir das cadeias MCMC.	35
3.11	Gráfico ilustrando a classificação dos animais com base nos seus valores genéticos, representados por a_i	36
3.12	Gráfico ilustrando a classificação dos animais com base nos seus valores genéticos, representados por a_i	37
3.13	Gráfico ilustrando a classificação dos animais com base nos seus valores genéticos, representados por a_i	38
3.14	Gráfico ilustrando a classificação dos animais com base nos seus valores genéticos, representados por a_i	39
3.15	Gráfico ilustrando a classificação dos animais com base nos seus valores genéticos, representados por a_i	40

LISTA DE TABELAS

3.1	Resultados do Diagnóstico de Gelman & Rubin.	23
3.2	Resultados do Diagnóstico de Raftery & Lewis.	24
3.3	Resultados do Diagnóstico de Geweke.	25
3.4	Resultados do Intervalo HPD.	28

1. INTRODUÇÃO

A criação de gado leiteiro é uma das atividades agropecuárias mais importantes em diversas regiões do mundo, tendo um impacto significativo tanto na economia quanto na sociedade. Nos últimos anos, a pecuária leiteira tem enfrentado desafios relacionados à necessidade de aumentar a produção de leite, melhorar a qualidade do produto e reduzir os impactos ambientais. Para superar esses desafios, é fundamental implementar práticas eficientes de manejo e estratégias de melhoramento genético. Indicadores como a produção de leite em 305 dias de lactação (L305), grau de sangue (GS) e a idade ao primeiro parto (IPP) são variáveis-chave que auxiliam na avaliação do desempenho dos animais, ajudando os produtores a tomar decisões mais assertivas sobre a gestão dos rebanhos.

Estudos têm demonstrado que a IPP, idealmente situada entre 22 e 25 meses, está associada a uma maior produção de leite durante a primeira lactação, melhor retorno econômico e maior longevidade produtiva dos rebanhos [2][38]. Além disso, a IPP exerce uma forte influência sobre a saúde reprodutiva e a performance subsequente das vacas, impactando a longevidade e a sobrevivência até as paridades posteriores [9]. Em virtude de sua importância, o entendimento detalhado de como otimizar a IPP é essencial, equilibrando precocidade de produção com a eficiência global do rebanho.

Outro aspecto crucial na pecuária leiteira é o grau de sangue dos animais. O Girolando é uma raça híbrida, originada do cruzamento entre a raça Holandesa e o Zebu (geralmente a raça Gir). O grau de sangue, que representa a proporção de sangue de cada uma dessas raças, influencia diretamente as características produtivas e reprodutivas dos animais. Um maior grau de sangue Holandês, por exemplo, tende a aumentar a produção de leite, mas pode afetar a resistência a doenças típicas do clima quente, enquanto um maior grau de sangue Zebu pode melhorar a rusticidade e a adaptação ao ambiente tropical. Dessa forma, o grau de sangue é uma variável genética importante que pode ser otimizada para equilibrar a produção de leite com a robustez e a longevidade dos animais [16] [38].

Nos estudos genéticos aplicados aos modelos animais, os métodos tradicionais para estimar os parâmetros genéticos como herdabilidade e valores genéticos aditivos baseavam-se principalmente em técnicas frequentistas, como a Máxima Verossimilhança Restrita (REML) e os Mínimos Quadrados Generalizados (GLS). O método REML, por exemplo, foi amplamente utilizado para estimar a variância dos efeitos genéticos e ambientais, uma vez que fornece estimativas consistentes para a variância dos componentes do modelo, mesmo na presença de efeitos fixos mal especificados [32].

No entanto, esses métodos apresentam algumas deficiências críticas. Primeiramente, os métodos de máxima verossimilhança e mínimos quadrados assumem que os efeitos genéticos são normalmente distribuídos e independentes, o que pode ser uma simplificação excessiva, visto que efeitos não aditivos, como dominância e interações gênicas, não são facilmente modelados [15]. Além disso, os modelos tradicionais frequentemente não consideram a incerteza associada às estimativas, limitando a profundidade da análise estatística. Em contraste, o método bayesiano permite incorporar distribuições a priori que podem refletir o conhecimento prévio, modelando a incerteza dos parâmetros e oferecendo uma representação mais rica dos efeitos genéticos.

Outro desafio com os métodos tradicionais é que eles frequentemente não conseguem capturar a complexidade das dependências entre os indivíduos de uma população, especialmente quando as informações de pedigree são incompletas ou quando os dados são desbalanceados. A matriz de parentesco, que é essencial para representar a relação genética entre os indivíduos, é comumente usada em modelos bayesianos para melhorar as estimativas, mas é difícil de ser incorporada diretamente em modelos frequentistas sem simplificações substanciais [32].

A inferência bayesiana oferece vantagens claras sobre os métodos tradicionais, especialmente em contextos complexos como o melhoramento genético animal. A principal diferença é que a abordagem bayesiana incorpora a incerteza de forma explícita, atualizando as distribuições a priori com dados observacionais para gerar distribuições a posteriori. Além disso, a flexibilidade dos modelos bayesianos permite a incorporação de distribuições mais complexas para os parâmetros, como efeitos genéticos não aditivos, o que não é facilmente alcançado com os métodos tradicionais [15]. Ao integrar tanto dados fenotípicos quanto genômicos, a inferência bayesiana possibilita a construção de distribuições posteriores detalhadas, essenciais para estimar valores genéticos e variâncias associadas.

Por meio do uso de amostragem MCMC (Monte Carlo via Cadeias de Markov), a inferência bayesiana pode aproximar distribuições a posteriori complexas que envolvem dependências entre múltiplos parâmetros, o que seria intratável usando métodos analíticos tradicionais. Isso é particularmente importante em modelos hierárquicos, como os modelos animais, onde é necessário estimar os efeitos genéticos de indivíduos, mesmo sem uma observação direta da variável de interesse. A aplicação de modelos animais na análise genética do gado leiteiro é amplamente justificada, pois permite a decomposição da variância fenotípica em componentes genéticos e ambientais, proporcionando uma compreensão mais clara dos fatores que influenciam o desempenho reprodutivo e produtivo dos animais. A estimativa precisa da herdabilidade e dos valores genéticos é essencial para direcionar programas de melhoramento genético e selecionar reprodutores de forma mais eficiente [15].

Neste contexto, o objetivo deste estudo é utilizar a inferência bayesiana para estimar parâmetros genéticos relacionados à produção de leite, à idade ao primeiro parto (IPP) e ao grau de sangue em rebanhos de gado Girolando, com base em dados da Fazenda Cruzeiro do Sul. A análise de dados fenotípicos e genômicos permitirá gerar estimativas mais precisas e fornecer subsídios para uma seleção genética mais eficiente e um manejo reprodutivo mais assertivo, com vistas a melhorar a eficiência produtiva e a sustentabilidade da pecuária leiteira na região

do Triângulo Mineiro.

1.1 OBJETIVOS

1.1.1 OBJETIVO GERAL

O objetivo geral deste trabalho é utilizar a inferência bayesiana e o algoritmo de Monte Carlo Hamiltoniano (*Hamiltonian Monte Carlo - HMC*) para estimar parâmetros genéticos, como herdabilidade, valores genéticos (breeding values), o efeito da idade ao primeiro parto (IPP) e o efeito do grau de sangue, relacionados à produção de leite em rebanhos bovinos da raça Girolando, utilizando dados da Fazenda Cruzeiro do Sul, localizada no Triângulo Mineiro.

1.1.2 OBJETIVOS ESPECÍFICOS

- Estimar os parâmetros genéticos, como valores genéticos e herdabilidade, além dos efeitos da idade ao primeiro parto (IPP) e do grau de sangue, relacionados à produção de leite, utilizando dados da Fazenda Cruzeiro do Sul;
- Avaliar a influência da IPP e do grau de sangue na produção de leite das vacas leiteiras, com base na inferência bayesiana;
- Considerar a matriz de parentesco na modelagem bayesiana dos efeitos genéticos aditivos, com o intuito de melhorar a precisão das estimativas e levar em conta a estrutura de parentesco dos indivíduos na população;
- Implementar o modelo de inferência bayesiana no Stan, utilizando a interface rstan para realizar a amostragem.

Para atingir esses objetivos, será utilizada a inferência bayesiana. Esta abordagem estatística, especificamente para o problema de pesquisa desta monografia, será detalhada na Metodologia o que corresponde ao Capítulo 2 deste trabalho. Essa metodologia permitirá a modelagem da incerteza associada aos parâmetros genéticos e não às suas estimativas como é o caso na inferência clássica. Este enfoque também permite fornecer estimativas dos parâmetros de interesse, superando as limitações dos métodos clássicos. O uso da inferência bayesiana proporcionará uma análise mais precisa ao considerar tanto os efeitos fixos quanto genéticos, pois, enquanto a inferência clássica também pode modelar ambos, a abordagem bayesiana se destaca ao incorporar a incerteza associada aos parâmetros. A inferência bayesiana é mais precisa porque ela atualiza as distribuições a priori (baseadas no conhecimento prévio sobre os parâmetros) com a informação oriunda dos dados experimentais, gerando uma distribuição a posteriori que reflete a incerteza de forma mais completa. Esse processo é realizado por meio do Teorema de Bayes, que combina essas duas fontes de informação, tornando a análise mais ajustada aos dados observados.

2. METODOLOGIA

A inferência sobre os parâmetros do modelo utilizado neste trabalho foi realizada sob o enfoque bayesiano. Esta abordagem permite determinar a distribuição a posteriori dos parâmetros de interesse por meio do Teorema de Bayes. Este importante teorema permite acrescentar a informação oriunda dos dados observados ou experimentais, alocada na função de verossimilhança, ao conhecimento prévio alocado nas distribuições a priori.

A utilização da inferência bayesiana neste estudo justifica-se pela sua capacidade de lidar com cenários estatísticos complexos, como dados desbalanceados ou amostras de tamanho reduzido, situações frequentemente encontradas em experimentos genéticos com bovinos leiteiros [32].

O desbalanceamento de dados refere-se a situações em que diferentes grupos ou categorias possuem quantidades desiguais de observações, o que pode dificultar a estimação precisa dos efeitos associados [15]. Em estudos genéticos, por exemplo, certos indivíduos podem ter muitas medições enquanto outros têm poucas, comprometendo abordagens tradicionais baseadas em inferência frequentista.

A inferência bayesiana oferece flexibilidade porque não depende exclusivamente de grandes amostras ou de dados balanceados para gerar estimativas confiáveis [31]. Ela combina informações a priori — baseadas em conhecimentos prévios sobre os parâmetros — com a informação extraída dos dados observados, resultando na chamada distribuição a posteriori.

Toda a inferência no paradigma bayesiano é construída a partir da distribuição a posteriori [11]. Portanto, o objetivo central é a determinação desta distribuição, pois ela descreve completamente o que se sabe sobre os parâmetros de interesse à luz dos dados e das crenças prévias. A partir da distribuição a posteriori, podem ser feitas previsões, tomadas de decisão e avaliações de incerteza de maneira formal e coerente.

A inferência bayesiana tem sua origem formal associada ao matemático e ministro presbiteriano Thomas Bayes (1701–1761). Sua principal contribuição é o trabalho póstumo intitulado *An Essay towards solving a Problem in the Doctrine of Chances* [1], publicado em 1763 por seu amigo Richard Price na *Philosophical Transactions of the Royal Society of London*. Nesse ensaio, Bayes propôs um método para atualizar a probabilidade de uma hipótese com base em novas evidências, dando origem ao que hoje conhecemos como Teorema de Bayes.

O Teorema de Bayes permite determinar a chamada distribuição a posteriori, que é a distribuição de probabilidade dos parâmetros de interesse após a observação dos dados. Essa distribuição combina a informação prévia, expressa pela distribuição a priori, com a informa-

ção contida nos dados, representada pela verossimilhança [21]. De forma geral, toda inferência realizada sob o enfoque bayesiano é fundamentada na análise dessa distribuição a posteriori [33].

Apesar da formulação inicial de Bayes, o desenvolvimento e a aplicação mais ampla da inferência bayesiana ocorreram posteriormente, graças a Pierre-Simon Laplace (1749–1827). Laplace generalizou o teorema e aplicou-o a uma variedade de problemas científicos e práticos, desde questões de astronomia até demografia e ciências atuariais [21]. Laplace não apenas generalizou o teorema, mas também o aplicou a uma ampla variedade de problemas científicos e práticos. Entre suas principais contribuições, destacam-se o uso da inferência bayesiana para prever órbitas planetárias a partir de dados astronômicos, estimar a massa de Saturno e da Terra, calcular a probabilidade de sucessos em julgamentos por júri e analisar questões demográficas, como a expectativa de vida e a sobrevivência de populações [23][8]. Essas aplicações pioneiras reforçaram a credibilidade e a utilidade do raciocínio bayesiano em contextos práticos, muito antes da formalização moderna da estatística, estabelecendo uma base sólida para seu uso posterior nas ciências e engenharias. Em razão dessa sistematização e aplicação extensa, Laplace é frequentemente considerado um dos principais responsáveis pela popularização do raciocínio bayesiano no século XIX [34].

Toda a inferência bayesiana é realizada a partir da distribuição conjunta a posteriori dos parâmetros, a qual incorpora simultaneamente a informação prévia e a evidência fornecida pelos dados. Entretanto, a determinação analítica dessa distribuição é frequentemente inviável, especialmente em problemas de alta complexidade. Isso ocorre porque a obtenção da distribuição marginal de interesse requer a resolução de integrais múltiplas de alta dimensão, que, além de serem computacionalmente custosas, podem envolver funções de forma altamente irregular [30][11].

Em cenários de grande dimensionalidade, o cálculo dessas integrais se torna matematicamente intratável com métodos clássicos de integração numérica, exigindo o uso de métodos de aproximação estocástica, como os algoritmos MCMC (Monte Carlo via Cadeias de Markov). Esses métodos possibilitam a amostragem da distribuição a posteriori sem a necessidade de calcular explicitamente as integrais associadas, tornando a inferência viável mesmo em modelos de grande complexidade [6].

Graças ao desenvolvimento de métodos computacionais, como o algoritmo de Metropolis-Hastings [24][18], o amostrador de Gibbs [13] e o No-U-Turn Sampler (NUTS) [19], tornou-se possível aproximar de forma satisfatória a distribuição conjunta a posteriori dos parâmetros de interesse.

Esses métodos baseados em amostragem estocástica permitem gerar sequências de amostras que, sob condições adequadas, convergem para a distribuição a posteriori desejada, viabilizando a inferência bayesiana mesmo em cenários de alta complexidade computacional. A partir das amostras obtidas, é possível realizar inferências sobre parâmetros individuais, estimar intervalos de credibilidade e calcular medidas de associação entre parâmetros, superando assim as limitações impostas pelas integrais intratáveis presentes na formulação teórica da inferência

bayesiana.

Sob o enfoque bayesiano, é possível realizar inferência sobre os parâmetros do modelo utilizado neste trabalho cuja expressão é apresentada na Eq. 2.1. Este é um modelo que permite incorporar efeitos fixos e genéticos para explicar a variável resposta,

$$y = Xb + Za + e. \quad (2.1)$$

Os elementos que compõem o modelo apresentado na Eq. 2.1 são descritos nos itens que seguem, em que:

- y é o vetor de dados de ordem n ,
- b é o vetor de efeitos fixos de ordem p ,
- a é o vetor de valores genéticos aditivos de ordem q ,
- e é o vetor de erros de ordem n ,
- X e Z são matrizes de incidência que associam b e a aos dados y .

Um registro fenotípico para uma determinada característica é modelado como uma combinação linear dos efeitos de algumas variáveis explicativas. Assume-se que a distribuição dos dados y (vetor de ordem n) para essa característica, dado alguns parâmetros β , a e σ_e^2 , segue o processo normal multivariado como apresentado na Eq. 2.3

$$\begin{aligned} y|\beta, a, \sigma_e^2 &\sim \mathcal{N}(X\beta + Za, I\sigma_e^2) \\ e &\sim \mathcal{N}(\mathbf{0}, I\sigma_e^2) \end{aligned} \quad (2.2)$$

onde I é a matriz identidade e σ_e^2 é a variância residual.

Na modelagem estatística de dados genéticos, a matriz de parentesco (ou matriz de relacionamento aditivo) desempenha um papel central, especialmente na inferência bayesiana aplicada à genética quantitativa. Essa matriz, geralmente representada pela letra A , descreve o grau de parentesco aditivo entre indivíduos, ou seja, a proporção esperada de alelos idênticos por descendência compartilhados entre pares de indivíduos em uma população [32].

A matriz de parentesco é crucial para especificar a estrutura de covariância dos efeitos genéticos aditivos em modelos lineares mistos e modelos bayesianos hierárquicos. No contexto bayesiano, ela é frequentemente incorporada na especificação da distribuição a priori dos efeitos genéticos. Por exemplo, os efeitos genéticos podem ser assumidos como provenientes de uma distribuição normal multivariada com média zero e matriz de variância-covariância proporcional à matriz de parentesco A multiplicada pela variância aditiva genética (σ_A^2). Matematicamente, temos:

$$a|A, \sigma_a^2 \sim \mathcal{N}(O, A\sigma_a^2) \quad (2.3)$$

onde a representa o vetor de efeitos genéticos aditivos e A é a matriz de parentesco genético aditivo (de dimensão $q \times q$) e σ_a^2 é a variância genética aditiva na população base. Em uma perspectiva bayesiana a distribuição apresentada em 2.3 representa a incerteza sobre os efeitos genéticos antes que os dados sejam observados ou, em outras palavras, é a distribuição *a priori* de tais efeitos. Um valor alto de σ_a^2 implica uma grande incerteza. Os parâmetros que são o foco da inferência são β , a , σ_a^2 e σ_e^2 , e, possivelmente, funções deles, como o coeficiente de herdabilidade $\frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}$. Para realizar uma análise bayesiana, distribuições *a priori* devem ser atribuídas a cada um dos parâmetros β , a , σ_a^2 e σ_e^2 .

A matriz A reflete o pedigree da população e sua correta especificação é fundamental para capturar as dependências entre os indivíduos, contribuindo para estimativas mais precisas dos componentes de variância e dos valores genéticos preditos.

A inferência bayesiana fundamenta-se na determinação da distribuição conjunta a posteriori dos parâmetros de interesse, combinando informações prévias e evidências observacionais. No contexto do modelo animal bayesiano, envolvendo os efeitos fixos β , os efeitos genéticos aditivos a , a variância genética aditiva σ_a^2 e a variância residual σ_e^2 , o Teorema de Bayes possibilita expressar a distribuição a posteriori desses parâmetros como proporcional ao produto da verossimilhança pelas distribuições *a priori* associadas a cada parâmetro.

Em particular, a posteriori conjunta pode ser escrita conforme a Eq. 2.4, que apresenta a distribuição a posteriori proporcional ao produto da verossimilhança do modelo com as prioris dos efeitos genéticos, dos efeitos fixos e das variâncias. Esta formulação é suficiente para realizar a inferência bayesiana, uma vez que a constante de normalização, comum a todos os parâmetros, não interfere na análise dos parâmetros de interesse [30][11].

$$p(\beta, \mathbf{a}, \sigma_a^2, \sigma_e^2 | \mathbf{y}) \propto p(\mathbf{y} | \beta, \mathbf{a}, \sigma_e^2) \cdot p(\mathbf{a} | \sigma_a^2) \cdot p(\sigma_a^2) \cdot p(\sigma_e^2) \cdot p(\beta) \quad (2.4)$$

Para o problema ou modelo abordado neste trabalho, a distribuição conjunta a posteriori é decomposta ou foi construída a partir da função de verossimilhança apresentada na Eq. 2.5 e das distribuições *a priori* para β , a , σ_a^2 e σ_e^2 apresentadas, respectivamente, nas Eqs. 2.6, 2.7, 2.9 e 2.10.

A função de verossimilhança do modelo animal considerado neste trabalho é construída com base no modelo linear geral $\mathbf{y} = \mathcal{X}\beta + \mathcal{Z}\mathbf{a} + \mathbf{e}$, onde $\mathbf{e} \sim \mathcal{N}(0, \sigma_e^2 \mathcal{I})$. Dessa forma, a verossimilhança dos dados y , condicionada aos parâmetros do modelo, é proporcional a:

$$p(\mathbf{y} | \beta, \mathbf{a}, \sigma_e^2) \propto (\sigma_e^2)^{-n/2} \exp \left(-\frac{1}{2\sigma_e^2} (\mathbf{y} - \mathcal{X}\beta - \mathcal{Z}\mathbf{a})^\top (\mathbf{y} - \mathcal{X}\beta - \mathcal{Z}\mathbf{a}) \right) \quad (2.5)$$

A função de verossimilhança modela, portanto, a distribuição dos erros residuais do modelo, assumindo que são independentes e homocedásticos.

Para os efeitos fixos β , foi assumida uma distribuição *a priori* normal multivariada informativa, baseada em conhecimento prévio, conforme discutido anteriormente. Especificamente,

temos:

$$\boldsymbol{\beta} \propto \mathcal{N}(\mu_0, \Sigma_0) \quad (2.6)$$

onde μ_0 é um vetor de médias previamente especificadas e Σ_0 é a matriz de covariância correspondente, refletindo a incerteza associada a essas expectativas. Para o vetor de efeitos genéticos aditivos \mathbf{a} , foi alocada uma informação prévia por meio de uma distribuição normal multivariada, apresentada na forma:

$$p(\mathbf{a}|\sigma_a^2) \propto (\sigma_a^2)^{-n_a/2} |\mathcal{A}|^{-1/2} \exp\left(-\frac{1}{2\sigma_a^2} \mathbf{a}^\top \mathcal{A}^{-1} \mathbf{a}\right), \quad (2.7)$$

Em notação compacta, essa distribuição é expressa como:

$$\mathbf{a}|\sigma_a^2 \sim \mathcal{N}(0, \sigma_a^2 \mathcal{A}). \quad (2.8)$$

onde \mathcal{A} é a matriz de parentesco entre os animais, responsável por capturar a estrutura genética populacional. Essa matriz é fundamental para a correta modelagem da transmissão dos valores genéticos aditivos.

As variâncias genética aditiva σ_a^2 e residual σ_e^2 receberam distribuições a priori do tipo Inversa-Gama, que controlam a variabilidade genética e ambiental, respectivamente. A priori para σ_a^2 é dada por:

$$\sigma_a^2 \sim \text{Inv-}\chi^2(\nu_a, S_a) \quad (2.9)$$

e, analogamente, para σ_e^2 tem-se:

$$\sigma_e^2 \sim \text{Inv-}\chi^2(\nu_e, S_e) \quad (2.10)$$

A forma funcional da densidade de probabilidade das equações 2.9 e 2.10 é, respectivamente:

$$p(\sigma_a^2) \propto (\sigma_a^2)^{-(\nu_a/2+1)} \exp\left(-\frac{S_a}{2\sigma_a^2}\right) \quad (2.11)$$

e,

$$p(\sigma_e^2) \propto (\sigma_e^2)^{-(\nu_e/2+1)} \exp\left(-\frac{S_e}{2\sigma_e^2}\right) \quad (2.12)$$

onde ν_a e ν_e representam os graus de liberdade associados, e S_a e S_e os parâmetros de escala, definidos com base em conhecimento prévio sobre a magnitude esperada dessas variâncias no contexto da produção leiteira.

Essas especificações são essenciais para a construção da distribuição a posteriori completa, pois cada elemento da função é incorporado no produto da verossimilhança pelas prioris, con-

forme determinado pelo Teorema de Bayes.

Portanto, a partir das equações 2.5, 2.6, 2.7, 2.11 e 2.12, a expressão da posteriori completa é dada por:

$$\begin{aligned}
 p(\boldsymbol{\beta}, \mathbf{a}, \sigma_a^2, \sigma_e^2 | \mathbf{y}) &\propto (\sigma_e^2)^{-n/2} \exp\left(-\frac{1}{2\sigma_e^2}(\mathbf{y} - \mathcal{X}\boldsymbol{\beta} - \mathcal{Z}\mathbf{a})^\top(\mathbf{y} - \mathcal{X}\boldsymbol{\beta} - \mathcal{Z}\mathbf{a})\right) \\
 &\quad \times (\sigma_a^2)^{-n_a/2} |\mathcal{A}|^{-1/2} \exp\left(-\frac{1}{2\sigma_a^2} \mathbf{a}^\top \mathcal{A}^{-1} \mathbf{a}\right) \\
 &\quad \times (\sigma_a^2)^{-(\nu_a/2+1)} \exp\left(-\frac{S_a}{2\sigma_a^2}\right) \\
 &\quad \times (\sigma_e^2)^{-(\nu_e/2+1)} \exp\left(-\frac{S_e}{2\sigma_e^2}\right).
 \end{aligned} \tag{2.13}$$

Uma das grandes dificuldades da inferência bayesiana é a complexidade em se obter medidas que resumam adequadamente a distribuição a posteriori de cada parâmetro [32]. Em contraste com métodos frequentistas, que muitas vezes fornecem estimativas pontuais e intervalos de confiança diretamente das soluções analíticas, o enfoque bayesiano lida com distribuições completas e, frequentemente, de alta dimensionalidade e forma não trivial [11].

Quando as distribuições a posteriori são de formas conhecidas e simples, como normais ou outras distribuições bem comportadas, pode-se calcular diretamente medidas como a média, a mediana e intervalos de credibilidade. Porém, em modelos mais realistas e hierárquicos, como o modelo animal, as posterioris resultantes podem apresentar forte assimetria, múltiplas modas ou dependência complexa entre parâmetros [32].

Além disso, as integrais necessárias para obter medidas marginais raramente são resolvíveis de forma analítica. Isso torna essencial o uso de métodos numéricos, como os algoritmos MCMC, para amostrar da distribuição posterior. Mesmo após a geração de amostras via métodos MCMC, persiste o desafio de garantir que essas amostras sejam representativas da distribuição estacionária desejada. A principal preocupação refere-se à convergência da cadeia de Markov para sua distribuição limite, que pode ser prejudicada por diversos fatores, como a influência dos valores iniciais escolhidos (efeito de inicialização) e a presença de alta autocorrelação entre amostras sucessivas [6][30].

O efeito do valor inicial pode fazer com que a cadeia leve muito tempo para "esquecer" o estado inicial, retardando o alcance da distribuição estacionária. Já a alta autocorrelação reduz a eficiência da amostragem, pois gera amostras altamente dependentes, que representam menos informação efetiva sobre a posteriori do que um conjunto de amostras aproximadamente independentes.

Por isso, é essencial realizar diagnósticos de convergência adequados, como o critério de Gelman-Rubin, o critério de Geweke, o critério de Raftery & Lewis e as análises de autocorrelação, para assegurar que a cadeia tenha atingido um comportamento estacionário e que a amostragem seja suficientemente eficiente. Apenas após essa validação torna-se apropriado extrair medidas resumo da posteriori, como a média, a mediana e os intervalos de credibilidade, que sintetizam a informação inferida a partir do modelo.

Os métodos MCMC tornaram-se uma ferramenta computacional fundamental na estatística bayesiana, especialmente devido à necessidade de obter amostras da distribuição a posteriori, que é a base de toda a inferência sob a perspectiva bayesiana. Em muitos problemas reais, essa distribuição assume formas complexas e de alta dimensão, tornando inviável o uso de técnicas analíticas ou numéricas tradicionais de integração [33]. Assim, os métodos MCMC surgiram como uma solução prática e eficiente para amostrar dessa distribuição de maneira iterativa e probabilística.

O princípio central dos métodos MCMC consiste em construir uma cadeia de Markov cuja distribuição estacionária corresponda, em um sentido assintótico, à distribuição a posteriori desejada [37]. Entre os métodos mais relevantes estão o algoritmo de Metropolis-Hastings e o Gibbs Sampling, que desempenham papéis centrais na prática da inferência bayesiana moderna.

O algoritmo Metropolis-Hastings [24][18] é um dos métodos mais versáteis e amplamente utilizados. Ele se baseia em uma estratégia de aceitação-rejeição que permite gerar amostras mesmo em distribuições com estruturas complexas e espaços de parâmetros de difícil acesso. Por outro lado, o Gibbs Sampling, introduzido por Geman e Geman (1984) [13], é um caso particular do Metropolis-Hastings e tem se mostrado particularmente útil em modelos hierárquicos, onde as variáveis podem ser amostradas sequencialmente de suas distribuições condicionais completas [33].

A necessidade de métodos como Gibbs e Metropolis-Hastings se intensifica em contextos onde o número de parâmetros do modelo é elevado ou onde dependências complexas entre os parâmetros tornam a amostragem direta da posteriori impraticável. Sem essas ferramentas, a inferência bayesiana seria limitada a modelos simplificados ou requereria aproximações grosseiras.

O desenvolvimento posterior, como o algoritmo de salto reversível de Green (1995) [17], ampliou ainda mais as aplicações, permitindo que distribuições a posteriori em espaços de dimensão variável fossem tratadas eficientemente. No entanto, em muitos cenários aplicados, como na genética quantitativa ou em modelos de regressão hierárquica, os algoritmos de Gibbs e Metropolis-Hastings continuam sendo as abordagens mais empregadas para a geração de amostras da distribuição a posteriori [33].

Embora métodos como o Gibbs Sampling e o Metropolis-Hastings sejam amplamente utilizados em inferência bayesiana, especialmente em genética quantitativa, eles podem apresentar limitações quando aplicados a modelos com alta dimensionalidade ou com distribuições a posteriori altamente correlacionadas e complexas [33]. Nesse cenário, o Hamiltonian Monte Carlo (HMC) surge como uma alternativa poderosa.

O HMC é um algoritmo de amostragem baseado nos princípios da mecânica hamiltoniana. A ideia central é tratar os parâmetros de interesse como “posições” e introduzir variáveis auxiliares chamadas de “momentos”, que simulam um sistema físico em movimento. A energia total do sistema, descrita pela função Hamiltoniana, é composta pela soma da energia potencial (associada à função negativa do logaritmo da distribuição a posteriori) e da energia cinética (associada aos momentos auxiliares) [26].

A simulação da dinâmica hamiltoniana permite que o algoritmo explore o espaço de parâmetros de maneira direcionada, evitando o comportamento errático de “passeio aleatório” típico de algoritmos como o Metropolis-Hastings.

O algoritmo *Hamiltonian Monte Carlo* (HMC) é inspirado na mecânica clássica, especificamente na dinâmica Hamiltoniana, que modela a evolução de sistemas físicos por meio de duas variáveis principais: a **posição** q e o **momento** p .

A **função Hamiltoniana**, denotada por $H(q, p)$, representa a *energia total* de um sistema físico e é definida como:

$$H(q, p) = U(q) + K(p)$$

onde:

- $U(q)$ é a **energia potencial**, dependente apenas da posição q ;
- $K(p)$ é a **energia cinética**, dependente apenas do momento p .

No contexto estatístico do HMC, essa formulação é adaptada da seguinte forma:

- A posição q representa os parâmetros do modelo estatístico;
- O momento p é uma variável auxiliar introduzida para facilitar a amostragem;
- A energia potencial é dada por $U(q) = -\log \pi(q)$, onde $\pi(q)$ é a densidade da distribuição posterior (desconsiderando a constante de normalização);
- A energia cinética normalmente assume a forma $K(p) = \frac{1}{2}p^\top M^{-1}p$, com $p \sim \mathcal{N}(0, M)$, sendo M a matriz de massa (geralmente identidade).

Assim, a Hamiltoniana que interessa no contexto de HMC é:

$$H(q, p) = -\log \pi(q) + \frac{1}{2}p^\top M^{-1}p$$

A evolução do sistema é regida pelas chamadas **equações de Hamilton**, dadas por:

$$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i}, \quad \frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}$$

Essas equações descrevem como os parâmetros e os momentos evoluem ao longo do tempo, de forma a preservar a energia total do sistema. No algoritmo HMC, essa simulação é realizada numericamente (por exemplo, por meio do método *leapfrog*), permitindo propostas mais eficientes para a cadeia de Markov com menor autocorrelação entre amostras.

No HMC, uma trajetória é definida como o caminho percorrido pelos parâmetros e seus momentos através do espaço, conforme a simulação das equações diferenciais de Hamilton. Essa trajetória é calculada utilizando o método numérico chamado leapfrog, que alterna atualizações das posições e dos momentos em pequenos passos discretos [4].

De forma resumida, o algoritmo HMC segue os seguintes passos:

- **Início:** Defina a posição inicial $\theta^{(1)}$ e o tamanho dos passos ϵ .

- **Passo generalizado:** Para $t = 1, 2, \dots$, faça:

- **Reamostragem do momento r :**

$$r^{(t)} \sim N(0, M)$$

Atualize $(\theta_0, r_0) = (\theta^{(t)}, r^{(t)})$.

- **Simulação da discretização da dinâmica hamiltoniana:**

$$r_0 \leftarrow r_0 - \frac{\epsilon}{2} \nabla U(\theta_0)$$

onde $U(\theta)$ é a função potencial associada à posição θ .

- **Iteração para $i = 1, 2, \dots, L$:**

$$\theta_i \leftarrow \theta_{i-1} + \epsilon M^{-1} r_{i-1}$$

$$r_i \leftarrow r_{i-1} - \epsilon \nabla U(\theta_i)$$

onde M^{-1} é a matriz inversa de M , a matriz de massa associada.

- **Atualização final do momento:**

$$r_L \leftarrow r_L - \frac{\epsilon}{2} \nabla U(\theta_L)$$

- **Resultado da amostragem:**

$$(\hat{\theta}, \hat{r}) = (\theta_L, r_L)$$

- **Amostragem de Metropolis-Hastings:** Gere $u \sim \text{Uniforme}[0, 1]$. Calcule a razão de aceitação:

$$\rho = \exp(H(\hat{\theta}, \hat{r}) - H(\theta^{(t)}, r^{(t)}))$$

onde $H(\theta, r)$ é a função hamiltoniana associada à energia do sistema. Se $u < \min(1, \rho)$, então aceite a amostra:

$$\theta^{(t+1)} = \hat{\theta}$$

A Figura 2.1 ilustra o movimento de um objeto em uma curva suave e sem atrito, conforme a dinâmica do HMC. Esse exemplo unidimensional ajuda a visualizar como o algoritmo explora o espaço de parâmetros.

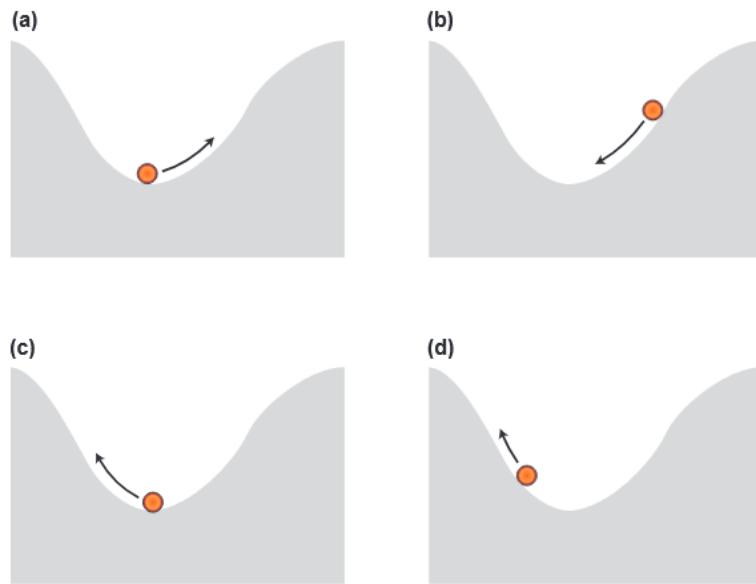


Figura 2.1: Exemplo unidimensional de HMC - movimento de um objeto em uma curva suave e sem atrito. [36]

- (a) Aplica-se uma força com direção e intensidade aleatórias sobre o objeto.
- (b) O objeto se move ao longo da curva, com sua energia cinética sendo convertida em energia potencial.
- (c) No ponto mais baixo da curva, toda a energia está na forma de cinética.
- (d) A energia cinética é novamente convertida em energia potencial, e o movimento continua em direção oposta.

Apesar de suas vantagens, o desempenho do HMC depende criticamente da escolha dos hiperparâmetros, como o número de passos da trajetória (L) e o tamanho do passo (ϵ). Escolhas inadequadas podem levar a trajetórias muito curtas, que não exploram bem o espaço, ou muito longas, que desperdiçam computação. Para solucionar essa questão, foi desenvolvido o No-U-Turn Sampler (NUTS), uma extensão adaptativa do HMC que ajusta automaticamente o comprimento da trajetória, parando o movimento antes que ele comece a “dar voltas” desnecessárias, otimizando a eficiência da amostragem [19].

O NUTS elimina a necessidade de definir o número de passos L de forma manual, construindo trajetórias dinâmicas até que um critério de parada (a “não-volta” ou no U-turn) seja atingido [19]. Esse critério detecta quando a trajetória proposta começa a voltar em direção ao ponto inicial, o que é um sinal de que o espaço de parâmetros já foi suficientemente explorado naquela direção. Assim, o NUTS adapta automaticamente o comprimento das trajetórias durante a amostragem, otimizando o equilíbrio entre eficiência computacional e exploração do espaço amostral [19].

Além disso, o NUTS é frequentemente combinado com esquemas de ajuste automático do tamanho de passo (ϵ) durante as fases iniciais da simulação (warm-up), resultando em um método robusto para problemas com alta dimensionalidade e topologias complexas da posteriori.

Nos últimos anos, tanto o HMC quanto o NUTS têm sido implementados em softwares modernos de inferência bayesiana, como Stan [10], sendo aplicados em uma ampla gama de

problemas em estatística, biologia, genética quantitativa e machine learning, principalmente em modelos hierárquicos complexos onde os métodos MCMC tradicionais apresentam dificuldades [26][19]. Diante dessas vantagens, neste trabalho, será utilizado o algoritmo NUTS como método de amostragem da distribuição a posteriori.

Embora os métodos MCMC, como o Gibbs Sampling e o Metropolis-Hastings, sejam fundamentais para amostrar da distribuição a posteriori, a confiabilidade dessas amostras depende da garantia de que a cadeia de Markov atingiu sua distribuição estacionária. Ou seja, é fundamental assegurar que as amostras coletadas refletem adequadamente a distribuição a posteriori de cada parâmetro do modelo [37]. Caso contrário, as inferências podem ser enviesadas ou imprecisas. A convergência garante que os valores amostrados representam, de fato, a distribuição a posteriori dos parâmetros de interesse, permitindo interpretações confiáveis dos resultados.

Para isso, foram utilizados diferentes critérios de diagnóstico de convergência, aplicados às cadeias geradas por meio do algoritmo padrão do Stan, o NUTS. No total foram obtidas três cadeias amostradas de forma independentes. O Stan, por padrão, combina as cadeias geradas em uma cadeia única, construída a partir da junção das anteriores e portanto, neste estudo, obteve-se quatro cadeias para cada parâmetro. Os critérios empregados foram:

1. Critério de Gelman & Rubin (1992): Esse método é amplamente utilizado por sua simplicidade e eficácia. O diagnóstico compara a variância entre múltiplas cadeias MCMC iniciadas em pontos diferentes com a variância dentro de cada cadeia. A métrica resultante, conhecida como \hat{R} (R-hat), deve se aproximar de 1 para indicar convergência. Valores de \hat{R} expressivamente maiores que 1 sugerem que as cadeias ainda não estão bem misturadas ou não atingiram a distribuição estacionária [12].
2. Diagnóstico de Geweke (1992): O critério de Geweke avalia a estacionariedade de uma única cadeia, comparando a média das amostras da parte inicial da cadeia (geralmente o primeiro 10%) com a média da parte final (os últimos 50%). A comparação é realizada obtendo um escore z , em que quando o valor deste escore é maior que 1,96 ou menor que -1,96, por exemplo, assume-se que existem evidências suficientes para a rejeição da hipótese de convergência da respectiva cadeia ou que a cadeia ainda não convergiu. Esse método é útil para avaliar a estabilidade de parâmetros ao longo da cadeia [14].
3. Diagnóstico de Raftery & Lewis (1992): Este método fornece uma estimativa do número mínimo de iterações necessárias para garantir uma precisão desejada em termos de quantis da distribuição a posteriori. Além disso, sugere o tamanho do burn-in (as iterações iniciais a serem descartadas) e o espaçamento ideal entre amostras (thinning) para reduzir autocorrelação. O método é particularmente útil quando há limitação computacional ou quando se deseja garantir eficiência no uso das amostras [29].

A utilização combinada desses diagnósticos é altamente recomendada, pois eles oferecem diferentes perspectivas sobre a convergência das cadeias. Garantir que as amostras sejam representativas da distribuição a posteriori é um passo essencial para que as inferências bayesianas baseadas em MCMC sejam robustas e confiáveis [33][37].

Além dos critérios formais, é prática comum na análise bayesiana empregar ferramentas gráficas como suporte à avaliação da convergência das cadeias MCMC. Esses métodos informais são especialmente úteis quando os testes formais indicam possíveis incongruências. Entre os recursos gráficos mais utilizados estão os gráficos de traços (traceplots), que representam a evolução das amostras ao longo das iterações. Idealmente, o gráfico de traço deve apresentar um padrão semelhante a um ruído branco. Cadeias com deslocamentos sistemáticos, patamares distintos ou separação visível entre cadeias indicam que a convergência pode não ter sido atingida. Em contrapartida, cadeias que se sobreponem bem, oscilam em torno de uma média comum e apresentam boa mistura visual são indícios favoráveis de convergência [5, 11].

Outros gráficos importantes incluem histogramas e boxplots das amostras posteriores, utilizados para verificar a estabilidade e a forma das distribuições marginais dos parâmetros entre cadeias independentes. A consistência visual entre esses gráficos para diferentes cadeias reforça a evidência de que a amostragem atingiu a distribuição estacionária. Adicionalmente, a função de autocorrelação (ACF) permite avaliar a dependência entre amostras consecutivas. Altos valores de autocorrelação indicam que a cadeia percorre o espaço de maneira lenta, reduzindo a eficiência da amostragem. Dessa forma, mesmo que algum parâmetro apresente resultados limítrofes nos testes formais, a inspeção gráfica pode justificar sua aceitação, desde que os padrões visuais observados estejam compatíveis com convergência adequada [5, 7].

Uma vez assegurada a convergência das cadeias e a qualidade das amostras, a inferência sobre os parâmetros pode ser feita a partir de estatísticas-resumo da distribuição a posteriori, como a média, a mediana ou os intervalos de credibilidade. Dentre os tipos de intervalos, destaca-se o intervalo de máxima densidade a posteriori (HPD - Highest Posterior Density), que representa a região mais provável da distribuição onde a densidade é mais alta e que contém uma determinada proporção da massa de probabilidade, geralmente 95%. Diferente de intervalos simétricos em torno da média, os intervalos HPD são mais informativos, especialmente quando a distribuição a posteriori é assimétrica ou multimodal [20]. A interpretação de um intervalo HPD de 95% é que há 95% de probabilidade de o parâmetro estar dentro daquele intervalo, dado o modelo e os dados observados.

2.1 OS DADOS

A base de dados utilizada neste trabalho é composta por informações sobre as vacas Girolando da Fazenda Cruzeiro do Sul, obtidas por meio do Portal Web do produtor da Associação Brasileira dos Criadores de Girolando. Essas informações são coletadas e registradas periodicamente por um técnico de leite, que realiza medições e calcula a média diária da produção de leite das vacas. A base de dados do portal contém uma ampla gama de variáveis, que são constantemente atualizadas e permitem uma análise detalhada dos rebanhos.

Para a realização deste estudo, foram extraídas informações a partir do cruzamento de três bases de dados presentes no portal, contemplando dados fenotípicos e genômicos dos animais. As variáveis selecionadas para esta análise incluem: o nome do animal, nome dos pais, número

de registro, grau de sangue, PTA (Predicted Transmitting Ability, ou Habilidade Prevista de Transmissão) [27], data de nascimento, idade no parto, idade ao primeiro parto (IPP), mês e data do parto, data de encerramento da lactação, média diária de produção de leite, dias de lactação e produção total de leite.

É importante ressaltar que a base de dados utilizada neste estudo é composta por 160 vacas distintas, totalizando 350 registros de lactações, uma vez que para algumas vacas foram registradas múltiplas lactações ao longo do período de observação. Para fortalecer a estrutura genética e a matriz de parentesco do estudo, também foram incluídas informações de 59 pais e 76 mães dessas vacas, correspondendo a 135 registros adicionais. Assim, a base final compreende 485 observações relativas a 295 animais únicos.

Dentre esses 295 animais, 160 correspondem às vacas com registros de lactação, enquanto os demais referem-se aos seus ascendentes, herdabilidadequais, embora não haja observações fenotípicas diretas de produção de leite, é possível estimar seus valores genéticos aditivos por meio da matriz de parentesco. Essa capacidade é uma das principais vantagens do modelo animal: permitir a estimativa dos valores genéticos mesmo para indivíduos sem registros diretos, utilizando as informações genéticas herdadas de seus descendentes e ancestrais [25].

Na modelagem proposta, portanto, foram estimados 295 parâmetros associados aos efeitos genéticos aditivos (a) — um para cada animal incluído na matriz de parentesco.

Além disso, a base original contém diversas variáveis. No entanto, para este estudo foram selecionadas apenas três variáveis: a produção média diária de leite, o grau de sangue e a idade ao primeiro parto (IPP). A produção média diária foi escolhida por ser um indicador direto da eficiência produtiva da vaca, essencial para programas de melhoramento genético. O grau de sangue foi incluído porque a composição racial influencia diretamente o desempenho produtivo e reprodutivo dos animais, principalmente na raça Girolando, que é fruto do cruzamento entre Gir e Holandês [22]. Por fim, a idade ao primeiro parto foi selecionada devido à sua forte relação com a longevidade produtiva e a eficiência reprodutiva das vacas leiteiras, sendo considerada uma característica econômica de alta relevância [3].

Os dados presentes nessa base oferecem uma oportunidade para uma análise detalhada e precisa dos parâmetros genéticos, permitindo um estudo aprofundado das características reprodutivas e produtivas das vacas Girolando. A utilização dessas informações possibilita uma compreensão mais clara dos fatores que influenciam o desempenho do rebanho, fornecendo informações valiosas para a implementação de estratégias de melhoramento genético e otimização da produção leiteira.

2.2 IMPLEMENTAÇÃO

A implementação das análises foi realizada no software R [35], uma linguagem amplamente utilizada para análise estatística e científica. A análise envolveu a utilização de dados genéticos em formato de pedigree, a fim de modelar a herança da característica média diária de produção de leite, levando em consideração tanto a variabilidade genética quanto a ambiental .

O banco de dados utilizado, conforme descrito na seção 2.1, contém informações sobre os indivíduos, seus pais, a variável de interesse (média diária de produção de leite por vaca) e as variáveis explicativas, grau de sangue (GS) e idade ao primeiro parto (IPP). Essas variáveis foram selecionadas para a modelagem, e a matriz de parentesco foi incorporada como informação prévia, sendo alocada na distribuição a priori dos efeitos genéticos aditivos.

A produção de leite de algumas vacas foi registrada em mais de um período de lactação. No entanto, assim como para aquelas com apenas um registro, a variável resposta final considerada foi a média diária de produção de leite. Mesmo com a agregação dos dados em médias diárias, as vacas que tiveram múltiplos registros contribuíram mais de uma vez para a variável resposta y . Essa repetição de observações para os mesmos animais é importante, pois permite uma melhor estimativa da variância residual do modelo, refletindo a variabilidade não explicada pelos efeitos genéticos e fixos. Portanto, embora a modelagem tenha utilizado a média diária de produção, as repetições contribuíram para melhorar a precisão da estimativa da variância residual, que é crucial para avaliar a incerteza associada à produção de leite.

Uma parte crucial na implementação do modelo apresentado na Eq. 2.1 é a correta especificação das matrizes X e Z , com dimensões $n \times p$ e $n \times q$, respectivamente. Na matriz de incidência X , foram alocadas as variáveis de efeito fixo, grau de sangue GS e idade ao primeiro parto IPP. A variável GS é categórica, por isso, foram criadas 9 variáveis dummy para representar os diferentes graus sanguíneos. Os parâmetros associados a essas variáveis representam a média de produção de leite em cada categoria de GS. O parâmetro relacionado à IPP, por sua vez, reflete o efeito dessa variável na produção média diária de leite. Esses 10 parâmetros compõe o β .

Dessa forma, a matriz X tem dimensão 350×10 , pois o banco de dados possui 350 observações de lactação média diária por período, com 10 parâmetros associados a essa matriz. A matriz Z , por sua vez, indica a associação entre o valor de produção média de leite (y) e o efeito genético aditivo (a) de cada animal. Essa matriz, com dimensão 350×295 , reflete a relação dos dados de produção com os efeitos genéticos aditivos, sendo que a quantidade de animais diferentes no banco de dados é 295. Importante destacar que, mesmo para indivíduos que não possuem observações diretas de produção de leite, como no caso dos touros, seus valores genéticos aditivos podem ser estimados a partir das informações de pedigree.

A distribuição a priori para β , foi uma normal multivariada (Eq. 2.6) com vetor de médias 18, 15, 20, 17, 17, 19, 20, 16, 19 e 0, respectivamente para os parâmetros associados aos graus de sangue $\frac{1}{2}H$, $\frac{1}{4}H$, $\frac{3}{4}H$, $\frac{3}{8}H$, $\frac{5}{16}H$, $\frac{5}{8}H$, $\frac{7}{8}H$, GIR, PS e IPP. Essas médias para os diferentes graus de sangue refletem o conhecimento prévio dos produtores rurais da região. O valor 0 para o parâmetro associado à IPP indica a falta do conhecimento do efeito desta variável sobre a produção de leite. A incerteza a priori foi retratada por uma matriz de variâncias e covariâncias em que as covariâncias são todas nulas e as variâncias são todas iguais a 4 para os parâmetros associados aos graus de sangue e igual a 9 para o parâmetro associado ao IPP.

Para as variâncias σ_e^2 e σ_a^2 , foram atribuídas distribuições gama inversa, com parâmetros (2, 14) e (2, 9), respectivamente. Assim, o valor esperado e a moda a priori para σ_e^2 é respecti-

vamente, 14 e 4,6, enquanto para σ_a^2 é respectivamente, 9 e 3. E desta forma uma estimativa a priori para h^2 seria $\frac{9}{9+14} = 0,3913$, usando o valor esperado e $\frac{3}{3+4,6} = 0,3947$, usando a moda.

Para os efeitos genéticos aditivos (a) a distribuição a priori foi normal-multivariada (Eq. 2.8) com média 0 e matriz de variâncias e covariâncias, $\sigma_a^2 \mathcal{A}$.

A análise estatística foi conduzida utilizando o Stan [10], uma linguagem de programação probabilística amplamente usada para inferência bayesiana. O modelo **Stan** utilizado está disponibilizado no Apêndice A deste trabalho.

O Stan foi acessado a partir do R [35] por meio da interface rstan, que conecta diretamente o R ao Stan. O código R utilizado é apresentado no Apêndice B. Para a amostragem MCMC dos parâmetros, configurou-se o modelo para realizar 22.000 iterações com 3 cadeias. As 2.000 iterações iniciais foram reservadas para o aquecimento (warmup), etapa necessária para ajustar os valores à distribuição de interesse, já que os primeiros valores podem ainda ser influenciados pela escolha inicial da cadeia. Após essa fase de aquecimento, foram salvos um valor a cada 5 iterações, resultando em uma amostra final de 4.000 valores por cadeia. O espaçamento de 5 iterações foi adotado para garantir que as amostras coletadas fossem independentes, o que é crucial para assegurar a qualidade e confiabilidade da inferência. Na implementação, utilizou-se o algoritmo NUTS para otimizar o processo de amostragem.

A convergência das cadeias para a distribuição de equilíbrio, condição necessária para a validade das inferências, foi verificada utilizando o gráfico traceplot e os critérios de Gelman-Rubin, Raftery & Lewis e Geweke para assegurar a confiabilidade das amostras obtidas.

3. RESULTADOS

Neste capítulo, antes da apresentação dos resultados que permitem inferir sobre os parâmetros do modelo (Eq. 2.1), serão discutidos os resultados dos diagnósticos de convergência. Inicialmente, a convergência das cadeias foi avaliada de maneira informal por meio de gráficos, examinando o comportamento das cadeias para todos os parâmetros. Propositalmente, nas Figuras 3.1, 3.2, 3.3 e 3.4, é apresentado o comportamento das cadeias para os seguintes parâmetros: a_{50} , a_{64} , a_{81} , a_{141} , a_{146} , a_{163} , a_{195} , a_{215} , a_{228} , a_{251} , a_{258} , β_2 , h_2 e σ_a^2 .

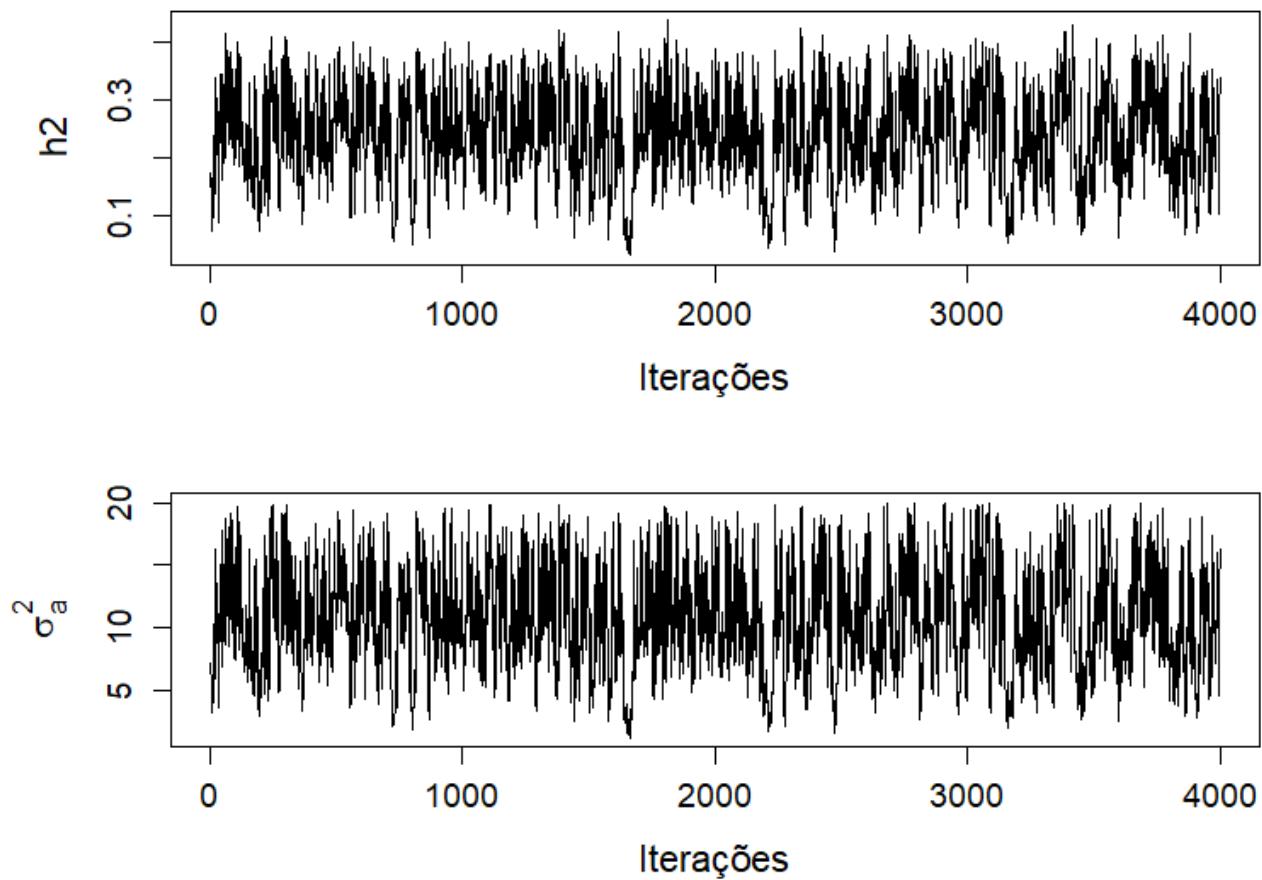


Figura 3.1: Gráfico de traço das cadeias MCMC dos parâmetros h_2 e σ_a^2 .

A análise dos critérios de convergência, como Raftery & Lewis e Geweke, indica que a convergência não é completamente satisfeita para todos os parâmetros. No entanto, ao observar o comportamento das cadeias nos gráficos (Figuras 3.1, 3.2, 3.3 e 3.4), além dos diagnósticos

formais apresentados nas Tabelas 3.1, 3.2 e 3.3, pode-se concluir que, para os fins deste trabalho, as cadeias são consideradas convergentes.

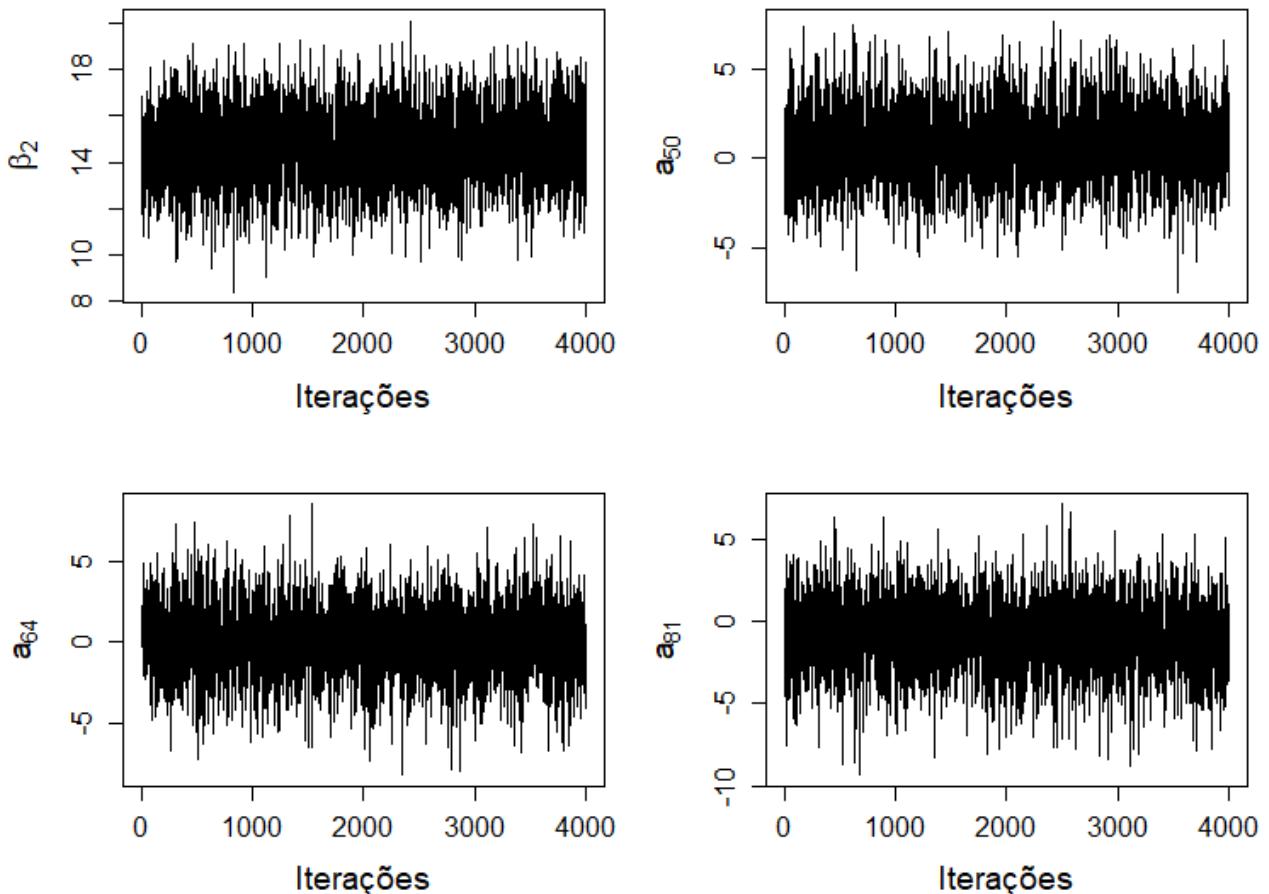


Figura 3.2: Gráfico de traço das cadeias MCMC dos parâmetros β_2 , a_{50} , a_{64} e a_{81} .

A convergência das cadeias foi confirmada pelo valor de \hat{R} , do critério de Gelman & Rubin, que foi igual a 1 para a grande maioria dos parâmetros, conforme mostrado na Tabela 3.1, e levemente superior a 1 para alguns parâmetros. Isso indica que, de forma geral, as cadeias atingiram a distribuição estacionária. Além disso, os valores da coluna \hat{R}_{sup} , que representam o limite superior do intervalo de confiança para \hat{R} , também se mostraram próximos de 1, reforçando a evidência de convergência satisfatória. Portanto, os resultados do diagnóstico de Gelman & Rubin sugerem que as iterações realizadas foram suficientes para garantir que o processo de amostragem explorou adequadamente o espaço dos parâmetros e gerou amostras representativas.

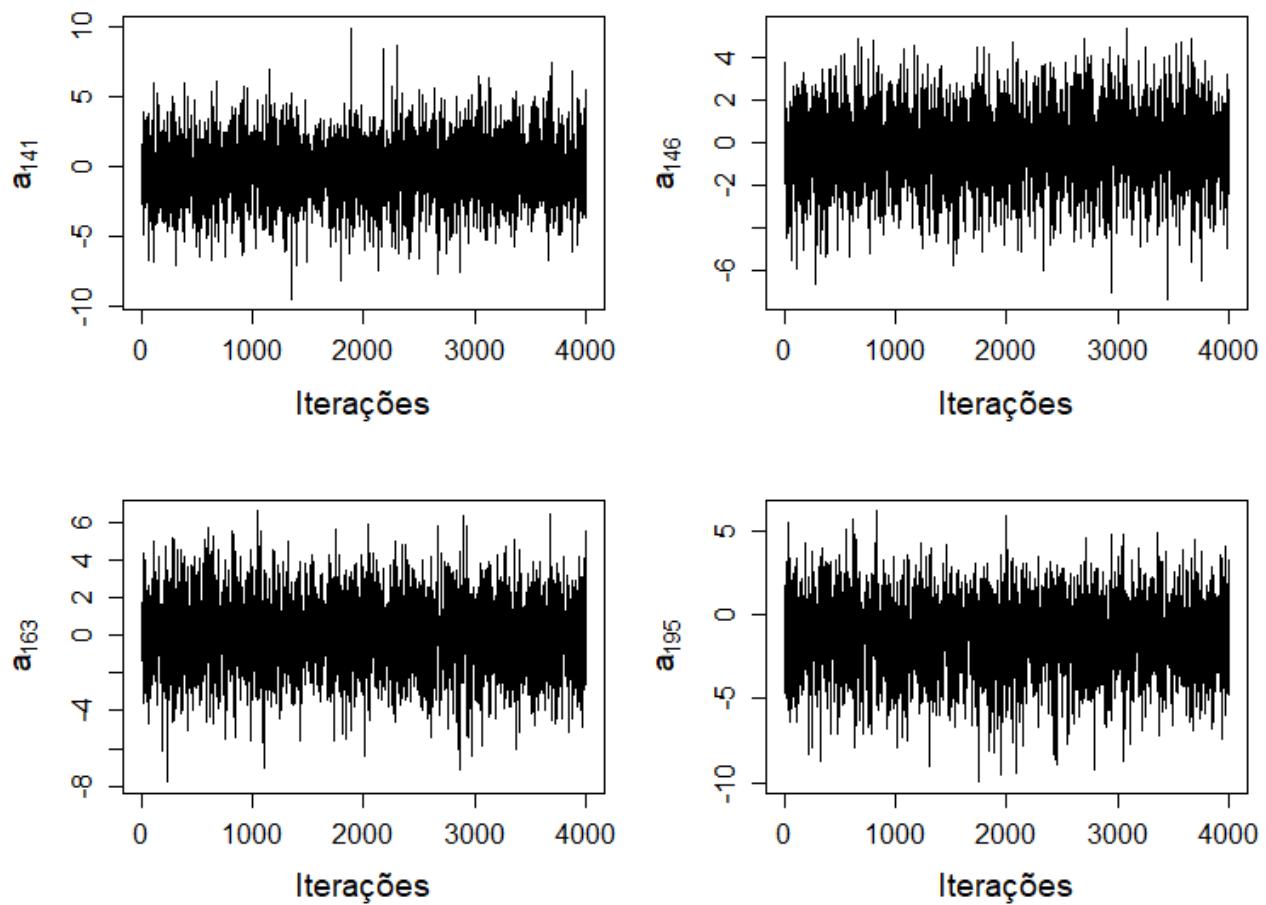


Figura 3.3: Gráfico de traço das cadeias MCMC dos parâmetros a_{141} , a_{146} , a_{163} e a_{195} .

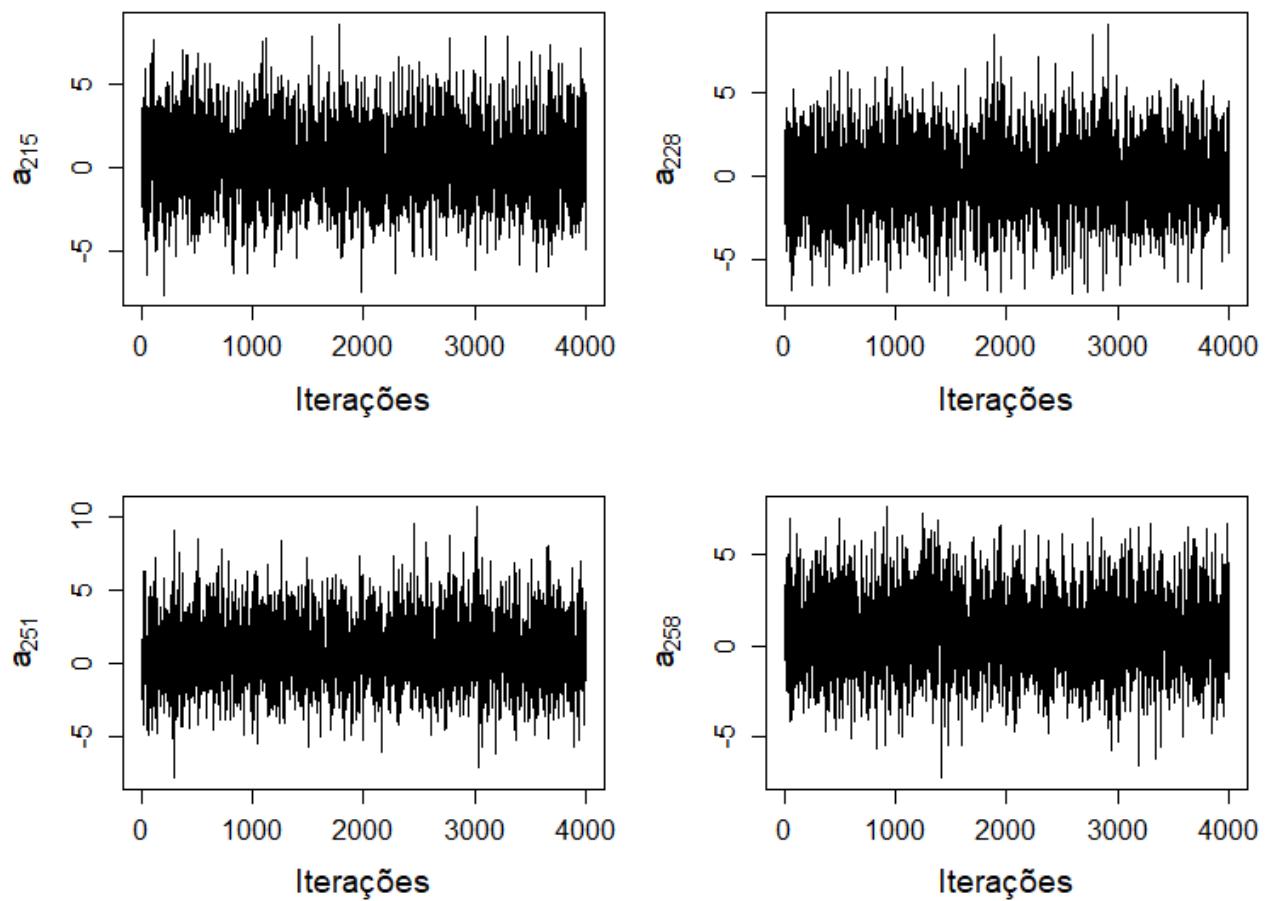


Figura 3.4: Gráfico de traço das cadeias MCMC dos parâmetros a_{215} , a_{228} , a_{251} e a_{258} .

Tabela 3.1: Resultados do Diagnóstico de Gelman & Rubin.

Parâmetro	\bar{R}	\bar{R}_{sup}	Parâmetro	\bar{R}	\bar{R}_{sup}	Parâmetro	\bar{R}	\bar{R}_{sup}	Parâmetro	\bar{R}	\bar{R}_{sup}
$\beta_1.$	1,00	1,00	$a_{69.}$	1,00	1,01	$a_{147.}$	1,00	1,00	$a_{225.}$	1,00	1,00
$\beta_2.$	1,00	1,00	$a_{70.}$	1,00	1,01	$a_{148.}$	1,00	1,00	$a_{226.}$	1,00	1,00
$\beta_3.$	1,00	1,00	$a_{71.}$	1,00	1,00	$a_{149.}$	1,00	1,00	$a_{227.}$	1,00	1,00
$\beta_4.$	1,00	1,00	$a_{72.}$	1,00	1,00	$a_{150.}$	1,00	1,00	$a_{228.}$	1,00	1,00
$\beta_5.$	1,00	1,00	$a_{73.}$	1,00	1,00	$a_{151.}$	1,00	1,00	$a_{229.}$	1,00	1,00
$\beta_6.$	1,00	1,00	$a_{74.}$	1,00	1,00	$a_{152.}$	1,00	1,00	$a_{230.}$	1,00	1,00
$\beta_7.$	1,00	1,00	$a_{75.}$	1,00	1,00	$a_{153.}$	1,00	1,00	$a_{231.}$	1,00	1,00
$\beta_8.$	1,00	1,00	$a_{76.}$	1,00	1,00	$a_{154.}$	1,00	1,00	$a_{232.}$	1,00	1,00
$\beta_9.$	1,00	1,00	$a_{77.}$	1,00	1,00	$a_{155.}$	1,00	1,00	$a_{233.}$	1,00	1,00
$\beta_{10.}$	1,00	1,00	$a_{78.}$	1,00	1,00	$a_{156.}$	1,00	1,01	$a_{234.}$	1,00	1,00
$a_1.$	1,00	1,00	$a_{79.}$	1,00	1,01	$a_{157.}$	1,00	1,00	$a_{235.}$	1,00	1,00
$a_2.$	1,00	1,00	$a_{80.}$	1,00	1,00	$a_{158.}$	1,00	1,00	$a_{236.}$	1,00	1,00
$a_3.$	1,00	1,00	$a_{81.}$	1,00	1,00	$a_{159.}$	1,00	1,00	$a_{237.}$	1,00	1,00
$a_4.$	1,00	1,00	$a_{82.}$	1,00	1,00	$a_{160.}$	1,00	1,00	$a_{238.}$	1,00	1,00
$a_5.$	1,00	1,00	$a_{83.}$	1,00	1,00	$a_{161.}$	1,00	1,00	$a_{239.}$	1,00	1,00
$a_6.$	1,00	1,00	$a_{84.}$	1,00	1,01	$a_{162.}$	1,00	1,00	$a_{240.}$	1,00	1,01
$a_7.$	1,00	1,01	$a_{85.}$	1,00	1,00	$a_{163.}$	1,00	1,00	$a_{241.}$	1,00	1,00
$a_8.$	1,00	1,01	$a_{86.}$	1,00	1,00	$a_{164.}$	1,00	1,00	$a_{242.}$	1,00	1,00
$a_9.$	1,00	1,00	$a_{87.}$	1,00	1,00	$a_{165.}$	1,00	1,01	$a_{243.}$	1,00	1,00
$a_{10.}$	1,00	1,00	$a_{88.}$	1,00	1,00	$a_{166.}$	1,00	1,00	$a_{244.}$	1,00	1,00
$a_{11.}$	1,00	1,00	$a_{89.}$	1,00	1,00	$a_{167.}$	1,00	1,00	$a_{245.}$	1,00	1,01
$a_{12.}$	1,00	1,00	$a_{90.}$	1,00	1,00	$a_{168.}$	1,00	1,00	$a_{246.}$	1,00	1,00
$a_{13.}$	1,00	1,00	$a_{91.}$	1,00	1,00	$a_{169.}$	1,00	1,00	$a_{247.}$	1,00	1,00
$a_{14.}$	1,00	1,00	$a_{92.}$	1,00	1,00	$a_{170.}$	1,00	1,00	$a_{248.}$	1,00	1,00
$a_{15.}$	1,00	1,01	$a_{93.}$	1,00	1,00	$a_{171.}$	1,00	1,00	$a_{249.}$	1,00	1,00
$a_{16.}$	1,00	1,00	$a_{94.}$	1,00	1,00	$a_{172.}$	1,00	1,00	$a_{250.}$	1,00	1,00
$a_{17.}$	1,00	1,00	$a_{95.}$	1,00	1,00	$a_{173.}$	1,00	1,00	$a_{251.}$	1,00	1,00
$a_{18.}$	1,00	1,00	$a_{96.}$	1,00	1,00	$a_{174.}$	1,00	1,00	$a_{252.}$	1,00	1,00
$a_{19.}$	1,00	1,01	$a_{97.}$	1,00	1,00	$a_{175.}$	1,00	1,01	$a_{253.}$	1,00	1,00
$a_{20.}$	1,00	1,00	$a_{98.}$	1,00	1,01	$a_{176.}$	1,00	1,00	$a_{254.}$	1,00	1,00
$a_{21.}$	1,00	1,00	$a_{99.}$	1,00	1,00	$a_{177.}$	1,00	1,00	$a_{255.}$	1,00	1,00
$a_{22.}$	1,00	1,00	$a_{100.}$	1,00	1,00	$a_{178.}$	1,00	1,01	$a_{256.}$	1,00	1,01
$a_{23.}$	1,00	1,00	$a_{101.}$	1,00	1,00	$a_{179.}$	1,00	1,00	$a_{257.}$	1,00	1,00
$a_{24.}$	1,00	1,00	$a_{102.}$	1,00	1,00	$a_{180.}$	1,00	1,00	$a_{258.}$	1,00	1,01
$a_{25.}$	1,00	1,00	$a_{103.}$	1,00	1,00	$a_{181.}$	1,00	1,00	$a_{259.}$	1,00	1,00
$a_{26.}$	1,00	1,00	$a_{104.}$	1,00	1,00	$a_{182.}$	1,00	1,00	$a_{260.}$	1,00	1,00
$a_{27.}$	1,00	1,00	$a_{105.}$	1,00	1,00	$a_{183.}$	1,00	1,01	$a_{261.}$	1,00	1,00
$a_{28.}$	1,00	1,00	$a_{106.}$	1,00	1,00	$a_{184.}$	1,00	1,00	$a_{262.}$	1,00	1,00
$a_{29.}$	1,00	1,01	$a_{107.}$	1,00	1,00	$a_{185.}$	1,00	1,00	$a_{263.}$	1,00	1,00
$a_{30.}$	1,00	1,00	$a_{108.}$	1,00	1,00	$a_{186.}$	1,00	1,00	$a_{264.}$	1,00	1,00
$a_{31.}$	1,00	1,01	$a_{109.}$	1,00	1,00	$a_{187.}$	1,00	1,00	$a_{265.}$	1,00	1,00
$a_{32.}$	1,00	1,00	$a_{110.}$	1,00	1,01	$a_{188.}$	1,00	1,00	$a_{266.}$	1,00	1,00
$a_{33.}$	1,00	1,00	$a_{111.}$	1,00	1,00	$a_{189.}$	1,00	1,00	$a_{267.}$	1,00	1,00
$a_{34.}$	1,00	1,00	$a_{112.}$	1,00	1,00	$a_{190.}$	1,00	1,00	$a_{268.}$	1,00	1,00
$a_{35.}$	1,00	1,00	$a_{113.}$	1,00	1,00	$a_{191.}$	1,00	1,00	$a_{269.}$	1,00	1,00
$a_{36.}$	1,00	1,00	$a_{114.}$	1,00	1,00	$a_{192.}$	1,00	1,01	$a_{270.}$	1,00	1,00
$a_{37.}$	1,00	1,01	$a_{115.}$	1,00	1,00	$a_{193.}$	1,00	1,00	$a_{271.}$	1,00	1,00
$a_{38.}$	1,00	1,00	$a_{116.}$	1,00	1,00	$a_{194.}$	1,00	1,01	$a_{272.}$	1,00	1,00
$a_{39.}$	1,00	1,00	$a_{117.}$	1,00	1,00	$a_{195.}$	1,00	1,00	$a_{273.}$	1,00	1,00
$a_{40.}$	1,00	1,00	$a_{118.}$	1,00	1,00	$a_{196.}$	1,00	1,00	$a_{274.}$	1,00	1,00
$a_{41.}$	1,00	1,01	$a_{119.}$	1,00	1,01	$a_{197.}$	1,00	1,00	$a_{275.}$	1,00	1,00
$a_{42.}$	1,00	1,00	$a_{120.}$	1,00	1,00	$a_{198.}$	1,00	1,00	$a_{276.}$	1,00	1,00
$a_{43.}$	1,00	1,00	$a_{121.}$	1,00	1,00	$a_{199.}$	1,00	1,00	$a_{277.}$	1,00	1,00
$a_{44.}$	1,00	1,00	$a_{122.}$	1,00	1,00	$a_{200.}$	1,00	1,00	$a_{278.}$	1,00	1,00
$a_{45.}$	1,00	1,00	$a_{123.}$	1,00	1,00	$a_{201.}$	1,00	1,00	$a_{279.}$	1,00	1,00
$a_{46.}$	1,00	1,01	$a_{124.}$	1,00	1,00	$a_{202.}$	1,00	1,00	$a_{280.}$	1,00	1,00
$a_{47.}$	1,00	1,00	$a_{125.}$	1,00	1,01	$a_{203.}$	1,00	1,00	$a_{281.}$	1,00	1,00
$a_{48.}$	1,00	1,00	$a_{126.}$	1,00	1,00	$a_{204.}$	1,00	1,00	$a_{282.}$	1,00	1,00
$a_{49.}$	1,00	1,00	$a_{127.}$	1,00	1,00	$a_{205.}$	1,00	1,00	$a_{283.}$	1,00	1,01
$a_{50.}$	1,00	1,00	$a_{128.}$	1,00	1,00	$a_{206.}$	1,00	1,00	$a_{284.}$	1,00	1,00
$a_{51.}$	1,00	1,01	$a_{129.}$	1,00	1,01	$a_{207.}$	1,00	1,00	$a_{285.}$	1,00	1,00
$a_{52.}$	1,00	1,00	$a_{130.}$	1,00	1,00	$a_{208.}$	1,00	1,00	$a_{286.}$	1,00	1,00
$a_{53.}$	1,00	1,01	$a_{131.}$	1,00	1,00	$a_{209.}$	1,00	1,00	$a_{287.}$	1,00	1,00
$a_{54.}$	1,00	1,00	$a_{132.}$	1,00	1,00	$a_{210.}$	1,00	1,00	$a_{288.}$	1,00	1,00
$a_{55.}$	1,00	1,00	$a_{133.}$	1,00	1,01	$a_{211.}$	1,00	1,00	$a_{289.}$	1,00	1,00
$a_{56.}$	1,00	1,00	$a_{134.}$	1,00	1,00	$a_{212.}$	1,00	1,00	$a_{290.}$	1,00	1,00
$a_{57.}$	1,00	1,00	$a_{135.}$	1,00	1,01	$a_{213.}$	1,00	1,01	$a_{291.}$	1,00	1,00
$a_{58.}$	1,00	1,00	$a_{136.}$	1,00	1,00	$a_{214.}$	1,00	1,00	$a_{292.}$	1,00	1,00
$a_{59.}$	1,00	1,00	$a_{137.}$	1,00	1,00	$a_{215.}$	1,00	1,00	$a_{293.}$	1,00	1,00
$a_{60.}$	1,00	1,00	$a_{138.}$	1,00	1,00	$a_{216.}$	1,00	1,00	$a_{294.}$	1,00	1,00
$a_{61.}$	1,00	1,00	$a_{139.}$	1,00	1,00	$a_{217.}$	1,00	1,00	$a_{295.}$	1,00	1,00
$a_{62.}$	1,00	1,01	$a_{140.}$	1,00	1,00	$a_{218.}$	1,00	1,00	σ_a^2	1,01	1,02
$a_{63.}$	1,00	1,00	$a_{141.}$	1,00	1,01	$a_{219.}$	1,00	1,00	σ_e^2	1,00	1,00
$a_{64.}$	1,00	1,01	$a_{142.}$	1,00	1,00	$a_{220.}$	1,00	1,00	σ_e	1,00	1,00
$a_{65.}$	1,00	1,00	$a_{143.}$	1,00	1,01	$a_{221.}$	1,00	1,00	$h2$	1,01	1,02
$a_{66.}$	1,00	1,00	$a_{144.}$	1,00	1,00	$a_{222.}$	1,00	1,00	lp	1,01	1,03
$a_{67.}$	1,00	1,00	$a_{145.}$	1,00	1,00	$a_{223.}$	1,00	1,00			
$a_{68.}$	1,00	1,00	$a_{146.}$	1,00	1,00	$a_{224.}$	1,00	1,00			

Tabela 3.2: Resultados do Diagnóstico de Raftery & Lewis.

Parâmetro	\mathcal{I}	Parâmetro	\mathcal{I}	Parâmetro	\mathcal{I}	Parâmetro	\mathcal{I}
$h2$	9,66	a_{277}	1,14	a_{165}	1,08	a_{115}	1,03
σ_a^2	7,08	a_{279}	1,14	a_{167}	1,08	a_{120}	1,03
a_{255}	4,80	a_{289}	1,14	a_{181}	1,08	a_{131}	1,03
a_{257}	3,63	a_{65}	1,12	a_{193}	1,08	a_{141}	1,03
a_{77}	2,94	a_{68}	1,12	a_{194}	1,08	a_{143}	1,03
a_{260}	2,73	a_{76}	1,12	a_{207}	1,08	a_{151}	1,03
a_{18}	2,60	a_{109}	1,12	a_{210}	1,08	a_{158}	1,03
a_{237}	2,60	a_{111}	1,12	a_{215}	1,08	a_{174}	1,03
a_{238}	2,57	a_{116}	1,12	a_{223}	1,08	a_{177}	1,03
lp	2,55	a_{121}	1,12	a_{228}	1,08	a_{201}	1,03
a_{198}	2,54	a_{126}	1,12	a_{233}	1,08	a_{205}	1,03
a_9	2,51	a_{127}	1,12	a_{258}	1,08	a_{214}	1,03
a_{202}	2,50	a_{142}	1,12	a_{262}	1,08	a_{217}	1,03
a_{203}	2,50	a_{185}	1,12	a_{265}	1,08	a_{224}	1,03
a_{123}	2,47	a_{199}	1,12	a_{270}	1,08	a_{235}	1,03
a_{43}	2,44	a_{216}	1,12	a_{276}	1,08	a_{256}	1,03
a_{134}	2,40	a_{252}	1,12	a_{280}	1,08	a_{286}	1,03
a_{180}	2,40	a_{261}	1,12	a_{283}	1,08	a_{288}	1,03
a_{191}	2,25	a_{271}	1,12	a_{294}	1,08	β_6	1,01
a_{106}	2,13	a_{273}	1,12	β_5	1,05	β_8	1,01
a_{178}	2,12	a_{274}	1,12	β_9	1,05	a_3	1,01
a_{208}	2,12	a_{282}	1,12	a_1	1,05	a_7	1,01
a_{269}	1,94	a_{287}	1,12	a_{17}	1,05	a_{11}	1,01
a_6	1,84	a_4	1,10	a_{19}	1,05	a_{24}	1,01
a_{94}	1,35	a_{14}	1,10	a_{23}	1,05	a_{27}	1,01
a_{156}	1,32	a_{35}	1,10	a_{26}	1,05	a_{30}	1,01
a_{188}	1,27	a_{63}	1,10	a_{29}	1,05	a_{31}	1,01
a_{295}	1,27	a_{67}	1,10	a_{39}	1,05	a_{50}	1,01
a_{266}	1,24	a_{81}	1,10	a_{40}	1,05	a_{52}	1,01
a_{54}	1,22	a_{89}	1,10	a_{42}	1,05	a_{58}	1,01
a_{241}	1,22	a_{107}	1,10	a_{46}	1,05	a_{60}	1,01
a_{290}	1,22	a_{122}	1,10	a_{53}	1,05	a_{61}	1,01
a_{48}	1,20	a_{130}	1,10	a_{57}	1,05	a_{66}	1,01
a_{218}	1,20	a_{135}	1,10	a_{59}	1,05	a_{70}	1,01
a_{37}	1,19	a_{144}	1,10	a_{62}	1,05	a_{71}	1,01
a_{49}	1,19	a_{161}	1,10	a_{69}	1,05	a_{78}	1,01
a_{64}	1,19	a_{166}	1,10	a_{72}	1,05	a_{91}	1,01
a_{79}	1,19	a_{168}	1,10	a_{75}	1,05	a_{96}	1,01
a_{105}	1,19	a_{172}	1,10	a_{83}	1,05	a_{97}	1,01
a_{153}	1,19	a_{184}	1,10	a_{90}	1,05	a_{108}	1,01
a_{159}	1,19	a_{190}	1,10	a_{92}	1,05	a_{117}	1,01
a_{160}	1,19	a_{204}	1,10	a_{98}	1,05	a_{163}	1,01
a_{164}	1,19	a_{211}	1,10	a_{102}	1,05	a_{169}	1,01
a_{278}	1,19	a_{212}	1,10	a_{104}	1,05	a_{171}	1,01
a_{285}	1,19	a_{221}	1,10	a_{118}	1,05	a_{175}	1,01
a_5	1,17	a_{222}	1,10	a_{124}	1,05	a_{196}	1,01
a_{51}	1,17	a_{225}	1,10	a_{136}	1,05	a_{197}	1,01
a_{82}	1,17	a_{227}	1,10	a_{137}	1,05	a_{213}	1,01
a_{86}	1,17	a_{229}	1,10	a_{138}	1,05	a_{220}	1,01
a_{99}	1,17	a_{230}	1,10	a_{140}	1,05	a_{231}	1,01
a_{129}	1,17	a_{232}	1,10	a_{179}	1,05	a_{242}	1,01
a_{147}	1,17	a_{236}	1,10	a_{182}	1,05	a_{244}	1,01
a_{170}	1,17	a_{239}	1,10	a_{195}	1,05	a_{249}	1,01
a_{183}	1,17	a_{254}	1,10	a_{226}	1,05	a_{251}	1,01
a_{192}	1,17	a_{263}	1,10	a_{234}	1,05	a_{281}	1,01
a_{219}	1,17	a_{267}	1,10	a_{240}	1,05	β_3	0,99
a_{246}	1,17	a_{25}	1,09	a_{247}	1,05	β_7	0,99
a_{253}	1,17	β_4	1,08	a_{250}	1,05	a_{12}	0,99
a_{259}	1,17	a_{13}	1,08	a_{264}	1,05	a_{16}	0,99
σ_e^2	1,17	a_{15}	1,08	a_{272}	1,05	a_{44}	0,99
σ_e	1,17	a_{20}	1,08	a_{284}	1,05	a_{87}	0,99
β_1	1,14	a_{28}	1,08	a_{291}	1,05	a_{88}	0,99
a_{100}	1,14	a_{33}	1,08	a_{292}	1,05	a_{114}	0,99
a_{101}	1,14	a_{36}	1,08	β_{10}	1,03	a_{119}	0,99
a_{133}	1,14	a_{38}	1,08	a_2	1,03	a_{139}	0,99
a_{146}	1,14	a_{45}	1,08	a_8	1,03	a_{187}	0,99
a_{148}	1,14	a_{47}	1,08	a_{10}	1,03	a_{206}	0,99
a_{149}	1,14	a_{55}	1,08	a_{21}	1,03	a_{248}	0,99
a_{154}	1,14	a_{74}	1,08	a_{22}	1,03	a_{268}	0,99
a_{155}	1,14	a_{84}	1,08	a_{32}	1,03	a_{293}	0,99
a_{162}	1,14	a_{93}	1,08	a_{34}	1,03	β_2	0,97
a_{173}	1,14	a_{103}	1,08	a_{41}	1,03	a_{56}	0,97
a_{176}	1,14	a_{112}	1,08	a_{73}	1,03	a_{132}	0,97
a_{186}	1,14	a_{125}	1,08	a_{80}	1,03	a_{150}	0,97
a_{200}	1,14	a_{128}	1,08	a_{85}	1,03	a_{189}	0,97
a_{209}	1,14	a_{145}	1,08	a_{95}	1,03	a_{243}	0,97
a_{245}	1,14	a_{152}	1,08	a_{110}	1,03		
a_{275}	1,14	a_{157}	1,08	a_{113}	1,03		

Tabela 3.3: Resultados do Diagnóstico de Geweke.

Parâmetro	z	Parâmetro	z	Parâmetro	z	Parâmetro	z
$a_{215}.$	2,25	$a_{276}.$	0,63	$a_{71}.$	-0,07	$a_{124}.$	-0,73
$a_{258}.$	2,20	$a_{45}.$	0,63	$a_{25}.$	-0,08	$a_{69}.$	-0,74
$a_{163}.$	2,14	$a_{173}.$	0,61	$a_{111}.$	-0,08	$a_{152}.$	-0,75
$a_{81}.$	2,08	$a_{23}.$	0,60	$a_{143}.$	-0,08	$a_{200}.$	-0,75
$a_{195}.$	2,05	$a_3.$	0,59	$a_{67}.$	-0,09	$a_{189}.$	-0,75
$a_{64}.$	2,05	$a_{171}.$	0,58	$a_{272}.$	-0,09	$a_{33}.$	-0,77
$a_{41}.$	1,96	$a_{227}.$	0,57	$a_{16}.$	-0,09	$a_{274}.$	-0,77
$a_{108}.$	1,95	$a_{278}.$	0,57	$a_{168}.$	-0,10	$a_{122}.$	-0,79
$a_{249}.$	1,82	$a_{85}.$	0,56	$a_{89}.$	-0,11	$a_{114}.$	-0,81
$a_9.$	1,76	$a_{241}.$	0,56	$a_{95}.$	-0,12	$a_{201}.$	-0,82
$a_{236}.$	1,75	$a_{248}.$	0,53	$a_{129}.$	-0,13	$a_{97}.$	-0,82
$a_{208}.$	1,74	$a_{137}.$	0,49	$a_{75}.$	-0,15	$a_{295}.$	-0,83
$a_{264}.$	1,72	$a_{270}.$	0,47	$a_{222}.$	-0,16	$a_{58}.$	-0,84
$a_{47}.$	1,63	$a_{62}.$	0,47	$a_{154}.$	-0,16	$a_{174}.$	-0,84
$a_{142}.$	1,61	$a_{99}.$	0,46	$a_{107}.$	-0,17	$a_{282}.$	-0,85
$a_{225}.$	1,59	$a_{210}.$	0,46	$a_{30}.$	-0,17	$a_{283}.$	-0,86
$a_{59}.$	1,58	$a_{103}.$	0,46	$a_{54}.$	-0,17	$a_{242}.$	-0,86
$a_{79}.$	1,57	$a_{96}.$	0,45	$a_{279}.$	-0,18	$\beta_6.$	-0,87
$a_{209}.$	1,56	$a_{160}.$	0,43	$a_{106}.$	-0,18	$a_{117}.$	-0,88
$a_{178}.$	1,53	$a_{74}.$	0,42	$a_{88}.$	-0,19	$a_8.$	-0,89
$a_{252}.$	1,51	$a_{22}.$	0,41	$a_7.$	-0,19	$a_{266}.$	-0,89
$a_{60}.$	1,46	$a_{10}.$	0,40	$a_{240}.$	-0,19	$a_{239}.$	-0,90
$a_{247}.$	1,41	$a_{11}.$	0,40	$a_{120}.$	-0,20	$a_{156}.$	-0,90
$a_4.$	1,34	$a_{269}.$	0,39	$a_{18}.$	-0,21	$\beta_8.$	-0,91
$a_{140}.$	1,33	$a_{262}.$	0,38	$a_{194}.$	-0,22	$a_{165}.$	-0,91
$a_{192}.$	1,31	$a_{291}.$	0,37	$\beta_9.$	-0,22	$a_{113}.$	-0,91
$a_{150}.$	1,31	$a_{78}.$	0,37	$a_{13}.$	-0,22	$a_{212}.$	-0,92
$a_{253}.$	1,29	$a_{126}.$	0,36	$a_{223}.$	-0,22	$a_{275}.$	-0,92
$a_{155}.$	1,29	$a_{256}.$	0,36	$a_{185}.$	-0,22	$a_{19}.$	-0,93
$a_{233}.$	1,27	$a_{131}.$	0,36	$a_{100}.$	-0,23	$a_{190}.$	-0,94
$\beta_{10}.$	1,26	$a_{191}.$	0,35	$a_{277}.$	-0,23	$\beta_4.$	-0,96
$a_{68}.$	1,25	$a_{87}.$	0,35	$a_{52}.$	-0,24	$a_{293}.$	-0,96
$a_{72}.$	1,23	$a_{26}.$	0,35	$a_{136}.$	-0,24	$a_{151}.$	-0,98
$a_{219}.$	1,21	$a_{280}.$	0,33	$a_{294}.$	-0,25	$a_{34}.$	-0,98
$a_{164}.$	1,18	$a_{55}.$	0,32	$a_{197}.$	-0,25	$a_{218}.$	-1,00
$a_{153}.$	1,16	$a_{181}.$	0,31	$a_{196}.$	-0,26	$a_{39}.$	-1,00
$a_1.$	1,12	$a_{157}.$	0,30	$a_{254}.$	-0,27	$a_{132}.$	-1,02
$a_{119}.$	1,11	$a_{149}.$	0,29	$a_{167}.$	-0,29	$a_{112}.$	-1,04
$a_{206}.$	1,05	$\beta_5.$	0,28	$a_{235}.$	-0,30	$a_{229}.$	-1,09
$a_{110}.$	1,04	$a_{224}.$	0,28	$a_{231}.$	-0,30	$a_{204}.$	-1,12
$a_{238}.$	1,03	$a_{232}.$	0,26	$a_{170}.$	-0,33	$a_{188}.$	-1,14
$a_{20}.$	1,02	$a_{147}.$	0,25	$a_{263}.$	-0,34	$a_{213}.$	-1,14
$a_{193}.$	1,00	$a_{98}.$	0,25	$a_{86}.$	-0,36	$a_{175}.$	-1,16
$a_{83}.$	0,98	$a_{133}.$	0,24	$a_{257}.$	-0,38	$a_{31}.$	-1,18
$a_{148}.$	0,97	$a_5.$	0,24	$a_{144}.$	-0,38	$a_{42}.$	-1,20
$a_{115}.$	0,96	$a_{199}.$	0,23	$a_{159}.$	-0,38	$\beta_7.$	-1,20
$a_{127}.$	0,94	$a_{77}.$	0,23	$\beta_3.$	-0,39	$a_{145}.$	-1,22
$a_{82}.$	0,94	$a_{61}.$	0,22	$a_{94}.$	-0,39	$a_{244}.$	-1,23
$a_{230}.$	0,93	$a_{138}.$	0,20	lp	-0,40	$a_{217}.$	-1,23
$a_{29}.$	0,92	$a_{21}.$	0,19	$a_{177}.$	-0,40	$a_{73}.$	-1,26
$a_{51}.$	0,91	$a_{186}.$	0,18	$a_{281}.$	-0,40	$a_{56}.$	-1,28
$a_{93}.$	0,89	$a_{166}.$	0,17	$a_{130}.$	-0,41	$a_{53}.$	-1,30
$a_{287}.$	0,87	$a_{105}.$	0,15	$a_{243}.$	-0,41	$a_{125}.$	-1,31
$a_{250}.$	0,87	$a_{161}.$	0,14	$a_{255}.$	-0,42	$a_{37}.$	-1,40
$a_{286}.$	0,86	h_2	0,13	$a_{211}.$	-0,42	$a_{14}.$	-1,41
$a_{158}.$	0,86	$a_{28}.$	0,13	$a_{102}.$	-0,43	$a_{179}.$	-1,41
$a_{261}.$	0,85	$a_{271}.$	0,13	$a_{12}.$	-0,44	$a_{49}.$	-1,43
$a_{80}.$	0,85	$a_{15}.$	0,13	$a_{43}.$	-0,44	$a_{290}.$	-1,44
$a_{226}.$	0,82	$a_{128}.$	0,12	$a_{220}.$	-0,46	$a_{70}.$	-1,47
$a_{63}.$	0,81	$a_{46}.$	0,11	$a_{202}.$	-0,46	$a_{205}.$	-1,52
$a_{40}.$	0,80	$a_{288}.$	0,10	$a_{65}.$	-0,47	$a_{118}.$	-1,55
$a_{216}.$	0,79	$a_{285}.$	0,10	$a_{184}.$	-0,50	$a_{104}.$	-1,58
$a_{121}.$	0,78	$a_{35}.$	0,10	$a_{27}.$	-0,53	$a_{17}.$	-1,59
$a_6.$	0,77	$a_{234}.$	0,06	$a_{183}.$	-0,54	$a_{57}.$	-1,61
$a_{116}.$	0,76	$a_{66}.$	0,06	$a_{268}.$	-0,55	$a_{90}.$	-1,64
$a_{135}.$	0,75	$a_{214}.$	0,05	$a_{24}.$	-0,56	$a_{176}.$	-1,65
$a_{182}.$	0,73	σ_a^2	0,04	$a_{289}.$	-0,56	$a_{284}.$	-1,73
$a_{246}.$	0,73	$a_{162}.$	0,04	σ_e	-0,58	$a_{292}.$	-1,74
$a_{198}.$	0,73	$a_{36}.$	-0,01	σ_e^2	-0,58	$a_{84}.$	-1,92
$a_{123}.$	0,72	$a_{48}.$	-0,01	$a_{203}.$	-0,59	$a_{207}.$	-1,92
$a_{273}.$	0,71	$a_{91}.$	-0,01	$a_{187}.$	-0,60	$a_{146}.$	-2,09
$a_{38}.$	0,71	$a_{109}.$	-0,02	$a_{32}.$	-0,60	$a_{228}.$	-2,23
$a_{169}.$	0,68	$a_{76}.$	-0,03	$a_{250}.$	-0,64	$\beta_2.$	-2,23
$a_{172}.$	0,68	$a_{180}.$	-0,03	$a_{237}.$	-0,65	$a_{251}.$	-2,56
$a_{260}.$	0,67	$\beta_1.$	-0,05	$a_{92}.$	-0,65	$a_{50}.$	-2,66
$a_2.$	0,65	$a_{101}.$	-0,06	$a_{265}.$	-0,66	$a_{141}.$	-3,03
$a_{134}.$	0,65	$a_{44}.$	-0,07	$a_{245}.$	-0,69		
$a_{267}.$	0,64	$a_{221}.$	-0,07	$a_{139}.$	-0,73		

A partir dos resultados obtidos pelo critério de Raftery & Lewis, observa-se que para a maioria das cadeias, os valores da estatística I indicam que o número de iterações necessário para atingir a precisão desejada está dentro do intervalo recomendado, ou seja, menor que 5. De um total de 310 parâmetros no modelo, I superou 5 apenas para h_2 e σ_a^2 . O Fator de Dependência (I), conforme descrito por Raftery & Lewis [28], quantifica a quantidade de iteração necessária para reduzir a autocorrelação nas cadeias MCMC, e valores superiores a 5 podem indicar que a cadeia ainda lida com autocorrelação significativa. Embora esses valores possam sugerir que o número de iterações para essas variáveis ou parâmetros não foi suficiente, a análise gráfica detalhada, apresentada na Figura 3.1, demonstra que as cadeias para esses dois parâmetros estão se estabilizando ou já atingiram a distribuição de equilíbrio.

Portanto, embora dois valores de I estejam acima de 5, especificamente para os parâmetros h_2 e σ_a^2 , a análise gráfica confirma que as cadeias estão convergindo adequadamente, atingindo, assim, as distribuições estacionárias. Ressalta-se que esse comportamento foi observado não apenas nas cadeias desses dois parâmetros, mas em todas as cadeias analisadas.

O diagnóstico de Geweke também foi utilizado para avaliar a convergência das cadeias. Dentro os 310 parâmetros analisados, apenas 12 apresentaram valores de z fora do intervalo recomendado de -1,96 a 1,96. Esses parâmetros são: a_{50} , a_{64} , a_{81} , a_{141} , a_{146} , a_{163} , a_{195} , a_{215} , a_{228} , a_{251} , a_{258} e β_2 . Isso indica que essas variáveis podem não ter alcançado uma distribuição estacionária ideal. No entanto, essa diferença não implica necessariamente em falta de convergência. A análise gráfica dessas variáveis, apresentada nas Figuras 3.2, 3.3 e 3.4, sugere que, apesar dessas flutuações iniciais, as cadeias estão se estabilizando ao longo das iterações. Além disso, os diagnósticos de Gelman & Rubin e Raftery & Lewis indicam que, para os parâmetros mencionados, as cadeias estão convergindo adequadamente, como evidenciado pelos valores de \hat{R} próximos de 1 e pelos resultados satisfatórios de I .

A análise dos intervalos de densidade posterior com 95% de credibilidade (HPD), conforme a Tabela 3.4, revelou informações significativas sobre os parâmetros estimados no modelo. Para a grande maioria dos parâmetros (Fig. 3.6, 3.7, 3.8, 3.9 e , 3.7) os intervalos HPD incluíram o valor zero, o que sugere que não há evidência estatística suficiente para afirmar que esses parâmetros são significativamente diferentes de zero. Isso indica incerteza quanto à importância de algumas variáveis para a explicação do fenômeno modelado. Especificamente, como mostrado na Fig. 3.5, os coeficientes de regressão β_1 a β_9 , que representam, respectivamente, os graus de sangue $\frac{1}{2}H$, $\frac{1}{4}H$, $\frac{3}{4}H$, $\frac{3}{8}H$, $\frac{5}{16}H$, $\frac{5}{8}H$, $\frac{7}{8}H$, GIR e PS, apresentaram médias posteriores variando entre aproximadamente 14,63 e 21,55, com intervalos HPD positivos. Isso indica que há uma associação positiva significativa entre as variáveis explicativas representadas por esses coeficientes e a variável resposta, sugerindo que estas variáveis desempenham um papel importante no fenômeno estudado.

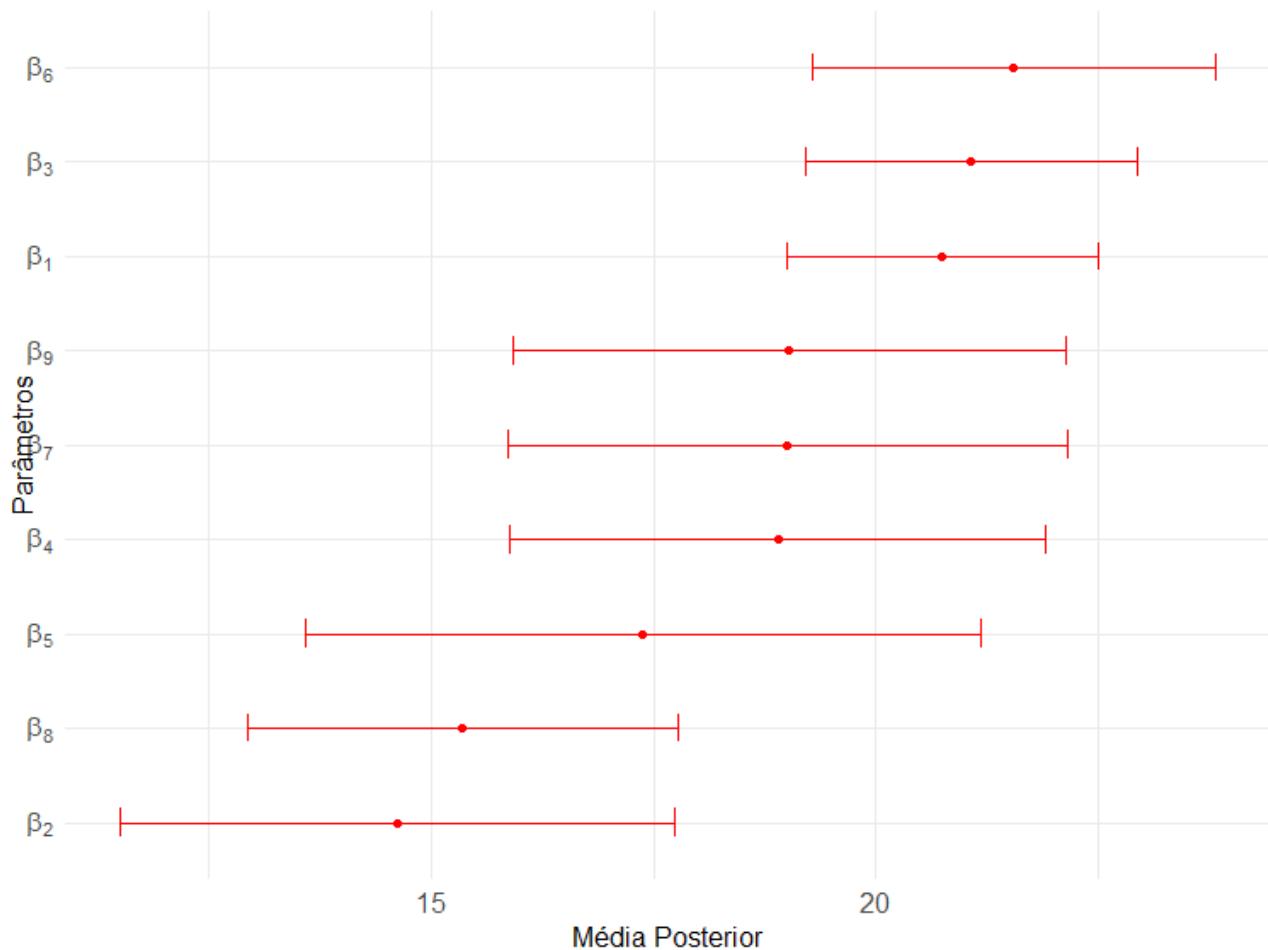


Figura 3.5: Gráfico de floresta das estimativas dos efeitos fixos β_1 a β_9 genéticos e seus intervalos HPD (95%) obtidos a partir das cadeias MCMC.

Tabela 3.4: Resultados do Intervalo HPD.

Parâmetro: [Inf.;Sup.]	Parâmetro: [Inf.;Sup.]	Parâmetro: [Inf.;Sup.]	Parâmetro: [Inf.;Sup.]
β_1 : [18, 99; 22, 49]	a_{69} : [-0, 39; 6, 14]	a_{147} : [-4, 57; 3, 57]	a_{225} : [-5, 28; 4, 09]
β_2 : [11, 51; 17, 74]	a_{70} : [-2, 38; 5, 64]	a_{148} : [-4, 66; 2, 07]	a_{226} : [-2, 41; 6, 10]
β_3 : [19, 21; 22, 93]	a_{71} : [-2, 09; 5, 97]	a_{149} : [-2, 14; 4, 52]	a_{227} : [-4, 40; 4, 39]
β_4 : [15, 88; 21, 91]	a_{72} : [-3, 55; 4, 11]	a_{150} : [-3, 73; 4, 71]	a_{228} : [-4, 58; 4, 21]
β_5 : [13, 59; 21, 17]	a_{73} : [-4, 11; 4, 49]	a_{151} : [-4, 51; 3, 95]	a_{229} : [-5, 01; 4, 00]
β_6 : [19, 28; 23, 81]	a_{74} : [-3, 61; 3, 70]	a_{152} : [-3, 98; 4, 18]	a_{230} : [-4, 35; 4, 58]
β_7 : [15, 86; 22, 15]	a_{75} : [-3, 74; 4, 08]	a_{153} : [-3, 28; 4, 01]	a_{231} : [-4, 27; 4, 73]
β_8 : [12, 95; 17, 77]	a_{76} : [-4, 50; 2, 68]	a_{154} : [-3, 26; 5, 13]	a_{232} : [-4, 64; 4, 14]
β_9 : [15, 92; 22, 13]	a_{77} : [-6, 30; 1, 73]	a_{155} : [-4, 05; 2, 92]	a_{233} : [-4, 10; 4, 78]
β_{10} : [-0, 02; 0, 07]	a_{78} : [-1, 61; 4, 85]	a_{156} : [-5, 26; 3, 39]	a_{234} : [-4, 43; 4, 84]
a_1 : [-1, 33; 5, 69]	a_{79} : [-4, 01; 3, 31]	a_{157} : [-5, 11; 2, 84]	a_{235} : [-4, 77; 4, 03]
a_2 : [-1, 43; 5, 11]	a_{80} : [-3, 88; 4, 58]	a_{158} : [-5, 77; 2, 38]	a_{236} : [-2, 28; 5, 75]
a_3 : [-1, 98; 5, 21]	a_{81} : [-5, 11; 3, 18]	a_{159} : [-4, 97; 3, 34]	a_{237} : [-3, 26; 3, 92]
a_4 : [-2, 67; 3, 51]	a_{82} : [-6, 53; 1, 55]	a_{160} : [-8, 03; 0, 57]	a_{238} : [-3, 82; 5, 19]
a_5 : [-5, 34; 2, 42]	a_{83} : [-3, 62; 4, 38]	a_{161} : [-3, 86; 4, 10]	a_{239} : [-3, 74; 4, 23]
a_6 : [-2, 88; 4, 95]	a_{84} : [-3, 84; 3, 99]	a_{162} : [-3, 38; 5, 15]	a_{240} : [-1, 45; 6, 24]
a_7 : [-1, 29; 6, 02]	a_{85} : [-2, 60; 5, 26]	a_{163} : [-3, 81; 3, 42]	a_{241} : [-4, 11; 4, 71]
a_8 : [-2, 01; 5, 41]	a_{86} : [-3, 38; 4, 17]	a_{164} : [-3, 78; 5, 39]	a_{242} : [-4, 65; 4, 03]
a_9 : [-3, 10; 5, 29]	a_{87} : [-2, 37; 4, 62]	a_{165} : [-1, 80; 6, 54]	a_{243} : [-2, 11; 5, 20]
a_{10} : [-4, 51; 3, 83]	a_{88} : [-2, 59; 5, 62]	a_{166} : [-4, 47; 4, 52]	a_{244} : [-4, 33; 3, 89]
a_{11} : [-2, 40; 4, 73]	a_{89} : [-2, 43; 4, 52]	a_{167} : [-4, 21; 4, 26]	a_{245} : [-4, 34; 4, 55]
a_{12} : [-1, 53; 5, 94]	a_{90} : [-3, 39; 3, 85]	a_{168} : [-4, 95; 3, 86]	a_{246} : [-6, 19; 1, 69]
a_{13} : [-4, 67; 3, 17]	a_{91} : [-3, 89; 4, 16]	a_{169} : [-3, 50; 4, 62]	a_{247} : [-4, 86; 3, 21]
a_{14} : [-2, 19; 5, 57]	a_{92} : [-2, 61; 5, 17]	a_{170} : [-3, 94; 4, 32]	a_{248} : [-4, 44; 4, 42]
a_{15} : [-1, 57; 6, 14]	a_{93} : [-4, 45; 3, 05]	a_{171} : [-4, 31; 4, 57]	a_{249} : [-4, 37; 4, 39]
a_{16} : [-3, 01; 4, 27]	a_{94} : [-3, 95; 3, 97]	a_{172} : [-4, 44; 4, 50]	a_{250} : [-2, 50; 4, 93]
a_{17} : [-2, 31; 5, 00]	a_{95} : [-3, 65; 4, 86]	a_{173} : [-4, 62; 4, 33]	a_{251} : [-3, 35; 5, 42]
a_{18} : [-6, 63; 1, 25]	a_{96} : [-1, 53; 5, 94]	a_{174} : [-4, 43; 4, 51]	a_{252} : [-4, 53; 4, 21]
a_{19} : [-1, 89; 5, 31]	a_{97} : [-3, 28; 4, 38]	a_{175} : [-3, 63; 4, 96]	a_{253} : [-4, 87; 4, 08]
a_{20} : [-2, 47; 4, 69]	a_{98} : [-4, 48; 3, 80]	a_{176} : [-4, 38; 4, 61]	a_{254} : [-4, 51; 4, 19]
a_{21} : [-3, 35; 4, 55]	a_{99} : [-6, 12; 2, 35]	a_{177} : [-4, 93; 3, 93]	a_{255} : [-5, 05; 4, 15]
a_{22} : [-3, 23; 4, 59]	a_{100} : [-4, 14; 4, 20]	a_{178} : [-5, 02; 3, 97]	a_{256} : [-2, 91; 5, 25]
a_{23} : [-3, 07; 5, 07]	a_{101} : [-4, 02; 4, 10]	a_{179} : [-2, 90; 5, 50]	a_{257} : [-2, 83; 5, 97]
a_{24} : [-2, 74; 5, 53]	a_{102} : [-2, 16; 5, 33]	a_{180} : [-2, 70; 6, 16]	a_{258} : [-3, 15; 4, 93]
a_{25} : [-3, 64; 4, 09]	a_{103} : [-3, 37; 4, 58]	a_{181} : [-4, 85; 3, 78]	a_{259} : [-4, 48; 4, 16]
a_{26} : [-2, 81; 5, 30]	a_{104} : [-4, 69; 3, 10]	a_{182} : [-3, 38; 5, 32]	a_{260} : [-4, 22; 4, 44]
a_{27} : [-3, 26; 4, 43]	a_{105} : [-5, 62; 2, 99]	a_{183} : [-4, 43; 4, 75]	a_{261} : [-3, 79; 4, 72]
a_{28} : [-4, 49; 3, 51]	a_{106} : [-3, 69; 3, 96]	a_{184} : [-4, 61; 4, 01]	a_{262} : [-4, 15; 4, 33]
a_{29} : [-2, 04; 5, 79]	a_{107} : [-3, 79; 4, 45]	a_{185} : [-4, 09; 5, 03]	a_{263} : [-4, 76; 3, 83]
a_{30} : [-1, 97; 5, 23]	a_{108} : [-3, 34; 4, 51]	a_{186} : [-4, 24; 4, 36]	a_{264} : [-5, 62; 3, 15]
a_{31} : [-4, 52; 3, 85]	a_{109} : [-5, 73; 1, 78]	a_{187} : [-3, 26; 5, 39]	a_{265} : [-4, 51; 3, 99]
a_{32} : [-4, 35; 4, 01]	a_{110} : [-1, 62; 6, 01]	a_{188} : [-4, 67; 3, 56]	a_{266} : [-3, 35; 4, 71]
a_{33} : [-3, 53; 3, 90]	a_{111} : [-1, 76; 5, 42]	a_{189} : [-4, 97; 3, 47]	a_{267} : [-5, 19; 3, 33]
a_{34} : [-3, 12; 4, 85]	a_{112} : [-6, 15; 1, 93]	a_{190} : [-5, 16; 2, 78]	a_{268} : [-2, 30; 5, 46]
a_{35} : [-4, 52; 3, 15]	a_{113} : [-2, 83; 4, 22]	a_{191} : [-3, 32; 5, 55]	a_{269} : [-4, 32; 4, 33]
a_{36} : [-3, 71; 4, 10]	a_{114} : [-2, 30; 4, 72]	a_{192} : [-5, 12; 3, 34]	a_{270} : [-4, 38; 4, 05]
a_{37} : [-6, 13; 2, 65]	a_{115} : [-2, 06; 5, 87]	a_{193} : [-4, 35; 4, 09]	a_{271} : [-5, 78; 3, 20]
a_{38} : [-4, 89; 3, 35]	a_{116} : [-2, 49; 3, 55]	a_{194} : [-4, 87; 3, 46]	a_{272} : [-3, 43; 5, 41]
a_{39} : [-2, 85; 5, 27]	a_{117} : [-2, 84; 4, 34]	a_{195} : [-5, 56; 2, 86]	a_{273} : [-4, 71; 4, 27]
a_{40} : [-0, 86; 7, 20]	a_{118} : [-3, 00; 5, 22]	a_{196} : [-3, 72; 5, 13]	a_{274} : [-4, 35; 4, 46]
a_{41} : [-4, 84; 2, 26]	a_{119} : [-2, 33; 5, 02]	a_{197} : [-4, 23; 4, 78]	a_{275} : [-4, 31; 4, 39]
a_{42} : [-5, 26; 3, 23]	a_{120} : [-2, 39; 5, 46]	a_{198} : [-6, 19; 2, 59]	a_{276} : [-2, 28; 6, 52]
a_{43} : [-3, 77; 4, 08]	a_{121} : [-5, 33; 2, 37]	a_{199} : [-4, 49; 4, 49]	a_{277} : [-6, 01; 2, 86]
a_{44} : [-2, 69; 4, 90]	a_{122} : [-2, 00; 5, 58]	a_{200} : [-4, 44; 4, 36]	a_{278} : [-3, 48; 5, 34]
a_{45} : [-3, 72; 3, 67]	a_{123} : [-6, 40; 1, 47]	a_{201} : [-3, 82; 4, 74]	a_{279} : [-4, 76; 4, 02]
a_{46} : [-2, 12; 5, 54]	a_{124} : [-4, 25; 4, 20]	a_{202} : [-4, 84; 4, 13]	a_{280} : [-4, 61; 4, 28]
a_{47} : [-2, 16; 5, 20]	a_{125} : [-6, 24; 1, 58]	a_{203} : [-4, 88; 4, 33]	a_{281} : [-3, 67; 4, 90]
a_{48} : [-4, 03; 4, 41]	a_{126} : [-4, 22; 3, 89]	a_{204} : [-3, 96; 5, 05]	a_{282} : [-4, 98; 3, 97]
a_{49} : [-3, 33; 4, 07]	a_{127} : [-4, 89; 3, 27]	a_{205} : [-3, 68; 4, 99]	a_{283} : [-3, 86; 4, 75]
a_{50} : [-3, 32; 4, 72]	a_{128} : [-5, 72; 1, 44]	a_{206} : [-3, 71; 4, 71]	a_{284} : [-4, 12; 4, 42]
a_{51} : [-3, 27; 4, 90]	a_{129} : [-4, 71; 3, 18]	a_{207} : [-4, 67; 4, 29]	a_{285} : [-5, 62; 3, 21]
a_{52} : [-3, 02; 3, 45]	a_{130} : [-4, 51; 3, 92]	a_{208} : [-5, 57; 3, 52]	a_{286} : [-4, 33; 4, 58]
a_{53} : [-2, 48; 5, 77]	a_{131} : [-2, 22; 4, 89]	a_{209} : [-3, 96; 4, 82]	a_{287} : [-3, 70; 5, 30]
a_{54} : [-3, 62; 3, 95]	a_{132} : [-3, 89; 4, 53]	a_{210} : [-6, 78; 1, 81]	a_{288} : [-4, 89; 3, 97]
a_{55} : [-4, 42; 2, 76]	a_{133} : [-2, 32; 5, 04]	a_{211} : [-2, 93; 5, 62]	a_{289} : [-4, 18; 4, 99]
a_{56} : [-1, 91; 5, 56]	a_{134} : [-5, 37; 2, 96]	a_{212} : [-5, 49; 3, 57]	a_{290} : [-4, 65; 4, 43]
a_{57} : [-3, 01; 4, 21]	a_{135} : [-2, 27; 5, 52]	a_{213} : [-3, 94; 5, 00]	a_{291} : [-4, 64; 4, 77]
a_{58} : [-2, 36; 5, 50]	a_{136} : [-4, 10; 4, 22]	a_{214} : [-3, 90; 4, 82]	a_{292} : [-4, 11; 4, 81]
a_{59} : [-3, 74; 4, 47]	a_{137} : [-4, 87; 3, 56]	a_{215} : [-3, 81; 4, 94]	a_{293} : [-4, 45; 4, 48]
a_{60} : [-4, 69; 3, 20]	a_{138} : [-5, 47; 2, 54]	a_{216} : [-4, 59; 4, 39]	a_{294} : [-4, 72; 4, 10]
a_{61} : [-3, 51; 3, 58]	a_{139} : [-0, 51; 6, 18]	a_{217} : [-5, 59; 3, 24]	a_{295} : [-4, 77; 4, 29]
a_{62} : [-2, 87; 4, 91]	a_{140} : [-4, 54; 3, 59]	a_{218} : [-4, 40; 4, 65]	σ_{a2} : [2, 82; 18, 32]
a_{63} : [-5, 61; 1, 80]	a_{141} : [-4, 58; 3, 96]	a_{219} : [-5, 22; 3, 84]	σ_{e2} : [27, 72; 38, 72]
a_{64} : [-4, 04; 4, 35]	a_{142} : [-5, 35; 3, 42]	a_{220} : [-4, 06; 3, 85]	σ_e : [5, 28; 6, 24]
a_{65} : [-4, 13; 2, 35]	a_{143} : [-3, 99; 4, 38]	a_{221} : [-4, 71; 4, 30]	h^2 : [0, 09; 0, 38]
a_{66} : [-4, 12; 2, 26]	a_{144} : [-2, 49; 5, 52]	a_{222} : [-4, 05; 4, 99]	lp : [-1238, 98; -1012, 12]
a_{67} : [-5, 63; 1, 23]	a_{145} : [-2, 02; 5, 01]	a_{223} : [-4, 50; 4, 40]	
a_{68} : [-4, 72; 3, 55]	a_{146} : [-3, 82; 2, 99]	a_{224} : [-4, 45; 4, 72]	

O parâmetro β_{10} , que corresponde ao IPP, apresentou uma média de 0,025, com um intervalo de credibilidade (HPD) variando de -0,02 a 0,07, conforme mostrado na Tabela 3.4. O fato de o intervalo HPD incluir o valor zero sugere que não há uma evidência estatística forte de que o efeito do IPP sobre a produção de leite seja significativamente diferente de zero. Isso indica que, com base nos dados analisados, o IPP não tem um efeito claro e substancial sobre a produção de leite no modelo, evidenciando a incerteza quanto à sua contribuição para o fenômeno estudado.

Um ponto importante a ser destacado é que, ao observar os coeficientes β_1 , β_2 , β_3 , β_6 e β_8 , observa-se que, enquanto os intervalos HPD de β_1 , β_3 e β_6 não se sobreponem ao intervalo de β_8 e β_2 , isso indica que esses parâmetros são estatisticamente diferentes entre si, ou seja, eles têm efeitos distintos sobre a variável resposta. Isso sugere que, para esses parâmetros, existe uma associação positiva significativa com a variável dependente, enquanto os outros parâmetros apresentam uma variação com maior incerteza, dado que seus intervalos incluem zero, o que indica que esses efeitos podem não ser diferentes de zero.

Os coeficientes β associados ao grau de sangue (GS) representam a média de produção de leite em cada caso com intervalos HPD coerentemente positivos e à medida que o grau de sangue aumenta a estimativa da produção de leite também aumenta. É importante destacar que, no contexto deste estudo, o grau de sangue é uma medida da proximidade genética com a raça Holandesa, sendo que um maior grau de sangue está associado a uma maior contribuição genética desta raça. Assim, os animais com maior grau de sangue possuem um maior potencial genético para a produção de leite, o que justifica o aumento observado na produção à medida que o grau de sangue se eleva. O IPP, por outro lado, também apresentou um efeito positivo, embora com uma magnitude menor, sugerindo que, embora a relação com a produção de leite seja significativa, seu impacto não é tão forte quanto o do grau de sangue.

O parâmetro σ_a^2 representa a variância dos efeitos aleatórios de animais, e sua estimativa, através da média a posteriori foi 10,57, com um intervalo de credibilidade entre 2,82 e 18,32, conforme apresentado na Tabela 3.4. Este resultado sugere que há uma variação significativa atribuível aos efeitos genéticos no modelo, ou seja, uma parte substancial da variabilidade dos dados pode ser explicada por fatores aleatórios genéticos.

Da mesma forma, a variância residual σ_e^2 (com média de 33,22 e intervalo HPD de 27,72 a 38,72) e seu desvio-padrão σ_e (média de 5,76; intervalo de 5,28 a 6,24), conforme apresentado na Tabela 3.4, indicam uma variabilidade não explicada relevante. Esses resultados indicam que uma fração significativa da variabilidade dos dados ainda é atribuível a fontes de erro não modeladas.

Os valores do parâmetro lp (logverossimilhança posteriori) apresentados na tabela variam entre -1238,98 e -1012,12, conforme mostrado na Tabela 3.4. O parâmetro lp é um output do rstan [10] que reflete a adequação do modelo aos dados observados. Esse parâmetro é tipicamente utilizado para comparar modelos diferentes, o que não é o caso aqui, pois estamos trabalhando com um único modelo. Portanto, não será utilizado como métrica para avaliação do ajuste, dado que não há outros modelos para comparação.

A herdabilidade (h^2) foi estimada com uma média de 0,235, com intervalo HPD variando

entre 0,09 e 0,38. A herdabilidade é uma medida que indica a fração da variabilidade total de uma característica que pode ser atribuída a fatores genéticos, representando a parte da variabilidade que é transmitida de uma geração para a próxima. Com o valor estimado de 0,235, isso sugere que cerca de 23,5% da variabilidade na produção de leite é explicada por fatores genéticos.

É importante notar que a herdabilidade não explica completamente o modelo de produção de leite. Fatores ambientais, como alimentação, manejo, condições sanitárias e outros aspectos relacionados ao ambiente dos animais, desempenham um papel crucial na produção de leite e podem ser responsáveis pela maior parte da variabilidade observada. Portanto, embora os efeitos genéticos apresentem um impacto importante, ele não é exclusivo, e a gestão ambiental e de manejo continua sendo de grande relevância na produção de leite. A herdabilidade, portanto, deve ser vista como um indicador da influência genética, mas não como um fator determinante único para a produção de leite.

Após a análise dos intervalos HPD, que fornecem uma estimativa da incerteza associada a todos os parâmetros, os animais foram ranqueados conforme seus valores genéticos, representados por a_i nas Figuras 3.6, 3.7, 3.8, 3.9 e 3.10. Observou-se que os animais a_{40} , a_{69} , a_{139} , a_{110} e a_{165} se destacaram como os mais bem classificados no ranking. Além disso, ao se considerar o valor estimado da herdabilidade (h^2), observou-se que a contribuição genética para a variabilidade da produção de leite é substancial, reforçando a relevância dos programas de melhoramento genético dentro do rebanho. A análise do ranking complementa essa compreensão, destacando o impacto dos animais com maior valor genético sobre o desempenho produtivo do rebanho.

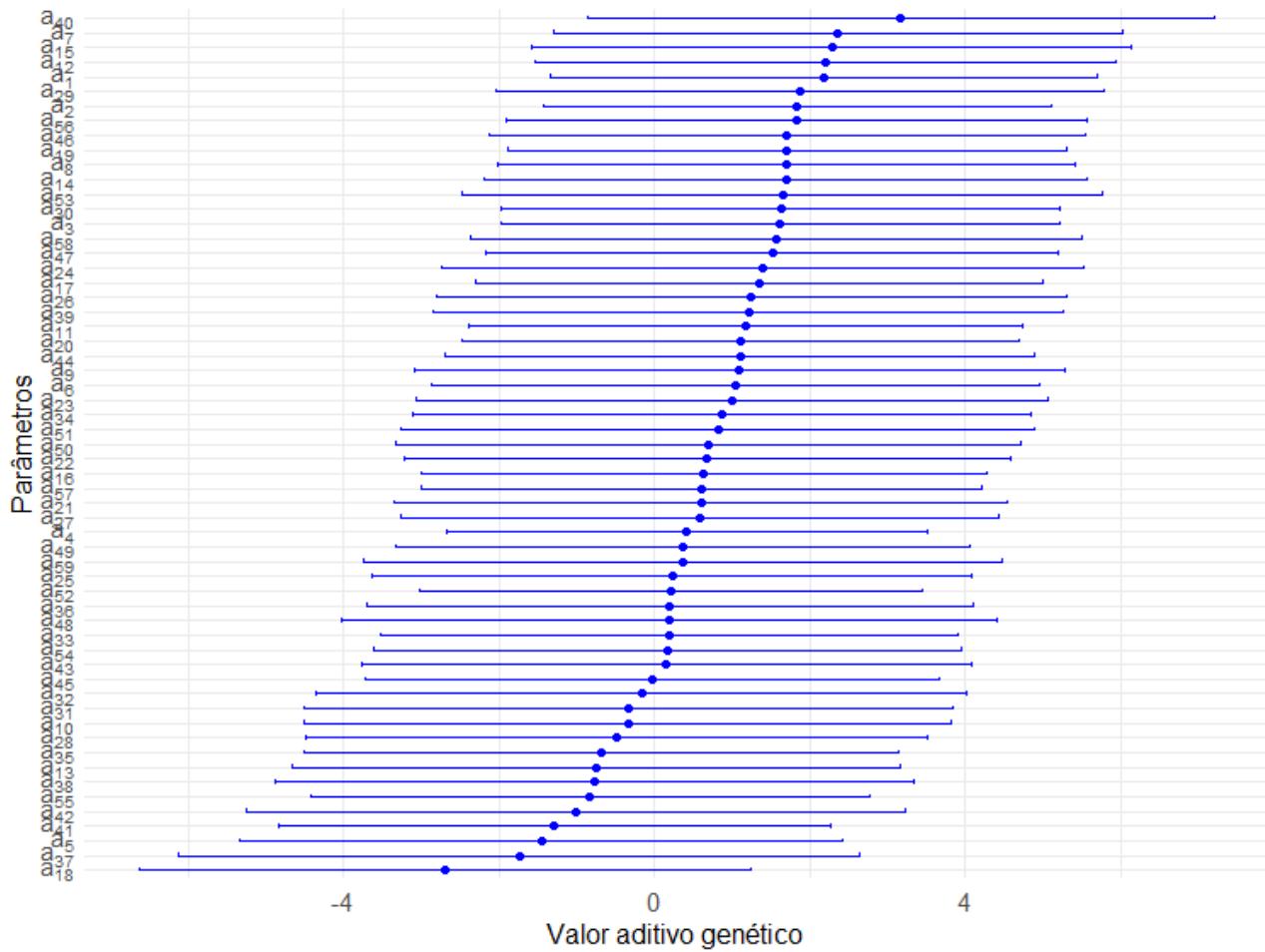


Figura 3.6: Gráfico de floresta das estimativas dos efeitos aditivos genéticos e seus intervalos HPD (95%) obtidos a partir das cadeias MCMC.

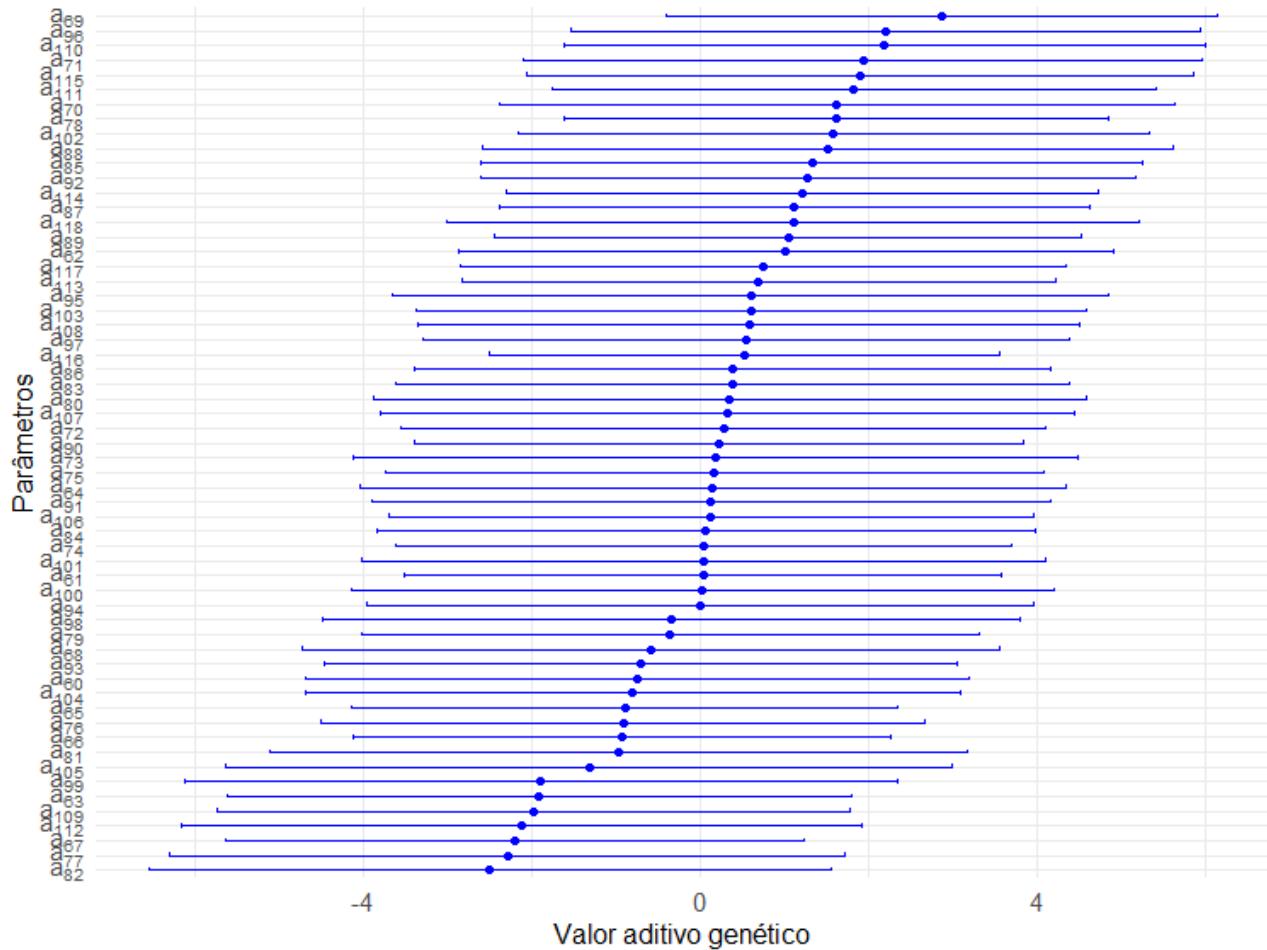


Figura 3.7: Gráfico de floresta das estimativas dos efeitos aditivos genéticos e seus intervalos HPD (95%) obtidos a partir das cadeias MCMC.

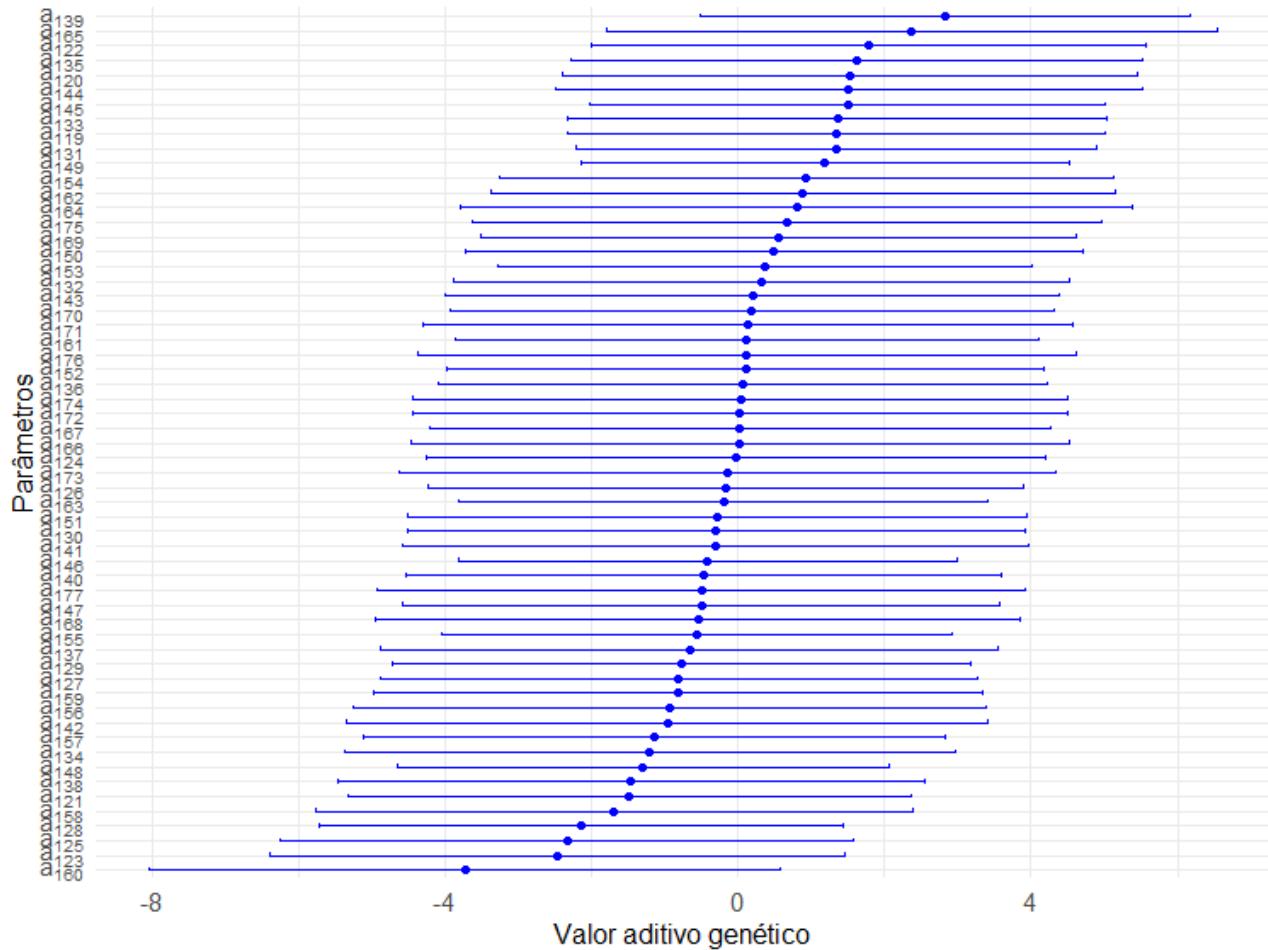


Figura 3.8: Gráfico de floresta das estimativas dos efeitos aditivos genéticos e seus intervalos HPD (95%) obtidos a partir das cadeias MCMC.

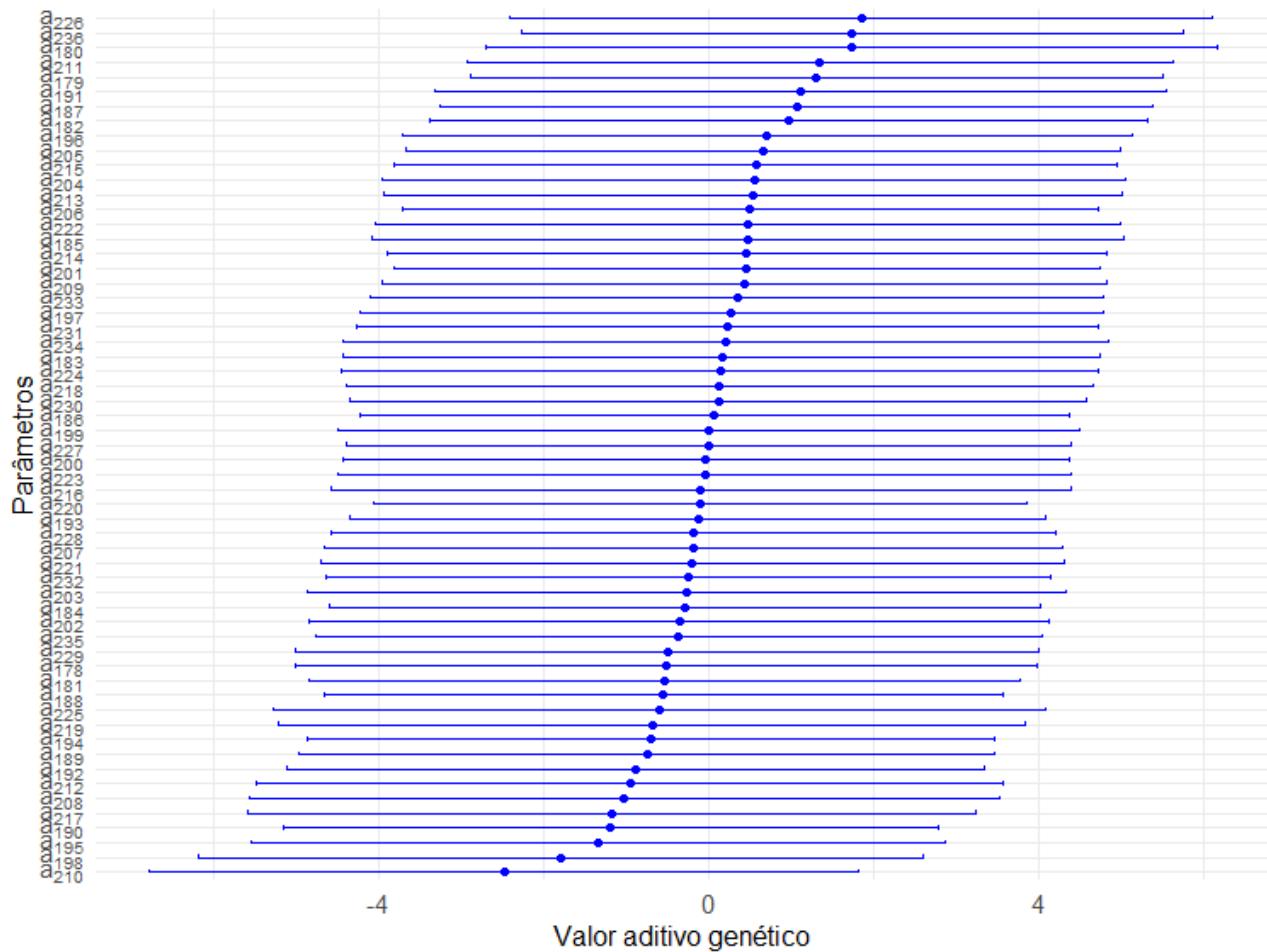


Figura 3.9: Gráfico de floresta das estimativas dos efeitos aditivos genéticos e seus intervalos HPD (95%) obtidos a partir das cadeias MCMC.

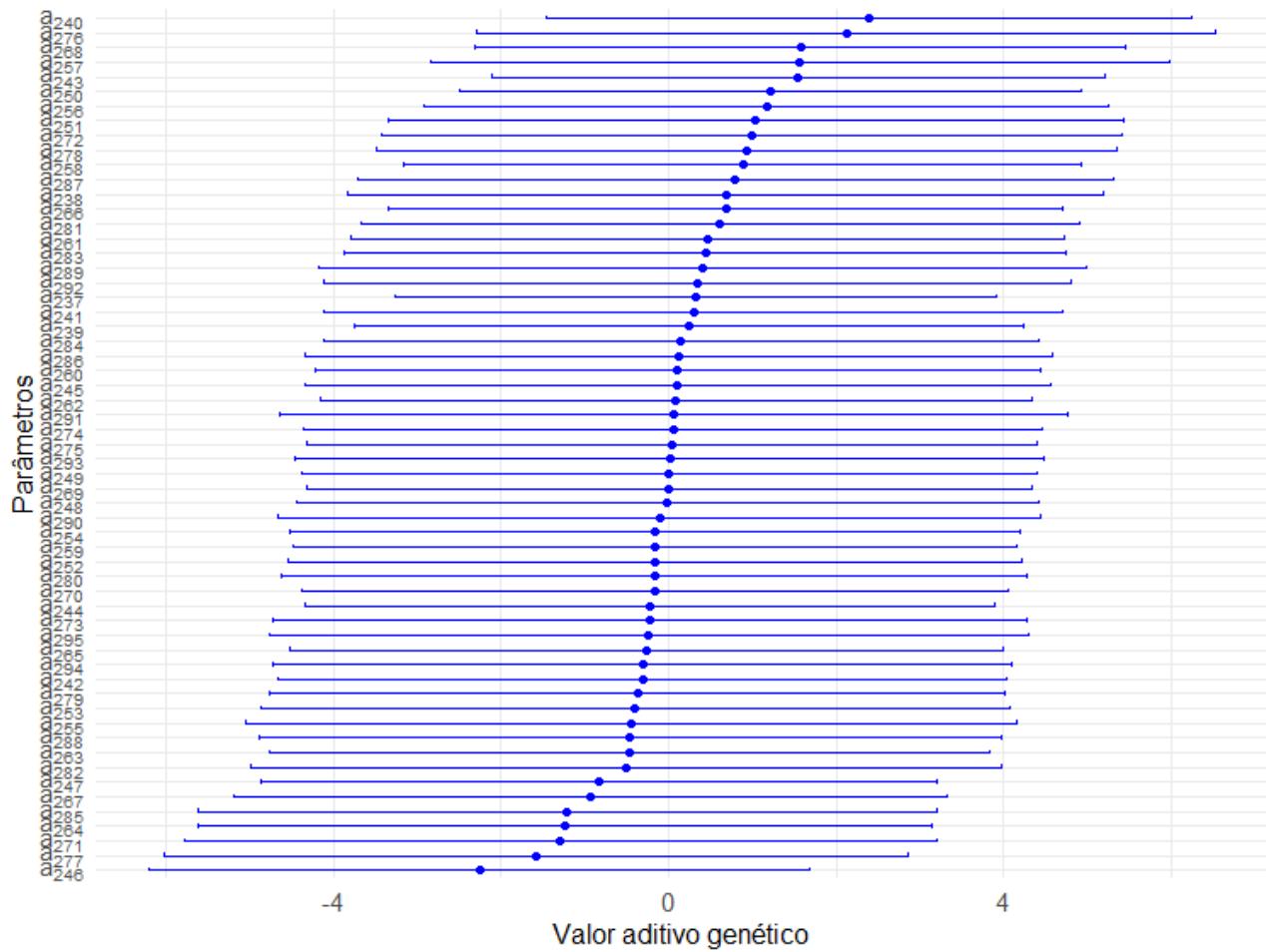


Figura 3.10: Gráfico de floresta das estimativas dos efeitos aditivos genéticos e seus intervalos HPD (95%) obtidos a partir das cadeias MCMC.

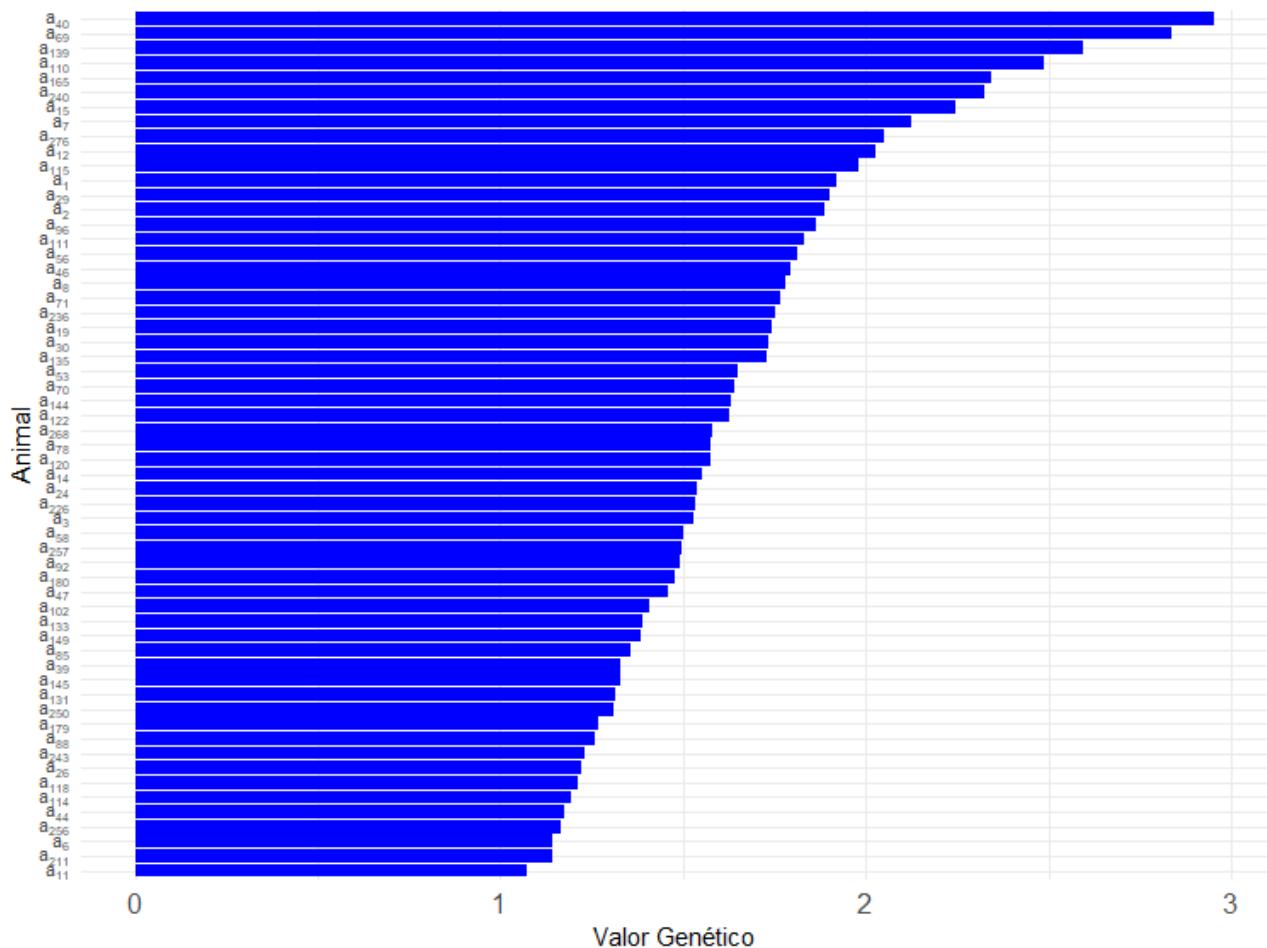


Figura 3.11: Gráfico ilustrando a classificação dos animais com base nos seus valores genéticos, representados por a_i

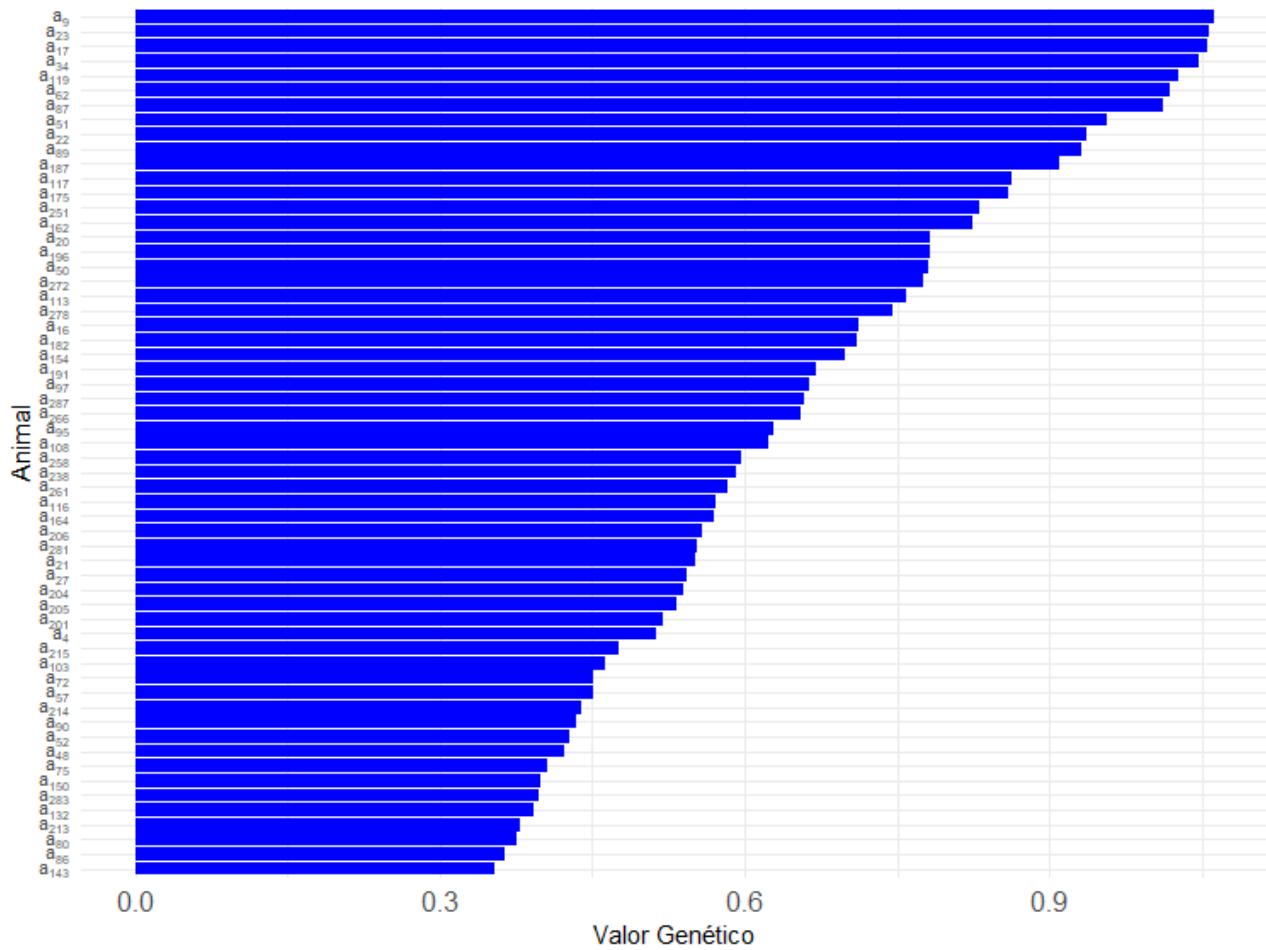


Figura 3.12: Gráfico ilustrando a classificação dos animais com base nos seus valores genéticos, representados por a_i .

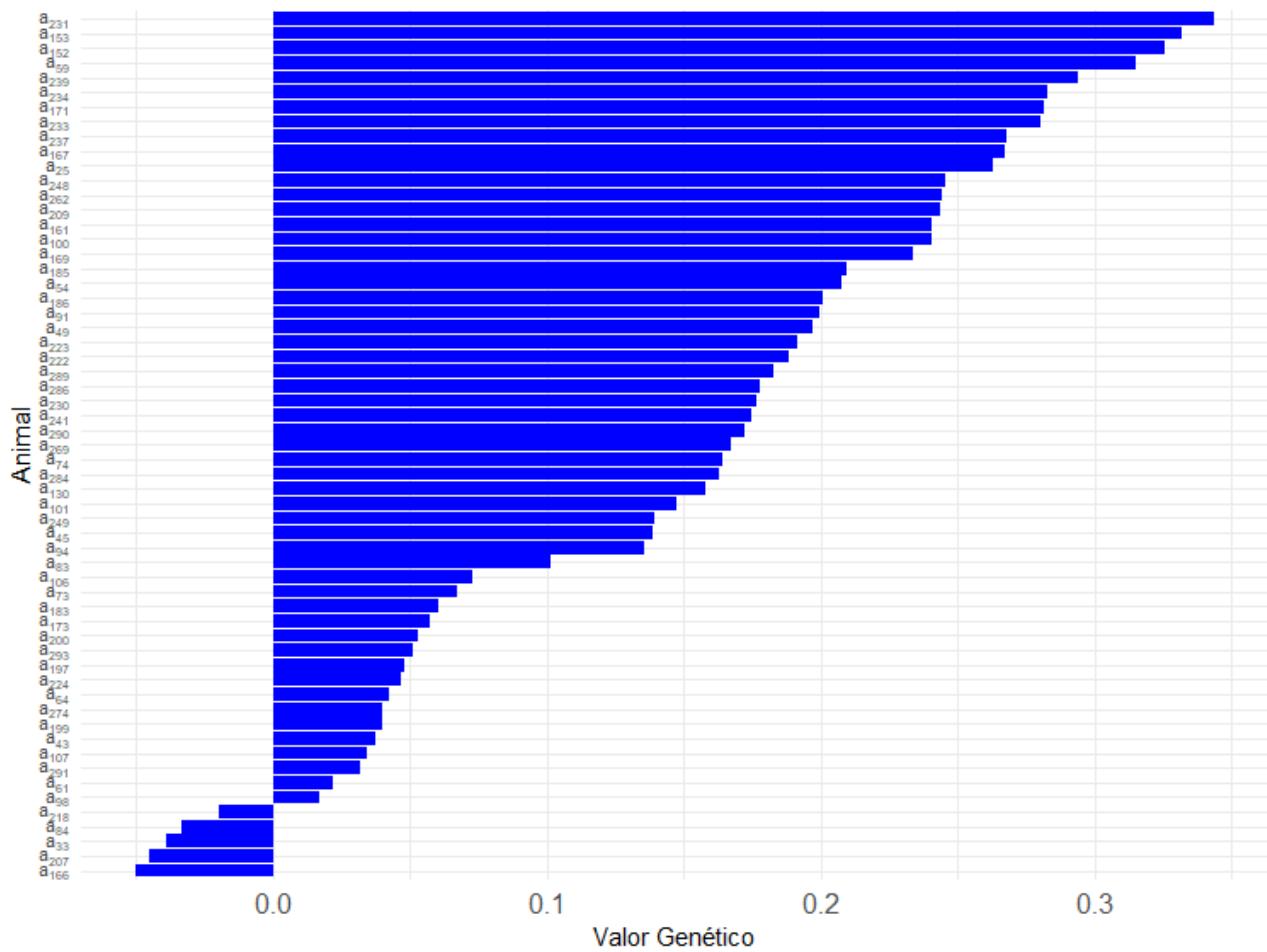


Figura 3.13: Gráfico ilustrando a classificação dos animais com base nos seus valores genéticos, representados por a_i .

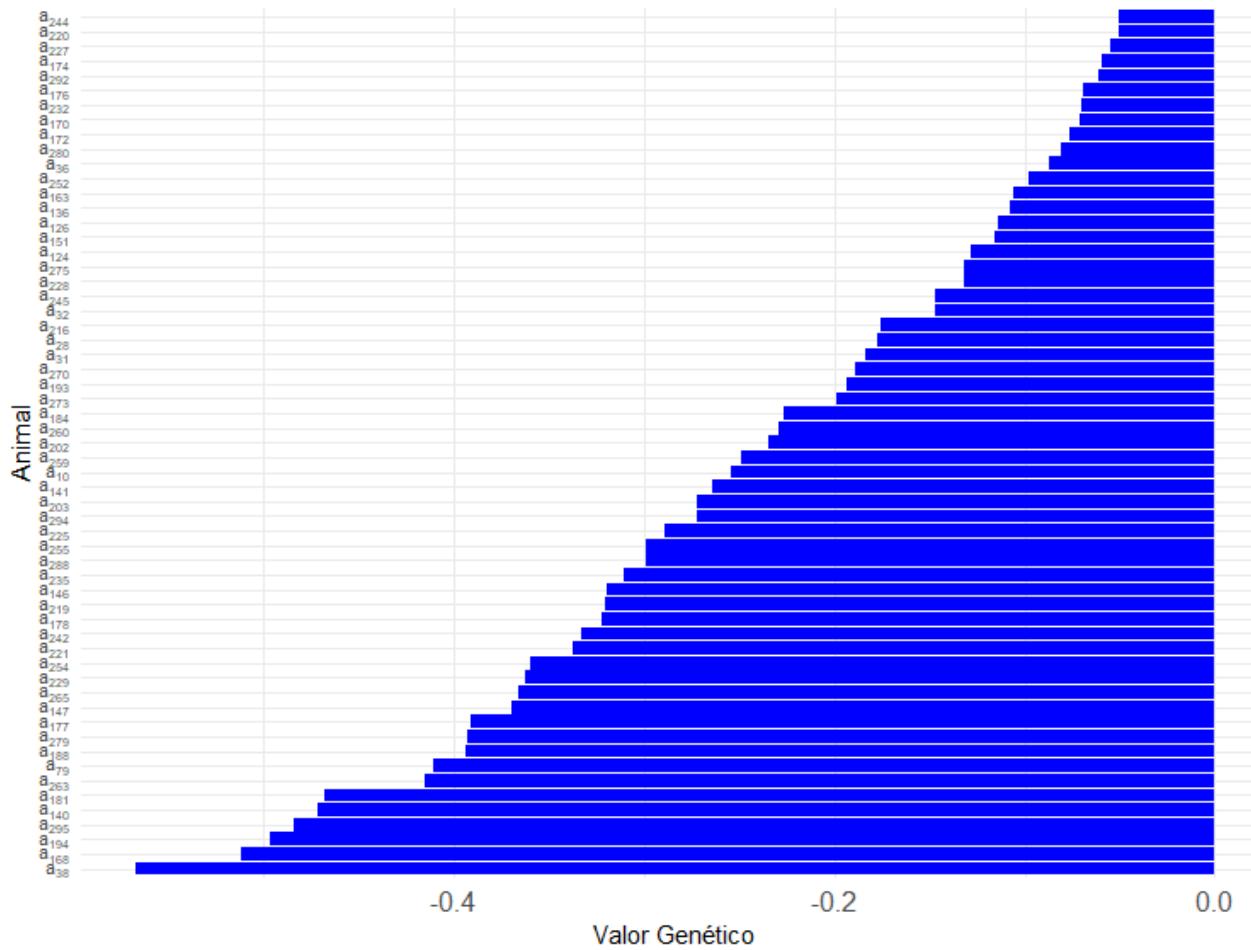


Figura 3.14: Gráfico ilustrando a classificação dos animais com base nos seus valores genéticos, representados por a_i .

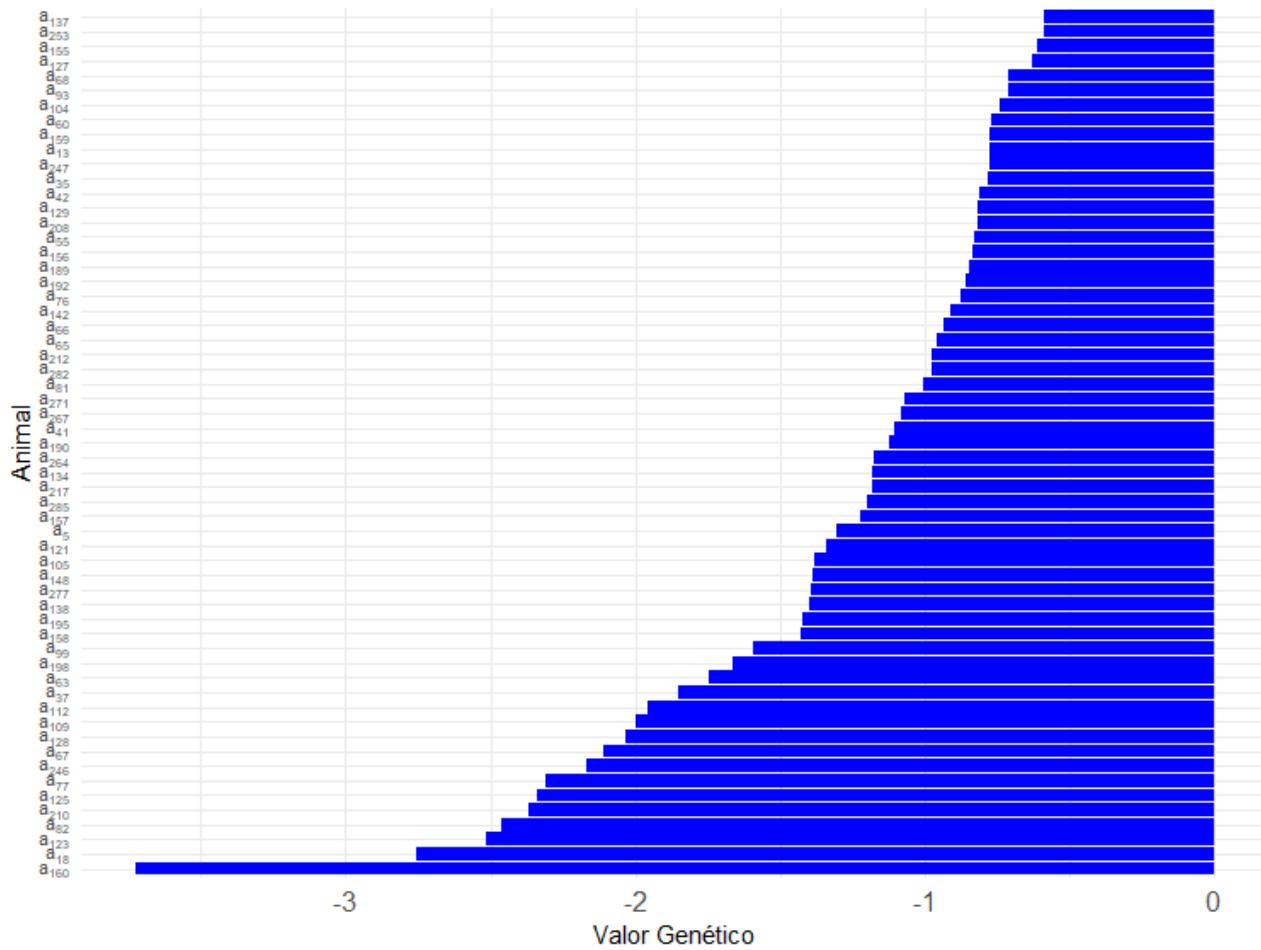


Figura 3.15: Gráfico ilustrando a classificação dos animais com base nos seus valores genéticos, representados por a_i .

4. CONCLUSÕES

Este trabalho teve como objetivo estimar parâmetros genéticos associados à produção leiteira em um rebanho bovino utilizando um modelo animal Bayesiano com auxílio do amostrador de Monte Carlo Hamiltoniano (HMC). A aplicação do modelo proposto permitiu estimar os componentes de variância genética aditiva e residual, além de calcular a herdabilidade da característica em questão.

Os resultados obtidos demonstraram que a herdabilidade estimada foi compatível com valores encontrados na literatura, situando-se em torno de 0,3. O intervalo de credibilidade obtido para h^2 reforça essa compatibilidade, indicando que há variabilidade genética suficiente para justificar a seleção com base em valores genéticos aditivos. Isso sustenta a viabilidade do uso de modelos mistos Bayesianos no contexto de melhoramento genético, especialmente quando se deseja incorporar de forma mais flexível o conhecimento prévio ou ajustar o modelo a pequenas populações.

Além disso, a análise dos efeitos fixos associados aos graus de sangue indicou maiores estimativas para os parâmetros β_6 , β_3 e β_1 , correspondendo, respectivamente, aos graus de sangue $\frac{5}{8}H$, $\frac{3}{4}H$ e $\frac{1}{2}H$. Isso sugere que, entre os níveis avaliados, esses graus de sangue apresentaram, em média, maior contribuição para a produção leiteira. Tal resultado pode refletir a combinação genética mais favorável desses grupos dentro da amostra estudada, sendo um indicativo da sua relevância no desempenho produtivo.

Com base nos objetivos específicos propostos — estimar a herdabilidade, quantificar os efeitos genéticos aditivos e interpretar os efeitos fixos de grau de sangue e idade ao parto — conclui-se que todos foram atendidos com sucesso. O modelo apresentou desempenho computacional satisfatório, e a metodologia utilizada mostrou-se adequada para a estrutura dos dados e o objetivo do estudo.

Como perspectiva futura, recomenda-se realizar avaliações e análises incluindo outras variáveis de efeito fixo que possam ser de interesse para o produtor, como fatores ambientais, de manejo e reprodutivos, além de expandir a modelagem para incluir mais informações sobre o parentesco dos animais. Isso permitirá uma análise mais robusta e detalhada do desempenho produtivo, considerando tanto os efeitos genéticos quanto os não genéticos, o que pode oferecer uma compreensão mais precisa para programas de seleção e estratégias de manejo adaptadas às características do rebanho.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Bayes, T.: *An Essay towards solving a Problem in the Doctrine of Chances*. Philosophical Transactions of the Royal Society of London, 53:370–418, 1763. Published posthumously by Richard Price.
- [2] Berry, D. e Cromie, A.: *Associations between age at first calving and subsequent performance in Irish spring calving Holstein-Friesian dairy cows*. Livestock Science, 123:44–54, 2009.
- [3] Berry, D. P. e Cromie, A. R.: *Associations between age at first calving and subsequent performance in Irish spring calving Holstein-Friesian dairy cows*. Livestock Science, 123(1):44–54, 2009.
- [4] Betancourt, M.: *A conceptual introduction to Hamiltonian Monte Carlo*. arXiv preprint arXiv:1701.02434, 2017.
- [5] Brooks, S. e Gelman, A.: *General Methods for Monitoring Convergence of Iterative Simulations*. Journal of Computational and Graphical Statistics, 7(4):434–455, 1998.
- [6] Brooks, S., Gelman, A., Jones, G. L. e Meng, X. L.: *Handbook of Markov Chain Monte Carlo*. CRC Press, 2011.
- [7] Cowles, M. K. e Carlin, B. P.: *Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review*. Journal of the American Statistical Association, 91(434):883–904, 1996.
- [8] Dale, A. I.: *A History of Inverse Probability: From Thomas Bayes to Karl Pearson*. Springer, 1999.
- [9] Eastham, N. T., Coates, A., Cripps, P., Richardson, H., Smith, R. e Oikonomou, G.: *Associations between age at first calving and subsequent lactation performance in UK Holstein and Holstein-Friesian dairy cows*. PLoS ONE, 13, 2018.
- [10] Gelman, A., Carlin, C. P., Stern, S. M., Dunson, D. B., Vehtari, A. e Rubin, R. L.: *Bayesian Data Analysis*. CRC Press, Boca Raton, FL, 3rd ed., 2013, ISBN 978-1-4398-0129-2.
- [11] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. e Rubin, D. B.: *Bayesian Data Analysis*. CRC Press, 3rd ed., 2013.

- [12] Gelman, A. e Rubin, D. B.: *Inference from Iterative Simulation Using Multiple Sequences*. Statistical Science, 7(4):457–472, 1992.
- [13] Geman, S. e Geman, D.: *Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images*. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-6(6):721–741, 1984.
- [14] Geweke, J.: *Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments*. pp. 169–193, 1992.
- [15] Gianola, D. e Fernando, R.: *Bayesian Methods in Animal Breeding Theory*. Journal of Animal Science, 63:217–244, 1986.
- [16] Girolando, A. B. dos Criadores de: *Sobre a Raça Girolando*, 2024. <https://www.girolando.com.br/girolando/sobre-a-raca>, Acessado em 22/04/2025.
- [17] Green, P. J.: *Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination*. Biometrika, 82(4):711–732, 1995.
- [18] Hastings, W. K.: *Monte Carlo Sampling Methods Using Markov Chains and Their Applications*. Biometrika, 57(1):97–109, 1970.
- [19] Hoffman, M. D. e Gelman, A.: *The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo*. Journal of Machine Learning Research, 15:1593–1623, 2014.
- [20] Kruschke, J. K.: *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. Academic Press, 2nd ed., 2015.
- [21] Laplace, P. S.: *Théorie analytique des probabilités*. Courcier, Paris, 1812.
- [22] Machado, M. A. D. et al.: *Genetic evaluation of Girolando cattle in Brazil*. Tropical Animal Health and Production, 45(5):997–1004, 2013.
- [23] McGraw, S. B.: *The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy*. Yale University Press, 2011.
- [24] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. e Teller, E.: *Equation of State Calculations by Fast Computing Machines*. Journal of Chemical Physics, 21(6):1087–1092, 1953.
- [25] Mrode, R. A.: *Linear Models for the Prediction of Animal Breeding Values*. CABI, 3rd ed., 2014.
- [26] Neal, R. M.: *MCMC using Hamiltonian dynamics*. Chapman Hall/CRC, 2011.
- [27] Pedrosa, V. B.: *Como Interpretar a PTA*, s.d. <https://canaldoleite.com/colunas/victor-breno-pedrosa/como-interpretar-a-pta/>.

- [28] Raftery, A. E., Lewis, S. *et al.*: *How many iterations in the Gibbs sampler.* Bayesian statistics, 4(2):763–773, 1992.
- [29] Raftery, A. E. e Lewis, S. M.: *How Many Iterations in the Gibbs Sampler?* pp. 763–773, 1992.
- [30] Robert, C. P. e Casella, G.: *Monte Carlo Statistical Methods.* Springer, 2nd ed., 2004.
- [31] Robert, C. P. e Casella, G.: *Monte Carlo Statistical Methods.* Springer, 2011.
- [32] Sorensen, D. e Gianola, D.: *Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics.* Statistics for Biology and Health. Springer-Verlag, 2002.
- [33] Sorensen, D. e Gianola, D.: *Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics.* Statistics for Biology and Health. Springer New York, 2007, ISBN 9780387954400.
- [34] Stigler, S. M.: *The History of Statistics: The Measurement of Uncertainty before 1900.* Harvard University Press, 1986.
- [35] Team, R. C.: *R: A Language and Environment for Statistical Computing*, 2022. <https://www.R-project.org/>.
- [36] Thomas, S. e Tu, W.: *Learning Hamiltonian Monte Carlo in R*, 2020. <https://arxiv.org/abs/2006.16194>.
- [37] Tierney, L.: *Markov Chains for Exploring Posterior Distributions.* Annals of Statistics, 22(4):1701–1728, 1994.
- [38] Wathes, D., Pollott, G. E., Johnson, K. F., Richardson, H. e Cooke, J.: *Heifer fertility and carry over consequences for life time production in dairy and beef cattle.* Animal: An International Journal of Animal Bioscience, 8 Suppl 1:91–104, 2014.

A. APÊNDICE: CÓDIGO DO MODELO BAYESIANO EM STAN.

```

data {
  int<lower=1> n;           // Numero total de observacoes
  int<lower=1> q;           // Numero de animais
  int<lower=1> p;           // Numero de efeitos fixos
  vector[n] y;             // Producao de leite
  matrix[q, q] A;          // Matriz de parentesco A
  matrix[n, p] X;          // Matriz de incidencia de efeitos fixos
  matrix[n, q] Z;          // Matriz de incidencia de animais
  vector[p] mu_prior;      // Media da distribuicao normal multivariada
  cov_matrix[p] Sigma_prior; // Matriz de covariancia
}

parameters {
  vector[p] beta;          // Efeitos fixos
  vector[q] a;              // Efeito genetico aditivo (aleatorio)
  real<lower=1e-10> sigma_a2; // Precisao do efeito genetico
  real<lower=1e-10> sigma_e2; // Precisao do erro residual
}

transformed parameters {
  real<lower=1e-10> sigma_e;
  sigma_e = sqrt(sigma_e2);
}

model {
  sigma_e2 ~ inv_gamma(2, 14);
  sigma_a2 ~ inv_gamma(2, 9);
  beta ~ multi_normal(mu_prior, Sigma_prior);
  a ~ multi_normal(rep_vector(0, q), sigma_a2 * A); // Efeitos geneticos
  y ~ normal(X * beta + Z * a, sigma_e); // Modelo de producao de leite
}

generated quantities {
  real h2;
  h2 = sigma_a2 / (sigma_a2 + sigma_e2); // Calculo da herdabilidade
}

```

B. APÊNDICE: CÓDIGO R.

```

# Limpa o ambiente
rm(list = c())

# Carrega pacote para estrutura de parentesco
library(kinship2)

# Carrega pacote para leitura de arquivos Excel
library(readxl)

# Le os dados do arquivo Excel
dados <- read_excel("TCC2.xlsx")

# Anexa os dados para facilitar o uso de colunas diretamente
attach(dados)

# Mostra os nomes das variaveis no conjunto de dados
names(dados)

# Remove individuos duplicados, mantendo apenas a primeira ocorrencia
unique_rows <- !duplicated(dados["ID"])
dados2 <- dados[unique_rows,]

# Cria o objeto pedigree com base nos IDs, pais, maes e sexo
tped <- with(dados2, pedigree(id = ID, dad = Pai, mom = Mae, sex = Sexo))

# Calcula a matriz de parentesco A (coeficientes de parentesco)
A <- kinship(tped)

# Verifica a dimensao da matriz A
dim(A)

# Seleciona as 350 primeiras observacoes para analise
df2 <- dados[1:350,]

# Define o vetor de resposta (producao de leite media diaria)
y <- df2$"Media_Diaria"

```

```

# Define o numero de observacoes
n <- length(y)

# Numero de individuos unicos na matriz A
q <- nrow(A)

# Modelo linear apenas para gerar a matriz do modelo X
modelo <- lm(y ~ df2$GS + as.numeric(df2$IPP) - 1)

# Extrai a matriz de incidencia dos efeitos fixos
X <- model.matrix(modelo)

# Numero de efeitos fixos
p <- ncol(X)

# Cria a matriz de incidencia Z para os efeitos aleatorios (animais)
Z <- matrix(0, nrow = n, ncol = q)
for (i in 1:n) {
  Z[i, df2$ID[i]] <- 1
}

# Carrega pacotes necessarios para usar o Stan
library("rstan")
library(rstudioapi)

# Configura o uso de multiplos nucleos para acelerar a amostragem
options(mc.cores = parallel::detectCores())
rstan_options(auto_write = TRUE)
options(mc.cores = 4)

# Compila o modelo Stan a partir do arquivo modelo2.stan
model <- stan_model("modelo2.stan")

# Define o vetor de medias a priori para os efeitos fixos (mu_prior)
b0 <- c(18, 15, 20, 17, 17, 19, 20, 16, 19, 0) # IPP nao tem efeito -> 0

# Define a matriz de covariancia a priori (Sigma_prior) dos efeitos fixos
sig_b0 <- diag(p) * 4 # Variancia padrao 4
sig_b0[10,10] <- 9 # Variancia maior para o efeito IPP

# Cria a lista com todos os dados a serem passados ao Stan
lista <- list(
  A = A,
  n = nrow(df2),
  q = ncol(Z),
  Z = Z,
  X = X,
  y = y,
  p = ncol(X),

```

```
mu_prior = b0,
Sigma_prior = sig_b0
)

# Executa a amostragem no Stan
fit <- sampling(
  model,
  lista ,
  iter = 22000,      # Total de iteracoes por cadeia
  chains = 3,         # Numero de cadeias
  warmup = 2000,      # Numero de iteracoes de aquecimento
  thin = 5            # Fator de thinning
)

# Extrai os parametros amostrados
params <- extract(fit)
```