OSVALDO MOISÉS ASSINDE

INFERÊNCIA SOBRE GRÁFICOS DE RADAR: UM EXEMPLO NO MELHORAMENTO VEGETAL

LAVRAS – MG

2025

OSVALDO MOISÉS ASSINDE

INFERÊNCIA SOBRE GRÁFICOS DE RADAR: UM EXEMPLO NO MELHORAMENTO VEGETAL

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, para obtenção do título de Mestre

Prof. Júlio Silvio de Sousa Bueno Filho
Orientador

LAVRAS – MG
2025

Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).

Assinde,Osvaldo Moisés

Inferência sobre gráficos de radar: um exemplo no melhoramento vegetal / Osvaldo Moisés Assinde. – 2025.

124 p. :

Dissertação(mestrado acadêmico)–Universidade Federal de Lavras, 2025.

Orientador: Prof. Júlio Silvio de Sousa Bueno Filho.

Coorientador: .

Bibliografia.

1. Inferência Bayesiana. 2. Gráficos de Radar. 3. Melhoramento Genético. I.Filho,Júlio Silvio de Sousa Bueno .

OSVALDO MOISÉS ASSINDE

INFERÊNCIA SOBRE GRÁFICOS DE RADAR: UM EXEMPLO NO MELHORAMENTO VEGETAL

INFERENCE ON RADAR CHARTS: AN EXAMPLE IN PLANT BREEDING

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, para obtenção do título de Mestre.

APROVADA em 31 de janeiro de 2025.

Prof. Dr. Jose Airton Rodrigues Nunes – UFLA

Prof. Dr. Paulo Henrique Sales Guimarães – UFLA

Prof. Dr. Danilo Machado Pires – UNIFAL

Prof. Júlio Silvio de Sousa Bueno Filho

Orientador

LAVRAS – MG

2025

Este trabalho de pesquisa é inteiramente dedicado aos meus pais F.C ASSINDE (in Memorian) e Fernanda Moisés,a minha esposa Lourena e aos meus filhos Osvaldo, Yumina e Franswesley, aos meus irmãos Luis, Abu, Valdemiro e Horacio, Ildefonso,Cosme (in memorian), as minhas irmãs Verônica, Mônica, Nilza e Fernanda (in Memorian). Os maiores incentivadores das realizações dos meus sonhos. Muito obrigado!

AGRADECIMENTOS

Primeiro, expressar minha profunda gratidão ao meu Bom Deus, cuja Sua presença sustentou-me ao longo desses dois anos de mestrado.

Aos meus pais, Francisco Cosme Assinde (*In Memorian*) e Fernanda Moisés, merecem uma gratidão especial por seu amor, incentivo e apoio incansável. Cada esforço para me dar o melhor que podiam não passou e nunca passará despercebido. Agradeço também aos meus irmãos e minhas irmãs, verdadeiros incentivadores dos meus estudos, por suas palavras, atitudes de amor e financeiros, por não me deixarem nunca me sentir sozinho.

Ao meu orientador, Prof. Júlio Silvio de Sousa Bueno Filho, que vai além do papel de orientador, sendo também um inspirador e referencia na area. Agradeço por compartilhar seus conhecimentose a honra de ser seu orientado.

À banca examinadora, Prof. Dr. José Airton Rodrigues Nunes (UFLA), Prof. Dr. Paulo Henrique Sales Guimarães (UFLA) e Prof. Dr. Danilo Machado Pires (UNIFAL), agradeço pelas valiosas contribuições que tornaram este trabalho digno de aprovação.

Aos colegas de pós-graduação, agradeço por compartilharem conhecimento, tirarem dúvidas e facilitarem minha adaptação a uma nova cidade e Pais. Que nossa amizade continue contribuindo para a ciência e que perdure por toda a vida.

Agradeço aos amigos, por todas as vezes que precisei de um ombro amigo e por estarem presente na minha vida.

Aos professores do Departamento de Estatística e Experimentação Agropecuária, meu reconhecimento pelas excelentes aulas e pela contribuição para minha formação.

A Universidade Federal de Lavras, na pessoa do Prof. Dr. Tales Jesus Fernandes, coordenador do Departamento de Estatistica e Experimentação Agropecuária, expresso a essa Instituição minha gratidão por permitir que trilhasse o árduo, mas satisfatório caminho acadêmico do mestrado.

Por fim, e não menos importante, agradeço o apoio financeiro da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) durante esses dois anos de estudo.

RESUMO

Este estudo teve como objetivo desenvolver uma ferramenta para a seleção multivariada baseada em inferência Bayesiana em cada caractere univariado para identificar genótipos superiores em programas de melhoramento vegetal, utilizando o gráfico de radar para avaliar a regularidade e equilíbrio dos genótipos, destacando tanto a consistência do desempenho quanto especializações vantajosas. Inicialmente, um experimento simulado avaliou o potencial da metodologia, que depois foi aplicada a um estudo real com dez características agronômicas e tecnológicas em 45 genótipos de sorgo sacarino. A técnica proposta pode ser ajustada conforme os interesses dos melhoristas, sendo aplicável a outras culturas e estudos de interação genótipo-ambiente. Embora a abordagem Bayesiana seja flexível e forneça distribuições completas para os parâmetros, resultados semelhantes poderiam ser obtidos com estimativas REML, ou métodos semelhantes de modelos mistos. Além de facilitar a seleção baseada em múltiplas características, o método proposto é de fácil aplicação e apresenta grande potencial no melhoramento vegetal.

Palavras-chave: análise multivariada; gráfico de radar; melhoramento vegetal; inferência; decisão estatística.

ABSTRACT

This study aimed to develop a tool for multivariate selection based on Bayesian inference from each trait to identify superior genotypes in plant breeding programs, using the radar chart to assess the regularity and balance of genotypes, highlighting both performance consistency and advantageous specializations. Initially, a simulated experiment evaluated the methodology's potential, which was later applied to a real study involving ten agronomic and technological traits in 45 sweet sorghum genotypes. The proposed technique can be adjusted according to breeders' interests, making it applicable to other crops and genotype-environment interaction studies. Although the Bayesian approach is flexible and provides complete distributions for the parameters, similar results could be obtained using REML estimates, or similar techniques from mixed models. In addition to facilitating selection based on multiple traits, the proposed method is easy to apply and shows great potential in plant breeding.

Keywords: multivariate analysis; radar chart; plant breeding; inference; statistical decision.

INDICADORES DE IMPACTO

O estudo apresenta um impacto potencial significativo na análise de experimentos com múltiplas variáveis de resposta, destacando a importância de abordagens visuais e estatísticas. O estudo proporciona uma visão clara e intuitiva da relação entre variáveis e tratamentos. Este método facilita a identificação de padrões e interações complexas entre variáveis de resposta, permitindo uma inferência mais detalhada e precisa. A visualização gráfica torna os resultados acessíveis a um público mais amplo, incluindo pesquisadores e tomadores de decisão, contribuindo para uma melhor comunicação dos achados experimentais.Os indicadores de impacto incluem Agricultura, Ciências da Saúde, Engenharia, Educação e Pesquisa, e mais. A aplicação de gráficos de radar em estudos nas áreas definidas pode melhorar a compreensão dos efeitos de diferentes tratamentos em múltiplas safras e variáveis ambientais simultaneamente. Além disso, podem ser utilizados para descrever respostas multivariadas de grupos e ajudar a visualizar o desempenho de várias propriedades materiais simultaneamente. Ao integrar gráficos de radar com a inferência estatística sobre tratamentos experimentais, o estudo fornece uma abordagem visual poderosa e eficaz para a tomada de decisões baseada em dados.

IMPACT INDICATORS

The study presents significant potential impact in the analysis of experiments with multiple response variables, highlighting the importance of visual and statistical approaches. The study provides a clear and intuitive view of the relationship between variables and treatments. This method facilitates the identification of patterns and complex interactions between response variables, allowing for more detailed and precise inference. The graphical displays render the results accessible to a broader audience, including researchers and decision-makers, contributing to better communication of experimental findings.The impact indicators include Agriculture, Health Sciences, Engineering, Education and Research, and more. The application of radar charts in studies from these areas can improve the understanding of the effects of different treatments across multiple crops and environmental variables simultaneously. Additionally, they can be used to describe multivariate responses of groups and help to discern the performance of various material properties simultaneously. By integrating radar charts with statistical infe-

rence on experimental treatments, the study provides a powerful and efficient visual approach for data-driven decision-making.

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO

1 INTRODUÇÃO

Há muitas situações em que a avaliação simultânea de duas ou mais características relacionadas é necessária. Monitorá-las e avaliá-las independentemente pode não ser recomendado (MONTGOMERY, 2009). Nessas situações, técnicas específicas podem ser empregadas para detectar, identificar e analisar as principais fontes de variabilidade em um processo (DUARTE et al., 2007).

A coleta de dados em pesquisas tem se tornado cada vez mais abrangente, abrangendo uma ampla variedade de tópicos. Quando esses dados são coletados em diferentes grupos e posteriormente comparados, o método usual envolve testes estatísticos, como χ^2, ANOVA, análise de variância multivariada ou testes t, seguidos da apresentação dos resultados em formato de texto ou tabela (HENRY, 1995).

Além de textos e tabelas, os resultados da pesquisa podem ser representados graficamente. No entanto, a visualização pode se tornar complexa quando há muitas variáveis ou grupos a serem analisados, especialmente quando as escalas de medição diferem (SAARY, 2008). Isso destaca a necessidade de ferramentas que facilitem a interpretação das interações entre múltiplas variáveis de resposta e os tratamentos experimentais.

O problema central deste estudo é como realizar inferência sobre tratamentos experimentais quando múltiplas variáveis de resposta são observadas simultaneamente no melhoramento vegetal. O uso de gráficos de radar, aliado a técnicas estatísticas apropriadas, tem potencial para aprimorar a capacidade de inferência sobre os efeitos dos tratamentos experimentais nessas condições.

Diante disso, este trabalho teve como objetivo combinar uma técnica de seleção multivariada (gráficos de radar) a resultados de inferência Bayesiana univariada para conjuntos e variáveis, de forma a permitir identificar genótipos superiores em programas de melhoramento vegetal. A metodologia proposta utiliza os gráficos de radar para diferenciar os tratamentos de um experimento. Especificamente, busca representar os valores genéticos (BLUP) para cada modelo univariado e comparar os tratamentos com base nessa abordagem visual

2 REFERENCIAL TEÓRICO

Neste capítulo, exploramos algumas literaturas que abordam sobre dados multivariados,aplicações e representações visuais dos gráficos de radare apresentamos os principais fundamentos teórico-conceituais da Inferencia bayesiana.

2.1 Dados multivariados

Muitos processos de experimentação são multivariados, pois envolvem a avaliação de diversas caracterısticas, ou variáveis respostas, em todas as unidades experimentais.

Para Saary (2008), os dados multivariados podem ser representados visualmente de várias maneiras bidimensionais, como glifos, perfis, curvas de Andrews, faces de Chernoff e caixas.

Assim, na busca de maneiras alternativas de examinar e exibir padrões nos dados, identificou uma forma de graficação chamada gráfico de radar proposto por Mayr (1877) (SOPHIA, 2023), mas, pouco difundido,sendo construidos a partir de análises isoladas (univariadas).

2.2 Gráfico de radar

Um gráfico de radar consiste num gráfico bidimensional usando eixos radiais para traçar um ou mais grupos de valores em várias variáveis.

A Figura 2.1 é um gráfico de radar (ou gráfico em teia de aranha), utilizado para comparar múltiplas variáveis de forma simultânea em diferentes categorias ou tratamentos. O gráfico possui 8 eixos radiais, cada um representando uma variável (Var 1 a Var 8). Os valores das variáveis são plotados ao longo desses eixos e conectados por linhas, formando um polígono para cada tratamento.

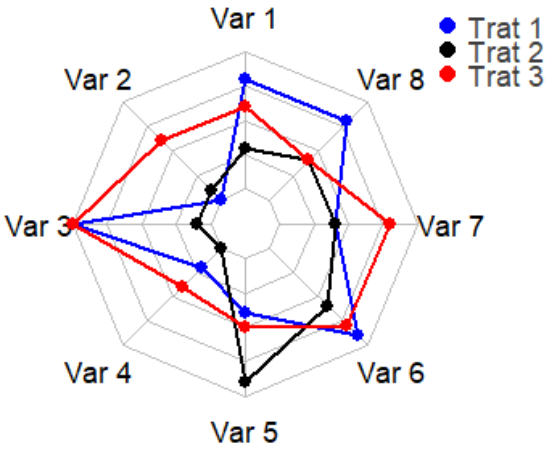
Cada tratamento é representado por uma cor distinta e um conjunto de pontos conectados. O Tratamento 1 (azul) apresenta valores elevados nas variáveis, Var 1 e Var 3. O Tra-

tamento 2 (preto) apresenta menores valores na maioria das variáveis, formando um polígono menor. O Tratamento 3 (vermelho) se destaca principalmente nas variáveis Var 3 e Var 7.

A forma dos polígonos indicam o desempenho relativo de cada tratamento. Quanto maior a área do polígono,melhor o desempenho geral do tratamento nas variáveis analisadas.

Este gráfico permite visualizar e comparar o desempenho dos tratamentos em diferentes variáveis de maneira intuitiva. A análise da área dos polígonos pode ser útil para decisões em experimentos científicos e melhoramento genético.

Figura 2.1 – exemplo de gráfico de radar.



Fonte:Do Autor (2024).

2.2.1 Características dos gráficos de radar

Os gráficos de radar possuem diversas características que os tornam úteis na análise de dados multivariados:

a) possibilitam a representação de múltiplas variáveis em um único gráfico, facilitando a comparação e a identificação de padrões;
b) permitem visualizar a distribuição dos valores em relação ao centro do gráfico, proporcionando uma análise intuitiva da variabilidade dos dados;
c) são eficazes para detectar discrepâncias e destacar variáveis que representam pontos fortes ou fracos dos genótipos analisados;
d) apresentam desafios na comparação entre diferentes polígonos, especialmente quando suas formas variam significativamente;

e) podem se tornar complexos e difíceis de interpretar quando há um grande número de variáveis ou quando os valores são muito próximos entre si;
f) são sensíveis a valores extremos, que podem distorcer a forma do polígono e dificultar a comparação entre os dados.

2.3 Estudos e Aplicações dos Gráficos de Radar

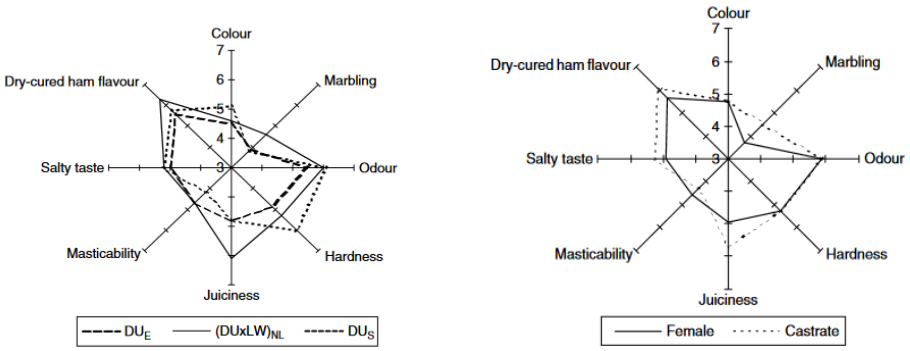
A visualização gráfica apresenta os resultados em termos de variáveis indiretas, como médias e coeficiente de variação (PIEPHO et al., 2008; CASTRO et al., 2016; FILHO et al., 2016).

Gráficos radiais em qualquer uma de suas diversas formas, ainda não são amplamente utilizados nos campos da Agricultura ou das Ciencias de Alimentos,não havendo muitos exemplos na literatura.

Alguns autores utilizaram a metodologia de selecção de indivíduos considerando um conjunto de características com base em gráficos de radar que não requer conhecimento prévio das covariâncias genéticas e fenotípicas. Este é um procedimento interessante, que fornece uma ferramenta analítica descritiva e utiliza um índice gráfico baseado na soma dos valores padronizados das características. Isso permite que o pesquisador visualize quais características da progênie possuem fenótipos favoráveis (REIS et al., 2011).

Soriano et al. (2005) usaram o gráficos de radar (aranha ou de teia de aranha) para a análise descritiva dos presuntos curados de fêmeas e castrados (em suinos). Os presuntos de castrados apresentaram mais marmoreio, suculência, sabor salgado e sabor característico de presunto curado em comparação aos presuntos de fêmeas

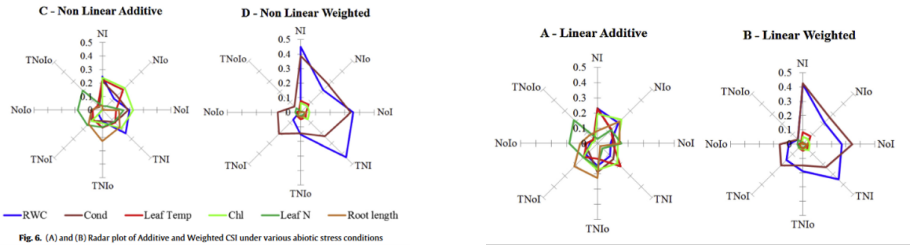
Figura 2.2 – gráficos de aranha (radar) para a comparação de três grupos (esquerda) ou dois grupos (direita).



Fonte: Soriano et al. (2005).

Banerjee et al. (2015) usaram um gráfico de radar para visualizar o Índice de Status da cultura de trigo,usado como indicador de condição de crescimento sob situações de estresse abiótico.

Figura 2.3 – índice de status da cultura de tabaco



Fonte: Banerjee et al. (2015).

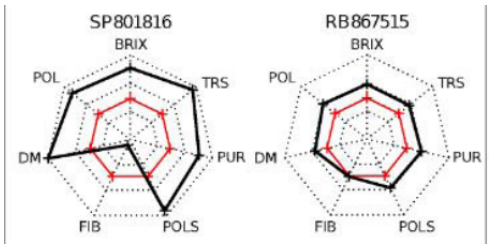
Kui et al. (2010) avaliaram características quantitativas e qualitativas do tabaco, avaliando sete variedades de tabaco no número de folhas, o peso da folha única, a produção por ha, preço médio, relação superior do tabaco, o teor de nicotina, o teor de açúcar e potássio reduzidos com base na consideração abrangente de muitos índices e coordenação em sítios experimentais de Zhucheng em 2007.

Figura 2.4 – avaliação de características quantitativas e qualitativas de tabaco

Fonte: Kui et al. (2010).

Silva *et al.* (2016) usaram para desenvolver uma abordagem de índice de seleção que considerou as informações de diversas características da cana-de-açúcar para identificar genótipos superiores utilizando a área do polígono de gráficos de radar tradicionais.

Figura 2.5 – índice de seleção para cana-de-açúcar



Fonte:Silva *et al.* (2016).

Nunes et al. (2005) classificaram as linhagens e decidiram quais recomendar a partir das estimativas de estabilidade e adaptabilidade, tendo em conta que haviam algumas variáveis ou características usando o gráfico polar onde são plotados os fenótipos inerentes a cada variável para dado genótipo gerando gráficos com formato “bola cheia” ou “bola murcha”.

Figura 2.6 – identificação de cultivares com maior estabilidade fenotípica

Fonte: Nunes et al. (2005).

O gráfico de radar pode ser construido de forma a ser mais claro e fácil para estabelecer comparações do que utilizando gráficos de barras. A vantagem seria a imédiata de onde se

localizam as diferenças importantes no quadro geral, constituindo uma maneira mais eficiente de resumir a variabilidade dos dados em uma única imagem. O grau de complexidade dos gráficos pode crescer de maneira controlada pelos interesses do pesquisador, adicionando ou removendo raios, conforme necessário e recalibrando a escala, se necessário (SAARY, 2008).

O uso de gráficos de radar para grandes volumes de informação é vista no trabalho de Ansari Ceri J. Phillips (2001), em que compara as opiniões de 427 participantes de quatro partes interessadas titulares: os agentes comunitários de saúde; o pessoal principal dos projetos; membros 'solo' da comunidade; e representantes de agências voluntárias, agências comunitárias e organizações não governamentais. Os dados de quatro grupos diferentes de partes interessadas são dispostos num gráfico de radar de 16 variáveis, cada uma variando de 1 a 7.

Dados de múltiplos grupos sobre múltiplas variáveis podem ser facilmente resumidos em um único gráfico inteligível, e as explicações textuais dos mesmos levam páginas para serem elaboradas. Isto não quer dizer que o gráfico simplifique demais os dados, mas apenas que os padrões são mais facilmente determinados.

2.4 Abordagem Bayesiana para obtenção de valores genéticos

Existem duas correntes de pensamento bem diferentes para a inferência em Estatística: a corrente frequentista e a corrente da chamada inferência Bayesiana. Segundo o enfoque frequentista, o conceito de probabilidade corresponde ao de frequência relativa, em que a população é descrita por um modelo probabilístico que é função de constantes (parâmetros) desconhecidas. Já a inferência Bayesiana é o processo pelo qual se encontra um modelo de probabilidade para um conjunto de dados e se resume o resultado por uma distribuição de probabilidade sobre os parâmetros do modelo e sobre quantidades não observadas, tais como predição para novas observações (GELMAN et al., 2000).

A escola Bayesiana foi, na prática, fundada pelo aristocrata e político francês Marquis Pierre-Simon Laplace através de várias obras publicadas de 1774 a 1812, e teve um papel preponderante na inferência científica durante o século XIX. No entanto, antes de Laplace, e aparentemente sem seu conhecimento, um artigo póstumo atribuído a um obscuro clérigo, o reverendo Thomas Bayes, foi apresentado na Royal Society em Londres, formalizando o mesmo princípio de inferência. O trabalho de Fisher sobre a verossimilhança na década de 1920 e o

da escola frequentista nas décadas de 1930 e 1940 fizeram a escola Bayesiana quase desaparecer, até que um renascimento começou em meados da década de 1950 (BLASCO, 1998), com grande desenvolvimento até os dias atuais.

Na abordagem Bayesiana, todos os efeitos do modelo, sejam fixos ou aleatórios conforme definidos sob o ponto de vista frequentista, são tratados como aleatórios. Em outras palavras, as definições de efeitos fixos e aleatórios na abordagem Bayesiana não surgem naturalmente como na abordagem frequentista (SORENSEN; GIANOLA, 2002).

A abordagem clássica ou frequentista diferencia-se da abordagem Bayesiana, pois os parâmetros são vistos como variáveis aleatórias cujo comportamento é regulado por uma distribuição de probabilidade que se assume sobre seus possíveis valores, traduzindo uma informação inicial (ou a priori) que se tenha sobre os parâmetros, antes mesmo de se obter os dados.

A seguir, essa informação a priori é combinada com aquela proveniente dos dados, produzindo uma informação a posteriori sobre os parâmetros. Essa informação a posteriori se expressa na chamada distribuição a posteriori dos parâmetros, que é utilizada para se fazer inferência sobre mesmos (COSTA et al., 2009).

2.4.1 Teorema de Bayes

Para se fazer afirmações sobre a probabilidade de θ dado que tem-se Y, deve-se começar com um modelo que forneça uma distribuição conjunta para o parâmetro e a amostra. A junção das informações que os dados (amostra) possuem com o conhecimento pré-existente (subjetivo) do pesquisador a respeito dos parâmetros só é possível através do Teorema de Bayes, daí o termo inferência Bayesiana (BOX; TIAO, 1973).

Seguindo a definição de Box e Tiao (1973), suponha $y = (y_1, \ldots, y_n)$ um vetor de n observações cuja distribuição de probabilidade $p(y|\theta)$ depende dos k valores do parâmetro $\theta' = (\theta_1, \ldots, \theta_k)$. Suponha também que a distribuição de probabilidade de θ é $p(\theta)$. Então,

$$p(\theta|y) = \frac{P(y|\theta)p(\theta)}{\int P(y|\theta)p(\theta)\,d\theta}$$

Em que $y = \{y_1, y_2, \ldots, y_n\}$. O denominador não traz nenhuma nova informação a respeito de θ, servindo antes como um *padronizador*, também chamado de *constante normalizadora*. Assim, o modelo pode ser reescrito como:

$$p(\theta|y) \propto p(y|\theta)p(\theta),$$

ou seja, a distribuição para θ posterior aos dados é proporcional ao produto da distribuição a priori para θ e a verossimilhança de θ, dado y (BOX; TIAO, 1992).

2.4.2 Distribuições a priori

Para DeGroot e Schervish (2012), a distribuição a priori é parte fundamental da inferência Bayesiana. Se não determinamos alguma priori específica, não conseguimos calcular a distribuição a posteriori e, portanto, a análise Bayesiana fica comprometida. Em geral, para uma mesma verossimilhança, diferentes escolhas de prioris podem nos levar a resultados ligeiramente diferentes. Isso é verdade especialmente quando temos uma grande quantidade de dados ou quando as prioris que estão sendo comparadas são muito dispersas.

Como exemplo, O'Hagan (1994) diz que na estatística clássica, se uma variável aleatória X tem distribuição binomial, o melhor estimador para θ, de acordo com algum critério, é X/n. Isto vale para todos os problemas em que podemos modelar a variável aleatória por uma distribuição binomial. No entanto, para a estatística Bayesiana, cada problema é único. De acordo com as informações disponíveis pelo especialista, a distribuição a priori é formulada e, por incorporar o conhecimento do investigador, ela pode diferir em cada problema. Mesmo que a verossimilhança seja a mesma, ao se utilizar diferentes distribuições a priori, as distribuições a posteriori foram diferentes, conduzindo assim a análises Bayesianas distintas. Para a escolha da priori, deve-se levar em consideração alguns aspectos importantes, tais como:

a) estar definida no espaço paramétrico;
b) conduzir a uma posteriori integrável;
c) refletir, de modo adequado, o conhecimento sobre o parâmetro obtido pelo especialista.

2.4.2.1 Distribuição a priori não informativa

Quando a verossimilhança é dominante ou deseja-se representar o desconhecimento sobre θ, a primeira ideia é pensar em todos os possíveis valores para θ como igualmente prová-

veis:

$$P(\theta) \propto k$$

em que θ é o parâmetro ou vetor de parâmetros limitado em um intervalo real, k uma constante qualquer.

Muitas vezes, a priori é a forma de se quantificar a incerteza sobre o experimento. Porém, a priori não-informativa não representa necessariamente o desconhecimento do pesquisador sobre o experimento, mas deve ser usada também de forma a viabilizar a inferência a posteriori (BOX; TIAO, 1973).

Utilizando uma generalização do teorema de Bayes, informações a priori sobre os parâmetros são utilizadas em associação com os dados amostrais (representados no teorema pela função de verossimilhança), possibilitando uma inferência a posteriori sobre aqueles (BOX; TIAO, 1973). Se existe informação a priori, a inferência Bayesiana pode determinar intervalos de credibilidade mais estreitos que os intervalos de Credibilidade, além de testes mais poderosos.

2.4.2.2 Distribuição a priori informativa

Quando o pesquisador possui informações prévias sobre o parâmetro desejado θ, utiliza-se de uma distribuição a priori informativa, ou seja, um modelo de probabilidade própria $p(\theta)$, por sua vez, especificada com o auxílio de constantes chamadas de hiperparâmetros (parâmetros da distribuição dos parâmetros). Inicialmente, os hiperparâmetros são considerados conhecidos e traduzem a informação que se tem sobre o parâmetro, antes da realização da amostra (TURKMAN et al., 2019).

A forma usual do Teorema de Bayes é:

$$p(\theta \mid Y) \propto p(Y \mid \theta)p(\theta).$$

Em outras palavras, o Teorema de Bayes diz que a distribuição de probabilidade de θ a posteriori para Y dado é proporcional ao produto da distribuição a priori para θ e a verossimilhança de θ dado Y (BOX; TIAO, 1973). Isto é:

Distribuição a posteriori $\propto$ verossimilhança $\times$ distribuição a priori.

Observando-se as quantidades y_i com n repetições independentes dado θ, ou seja, $p(y_i \mid \theta)$, segue que:

$$p(\theta \mid y_1, y_2, \ldots, y_n) \propto \prod_{i=1}^{n} p(y_i \mid \theta) p(\theta).$$

Obtida a distribuição a posteriori, é possível fazer inferências sobre as distribuições de probabilidade conjunta dos parâmetros. Na maioria dos casos, deseja-se encontrar uma distribuição para um parâmetro específico. A forma de encontrar tal distribuição, chamada distribuição marginal, está na integração da distribuição posteriori conjunta, obtendo uma função dessa integral para o parâmetro de interesse θ_i, isto é:

$$f(\theta_i \mid Y) = \int \cdots \int f(\theta \mid Y)\, d\theta_{(-i)},$$

em que $\theta_{(-i)}$ é o conjunto complementar de parâmetros para θ_i.

É comum em muitas situações a integração da equação ser inviável, ou até mesmo impossível, devido à complexidade da forma analítica das marginais.

Nessas situações, a amostragem Gibbs é uma boa alternativa para a obtenção de $f(\theta_i)$, método inicialmente utilizado por Gelfand e Smith (1990), na Estatística, e anteriormente elaborado por Geman e Geman (1984). Tal método consiste em um processo iterativo que gera valores convergentes para a distribuição marginal, sem que se conheça a sua expressão.

Além disso, ocorre que, se a distribuição condicional completa a posteriori não for uma densidade conhecida, podem-se utilizar outros métodos de simulação, tais como: a amostragem por importância (PAULINO et al., 2003), o *slice sampling* (NEAL, 2003), a amostragem por aceitação e rejeição e o Metropolis-Hastings (CHIB; GREENBERG, 1995; HASTINGS, 1970).

2.4.3 Algoritmo de amostragem

O método *Gibbs Sampling* (GS) é uma técnica de integração numérica por simulação, muito usual em situações nas quais a integração analítica completa é impossível. O GS é apli-

cável à estimação de componentes de variância e predição dos valores genéticos, permitindo, por suas propriedades, a inferência Bayesiana (CARNEIRO JÚNIOR et al., 2005).

A ideia do método de Monte Carlo via Cadeia de Markov (MCMC) consiste em simular um passeio aleatório no espaço do parâmetro θ, o qual converge para uma distribuição estacionária, que é a distribuição a posteriori $P(\theta \mid Y_v)$, em que Y_v é o vetor de observações.

Suponha-se que o vetor de parâmetros $\boldsymbol{\theta}$ seja dividido em p subvetores $p(\theta_1, \theta_2, \ldots, \theta_i)$ e que as distribuições condicionais de cada parâmetro θ_i dado todos os outros sejam conhecidas. Estas distribuições são chamadas distribuições condicionais completas e denotadas por:

$$p(\theta_1 \mid \theta_2, \theta_3, \ldots, \theta_p, Y),\ p(\theta_2 \mid \theta_1, \theta_3, \ldots, \theta_p, Y),\ \ldots,\ p(\theta_p \mid \theta_1, \theta_2, \ldots, \theta_{p-1}, Y)$$

O algoritmo de Gibbs pode ser descrito da seguinte maneira. Seja $\boldsymbol{\theta}$ o vetor de parâmetros de interesse e Y o vetor de dados:

a) defina $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_p^{(0)})$, um valor inicial arbitrário para o vetor de parâmetros $\boldsymbol{\theta}$, do qual $P(\boldsymbol{\theta}^{(0)} \mid Y) > 0$. Geralmente, amostramos de uma distribuição normal ou de uma distribuição especificada como a distribuição prévia de $\boldsymbol{\theta}$.

b) para $t = 1, 2, \ldots, L$, designe:

$$\begin{cases} \theta_1^{(t)} \sim p(\theta_1 \mid \theta_2^{(t-1)}, \theta_3^{(t-1)}, \ldots, \theta_p^{(t-1)}, Y) \\ \theta_2^{(t)} \sim p(\theta_2 \mid \theta_1^{(t)}, \theta_3^{(t-1)}, \ldots, \theta_p^{(t-1)}, Y) \\ \vdots \\ \theta_p^{(t)} \sim p(\theta_p \mid \theta_1^{(t)}, \theta_2^{(t)}, \ldots, \theta_{p-1}^{(t)}, Y) \end{cases}$$

c) mude o contador t para $t+1$ e retorne ao passo 2. O ciclo é repetido L vezes, produzindo assim $\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots, \boldsymbol{\theta}^{(L)}$.

Assumindo que a convergência foi atingida, à medida que o número t de iterações aumenta, a sequência de valores gerados se aproxima da distribuição de equilíbrio, assim, obtendo a densidade marginal desejada.

Gamerman (1996) afirma que, apesar de os resultados teóricos garantirem a convergência do Amostrador de Gibbs, sua utilização na prática pode ser dificultada em razão da eventual complexidade dos modelos. Essa complexidade faz com que a convergência do Amostrador de

Gibbs seja de difícil caracterização. A lentidão na convergência pode estar relacionada com a alta correlação entre os elementos da cadeia de um dado parâmetro.

Segundo Sorensen (1996), um ponto de partida para tal é a suposição de que a distribuição condicional dos dados y seja uma normal multivariada:

$$y \mid \beta, g, \sigma_e^2 \sim \mathcal{N}(X\beta + Zg, I\sigma_e^2)$$

A função de verossimilhança é dada por:

$$P(y \mid \beta, g, \sigma_g^2, \sigma_e^2) \propto \left(\frac{1}{\sigma_e^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma_e^2}(y - X\beta - Zg)'(y - X\beta - Zg)\right\}$$

3 MATERIAL E MÉTODOS

Nesta seção, descreve-se o procedimento de obtenção do experimento simulado e o material experimental que foi empregado no exemplo real. Apresenta-se também a metodologia da construção de gráficos de radar, a metodologia usada na amostragem para a inferencia bayesiana, os métodos de estimação dos componentes de variância e distribuições a posteriori condicionais completas para os parâmetros genéticos e fenotípicos. Por fim, descreve-se o procedimento de avaliação da metodologia de seleção proposta ao se usar os gráficos de radar.

3.1 Material de simulação

Foi feita uma simulação de um experimento em delineamento de blocos casualizados organizados em 20 tratamentos, 4 blocos e 80 unidades experimentais, usamos a inferência bayesiana com o objetivo de obter os valores dos parâmetros.

3.2 Material Experimental

Nos programas de melhoramento vegetal, é comum a análise de um grande número de tratamentos. Por essa razão, para este estudo foi usado como exemplo, um experimento em blocos incompletos organizados em alfa-látice 5×9 com três repetições utilizados na dissertação de Durães (2014) para Lavras.

3.3 Modelo da simulação do experimento

A variável resposta foi simulada na base de um vetor de tamanho n = 80. Todos os efeitos foram simulados em função do número de ordem do fator.

Para construir o vetor de efeito de tratamento foi utilizado o seguinte código R:

a) os efeitos de tratamentos foram simulados por:

```
t <- 20 : et <- qnorm((1:t/(t+1)),0,st))
```

onde: *et* - efeito de tratamento; *t* - número de tratamentos; *st* - desvio padrão dos tratamentos.

b) o mesmo tipo de simulação, controlando os efeitos médios, foi utilizado para os efeitos de blocos:

```
b <- 4 : eb <- qnorm((1:b/(b+1)),0,sb))
```

onde: *eb* - efeito de bloco; *b* - número de blocos; *sb* - desvio padrão dos blocos. Para a simulação, adotou-se $st = sb = 1$.

c) para construir o vetor de erro experimental, foi utilizado o seguinte código em R:

```
e  <- rnorm(n)*se
```

onde: *e* - erro experimental; *n* - número de unidades experimentais; *se* - desvio padrão do erro experimental.

Desta forma, garante-se que para os tratamentos e para os demais efeitos a distribuição têm sempre média zero e variância um,ou seja $\mathcal{N}(0, \sigma_G^2 = 1)$.

Os vetores simulados das variáveis resposta são dados por:

$$\boldsymbol{\eta} = \mathbb{E}[\boldsymbol{\eta} \mid \boldsymbol{\beta}, \boldsymbol{g}] + \boldsymbol{\varepsilon} \tag{3.1}$$

onde:

$\eta =$ Valor simulado;

$\boldsymbol{\beta} =$ vetor de efeitos de blocos;

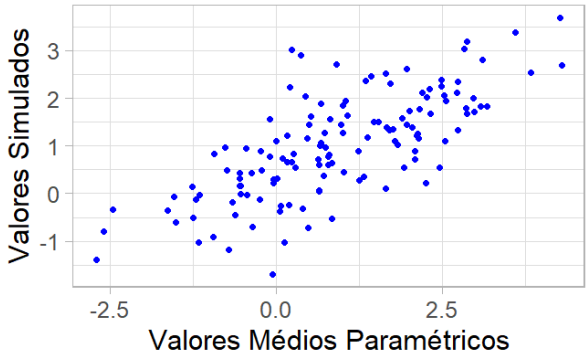
$\boldsymbol{g} =$ vetor de efeitos genéticos dos tratamentos;

$\boldsymbol{\varepsilon} =$ vetor de erros com distribuição normal padrão.

A correlação entre os valores paramétricos e os valores observados foi de $0,7539$.

A Figura 3.1 a seguir demonstra uma correlação positiva entre os valores simulados e os valores paramétricos de genótipos.

Figura 3.1 – gráfico de dispersão entre os valores simulados e sua esperança condicional aos valores de efeitos genéticos.



Fonte: Do Autor (2024).

A rotina (R) utilizada para simulação do experimento e obtenção da variável resposta simulada, encontrar-se-a nos Apêndices.

3.4 Modelo da análise e suas pressuposições

Conforme Lima e Bueno Filho (2006), consideremos o modelo individual univariado dado por :

$$\eta = \boldsymbol{X}\beta + Zg + \varepsilon \tag{3.2}$$

Onde:

η: é o vetor de dados, de ordem n.

$\boldsymbol{X}$: é a matriz $(n \times p)$ do delineamento dos blocos aplicada ao vetor de efeitos de blocos β.

β: é o vetor de dimensão $(p \times 1)$ das médias de blocos (tomados como efeitos fixos).

Z: é a matriz $(n \times q)$ de delineamento dos efeitos aleatórios de tratamentos (genótipos).

ε: é o vetor de resíduos $(n \times 1)$.

A partir das definições, os vetores definidos acima seguem as seguintes distribuições:

$$\eta \sim \mathcal{N}(X\beta + Zg, I\sigma_e^2)$$

$$g \sim \mathcal{N}(\phi, I\sigma_g^2)$$

$$\varepsilon \sim \mathcal{N}(\phi, I\sigma_e^2)$$

sendo que ϕ representa um vetor nulo de dimensão igual à do vetor considerado.

3.5 Metodologia empregada para a simulação dos dados

Para o banco de dados simulado,usou-se um experimento DBC. Os parâmetros estudados e distribuição a posteriori foram obtidos com o método de Monte Carlo via Cadeias de Markov (MCMC), com o amostrador de Gibbs. A análise computacional foi composta de 10 cadeias, cada uma com 4.000 iterações (as primeiras 1000 descartadas), com espaçamento dos pontos amostrados de 10 (thinning), e número de simulações igual a 1002 iterações salvas. Como o Gibbs sampler é um algoritmo iterativo, faz-se necessário a verificação de sua convergência pela estatística de Gelman-Rubin ($\hat{R}$) que mensura a razão da variância entre cadeias sobre a variância dentro de cadeias (GELMAN; RUBIN, 1992). Foram calculadas as médias para os genótipos considerando os métodos Bayesianos.

3.6 Estimação de Componentes de Variância e Valores Genéticos

Para a estimação das componentes de variância, optou-se em implementar algoritmos de amostragem Gibbs e inferência bayesiana. A análise foi realizada com a biblioteca `coda` e a função `mcmc()`.

3.7 Conjunto de parâmetros

O vetor dos parâmetros de interesse é dado por:

$$\theta = (\beta, g, \sigma_g^2, \sigma_e^2, \eta)$$

em que:

η é o vetor de dados, de ordem n;

β é o efeito de bloco;

g é o efeito de tratamento (genótipo); σ_g^2 é a variância genética; σ_e^2 é a variância residual;

3.8 Verossimilhança

A função de verossimilhança é dada por:

$$P(\eta \mid \beta, g, \sigma_g^2, \sigma_e^2) \propto \left(\frac{1}{\sigma_e^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma_e^2}(\eta - X\beta - Zg)'(\eta - X\beta - Zg)\right\}$$

3.9 Definição das distribuições a priori dos diferentes parâmetros do modelo

Admitindo que não haja informação prévia a respeito dos efeitos fixos, pode-se definir uma priori do tipo função constante no espaço paramétrico, o que é uma maneira de caracterizar total desconhecimento, segundo Sorensen (1996). Ou seja, $P(\cdot)$ constante.

Para as variâncias, pode-se considerar a distribuição χ^2 inversa escalada como a priori, diversificando os valores dos hiperparâmetros, pois são eles que vão fornecer a informação prévia que se tem em cada caso.

$$P(\sigma_g^2 \mid S_g^2, V_g) \propto \left(\frac{1}{\sigma_g^2}\right)^{\frac{V_g}{2}+1} \exp\left\{-\frac{V_g S_g^2}{2\sigma_g^2}\right\} \tag{3.3}$$

e

$$P(\sigma_e^2 \mid S^2, V) \propto \left(\frac{1}{\sigma_e^2}\right)^{\frac{V}{2}+1} \exp\left\{-\frac{VS^2}{2\sigma_e^2}\right\}, \tag{3.4}$$

onde σ_g^2 e σ_e^2 são as variâncias dos efeitos aleatórios e erro, respectivamente, S_g^2 e S^2 são os parâmetros escala da distribuição χ^2 inversa, e V_g e V são os graus de liberdade associados às distribuições χ^2 inversas.

Para Lima e Bueno Filho (2006) o uso de uma distribuição qui-quadrado inversa escalada não ocorreu por mero acaso e, sim, por razões matemáticas, já que ela conduz a uma integral de fácil resolução e a uma posteriori de natureza semelhante, ou seja, tem-se um caso de uma priori conjugada.

Para os efeitos aleatórios, empregou-se a distribuição normal, como se segue:

$$g \sim \mathcal{N}(\phi, A\sigma_g^2),$$

em que:

a) a distribuição $g \sim \mathcal{N}(\phi, A\sigma_g^2)$ pode ser reescrita como $g \sim \mathcal{N}(\phi, G)$, isto é, com $G = A\sigma_g^2$, onde $G^{-1} = \frac{A^{-1}}{\sigma_g^2}$ e, portanto, $A^{-1} = G^{-1}\sigma_g^2$.
b) o vetor ϕ é nulo e possui a mesma dimensão de g.
c) a matriz A representa a matriz de parentesco genético, cujos elementos descrevem os coeficientes de covariância entre os efeitos genéticos aleatórios. Neste caso, ela será considerada como a matriz identidade, ou seja, $A = I$.

3.9.1 Distribuição a posteriori conjunta

A distribuição a posteriori conjunta será definida conforme apresentado por Lima e Bueno Filho (2006). Portanto, a distribuição conjunta a posteriori, supondo-se independência entre os parâmetros, é dada pela seguinte expressão:

$$P(\beta, g, \sigma_e^2, \sigma_g^2 | y) \propto P(\beta)P(\sigma_g^2)P(\sigma_e^2)P(g|\sigma_g^2)P(y|\beta, g, \sigma_e^2, \sigma_g^2) \tag{3.5}$$

onde:

$P(y \mid \beta, g, \sigma_e^2, \sigma_g^2)$: verossimilhança das observações, dado o vetor de parâmetros;

$P(\beta)$: distribuição a priori dos efeitos fixos;

$P(\sigma_g^2)$: distribuição a priori da variância dos efeitos genéticos aditivos;

$P(\sigma_e^2)$: distribuição a priori da variância dos efeitos residuais;

$P(\mathbf{g} \mid \sigma_g^2)$: distribuição a priori do vetor de efeitos aleatórios genéticos.

A inferência sobre cada parâmetro é feita com as distribuições marginais.

A densidade posterior dos parâmetros é proporcional ao produto da densidade a priori e a função de verosimilhança. A distribuição a priori é construída com base na credibilidade acerca dos parâmetros ou a partir de resultados experimentais antecedentes. A verosimilhança é a função densidade de probabilidade das observações, dados os parâmetros desconhecidos (CAMPOS, 2013). Ou seja,

$$
\begin{aligned}
P(\beta, g, \sigma_g^2, \sigma_e^2 \mid y) \propto & \left(\frac{1}{\sigma_g^2}\right)^{q/2} \exp\left\{-\frac{1}{2\sigma_g^2}\left[g' A^{-1} g\right]\right\} \\
& \times (\sigma_g^2)^{-\left(\frac{\nu_g}{2}+1\right)} \exp\left\{-\frac{\nu_g s_g^2}{2\sigma_g^2}\right\} \\
& \times (\sigma_e^2)^{-\left(\frac{\nu}{2}+1\right)} \exp\left\{-\frac{\nu s^2}{2\sigma_e^2}\right\} \\
& \times \left(\frac{1}{\sigma_e^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma_e^2}\left[(y - X\beta - Zg)'(y - X\beta - Zg)\right]\right\}.
\end{aligned} \tag{3.6}
$$

3.9.2 Distribuições condicionais completas

A partir da equação 3.6, foram obtidas as distribuições condicionais a posteriori. A distribuição condicional completa para a componente da variância σ_e^2, é dada por:

$$
\begin{aligned}
P(\sigma_e^2 \mid \beta, g, \sigma_g^2, y) & \propto (\sigma_e^2)^{-\left(\frac{\nu}{2}+1\right)} \exp\left\{-\frac{\nu s^2}{2\sigma_e^2}\right\} \\
& \times \left(\frac{1}{\sigma_e^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma_e^2}\left[(y - X\beta - Zg)'(y - X\beta - Zg)\right]\right\} \\
& \propto (\sigma_e^2)^{-\left(\frac{n+\nu}{2}+1\right)} \exp\left\{-\frac{1}{2\sigma_e^2}\left[(y - X\beta - Zg)'(y - X\beta - Zg) + \nu s^2\right]\right\}
\end{aligned} \tag{3.7}
$$

ou seja, uma χ_{scal}^{-2} para a componente da variância, com $n+\nu$ graus de liberdade e parâmetro de escala dado por:

$$
\sigma_e^2 \mid (\beta, g, \sigma_g^2, y) \sim \chi^{-2}\left(n+\nu, \frac{(y - X\beta - Zg)'(y - X\beta - Zg) + \nu s^2}{n+\nu}\right)
$$

A variância genética também segue uma distribuição χ_{scal}^{-2}, dada por:

$$
\begin{aligned}
P(\sigma_g^2 \mid \beta, g, \sigma_e^2, y) &\propto \left(\frac{1}{\sigma_g^2}\right)^{q/2} \exp\left\{-\frac{1}{2\sigma_g^2}\left[g' A^{-1} g\right]\right\} \\
&\quad \times (\sigma_g^2)^{-\left(\frac{\nu_g}{2}+1\right)} \exp\left\{-\frac{\nu_g s_g^2}{2\sigma_g^2}\right\} \\
P(\sigma_g^2 \mid \beta, g, \sigma_e^2, y) &\propto (\sigma_g^2)^{-\left(\frac{q+\nu_g}{2}+1\right)} \exp\left\{-\frac{1}{2\sigma_g^2}\left(g' A^{-1} g + \nu_g s_g^2\right)\right\}
\end{aligned} \tag{3.8}
$$

ou seja,

$$
\sigma_g^2 \mid (\beta, g, \sigma_e^2, y) \sim \chi^{-2}\left(q+\nu_g, \frac{g' A^{-1} g + \nu_g s_g^2}{q+\nu_g}\right)
$$

Para o vetor β, tem-se:

$P(\beta \mid g, \sigma_g^2, \sigma_e^2, y)$

$$
\begin{aligned}
&\propto \exp\left\{-\frac{1}{2\sigma_e^2}(y - X\beta - Zg)'(y - X\beta - Zg)\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma_e^2}\left[(y-Zg)' - (X\beta)'\right]\left[(y-Zg) - X\beta\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma_e^2}\left[(y-Zg)'(y-Zg) - (y-Zg)'X\beta - \beta' X'(y-Zg) + \beta' X' X\beta\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma_e^2}\left[-(y-Zg)'X\beta - \beta' X'(y-Zg) + \beta' X' X\beta\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma_e^2}\left[\beta - (X'X)^{-1}X'(y-Zg)\right]'(X'X)\left[\beta - (X'X)^{-1}X'(y-Zg)\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma_e^2}\left[-(y-Zg)'X(X'X)^{-1}X'(y-Zg)\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma_e^2}\left[\beta - (X'X)^{-1}X'(y-Zg)\right]'(X'X)\left[\beta - (X'X)^{-1}X'(y-Zg)\right]\right\},
\end{aligned} \tag{3.9}
$$

portanto, temos:

$$
\beta \mid (g, \sigma_e^2, \sigma_g^2, y) \sim N\left((X'X)^{-1}X'(y-Zg), \sigma_e^2 (X'X)^{-1}\right).
$$

E, por último, para o vetor g, tem-se:

$$
P(g \mid \beta, \sigma_e^2, \sigma_g^2, y) \propto \exp\left\{-\frac{1}{2\sigma_e^2}\left(\frac{g' A^{-1} \sigma_e^2 g}{\sigma_g^2} + (y - X\beta - Zg)'(y - X\beta - Zg)\right)\right\}. \tag{3.10}
$$

Sendo $\delta = \frac{\sigma_e^2}{\sigma_g^2}$, tem-se:

$$
\begin{aligned}
&\propto \exp\left\{-\frac{1}{2\sigma_e^2}\left(g'\delta A^{-1}g + (y-X\beta-Zg)'(y-X\beta-Zg)\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma_e^2}\left(g'\delta A^{-1}g + \left[(y-X\beta)' - (Zg)'\right]\left[(y-X\beta)-(Zg)\right]\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma_e^2}\left(g'\delta A^{-1}g + (y-X\beta)'(y-X\beta) - (y-X\beta)'(Zg) - (Zg)'(y-X\beta) + (Zg)'(Zg)\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma_e^2}\left(g'\delta A^{-1}g - (y-X\beta)(Zg) - (Zg)'(y-X\beta) + (Zg)'(Zg)\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma_e^2}\left[g-(Z'Z+\delta A^{-1})^{-1}Z'(y-X\beta)\right]'(Z'Z+\delta A^{-1})\left[g-(Z'Z+\delta A^{-1})^{-1}Z'(y-X\beta)\right]\right\}.
\end{aligned}
\tag{3.11}
$$

Logo,

$$g \mid (\beta, \sigma_e^2, \sigma_g^2, y) \sim \mathcal{N}\left((Z'Z+\delta A^{-1})^{-1}Z'(y-X\beta), (Z'Z+\delta A^{-1})^{-1}\sigma_e^2\right)$$

3.10 Intervalos de Credibilidade e HPD's

Segundo Kinas (2008) uma alternativa ao intervalo de confiança clássico é o intervalo de credibilidade $[\theta_1, \theta_2]$, de tal forma que:

$$P(\theta_1 \leq \theta \leq \theta_2 | x) = 1 - \alpha.$$

Outra possibilidade é selecionar o subconjunto C de valores para θ com maiores probabilidades a posteriori (highest posterior density - HPD), tal que:

$$C = \{\theta \in \Theta : p(\theta|x) \geq k_\alpha\}$$

onde k_α é a maior constante que pode ser obtida, satisfazendo:

$$P(C|x) \geq 1 - \alpha.$$

Para distribuições unimodais e simétricas, o intervalo de credibilidade e o HPD são idênticos. Havendo assimetria, os dois critérios produzem intervalos diferentes; a diferença depende do grau de assimetria.

Se, no entanto, a distribuição a posteriori é multimodal, o conjunto C pode ser constituído de vários sub-intervalos em torno das modas mais destacadas. Neste caso, o HPD leva vantagem sobre o intervalo de credibilidade, fornecendo mais informações.

3.11 Correlação

O Coeficiente de Correlação de Pearson mede a intensidade e a direção da relação linear entre duas variáveis x e y (BRAVAIS, 1846; MARI; KOTZ, 2001; JOHNSON; BHATTACHARYYA, 2009) definido pela seguinte expressão :

$$\rho_P = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \tag{3.12}$$

onde (x_i, y_i) são pares de n observações, e $\bar{x}$ e $\bar{y}$ são as médias de x e y, respectivamente. O valor de ρ_P varia entre -1 e $+1$, sendo interpretado como:

a) $\rho_P \approx -1$: correlação negativa forte;
b) $\rho_P \approx +1$: correlação positiva forte;
c) $\rho_P \approx 0$: correlação fraca ou inexistente.

O coeficiente de Spearman é uma medida alternativa que substitui os valores das variáveis pelos seus postos. A fórmula é dada por (JOHNSON; BHATTACHARYYA, 2009):

$$\rho_S = \frac{\sum_{i=1}^{n}\left(R_i-\frac{n+1}{2}\right)\left(S_i-\frac{n+1}{2}\right)}{\frac{n(n^2-1)}{12}} \tag{3.13}$$

onde R_i e S_i são os postos das variáveis, e n é o número de observações. O coeficiente de Spearman também varia entre -1 e $+1$ e é apropriado para capturar relações *monotônicas*, mesmo que não sejam lineares.

3.12 Análises

O estudo foi conduzido para obter os valores dos parâmetros via inferência Bayesiana, as estimativas dos efeitos . Foi avaliada a precisão e o viés na estimação de parâmetros genéticos, como as componentes da variância, calculadas usando distribuições qui-quadradas com os graus

de liberdade estimados segundo a aproximação de Satterthwaite (SATTERTHWAITE, 1946), implementada na biblioteca `lmerTest` do `R`. Além disso, foram estimadas as herdabilidades na seleção entre parcelas:

$h^2 = \frac{\sigma_u^2}{\sigma_u^2+\sigma_e^2}$ e seleção entre médias de tratamentos, dada por $h_m^2 = \frac{\sigma_g^2}{\sigma_g^2+\frac{\sigma_e^2}{r}}$.

Foi considerado um delineamento experimental em Blocos Incompletos do tipo Alfa-látice .

3.12.1 Delineamento em Blocos Casualizados Completos (DBC)

Este delineamento foi adotado para garantir certa semelhança com o delineamento do experimnto real,que foi feito em blocos incompletos. No caso, foi simulado um DBC com 20 tratamentos em 4 blocos, com um total de 80 Unidades experimentais. O modelo linear utilizado para simular os efeitos e as realizações da variável observada foi:

$$\eta = \mu + \mathbf{Xb} + \mathbf{Zg} + \boldsymbol{\varepsilon} \tag{3.14}$$

sendo η o vetor das realizações da variável observada, μ é a média geral, sendo 5; $\mathbf{b}$ é o vetor dos efeitos fixos, com $\mathbf{b} \sim \mathcal{N}(0,1)$ e $\mathbf{X}$ é a matriz de delineamento dos efeitos fixos; $\mathbf{g}$ é o vetor dos efeitos aleatórios, simulado de uma distribuição normal, então $\mathbf{g} \sim \mathcal{N}(0,\sigma_g^2)$ e $\mathbf{Z}$ é a matriz de delineamento dos efeitos aleatórios. No caso do efeitos aleatórios, foi adotado para σ_g^2=1. Para o erro amostral ($\boldsymbol{\varepsilon}$) foi considerado que $\boldsymbol{\varepsilon} \sim \mathcal{N}(0,\sigma_e^2)$, onde $\sigma_e^2 = 1$. O vetor η gerado a partir de 3.14 é um vetor de uma variável aleatória contínua.

3.13 Amostragem da distribuição *a posteriori*

Após a obtenção de uma cadeia inicial (amostra piloto) de 4000 iterações MCMC, esta foi avaliada diagnóstico de Raftery e Lewis (RAFTERY; LEWIS, 1992) sendo estimado o descarte inicial necessário (*burn-in*) e a quebra da dependência da cadeia com os da amostragem intervalar (*jump*). Com estes valores se obteve outras cadeias com tamanho amostral 4000, para cada caractere, nas quais se realizou tanto o diagnóstico de Raftery e Lewis quanto o teste de Gelman e Rubin (GELMAN; RUBIN, 1992) para avaliar a convergência das cadeias.

3.14 Construção dos gráficos de radar

Os gráficos de radar foram desenvolvidos para facilitar a visualização de múltiplas variáveis simultaneamente, permitindo a identificação de padrões e diferenças entre genótipos . A presente metodologia baseia-se no uso do *Best Linear Unbiased Predictor* (BLUP), especificamente sua versão padronizada, denominada **t-BLUP**. Esse método possibilita a estimativa dos efeitos genéticos corrigidos para fatores ambientais, resultando em comparações mais precisas e interpretáveis.

O modelo BLUP utilizado pode ser expresso da seguinte forma:

$$y = X\beta + Zg + e, \tag{3.15}$$

onde:

y: vetor das observações fenotípicas;

X: matriz de delineamento dos efeitos fixos;

β: vetor de efeitos fixos;

Z: matriz de delineamento dos efeitos aleatórios;

g: vetor de efeitos genéticos aleatórios, estimados pelo BLUP;

e: vetor de erros aleatórios.

Os valores de g são padronizados para facilitar a comparação:

$$t\text{-BLUP} = \frac{g - \bar{g}}{\sigma_g}, \tag{3.16}$$

onde:

$\bar{g}$: média dos valores BLUP;

σ_g: desvio padrão dos valores BLUP.

A construção dos gráficos de radar teve como objetivo a visualização das variáveis em torno de um ponto central, com a possibilidade de calcular e exibir intervalos de credibilidade *(Highest Posterior Density, HPD)* para genótipos e médias gerais.

Inicialmente, foi determinado o número de variáveis a serem representadas, definindo, assim, o número de eixos do gráfico. Em seguida, os ângulos foram calculados de modo que os eixos permanecessem equidistantes. Todos os eixos partem do ponto central, com um espaçamento de 10% do tamanho do eixo em relação ao centro.

As observações dos genótipos são inseridas como polígonos no gráfico, convertendo as coordenadas cartesianas para coordenadas polares. Cada eixo do radar inicia-se no valor 0 e termina no valor 1. As médias, mínimos e máximos de cada variável são calculados para todos os genótipos, e os valores são normalizados para a escala do gráfico de radar.

Uma novidade proposta neste trabalho é a inclusão de uma região de confiança dentro do gráfico de radar, aplicável a um genótipo específico ou a grupos de genótipos. Para cada genótipo selecionado, calcula-se o polígono correspondente ao valor normalizado de cada variável, enquanto o intervalo de credibilidade (HPD) é calculado para o efeito médio das variáveis. Dessa forma, são gerados polígonos que representam os limites inferior e superior do HPD (CHEN; SHAO, 1999), permitindo a visualização da incerteza associada aos valores médios.

As figuras são geradas com as seguintes características:

a) os eixos são traçados a partir do ponto central, com linhas tracejadas e valores de referência para cada variável;
b) o polígono do genótipo selecionado é preenchido com uma cor semi-transparente;
c) a média geral das variáveis e o polígono médio do genótipo são representados com linhas contínuas;
d) os intervalos HPD são desenhados para o genótipo e para a média geral.

O gráfico de radar permite visualizar, de forma compacta e informativa, o desempenho de diferentes genótipos em várias variáveis, incluindo as incertezas expressas pelos intervalos HPD. A normalização dos valores e o ajuste das médias garantem uma visualização equilibrada, enquanto os parâmetros da função permitem ajustes conforme necessário.

Para validar a metodologia, foram simuladas cinco variáveis com distribuição normal, sendo que a primeira variável y_1 serviu como base para a geração das demais, com um grau decrescente de correlação.

O algoritmo para obtenção dessas variáveis é descrito a seguir: y_1: gerado como a soma de um termo médio (η) e um erro aleatório (e);

`y_1 <- round(eta + e, 1)` y_2: obtido a partir de y_1, adicionando uma pequena perturbação aleatória;

`y_2 <- round(y_1 + 0.1 * rnorm(n), 1)` y_3: combina pesos de y_1 e y_2, além de um ruído aleatório maior;

`y_3 <- round(0.7 * y_1 + 0.3 * y_2 + 0.5 * rnorm(n), 1)` y_4: derivado de y_3 com a adição de ruído;

`y_4 <- round(y_3 + 0.5 * rnorm(n), 1)` y_5: gerado a partir de uma média geral (μ) e um

termo de erro aleatório, sem correlação com as variáveis anteriores;

```
y_5 <- round(mu + rnorm(n), 1)
```

Os dados simulados encontram-se no Apêndice A, Tabela 7 .

3.15 Comparação dos gráficos de radar

Os gráficos de radar foram construidos padronizando-se as variáveis resposta usadas, que foram as médias das distribuições a posteriori (melhores preditores lineares) dos efeitos de genótipo. Estas médias foram apresentadas padronizadas pelo desvio padrão de cada efeito. Além disso se considerou apenas casos em que todas as variáveis estão em escalas do tipo quanto maior melhor (o que não traz perda de generalidade, pois pode ser facilmente invertido). O gráfico resultante deve ser do tipo quanto maior a área coberta pela figura, melhor o genótipo considerado (ou o grupo de genótipos que compõe o radar).

4 RESULTADOS E DISCUSSÃO

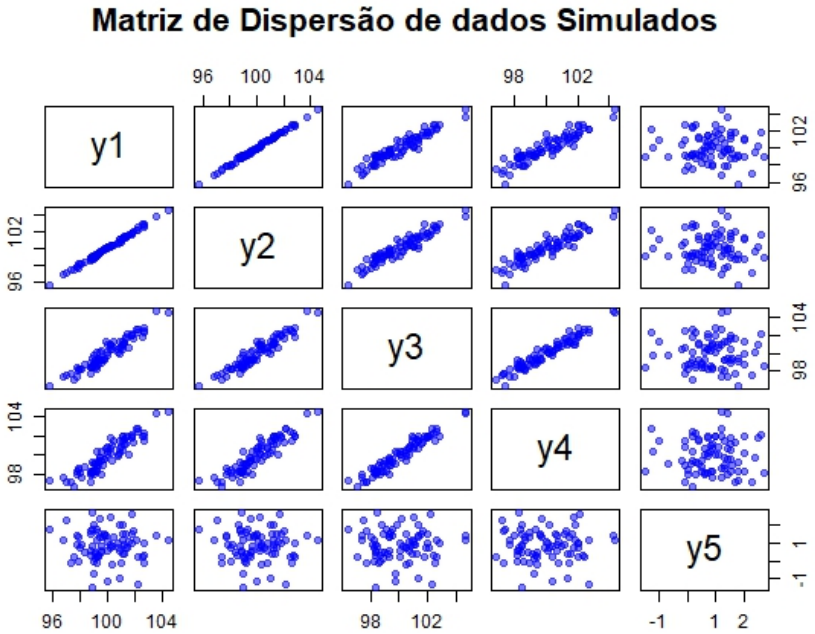
Nesta seção, serão apresentados e discutidos os resultados obtidos nas etapas de simulação e aplicação, com o objetivo de avaliar o desempenho da técnica de seleção de genótipos a partir dos diferentes critérios de seleção sob o método bayesiano.

4.1 Resultados das análises dos dados simulados

A Figura 4.1 exibe a relação entre cinco variáveis (y_1, y_2, y_3, y_4, y_5). Os gráficos de dispersão entre essas variáveis mostram relações lineares bem definidas, sugerindo uma correlação positiva forte. Isso indica que essas variáveis estão fortemente relacionadas. A variável y_5 apresenta um comportamento distinto,não há uma tendência clara nos gráficos de dispersão entre y_5 e as outras variáveis,os pontos estão distribuídos de forma aleatória, sugerindo ausência de correlação significativa.

As variáveis y_1, y_2, y_3, y_4 possuem uma forte relação linear, sugerindo uma dependência entre elas. A variável y_5 é independente das demais.

Figura 4.1 – matriz de dispersão dos dados simulados.



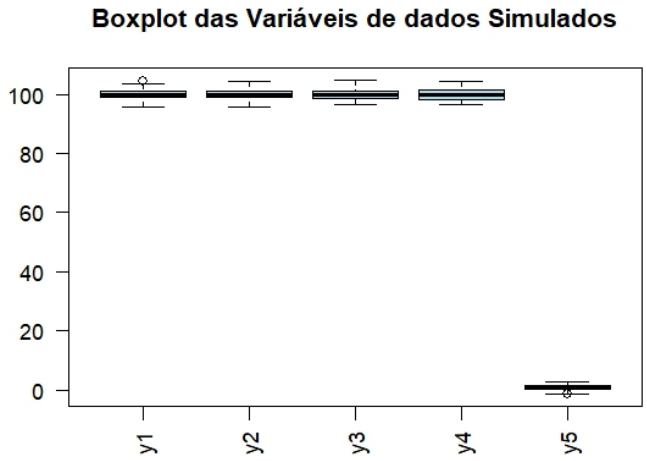
Fonte:Do Autor (2024).

A Figura 4.2 apresenta uma comparação de cinco variáveis y_1, y_2, y_3, y_4, y_5 em que a distribuição de y_1 a y_4, as medianas estão próximas de 100, as caixas são estreitas, indicando baixa variabilidade,existem alguns outliers acima dos bigodes superiores.

y_5 é significativamente diferente,sua mediana está próxima de 0, a sua dispersão é muito menor em comparação com as outras variáveis,existem outliers abaixo do bigode inferior.

As variáveis y_1 a y_4 possuem valores semelhantes, enquanto y_5 tem uma distribuição muito diferente.

Figura 4.2 – boxplot dos dados simulados.



Fonte: Do Autor (2024).

Com as componentes da variância obtidas, é possível calcular as estimativas de herdabilidade para dois cenários: seleção entre parcelas, e seleção entre médias de tratamentos.

Os resumos dos parâmetros genéticos são apresentados na Tabela 4.1 onde se observa que o experimento estimou adequadamente as componentes da variância genética e do erro experimental, bem como as herdabilidades.

Tabela 4.1 – Estimativas das componentes de variância e herdabilidades dos valores simulados, com seus respectivos intervalos de Credibilidade (IC) de 95% e valores paramétricos.

Parâmetro	**Valor Paramétrico**	**Estimativa**	**LI**	**LS**
σ_g^2	1	1.04	0.18	2.18
σ_e^2	1	0.84	0.41	1.27
h^2	0.5	0.52	0.26	0.80
h_m^2	0.75	0.75	0.51	0.92

Fonte: Do Autor (2024).

A Tabela 4.1 apresenta estimativas de parâmetros de variância e herdabilidade obtidos por análise Bayesiana, junto com seus respectivos intervalos de credibilidade (IC) de 95%. Os ICs foram calculados a partir dos quantis de 2,5% e 97,5% das distribuições a posteriori de cada parâmetro, refletindo a incerteza associada às estimativas. Verifica-se estimativas próximas aos valores paramétricos, com intervalos de credibilidade que refletem as incertezas associadas às

distribuições a posteriori. O parâmetro com maior incerteza foi σ_g^2 e a de menor incerteza foi h^2, enquanto h_m^2 apresentou a estimativa mais precisa e confiável.

Na Tabela 4.2 pode ser visualizado um resumo das Correlaçoes de Pearson e Spearman com seus respectivos Intervalo de Credibilidade (95%).

Tabela 4.2 – Correlação de Pearson e Spearman dos efeitos de tratamentos com seus respectivos efeitos paramétricos. Intervalos de Credibilidade (95%).

Método	**Correlação (%)**	LI	LS
Pearson	76.17	75.64	76.69
Spearman	75.39	74.85	75.92

Fonte: Do Autor (2024).

A correlação de Pearson sugere uma forte correlação linear entre os efeitos dos tratamentos, indicando que a relação entre as variáveis segue um padrão linear bem definido como também a correlação de Spearman indica uma forte correlação monotônica entre os efeitos dos tratamentos, ou seja, os efeitos dos tratamentos seguem uma ordem consistente, mas não necessariamente linear.

A Tabela 4.3 mostra correlações genotípicas elevadas entre a maioria dos parâmetros. Destacando-se as correlações positivas $\rho\{1,2\} = 0,99$ e $\rho\{2,3\} = 0,85$ e correlações baixas $\rho\{i,5\}$ (com o vetor $\mathbf{g_5}$). Isto permitirá discutir o efeito do processo seletivo em especial em $\mathbf{g_1}$ e verificar o que ocorre com a ausência de indicação em $\mathbf{g_5}$.

Tabela 4.3 – Correlações genotípicas entre os pares i, j de vetores de valores genéticos simulados para cada caractere $\rho\{\mathbf{g_i}, \mathbf{g_j}\}$.

	$\mathbf{g_1}$	$\mathbf{g_2}$	$\mathbf{g_3}$	$\mathbf{g_4}$	$\mathbf{g_5}$
$\mathbf{g_1}$	1.00	0.99	0.81	0.74	0.15
$\mathbf{g_2}$	–	1.00	0.85	0.76	0.08
$\mathbf{g_3}$	–	–	1.00	0.82	0.17
$\mathbf{g_4}$	–	–	–	1.00	0.01
$\mathbf{g_5}$	–	–	–	–	1.00

Fonte: Do Autor (2024).

A Tabela 4.4 apresenta uma maior variabilidade nas correlações entre os erros simulados, observando-se correlação forte entre $\mathbf{e_3}$ e $\mathbf{e_4}$ ($r = 0,82$), correlação moderada a baixa entre $\mathbf{e_1}$ e $\mathbf{e_3}$ ($r = 0,52$) e Correlação positiva forte entre $\mathbf{e_1}$ e $\mathbf{e_2}$ ($r = 0,98$) apresentam uma relação quase perfeita, sugerindo que possuem uma relação causal ou estrutural entre si e correlação negativa fraca entre $\mathbf{e_1}$ e $\mathbf{e_5}$ ($r = -0,10$), sugerindo uma possível tendência inversa .

Tabela 4.4 – Correlações dos vetores de erros simulados entre os pares i, j de vetores de valores genéticos simulados para cada caractere $\rho\{\mathbf{e_i}, \mathbf{e_j}\}$.

	$\mathbf{e_1}$	$\mathbf{e_2}$	$\mathbf{e_3}$	$\mathbf{e_4}$	$\mathbf{e_5}$
$\mathbf{e_1}$	1.00	0.98	0.52	0.30	-0.10
$\mathbf{e_2}$	–	1.00	0.53	0.31	-0.07
$\mathbf{e_3}$	–	–	1.00	0.82	0.04
$\mathbf{e_4}$	–	–	–	1.00	0.06
$\mathbf{e_5}$	–	–	–	–	1.00

Fonte: Do Autor (2024).

Ao comparar as correlações genotípicas com as dos erros simulados, observa-se que as correlações genotípicas são, em geral, mais elevadas, indicando uma forte coassociação entre os parâmetros genéticos. Isso sugere que fatores genéticos apresentam relações consistentes e previsíveis, o que pode ser essencial para a compreensão da hereditariedade e da transmissão de características entre gerações.

Por outro lado, as correlações dos erros simulados apresentam maior dispersão, com algumas correlações negativas e várias moderadas, evidenciando uma maior variabilidade aleatória. Essa dispersão pode ser atribuída a fatores externos ou ao ruído nos dados, dificultando a identificação de padrões claros e aumentando a incerteza nas estimativas derivadas dos modelos simulados.

A alta correlação genotípica indica uma dependência genética estruturada, o que é relevante para estratégias de seleção e melhoramento genético. Já a variabilidade observada nas correlações dos erros alerta para a necessidade de modelos mais robustos, capazes de lidar com essa dispersão e reduzir o impacto da variabilidade aleatória nos dados simulados.

Tabela 4.5 – Medidas de avaliação dos efeitos aleatórios de genótipos simulados.

Medidas	Viés	EQM	Pearson	Spearman
Valor	-3,47	9,47	0,80	0,82

Fonte: Do Autor (2024).

A Tabela 4.5 apresenta as medidas de avaliação dos efeitos aleatórios de genótipos simulados, incluindo o Viés, o EQM e as correlações de Pearson e Spearman.

Observa-se na Tabela 4.5 que o viés de $-3{,}47$ indica uma subestimação moderada nas estimativas em relação aos valores esperados, sugerindo uma tendência sistemática de o modelo gerar estimativas ligeramente menores do que o real.

O erro quadrático médio (EQM) de 9,47 revela uma variabilidade razoável nas predições, indicando uma precisão relativamente baixa nas estimativas, o que destaca a presença de erros consideraveis nas predições.

As correlações de Pearson $(0{,}80)$ e Spearman $(0{,}82)$ mostram uma forte relação entre os valores estimados e observados. A correlação de Pearson sugere uma relação linear significativa entre as estimativas e os valores verdadeiros. Já a correlação de Spearman, ligeiramente maior, indica uma forte concordância monotônica, mesmo com possíveis desvios da linearidade.

Esses resultados sugerem que, apesar do viés e da presença de erros, o modelo consegue capturar bem as tendências gerais dos efeitos aleatórios simulados, indicando uma boa associação entre os valores estimados e observados.

Na Tabela 4.6, encontram-se os resumos das Médias, desvio padrão e intervalos de credibilidade para a distribuição condicional a posteriori dos efeitos de genótipos, dadas as observações em cada uma das cinco variáveis simuladas $\boldsymbol{y_i}$.

As estimativas estão ordenadas em ordem crescente das Médias, destacamos os grupos de valores mais altos e mais baixos, onde verificaríamos a eventual correção da seleção (genótipos 1 a 4 deveriam estar no primeiro grupos destacado e genótipos 17 a 20 no segundo). Verifica-se pequenos desvios de estimativas (o que indicaria potencial inacurácia seletiva), mas em geral esta variável tem resposta bastante consistente com os valores paramétricos simulados.

Tabela 4.6 – Médias, desvio padrão e intervalos de credibilidade para a distribuição condicional a posteriori dos efeitos de genótipos, dadas as observações em cada uma das cinco variáveis simuladas $\mathbf{y_i}$.

Var.	Genótipo	$\bar{\mathbf{g}}_\mathbf{i}\|\mathbf{y_i}$	Dp	LI	LS
	1	-1.29	0.50	-2.24	-0.30
	7	-1.08	0.50	-2.03	-0.08
	5	-0.85	0.48	-1.79	0.08
y_1	⋮	⋮	⋮	⋮	⋮
	17	**0.74**	**0.48**	**0.18**	**1.66**
	15	**0.76**	**0.48**	**0.16**	**1.71**
	20	**0.85**	**0.48**	**0.04**	**1.86**
	1	-1.25	0.50	-2.23	-0.30
	7	-1.22	0.49	-2.18	-0.26
	2	-0.84	0.49	-1.85	0.05
y_2	⋮	⋮	⋮	⋮	⋮
	17	**0.77**	**0.48**	**-0.14**	**1.73**
	16	**0.81**	**0.48**	**-0.12**	**1.75**
	20	**0.83**	**0.48**	**-0.08**	**1.81**
	1	-1.63	0.55	-2.67	-0.53
	4	-1.30	0.54	-2.32	-0.23
	7	-1.06	0.54	-2.07	0.04
y_3	⋮	⋮	⋮	⋮	⋮
	15	**0.84**	**0.53**	**-0.17**	**1.93**
	17	**0.98**	**0.54**	**-0.09**	**2.03**
	20	**1.03**	**0.53**	**-0.00**	**2.06**
	2	-1.79	0.53	-2.80	-0.76
	1	-0.89	0.49	-1.87	0.07
	7	-0.76	0.48	-1.66	0.22
y_4	⋮	⋮	⋮	⋮	⋮
	18	**1.03**	0.49	0.07	1.95
	20	**1.03**	0.49	0.08	2.01
	3	**0.88**	0.49	-0.04	1.88
	9	-1.45	0.53	-2.54	-0.47
	1	-1.01	0.53	-2.02	0.06
	11	-1.05	0.53	-2.16	-0.04
y_5	⋮	⋮	⋮	⋮	⋮
	7	**1.97**	**0.55**	**-0.89**	**3.03**
	20	**1.80**	**0.55**	**-0.75**	**2.86**
	13	**1.16**	**0.54**	**-0.02**	**2.16**

Fonte:Do Autor (2024).

É importante destacar que a distribuiçção a posteriori dos valores genotípicos (resumos em **negrito** na Tabela 4.6) corroboram o esperado pela simulação. Para a variável **y1** apresentaram resultados com valor nulo improvável (IC não contém o zero) e para a variável **y5** o zero é provável (está no centro do IC). Dessa forma, independente dos objetivos da seleção envolverem quaisquer das variáveis, os IC estão coerentes com o que foi simulado.

Na seção que se segue produziremos resumos gráficos destes resultados.

4.1.1 Comparação de gráficos de radar simulados para cada dois genótipos

No caso da ilustração com o exemplo simulado, o que se pretende é verificar a possibilidade do radar servir como auxílio sinóptico para a seleção, de forma que se empregue critérios que levem em conta os vários caracteres em estudo simultaneamente. Isto porque, neste caso, conhecemos todos os valores reais.

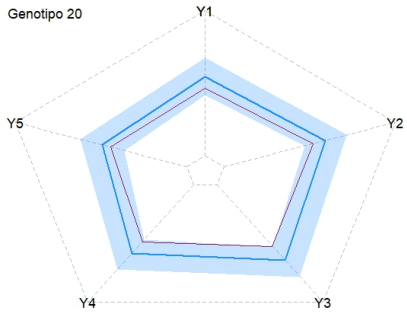
A Figura 4.3 mostra os gráficos de radar simulados para dois genótipos com "bola cheia" e "bola murcha" com base nas cinco características. Neste caso, quanto maior a média de cada genótipo (linha de cor azul), melhor é o desempenho do genótipo em relação às características consideradas, em comparação com a média geral dos genótipos (linha de cor violeta).

A linha violeta representa a média geral dos 20 genótipos para cada característica. A linha azul liga a média a posteriori do genótipo para cada característica, ou seja, é o estimador pontual do gráfico de radar específico do genótipo. A região de cor azul representa os intervalos HPD deste radar para cada genótipo.

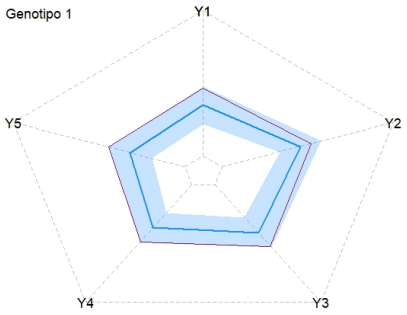
Nesta figura, obeserva-se que o genótipo 20 é "bola cheia", sendo superior à média para a maior parte das características, o genótipo 1 é "bola murcha", sendo inferior para a maior parte das características, os genótipos 12 e 4 têm resultados indefinidos, com manifestações intermediárias para várias das caracteristicas e discrepância disto para $\mathbf{y_5}$.

Estes resultados ilustram com precisão o que era esperado da simulação e evidenciam o potencial dos radares de visualizar diretamente a classificação geral ou dos genótipos em relação aos vários caracteres.

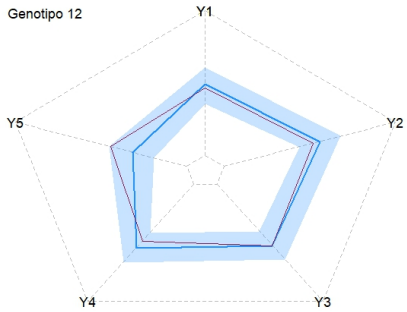
Figura 4.3 – gráficos de radar ilustrados para quatro dos genótipos simulados. A região hachurada em azul é construida com os limites inferiores e superiores de intervalos com 95% de credibilidade. Os genótipos têm valores ordenados do "pior" para o "melhor" (1 a 20), com correlações positivas decrescentes a partir do caráter 1 e correlação nula com o caráter 5.



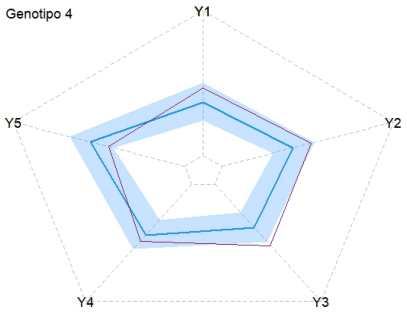
(a) Genótipo 20



(c) Genótipo 1



(b) Genótipo 12



(d) Genótipo 4

Fonte:Do Autor (2024).

4.1.2 Comparação de gráficos entre dois diferentes genótipos de dados simulados

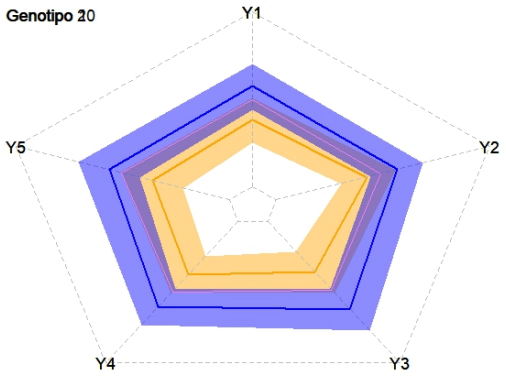
Na Figura 4.4, faz-se a comparação de dois genótipos na base do desempenho médio das suas características, em qual das cores de HPD há um desempenho melhor entre eles.

Observa-se que o genótipo 20 apresenta desempenho superior à média geral em todas as características, evidenciando seu status de "bola cheia". Por outro lado, o genótipo 1 é consistentemente inferior à média, sendo classificado como "bola murcha". Além disso, os genótipos 16 e 15 apresentam comportamentos contrastantes em características específicas.

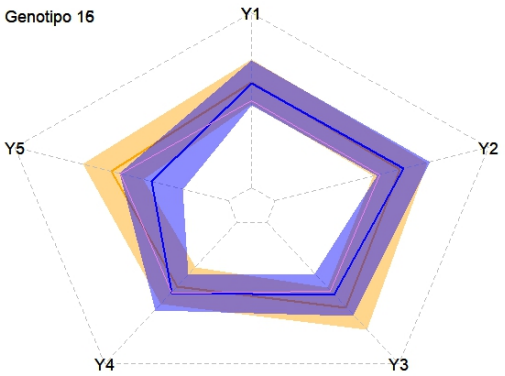
Para a característica $\mathbf{y_1}$, o genótipo 16 se destaca, superando tanto o genótipo 15 quanto a média geral. Já na característica $\mathbf{y_5}$, o genótipo 15 mostra desempenho superior ao genótipo

16 e à média geral. Nas demais características, os dois genótipos têm desempenhos similares, sem diferenças significativas.

Figura 4.4 – gráficos de dois diferentes genótipos onde a linha preta representa a média geral dos 20 genótipos para cada característica. Linha Azul: a média do genótipo 20 e 16. Cor Azul: HPD do genótipo 20 e 16 para cada característica. Linha laranja: a média do genótipo 1 e 15, enquanto que a cor laranja representa o HPD dos genótipos 1 e 15 para cada característica, respectivamente.



(a) Genótipos 20 vs 1



(b) Genótipos 16 vs 15

Fonte:Do Autor (2024).

É possível observar na Figura 4.4 que, para a cor azul, os genótipos 20 e 16 apresentaram melhores desempenhos do que os genótipos 1 e 15.

Esses resultados destacam como os gráficos de radar são eficazes para identificar e visualizar diretamente diferenças de desempenho, ilustrando tanto a classificação geral dos genótipos quanto os detalhes específicos de suas forças e fraquezas em relação às várias características avaliadas.

As sobreposições nos gráficos indicam similaridade no desempenho dos genótipos para determinadas características. Se a sobreposição for grande, significa que os genótipos possuem desempenhos similares, enquanto que pouca sobreposição indica diferenças significativas. No exemplo, os genótipos 20 e 16 (azul) teem um desempenho superior aos genótipos 1 e 15 (laranja) na maioria das características.

4.2 Aplicação

Os dados analisados nesta pesquisa para obtenção das estimativas de parâmetros foram extraídas da Dissertação de Durães (2014), com o título Heterose em Sorgo Sacarino, que tinha como objetivo avaliar o desempenho de linhagens e híbridos de sorgo sacarino e, especificamente, verificar e analisar o efeito heterótico relativo aos principais caracteres agronômicos e tecnológicos.

Foram avaliados 45 linhagens ou híbridos de sorgo sacarino, sendo 10 linhagens restauradoras de fertilidade (R) – machos de porte alto e alto teor de açúcares, três linhagens macho-estéreis (A) – fêmeas de porte baixo e baixo teor de açúcares (Tabela 4.7), 30 híbridos resultantes do cruzamento dialélico parcial entre as linhagens A e R, além de dois híbridos comerciais da empresa privada Monsanto.

Tabela 4.7 – Identificação das Linhagens de Sorgo Sacarino.

Linhagens	Pedigree
1	BR 007 A
2	BR 008 A
3	CMSXS222A
4	BR 500 R
5	BR 501 R
6	BR 504 R
7	BR 505 R
8	CMSXS 633 R
9	CMSXS 634 R
10	CMSXS 642 R
11	CMSXS 36643 R
12	CMSXS 644 R
13	CMSXS 647 R

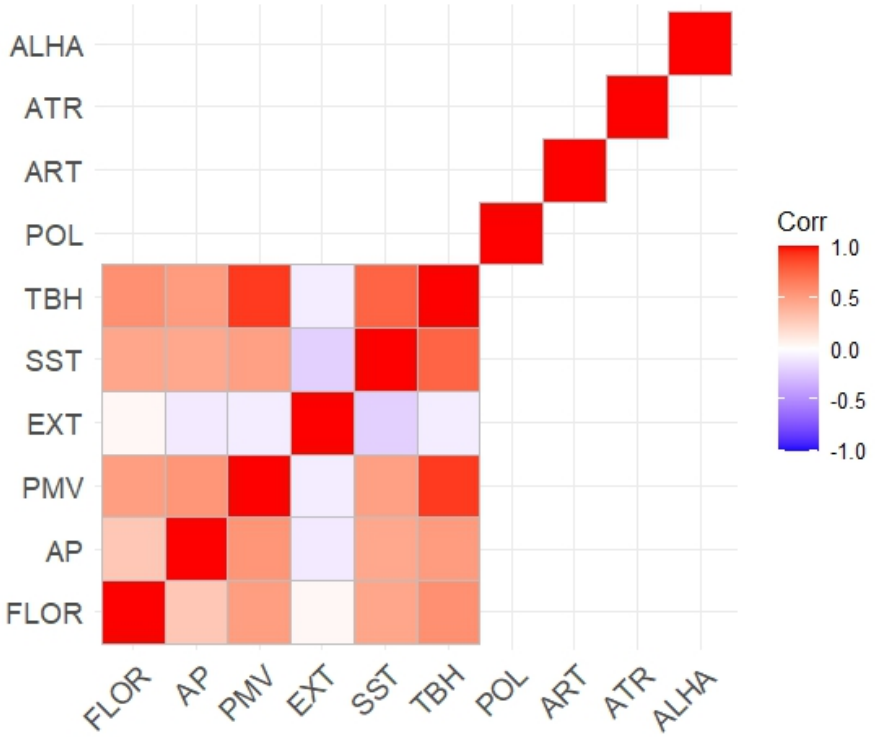
Fonte: Durães (2014).

Foram avaliados no total dez caracteres, sendo estes divididos em agronômicos, Florescimento (FLOR), Altura de planta (AP), Produção de massa verde (PMV); avaliados por ocasião da colheita, e tecnológicos,Extração de caldo (EXT),Sólidos solúveis totais % caldo (SST), Tonelada de Brix por hectare (TBH), Sacarose % colmo (POLc), Açúcares redutores totais % colmo (ARTc), Açúcares totais recuperáveis (ATR), Produção de álcool em litros por hectare (ALPHA) mensurados no Laboratório de Análises Tecnológicas da Embrapa Milho e Sorgo (DURÃES, 2014).

Na figura 4.5 apresenta uma matriz de correlação em que tons de vermelho que indicam correlações positivas, ou seja, quanto mais intenso, mais forte é a relação entre as variáveis. Enquanto tons de azul representam correlações negativas, o que significa que, quando uma variável aumenta, a outra tende a diminuir. Tons próximos ao branco indicam correlações fracas ou inexistentes, sugerindo pouca ou nenhuma relação entre as variáveis analisadas.

Observa-se que há fortes correlações positivas entre algumas variáveis, como PMV, SST, TBH e POL, indicando que essas características podem estar interligadas no crescimento ou na produção do sorgo. A variável FLOR apresenta correlações moderadas com algumas outras, o que pode sugerir que o florescimento impacta a produtividade. Por outro lado, variáveis como EXT e ART mostram correlações mais fracas com as demais, sugerindo menor dependência entre elas.

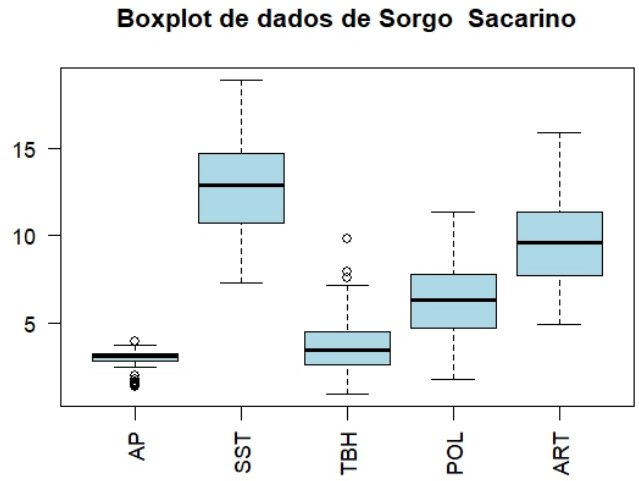
Figura 4.5 – matriz de correlação dos dados de Sorgo Sacarino.



Fonte: Do Autor (2024).

A Figura 4.6 apresenta um boxplot que ilustra a distribuição das linhagens para os caracteres SST, ART, AP, TBH e POL. Observa-se que as linhagens tendem a apresentar valores mais elevados para SST e ART, enquanto para o caráter AP os valores são mais baixos. Os caracteres TBH e POL exibem valores intermediários. A variabilidade dos dados entre os caracteres também difere, com o caráter SST apresentando a maior dispersão. A presença de outliers nos caracteres AP e TBH indica a ocorrência de valores extremos nesses grupos, o que sugere a existência de observações atípicas nas distribuições dessas variáveis.

Figura 4.6 – boxplot dos dados de Sorgo Sacarino.



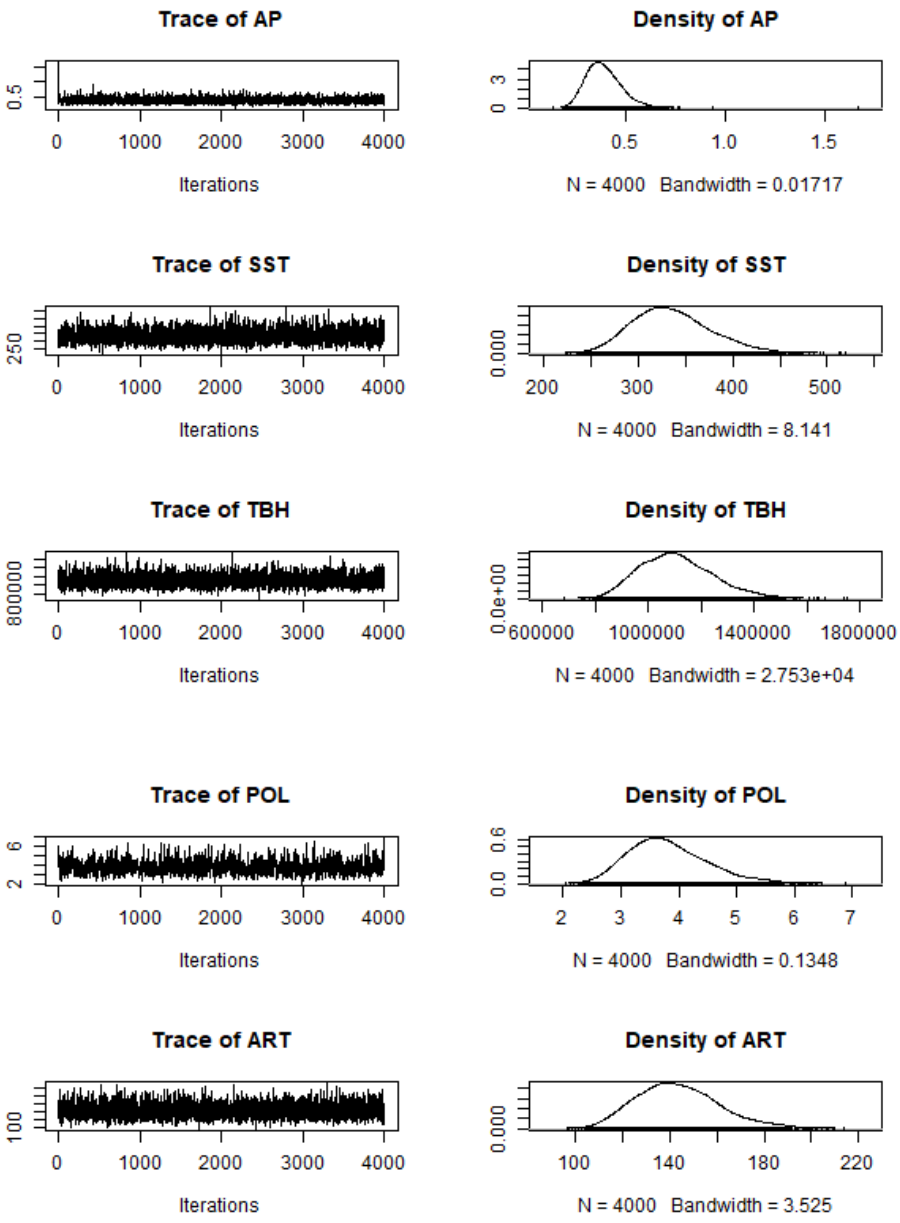
Fonte: Do Autor (2024).

4.2.1 Convergência e efeitos principais genótipos

A amostragem para o modelo deste trabalho foi realizada por meio do algoritmo de *Gibbs*. Os resultados obtidos pelos critério de Raftery e Lewis (1992) indicaram boas propriedades de convergência para todos os parâmetros. Além disso, todos os parâmetros passaram no teste de estacionariedade, indicando que a convergência foi alcançada.

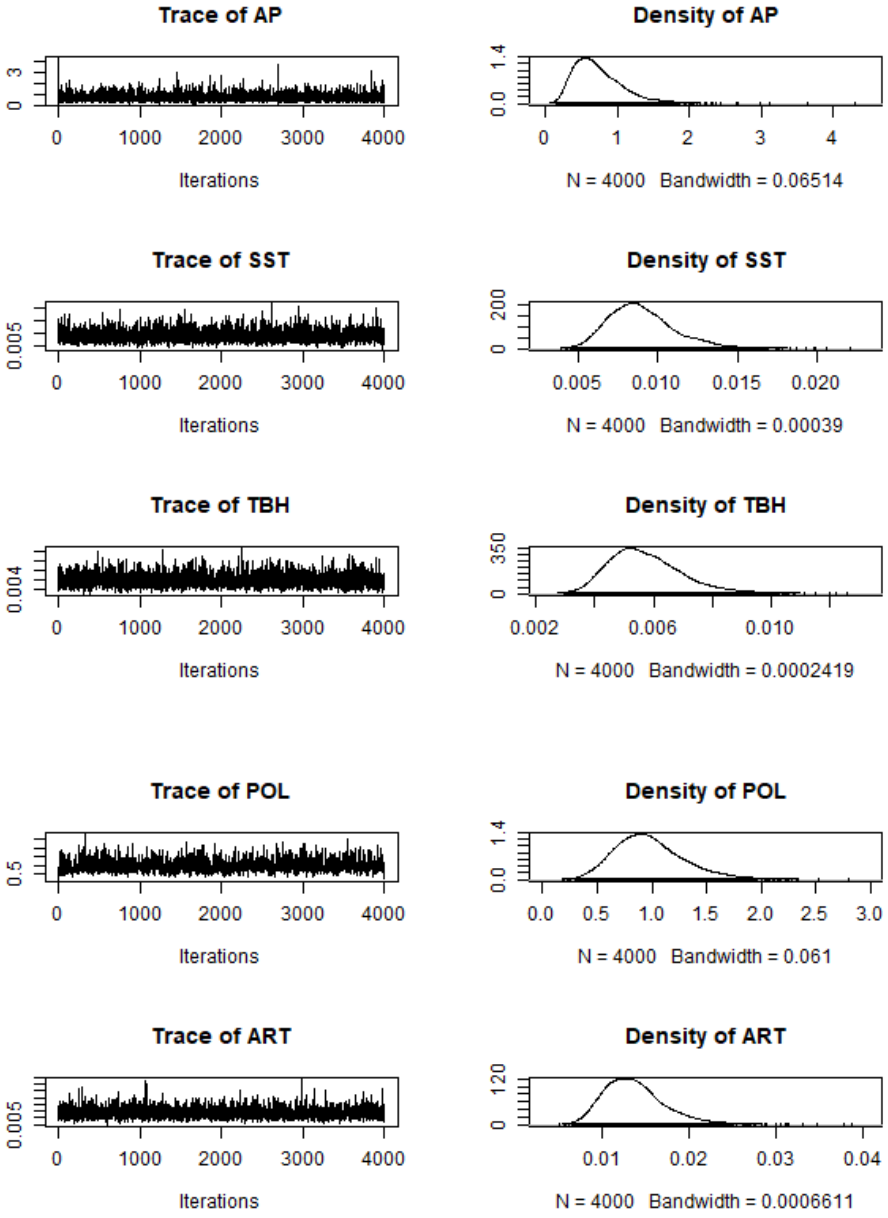
As Figuras 4.7 e 4.8 mostram os traços e densidades das cadeias para os componentes de variância das variáveis de interresse, que como pode ser observado apresenta um padrão de dispersão aparentemente uniforme em torno de um valor médio, corroborando os resultados dos critérios de convergência utilizados.

Figura 4.7 – traço e densidades das cadeias para os componentes de variância do genotipo.



Fonte: Do Autor (2024).

Figura 4.8 – traço e densidades das cadeias para os componentes de variância do Erro Experimental.



Fonte:Do Autor (2024).

Na Tabela 4.8 observam-se os resumos das estimativas (Médias, desvio padrão e intervalos de credibilidade para a distribuição condicional a posteriori dos efeitos de genótipos) dos 4 piores e 4 melhores efeitos de tratamentos da análise para cada característica dos genótipos. Os efeitos estão ordenados em ordem crescente das Médias, de modo que valores baixos estejam no primeiro grupo (especificamente, de 1 a 4) e valores altos no segundo grupo (especifica-

mente, de 42 a 45). Verifica-se que o genótipo CMSXS634R sempre se encontra em lugares de destaque em todas asa caracteristicas.

Tabela 4.8 – Médias, desvio padrão e intervalos de credibilidade para a distribuição condicional a posteriori dos efeitos de genótipos, dadas as observações em cada uma das cinco caracterstica (Continua)

Var.	Genótipo	$\bar{g}_i \mid y_i$	Dp	LI	LS
	008AxCMSXS647R	-1.18	0.58	-2.29	-0.03
	222AxCMSXS633R	-1.31	0.58	-2.49	-0.22
	008AxCMSXS634R	-0.85	0.58	-2.00	0.23
	007A x BR505R	-0.08	0.57	-1.22	0.98
AP	⋮	⋮	⋮	⋮	⋮
	007AxCMSXS647R	**0.75**	**0.57**	**-0.32**	**1.94**
	222AxBR500R	**1.47**	**0.58**	**0.37**	**2.60**
	222AxCMSXS642R	**2.12**	**0.58**	**0.99**	**3.29**
	CMSXS634R	**3.44**	**0.62**	**2.21**	**4.65**
	222AxCMSXS633R	-2.10	0.57	-3.18	-0.97
	CMSXS647R	-1.97	0.56	-3.12	-0.90
	008AxCMSXS634R	-1.81	0.58	-2.97	-0.72
	222AxBR504R	-1.19	0.55	-2.30	-0.15
SST	⋮	⋮	⋮	⋮	⋮
	007AxCMSXS647R	**0.75**	**0.57**	**-0.32**	**1.94**
	222AxBR500R	**1.47**	**0.58**	**0.37**	**2.60**
	222AxCMSXS642R	**2.12**	**0.58**	**0.99**	**3.29**
	CMSXS634R	**3.44**	**0.62**	**2.21**	**4.65**
	CMSXS647R	-1.87	0.54	-2.93	-0.81
	222AxBR504R	-1.59	0.51	-2.62	-0.63
	008AxCMSXS634R	-1.33	0.48	-2.25	-0.39
	007AxCMSXS647R	-1.31	0.48	-2.25	-0.39
TBH	⋮	⋮	⋮	⋮	⋮
	BR501R	**1.21**	**0.46**	**0.32**	**2.11**
	007A x BR501R	**1.22**	**0.48**	**0.28**	**2.11**
	CMSXS634R	**1.69**	**0.53**	**0.67**	**2.74**
	007AxCMSXS642R	**2.48**	**0.62**	**1.25**	**3.69**
	CMSXS647R	-1.55	0.52	-2.54	-0.56
	008AxCMSXS642R	-1.34	0.49	-2.28	-0.37
	222AxBR504R	-1.09	0.45	-1.98	-0.19
	008AxCMSXS633R	-0.94	0.43	-1.80	-0.13
ART	⋮	⋮	⋮	⋮	⋮
	008AxCMSXS634R	**1.38**	**0.57**	**0.34**	**2.50**
	CMSXS642R	**1.51**	**0.53**	**0.53**	**2.61**
	007AxBR501R	**1.53**	**0.52**	**0.57**	**2.56**
	222AxBR505R	**2.49**	**0.70**	**1.10**	**3.82**

Tabela 4.8– Médias, desvio padrão e intervalos de credibilidade para a distribuição condicional a posteriori dos efeitos de genótipos, dadas as observações em cada uma das cinco características (conclusão)

Var.	Genótipo	$\bar{g}_i \vert y_i$	Dp	LI	LS
	CMSXS647R	-1.87	0.53	-2.96	-0.89
	222AxBR504R	-1.58	0.52	-2.61	-0.56
	008AxCMSXS642R	-1.23	0.48	-2.20	-0.30
	008AxCMSXS633R	-0.86	0.48	-1.78	0.08
POL	⋮	⋮	⋮	⋮	⋮
	008A x BR505R	**1.22**	**0.49**	**0.31**	**2.20**
	CMSXS634R	**1.72**	**0.58**	**0.51**	**2.81**
	007A x BR501R	**1.81**	**0.53**	**0.76**	**2.80**
	007AxCMSXS642R	**2.63**	**0.61**	**1.40**	**3.79**

Fonte:Do Autor (2024).

A Tabela 4.9 apresenta os resultados das Análises de Variância (ANOVA) para diferentes variáveis: FLOR, AP, PMV, EXT, SST, TBH, POL, ART, ATR e ALHA. Os valores da variância genética (σ_g^2) e da variância do erro (σ_e^2) variam entre as variáveis, refletindo diferentes graus de influência genética e ambiental.

O p-valor associado a cada variável indica a significância estatística das diferenças entre os genótipos. Variáveis como FLOR, AP, PMV e TBH apresentam valores extremamente baixos, sugerindo diferenças altamente significativas entre os grupos analisados. As médias de cada variável fornecem uma referência central para os valores observados, sendo que algumas variáveis, como ALHA e ATR, possuem médias bem mais elevadas do que outras.

A herdabilidade estimada varia conforme a variável, indicando a proporção da variabilidade total que pode ser atribuída a fatores genéticos. Variáveis como AP, TBH e PMV apresentam alta herdabilidade (H_2-Cullis), sugerindo que a seleção genética pode ser eficiente nesses casos. A acurácia segue uma tendência semelhante, refletindo a confiabilidade das estimativas geradas pelo modelo.Valores mais elevados indicam maior confiabilidade na predição dos efeitos genotípicos. Variáveis como AP, TBH e PMV apresentam acurácia elevada, sugerindo que a seleção baseada nesses dados é mais precisa. Enquanto EXT, por apresentar um valor mais baixo, indica menor confiança na estimativa dos parâmetros genéticos.

O Coeficiente de Variação do Erro (CVe) também varia entre as variáveis, indicando a variabilidade relativa dentro dos grupos analisados. Variáveis como POL e ALHA apresentam CVe mais elevados, sugerindo maior dispersão dos dados em relação à média, enquanto ou-

tras, como FLOR e EXT, possuem CVe menores, indicando maior precisão na estimativa dos parâmetros.

As variáveis analisadas demonstram forte influência genética em algumas, com alta herdabilidade e acurácia, enquanto outras mostram maior variação ambiental, refletida em uma maior variância do erro e Coeficiente de Variação do Erro elevado.

Tabela 4.9 – Resultados das Análises de Variância (ANOVA) para as Variáveis.

Variável	σ_g^2	σ_e^2	P-valor	Média	H_2-Cullis	Acurácia	CVe
FLOR	14.98	8.78	1.60e-11	77.24	0.82	0.89	3.84
AP	0.15	0.05	3.87e-18	2.96	0.89	0.93	7.88
PMV	105.87	46.41	1.82e-14	40.83	0.86	0.92	16.68
EXT	5.53	16.66	0.01	69.11	0.48	0.68	5.91
SST	3.20	2.72	1.94e-09	12.91	0.78	0.87	12.79
TBH	1.74	0.66	5.90e-16	3.62	0.88	0.93	22.50
POL	2.49	2.34	2.47e-08	6.18	0.71	0.83	24.79
ART	3.26	3.00	1.63e-08	9.54	0.71	0.83	18.16
ATR	176.12	163.08	2.05e-08	71.01	0.71	0.83	17.98
ALHA	793955.20	391373.10	3.74e-13	2222.80	0.86	0.91	28.14

Fonte: Do Autor (2024).

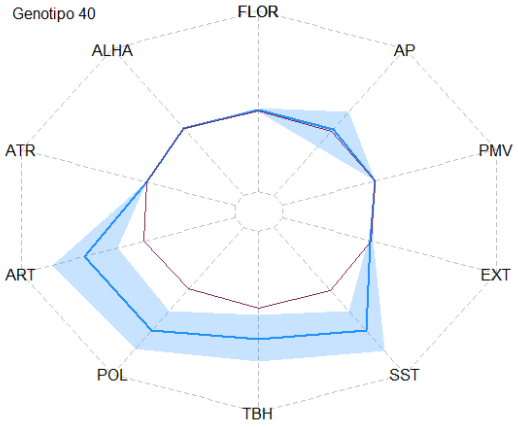
4.2.2 Comparação de gráficos de radar de três genótipos

Na Figura 4.9, a linha violeta representa a média geral dos 45 genótipos para cada característica. A linha azul indica a média de cada característica para um dado genótipo, enquanto a área sombreada em azul representa o intervalo de alta densidade posterior (HPD) de cada característica para esse genótipo.

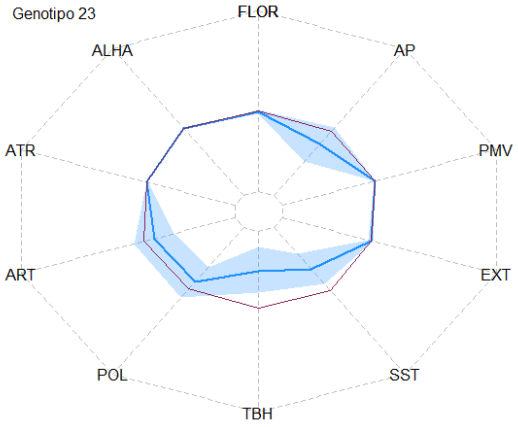
Comparando as médias dos genótipos e seus HPDs nos seus respectivos gráficos de radar (Figura 4.9), o genótipo 007AxCMSXS642R apresentou uma média alta e um intervalo bastante alto nas variáveis de interesse. Como mostrado na Figura 4.9a, pode-se notar que todas as cinco características estavam acima da média para este genótipo. Foi observado o contrario para a 4.9b,em que todas as características estavam abaixo da média para o 008AxCMSXS634R até a característica Altura de planta (AP) estava abaixo da média geral. No entanto, o 007AxCMSXS642R apresentou maior potencial para Açúcares redutores totais % colmo (ART) entre os dois genótipos.

Figura 4.9 – gráficos de radar de dois genótipos aleatórios de sorgo sacarino.

a) 007AxCMSXS642R



b) 008AxCMSXS634R



Fonte: Do Autor (2024).

4.2.3 Comparação de gráficos entre dois diferentes genótipos de Sorgo Sacarino

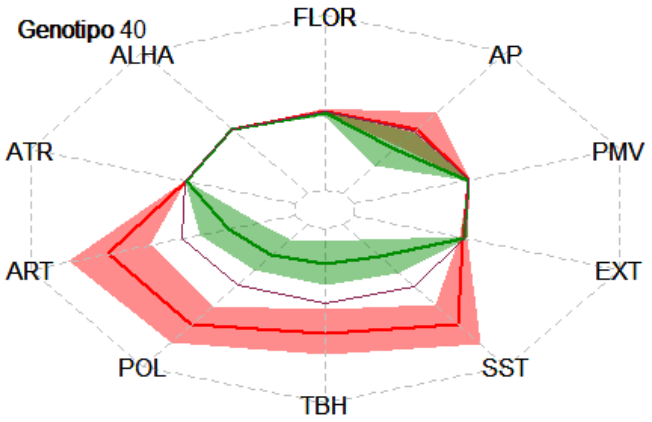
Na Figura 4.10 para cada característica, a linha violeta representa a média geral dos 45 genótipos; a cor Vermelha representa a média e a região do HPD para o genótipo 07AxCMSXS-642R; e a cor Verde, para os genótipos CMSXS634R e CMSXS647R.

Observando a Figura 4.10a, verifica-se que em todas as características de interesse (AP,SST, TBH,POL e ART) do genótipo 007AxCMSXS642R tem um maior desempenho, ao contrário do genótipo CMSXS647R, em que a sua média do genótipo está abaixo da média geral em todas as características de interesse.

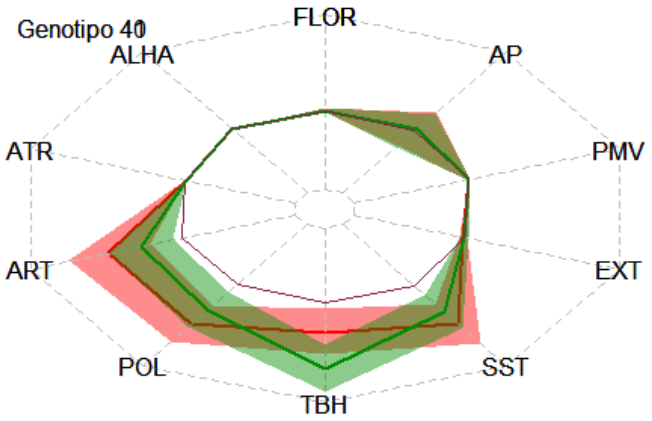
Na Figura 4.10b mais características do genótipo 007AxCMSXS642R, a saber: SST, POL, ART ,têm maior desempenho em relação as características do genótipo CMSXS634R, no entanto a característica TBH tem maior desempenho no genótipo CMSXS634R,que no genótipo 007AxCMSXS642R, para característica AP têm mesmo poder,nas outras cinco características, como tinha sido falado anteriormente, não têm relevância nenhuma.

Figura 4.10 – gráficos de diferentes genótipos de sorgo sacarino.

(a) 007AxCMSXS642R VS CMSXS647R



(b) 007AxCMSXS642R VS CMSXS634R



Fonte: Do Autor (2024).

4.3 Análise da técnica no auxílio à seleção

A técnica de visualização proposta neste estudo, com o uso da Inferência Bayesiana para obtenção da Média e HPD dos genótipos e construção do gráfico de radar, revelou-se um método eficiente porque permite uma visualização mais direta do conjunto dos efeitos de cada genótipo ou grupo de genótipos que se considerar no radar. Isto auxilia a eficácia do processo seletivo e a clareza na comunicação dos resultados.

É preciso notar que a implementação em R foi automatizada (e pacotes estatísticos derivados estão em vias de ser implementados), gerando gráficos de qualidade que consideramos desejável para programas de melhramento.

Esta visualização simultânea dos diversos caracteres, em especial dos mais relevantes, pode auxiliar diretamente a seleção nos diversos caracteres. É claro que um resultado inesperado também fica evidenciado: caracteres sem variabilidade genética aparecem como linhas sem expressão diferencial. Isto pode servir para se prestar menos atenção nestes caracteres ou mesmo construir outro radar sem estes caracteres, servindo como auxílio visual à inferênia na seleção.

A técnica tem grande potencial para aplicação em programas de melhoramento do Sorgo Sacarino, bem como em outros programas de melhoramento. Nunes et al. (2005) utilizaram o gráfico de radar para estudar a adaptabilidade e estabilidade de cultivares. Desde então, a metodologia tem sido aplicada a análises multicaracterísticas, incluindo índices de seleção (SILVA et al., 2016; LIMA et al., 2012; LIMA et al., 2015). Assim, a metodologia desenvolvida no presente estudo pode ser utilizada em cenários multicaracterísticas, mas também pode ser aplicada a um cenário multiambiente.

A partir da visualização do gráfico de radar proposto (Figura 4.9 e 4.10), o melhorista é capaz de identificar para quais características um determinado genótipo está abaixo ou acima da média desejada, permitindo a seleção ou descarte de materiais. Nunes et al. (2005) não comentaram sobre esse fato, uma vez que os gráficos foram avaliados apenas visualmente e baseadas em medições indiretas,baseando-se na noção de “bola cheia” e “bola murcha”, enquanto que (SILVA et al., 2016), usou equações de modelos mistos juntamente com componentes de variância estimados por REML/BLUP para a apresentação gráfica, com o teste de permutação.

Silva et al. (2016) afirmam que a maioria dos índices de seleção utilizados em programas de melhoramento são construídos com base em estimativas de parâmetros genéticos e médias

fenotípicas obtidas por meio de análises de variância. No entanto, o uso da Inferência Bayesiana é uma alternativa que pode ser utilizada na construção de índices que podem resultar em um processo de seleção mais preciso.

Essa alternativa não foi mencionada por Nunes et al. (2005) e nem Silva et al. (2016). Como apresentado na Figura 4.2 e 4.4, o genótipo foi identificado como tendo a maior Média e o maior HPD (genótipo 20 e 007AxCMSXS642R).

Embora os gráficos de radar possam fornecer uma boa interpretação visual quando um pequeno número de genótipos é avaliado, essa interpretação torna-se cada vez mais complicada com o aumento do número de genótipos. Liu et al. (2010) avaliaram apenas sete genótipos de tabaco e, consequentemente, os genótipos foram facilmente discriminados usando o gráfico de radar. No entanto, os melhoristas geralmente avaliam vários genótipos em programas de melhoramento, tornando impossível selecionar genótipos superiores por meio do gráfico de radar. Por exemplo, Silva et al. (2016) afirmam que Soriano et al. (2005) avaliaram 461 progênies de porcos, e Laino et al. (2015) avaliaram mais de 4000 acessões de trigo. Portanto, a aplicação de índices de seleção em programas de melhoramento que avaliam um número muito grande de genótipos é útil, uma vez que os genótipos podem ser classificados e selecionados com base em suas médias e intervalos de credibilidade. Entendemos, no entanto, que os radares podem auxiliar até mesmo na composição de tais índices, evidenciando a variabiidade para caracteres de interesse econômico variado.

Usando a Inferência Bayesiana, mesmo a univariada, e mostrando os intervalos de credibilidade para cada genótipo, distribuídos em seus respectivos gráficos de radar, o melhorista pode também identificar facilmente genótipos ou grupos de genótipos de interesse para vários caracteres econômicos simultaneamente.

5 CONCLUSÃO

Este trabalho descreveu a utilidade de uma técnica simples de visualização no apoio à seleção no melhoramento vegetal, que é o uso dos gráfico de radar com a implementação de intervalos de credibilidade para a visualização.

Os intervalos propostos são o padrão da inferência bayesiana univariada, mas poderiam ser obtidos intervalos semelhantes por técnicas de verossimilhança restrita usado modelos mistos (ou mesmo de efeitos fixos, embora seja pouco relevante para a inferência sobre genótipos).

Nossa conclusão é que tais gráficos podem ser facilmente implementados em linguagens como o R, gerando gráficos de boa qualidade.

Incorporar a inferência nos gráficos, ainda que de forma univariada, permite uma visualização global do processo de seleção e pode ser uma ferramenta auxiliar adequada para os melhoristas. O gráfico de radar permite a visualização imediata de genótipos ou grupos de genótipos promissores e inferiores.

O exemplo apresentado envolve sorgo sacarino, mas outras culturas e conjuntos de variáveis podem se beneficiar da técnica.

Como perspectiva futura, pretende-se implementar o gráfico derivado da inferência conjunta multivariada, levando em conta covariâncias genéticas e ambientais entre os caracteres.

REFERÊNCIAS

ANSARI CERI J. PHILLIPS, W. E. Interprofessional collaboration: a stakeholder approach to evaluation of voluntary participation in community partnerships. **Journal of Interprofessional Care**, Taylor & Francis, v. 15, n. 4, p. 351–368, 2001.

BANERJEE, V.; KRISHNAN, P.; DAS, B.; VERMA, A.; VARGHESE, E. Crop status index as an indicator of wheat crop growth condition under abiotic stress situations. **Field crops research**, Elsevier, v. 181, p. 16–31, 2015.

BLASCO, A. La controversia bayesiana en mejora animal. **ITEA**, v. 99, p. 5–41, 1998.

BOX, G. E.; TIAO, G. C. **Bayesian Inference in Statistical Analysis**. New York: John Wiley & Sons, 1992. 558 p.

BOX, G. E. P.; TIAO, G. C. **Bayesian Inference in Statistical Analysis**. Reading, MA: Wiley, 1973.

BRAVAIS, A. Sur les probabilités des erreurs de situation d'un point. **Memoirs de l'Académie Royale des Sciences de l'Institut de France**, v. 9, p. 255–332, 1846.

CAMPOS, B. M. Análise genética e comparação de modelos por inferência bayesiana e frequentista em características de crescimento de bovinos da raça tabapuã do estado da bahia. **Portuguese.) MSc Diss., Universidade Estadual do Sudoeste da Bahia, Itapetinga, Brazil**, 2013.

CARNEIRO JÚNIOR, J. M.; ASSIS, G. M. L. de; EUCLYDES, R. F.; LOPES, P. S. Influência da informação a priori na avaliação genética animal utilizando dados simulados. **Revista Brasileira de Zootecnia**, v. 34, n. 6, p. 1905–1913, 2005.

CASTRO, R.; PETERNELLI, L.; RESENDE, M.; MARINHO, C. et al. Selection between and within full-sib sugarcane families using the modified blupis method (blupism). **Genetics and Molecular Research**, v. 15, 2016. Disponível em: <http://dx.doi.org/10.4238/gmr.15017334>.

CHEN, M.-H.; SHAO, Q.-M. Monte carlo estimation of bayesian credible and hpd intervals. **Journal of computational and Graphical Statistics**, Taylor & Francis, v. 8, n. 1, p. 69–92, 1999.

CHIB, S.; GREENBERG, E. Understanding the metropolis-hastings algorithm. **The American Statistician**, Taylor & Francis, Alexandria, v. 49, n. 4, p. 327–335, Nov 1995.

COSTA, M. T. G. P.; SANCHES, A.; MUNARI, D. P. Estimação bayesiana de parâmetros genéticos de pesos corporais em um rebanho da raça guzerá. **Nucleus Animalium**, Fundação Educacional Ituverava, v. 1, n. 1, p. 1–13, 2009.

DEGROOT, M. H.; SCHERVISH, M. J. **Probability and Statistics**. 4th. ed. [S.l.]: Addison Wesley, 2012.

DUARTE, F. M.; POCOJESKI, E.; SILVA, L. S. d.; GRAUPE, F. A.; BRITZKE, D. Perdas de

nitrogênio por volatilização de amônia com aplicação de uréia em solo de várzea com diferentes níveis de umidade. **Ciência Rural**, SciELO Brasil, v. 37, p. 705–711, 2007.

DURÃES, N. N. L. **Heterose em Sorgo Sacarino**. 96 p. Dissertação (Dissertação (Mestrado em Genética e Melhoramento de Plantas)) — Universidade Federal de Lavras, Lavras, 2014.

FILHO, J. A.; TARDIN, F.; GUIMARÃES, J.; RESENDE, M. et al. Multi-trait blup model indicates sorghum hybrids with genetic potential for agronomic and nutritional traits. **Genetics and Molecular Research**, v. 15, 2016. Disponível em: <http://dx.doi.org/10.4238/gmr.15017071>.

GAMERMAN, D. **Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference**. London: Chapman & Hall/CRC, 1996. ISBN 978-0412607500.

GELFAND, A. E.; SMITH, A. F. Sampling-based approaches to calculating marginal densities. **Journal of the American statistical association**, Taylor & Francis, v. 85, n. 410, p. 398–409, 1990.

GELMAN, A.; CARLIN, J. B.; STERN, H. S.; RUBIN, D. B. **Bayesian Data Analysis**. London: Chapman and Hall/CRC, 2000. 526 p.

GELMAN, A.; RUBIN, D. B. Inference from iterative simulation using multiple sequences. **Statistical Science**, v. 7, n. 4, p. 457–472, 1992. Disponível em: <https://projecteuclid.org/euclid.ss/1177011136>.

GEMAN, S.; GEMAN, D. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. **IEEE Transactions on pattern analysis and machine intelligence**, IEEE, n. 6, p. 721–741, 1984.

HASTINGS, W. K. Monte carlo sampling methods using markov chains and their applications. **Biometrika**, Oxford University Press, London, v. 57, n. 1, p. 97–109, 1970.

HENRY, G. T. Graphing data: Techniques for display and analysis. **(No Title)**, 1995.

JOHNSON, R. A.; BHATTACHARYYA, G. K. **Statistics: Principles and Methods**. [S.l.]: John Wiley & Sons, 2009.

KINAS, P. G. **Análise Bayesiana da Decisão em FURG**. 2008. Acessado em 19 de fevereiro de 2025. Disponível em: <http://www.furg.br>.

KUI, L. et al. Application of radar chart analytical method in tobacco variety trial. **CHINESE TOBACCO SCIENCE**, v. 31, n. 6, p. 64–67, 2010.

LAINO, P.; LIMONTA, M.; GERNA, D.; VACCINO, P. Morpho-physiolological and qualitative traits of a bread wheat collection spanning a century of breeding in italy. **Biodiversity Data Journal**, Pensoft Publishers, n. 3, 2015.

LIMA, C. N. d. L.; BUENO FILHO, J. S. d. S. Ajuste de uma superfície de resposta no delineamento em blocos incompletos: análise de verossimilhança restrita e bayesiana em uma simulação de adubação em citros. **Revista de Matemática e Estatística**, n. 4, p. 133–157, 2006.

LIMA, D. C.; ABREU, F. B.; FERREIRA, R. A. D. C.; RAMALHO, M. A. P. Breeding common bean populations for traits using selection index. **Sci. Agric.**, v. 72, p. 132–137, 2015. Disponível em: <http://dx.doi.org/10.1590/0103-9016-2014-0130>.

LIMA, L. K.; RAMALHO, M. A. P.; ABREU, F. Implications of the progeny x environment interaction in selection index involving characteristics of the common bean. **Genet. Mol. Res.**, v. 11, p. 4093–4099, 2012. Disponível em: <http://dx.doi.org/10.4238/2012.September.19.5>.

LIU, K.; WANG, Y.; LUO, C.; CHEN, Z. et al. Application of radar chart analytical method in tobacco variety trial. **Chin. Tobacco Sci.**, v. 6, p. 016, 2010.

MARI, D. D.; KOTZ, S. **Correlation and Dependence**. [S.l.]: World Scientific, 2001.

MONTGOMERY, D. C. **ntroduction to Statistical Quality Control**. [S.l.]: United States of America, 2009.

NEAL, R. M. Slice sampling. **The annals of statistics**, Institute of Mathematical Statistics, v. 31, n. 3, p. 705–767, 2003.

NUNES, J. A. R.; RAMALHO, M. A.; ABREU, A. de F. Graphical method in studies of adaptability and stability of cultivars. **ANNUAL REPORT-BEAN IMPROVEMENT COOPERATIVE**, JAMES D. KELLY, ED. AND PUB, v. 48, p. 182, 2005.

O'HAGAN, A. **Kendall's Advanced Theory of Statistics: Volume 2B, Bayesian Inference**. London: Arnold, 1994.

PAULINO, C. D.; TURKMAN, M. A. A.; MURTEIRA, B. **Estatística Bayesiana**. Lisboa, Portugal: Edições Salamandra, 2003. 447 p.

PIEPHO, H.; MÖHRING, J.; MELCHINGER, A.; BÜCHSE, A. Blup for phenotypic selection in plant breeding and variety testing. **Euphytica**, v. 161, p. 209–228, 2008. Disponível em: <http://dx.doi.org/10.1007/s10681-007-9449-8>.

RAFTERY, A. E.; LEWIS, S. M. How many iterations in the gibbs sampler? **Bayesian Statistics**, Oxford University Press, p. 763–773, 1992.

REIS, C.; GONÇALVES, F.; RAMALHO, M.; ROSADO, A. Seleção de progênies de eucalipto pelo índice z por mqm e blup. **Pesquisa Agropecuária Brasileira**, v. 46, p. 516–522, 2011. Disponível em: <http://dx.doi.org/10.1590/S0100-204X2011000500009>.

SAARY, M. J. Radar plots: a useful way for presenting multivariate health care data. **Journal of clinical epidemiology**, Elsevier, v. 61, n. 4, p. 311–317, 2008.

SATTERTHWAITE, F. E. An approximate distribution of estimates of variance components. **Biometrics Bulletin**, JSTOR, v. 2, n. 6, p. 110–114, 1946.

SILVA, L. et al. Selection index using the graphical area applied to sugarcane breeding. Genetics and Molecular Research, 2016.

SOPHIA, A. **Gráficos de Radar**. 2023. Disponível em: <https://www.edrawsoft.com/pt/radar-chart.html>.

SORENSEN, D. **Gibbs Sampling in Quantitative Genetics**. Foulun: [s.n.], 1996. 186 p. Publicação sem editor nomeado.

SORENSEN, D.; GIANOLA, D. Likelihood, bayesian, and mcmc methods in quantitative genetics. **Statistics for Biology and Health**, Springer New York, 2002.

SORIANO, A.; QUILES, R.; MARISCAL, C.; RUIZ, A. G. Pig sire type and sex effects on carcass traits, meat quality and physicochemical and sensory characteristics of serrano dry-cured ham. **Journal of the Science of Food and Agriculture**, Wiley Online Library, v. 85, n. 11, p. 1914–1924, 2005.

TURKMAN, M. A. A.; PAULINO, C. D.; MÜLLER, P. **Computational Bayesian statistics: an introduction**. [S.l.]: Cambridge University Press, 2019. v. 11.

Referências

APENDICE A – Resumos das distribuições a posteriori para os efeitos de tratamentos

Tabela 1 – Médias, desvio padrão e intervalos de credibilidade para a distribuição conjunta a posteriori dos efeitos de genótipos, dadas as observações do caráter simulado $\mathbf{Y_1}$.

Caráter	Genótipo	$\bar{\mathbf{g}}_\mathbf{i}\|\mathbf{y_i}$	Dp	LI	LS
$\mathbf{Y_1}$	1	-0.85	0.48	-1.79	0.08
	2	-0.73	0.48	-1.67	0.18
	3	-0.47	0.48	-1.44	0.43
	4	-0.72	0.48	-1.65	0.23
	5	-1.29	0.50	-2.24	-0.30
	6	0.21	0.48	-0.73	1.12
	7	-1.08	0.50	-2.03	-0.08
	8	0.30	0.47	-0.62	1.25
	9	-0.32	0.48	-1.28	0.63
	10	-0.03	0.48	-0.94	0.96
	11	-0.40	0.48	-1.34	0.54
	12	0.20	0.47	-0.68	1.17
	13	0.24	0.48	-0.63	1.26
	14	0.74	0.48	-0.18	1.66
	15	0.76	0.48	0.16	1.71
	16	0.75	0.48	-0.20	1.69
	17	0.85	0.48	0.04	1.86
	18	0.54	0.48	-0.38	1.47
	19	0.70	0.48	-0.23	1.64
	20	0.56	0.48	0.40	1.48

Fonte: Do Autor (2024).

Tabela 2 – Médias, desvio padrão e intervalos de credibilidade para a distribuição conjunta a posteriori dos efeitos de genótipos, dadas as observações do caráter simulado $\mathbf{Y_2}$.

Caráter	Genótipo	$\bar{\mathbf{g}}_\mathbf{i}\|\mathbf{y_i}$	Dp	LI	LS
$\mathbf{Y_2}$	1	-1.25	0.50	-2.23	-0.30
	2	-0.84	0.49	-1.85	0.05
	3	-0.49	0.47	-1.38	0.43
	4	-0.85	0.48	-1.86	0.05
	5	-0.50	0.48	-1.38	0.49
	6	0.20	0.47	-0.77	1.10
	7	-1.22	0.49	-2.18	-0.26
	8	0.26	0.47	-0.67	1.19
	9	-0.44	0.48	-1.39	0.48
	10	0.06	0.47	-0.86	0.98
	11	-0.37	0.47	-1.29	0.55
	12	0.29	0.47	-0.66	1.18
	13	0.26	0.48	-0.64	1.22
	20	0.56	0.48	-0.36	1.48
	15	0.70	0.48	-0.23	1.68
	16	0.81	0.48	-0.12	1.75
	17	0.77	0.48	-0.14	1.73
	18	0.44	0.47	-0.43	1.43
	19	0.68	0.48	-0.24	1.61
	14	0.83	0.48	-0.08	1.81

Fonte: Do Autor (2024).

Tabela 3 – Médias, desvio padrão e intervalos de credibilidade para a distribuição conjunta a posteriori dos efeitos de genótipos, dadas as observações do caráter simulado $\mathbf{Y_3}$.

Caráter	Genótipo	$\bar{\mathbf{g}}_\mathbf{i}\|\mathbf{y_i}$	Dp	LI	LS
$\mathbf{Y_3}$	1	-1.63	0.55	-2.67	-0.53
	2	-0.92	0.54	-1.94	0.16
	3	0.08	0.53	-0.96	1.14
	4	-1.30	0.54	-2.32	-0.23
	5	-0.26	0.53	-1.31	0.82
	6	0.62	0.53	-0.42	1.63
	7	-1.06	0.54	-2.07	0.04
	8	0.64	0.52	-0.44	1.62
	9	-0.75	0.54	-1.83	0.25
	10	-0.63	0.53	-1.65	0.41
	11	0.30	0.53	-0.70	1.32
	12	0.00	0.53	-1.09	1.01
	13	0.19	0.54	-0.90	1.23
	14	0.62	0.53	-0.42	1.68
	15	0.98	0.54	-0.09	2.03
	16	0.17	0.53	-0.85	1.23
	17	0.84	0.53	-0.17	1.93
	18	0.41	0.54	-0.63	1.51
	19	0.69	0.53	-0.36	1.71
	20	1.03	0.53	-0.00	2.06

Fonte: Do Autor (2024).

Tabela 4 – Médias, desvio padrão e intervalos de credibilidade para a distribuição conjunta a posteriori dos efeitos de genótipos, dadas as observações do caráter simulado $\mathbf{Y_4}$.

Caráter	Genótipo	$\bar{\mathbf{g}}_\mathbf{i}\|\mathbf{y_i}$	Dp	LI	LS
$\mathbf{Y_4}$	1	-0.89	0.49	-1.87	0.07
	2	-1.79	0.53	-2.80	-0.76
	3	0.88	0.49	-0.04	1.88
	4	-0.41	0.47	-1.33	0.53
	5	-0.66	0.48	-1.60	0.29
	6	0.30	0.48	-0.69	1.21
	7	-0.76	0.48	-1.66	0.22
	8	0.74	0.49	-0.17	1.72
	9	-0.72	0.49	-1.68	0.19
	10	-0.48	0.49	-1.44	0.47
	11	-0.10	0.48	-1.08	0.78
	12	0.46	0.48	-0.51	1.37
	13	0.51	0.48	-0.39	1.48
	14	-0.41	0.47	-1.30	0.53
	15	-0.33	0.49	-1.34	0.59
	16	-0.01	0.48	-0.94	0.91
	17	0.82	0.49	-0.11	1.78
	18	1.03	0.49	0.07	1.95
	19	0.78	0.48	-0.15	1.74
	20	1.03	0.49	0.08	2.01

Fonte: Do Autor (2024).

Tabela 5 – Médias, desvio padrão e intervalos de credibilidade para a distribuição conjunta a posteriori dos efeitos de genótipos, dadas as observações do caráter simulado $\mathbf{Y_5}$.

Caráter	Genótipo	$\bar{\mathbf{g}}_\mathbf{i} \mid \mathbf{y_i}$	DP	LI	LS
$\mathbf{Y_5}$	1	-0.24	0.52	-1.22	0.82
	2	-0.47	0.53	-1.53	0.56
	3	-0.66	0.52	-1.66	0.36
	4	0.85	0.53	-0.14	1.94
	5	-1.01	0.53	-2.02	0.06
	6	0.17	0.52	-0.90	1.19
	7	1.97	0.55	-0.89	3.03
	8	-0.33	0.53	-1.36	0.66
	9	-1.45	0.53	-2.54	-0.47
	10	0.19	0.53	-0.84	1.22
	11	-1.05	0.53	-2.16	-0.04
	12	0.34	0.52	-0.66	1.39
	13	1.16	0.54	-0.02	2.16
	14	1.80	0.55	0.75	2.86
	15	-1.01	0.54	-2.04	0.06
	16	-0.95	0.52	-1.92	0.10
	17	0.32	0.52	-0.75	1.27
	18	-0.60	0.52	-1.63	0.41
	19	0.41	0.53	-0.60	1.45
	20	0.60	0.52	-0.38	1.65

Fonte: Do Autor (2024).

Tabela 6 – Vetores de valores paramétricos dos efeitos genéticos.

Genótipo	$\mathbf{g_1}$	$\mathbf{g_2}$	$\mathbf{g_3}$	$\mathbf{g_4}$	$\mathbf{g_5}$
1	-1.67	-1.35	-1.05	-1.39	-1.14
2	-1.31	-1.46	-2.09	-2.34	-0.14
3	-1.07	-1.11	-0.30	0.72	-0.17
4	-0.88	-0.90	-1.20	-0.37	0.75
5	-0.71	-0.65	0.51	-0.10	-0.32
6	-0.57	-0.54	0.10	-0.30	0.58
7	-0.43	-0.79	-1.06	-1.11	1.58
8	-0.30	-0.44	0.29	0.55	1.86
9	-0.18	-0.32	-0.59	-0.65	0.17
10	-0.06	-0.14	-0.63	-0.29	-1.90
11	0.06	-0.05	0.11	-0.43	0.74
12	0.18	0.20	-0.04	0.70	-0.65
13	0.30	0.36	0.07	0.69	-0.56
14	0.43	0.41	0.52	-0.22	1.77
15	0.57	0.54	0.60	-0.63	0.59
16	0.71	0.67	0.53	0.38	-0.34
17	0.88	0.87	1.32	1.44	-0.52
18	1.07	0.99	0.76	1.48	0.90
19	1.31	1.21	0.70	1.51	-0.58
20	1.67	1.62	1.52	1.59	1.06

Fonte: Do Autor (2024).

Tabela 7 – Dados simulados

(continua)

U.E.	Bloco	Genótipo	Y_1	Y_2	Y_3	Y_4	Y_5
1	1	1	97.40	97.76	98.14	98.42	98.15
2	2	1	98.23	98.55	98.33	98.61	98.71
3	3	1	100.41	101.02	98.92	98.27	99.61
4	4	1	99.59	100.09	100.56	100.43	100.19
5	1	2	98.39	98.06	97.29	96.83	98.49
6	2	2	98.41	98.26	97.78	97.70	100.76
7	3	2	99.80	99.87	98.22	98.12	99.13
8	4	2	99.66	99.32	99.36	98.49	100.80
9	1	3	100.43	100.50	100.76	101.80	98.56
10	2	3	98.61	98.49	99.52	101.27	99.18
11	3	3	98.17	98.00	100.27	100.92	100.42
12	4	3	100.39	100.35	100.11	100.91	100.15
13	1	4	96.45	96.31	97.45	98.87	101.38
14	2	4	99.84	99.50	98.33	98.81	101.89
15	3	4	98.62	98.47	98.33	99.49	101.17
16	4	4	101.37	101.26	100.16	101.12	101.03
17	1	5	96.68	96.84	98.60	98.57	99.85
18	2	5	98.08	98.16	98.96	97.52	99.26
19	3	5	98.36	98.58	100.55	100.43	100.35
20	4	5	100.20	99.79	101.02	100.47	100.78
21	1	6	98.17	97.94	99.02	99.47	99.07
22	2	6	101.13	101.08	101.14	100.51	100.39
23	3	6	100.66	100.56	101.05	100.30	100.94
24	4	6	101.11	101.46	101.96	101.65	101.82
25	1	7	97.13	96.91	97.75	97.80	100.36
26	2	7	98.62	98.46	97.42	96.88	102.72
27	3	7	99.19	98.99	99.59	100.65	101.94
28	4	7	99.43	99.22	100.58	101.09	104.83
29	1	8	99.19	99.30	100.42	101.44	103.11

Tabela 7 –Dados simulados

(continuação)

U.E.	Bloco	Genótipo	Y_1	Y_2	Y_3	Y_4	Y_5
30	2	8	100.38	100.09	100.76	100.12	102.45
31	3	8	99.28	99.23	99.94	100.02	101.98
32	4	8	102.68	102.72	102.20	102.61	103.18
33	1	9	98.64	98.57	98.66	99.22	99.02
34	2	9	100.91	100.75	99.76	99.67	99.49
35	3	9	98.00	97.67	98.52	97.71	100.24
36	4	9	100.76	100.74	99.86	100.03	101.13
37	1	10	98.29	98.80	97.95	98.21	98.68
38	2	10	100.08	100.06	99.09	99.37	97.94
39	3	10	100.89	100.90	99.65	99.24	99.14
40	4	10	100.62	100.62	100.67	101.04	98.86
41	1	11	98.52	98.47	99.72	98.97	99.43
42	2	11	97.86	97.91	99.80	99.78	100.94
43	3	11	100.07	100.19	101.52	101.27	101.51
44	4	11	101.51	101.48	100.61	99.80	100.41
45	1	12	99.26	99.23	100.45	101.59	97.32
46	2	12	99.75	99.86	100.07	101.21	99.55
47	3	12	100.28	100.36	98.75	98.53	99.56
48	4	12	101.78	102.07	101.01	101.36	100.03
49	1	13	99.63	99.52	99.02	99.08	98.86
50	2	13	100.09	99.98	99.89	100.27	100.29
51	3	13	101.59	101.83	101.34	102.64	101.25
52	4	13	99.88	100.00	100.84	101.03	102.64
53	1	14	100.11	100.54	100.26	99.63	100.42
54	2	14	100.27	100.29	100.68	98.74	102.20

Tabela 7 –Dados simulados

(conclusão)

U.E.	Bloco	Genótipo	Y_1	Y_2	Y_3	Y_4	Y_5
55	3	14	100.25	100.06	100.86	100.63	101.31
56	4	14	103.16	103.50	101.38	99.20	102.97
57	1	15	100.54	100.39	100.02	98.65	99.54
58	2	15	99.41	99.51	100.96	99.99	100.22
59	3	15	101.40	101.09	101.74	100.41	101.53
60	4	15	102.68	102.68	101.50	99.60	101.64
61	1	16	99.13	98.92	98.65	98.09	98.00
62	2	16	101.19	101.45	100.47	100.21	99.33
63	3	16	101.20	101.36	100.51	100.33	98.68
64	4	16	102.38	102.56	101.47	101.75	100.65
65	1	17	101.17	101.10	99.88	100.71	98.48
66	2	17	99.97	99.74	100.55	100.17	98.39
67	3	17	101.24	101.16	102.12	101.87	99.97
68	4	17	102.08	102.07	102.59	101.85	100.13
69	1	18	99.49	99.59	99.26	99.34	100.61
70	2	18	101.47	101.55	101.00	102.53	101.60
71	3	18	100.55	100.41	100.62	100.95	101.08
72	4	18	101.33	100.78	101.38	102.87	101.01
73	1	19	99.35	99.16	99.76	100.56	100.04
74	2	19	100.73	100.89	100.71	101.43	98.78
75	3	19	100.76	100.75	101.32	101.45	99.29
76	4	19	102.87	102.69	101.71	102.26	100.55
77	1	20	101.37	101.27	101.10	100.40	99.73
78	2	20	99.26	99.44	101.06	100.94	99.75
79	3	20	100.05	100.05	100.53	100.33	102.05
80	4	20	102.30	102.19	102.21	102.69	101.82

Fonte: Do Autor (2024).

APENDICE B – Códigos R para a simulação do experimento

```
# carrega biblioteca

library(ggplot2)

library(dplyr)

library(ggrepel)

library(ggthemes)

library(tidyverse)

library(xtable)

library(MASS)

library(mvtnorm)

# Obtencao dos valores dos parametros e comparacao
# dos estimadores de inferência Bayesiana

t <- 20 # n. tratamentos

b <- 4 # n. blocos

n <- 80 # n. unidades experimentais

r <-3 # n. de repeticoes
```

```
st <- 1 # sd dos tratamentos

sb <- 1 # sd dos blocos

se <- 1 # sd do erro de parcelas

(et <- qnorm((1:t/(t+1)),0,st)) # efeito de tratamento

(eb <- qnorm((1:b/(b+1)),0,sb)) # efeito de bloco

mu <- 100 # média

e <- rnorm(n)*se # erro experimental

base <- expand.grid(1:b,1:t) # É a matriz de delineamento

Bloc <- factor(base[,1]) # parâmetro do Modelo

Gen <- factor(base[,2])

X <- model.matrix(~ -1+Bloc) # modelo de médias de bloco

Z <- model.matrix(~ -1+Gen) # modelo de médias de Tratamento

eta_b <- mu+X%*%eb

# Fixando a simulação para ser sempre igual

set.seed(29112024)

# Simulando efeitos genéticos
```

```
g1_par <- round(et,3)

g2_par <- round(g1_par+0.1*rnorm(t),3)

g3_par <- round(0.7*g1_par+0.3*g2_par+0.5*rnorm(t),3)

g4_par <- round(g3_par+0.5*rnorm(t),3)

g5_par <- round(rnorm(t),3)

# Efeitos genéticos simulados e sua correlação

Ef_Gen_Sim <- cbind(g1_par,g2_par,g3_par,g4_par,g5_par)

t(Ef_Gen_Sim)

save(Ef_Gen_Sim,file="Ef_Gen_Sim.rda")

cor(Ef_Gen_Sim)

xtable(Ef_Gen_Sim)

xtable(cor(Ef_Gen_Sim))

# Simulando erros também correlacionados
# Mas com correlações diferentes

e1_par <- round(rnorm(n),3)

e2_par <- round(e1_par+0.2*rnorm(n),3)

e3_par <- round(0.1*e1_par+0.1*e2_par+0.5*rnorm(n),3)
```

```
e4_par <- round(e3_par+0.5*rnorm(n),3)

e5_par <- round(rnorm(n),3)

# Erros simulados e sua correlação

Erros_Sim <- cbind(e1_par,e2_par,e3_par,e4_par,e5_par)

cor(Erros_Sim)

xtable(Erros_Sim)

xtable(cor(Erros_Sim))

# Modelo da simulação do experimento para obtenção
# Experimento simulado

Y1 <- round(eta_b+Z\%*\%g1_par+e1_par,3)

Y2 <- round(eta_b+Z\%*\%g2_par+e2_par,3)

Y3 <- round(eta_b+Z\%*\%g3_par+e3_par,3)

Y4 <- round(eta_b+Z\%*\%g4_par+e4_par,3)

Y5 <- round(eta_b+Z\%*\%g5_par+e5_par,3)

Y <- cbind(Y1,Y2,Y3,Y4,Y5)

cor(Y)
```

```
# Correlação aparente nos dados

xtable(cor(Y))

dados <- data.frame(Bloc,Gen,Y1,Y2,Y3,Y4,Y5)

head(dados)

# Tabela com os dados

xtable(dados)

# Cabeçalho com os seis dados iniciais

xtable(rbind(head(dados),tail(dados)))

save(dados,file="dados.rda")

rm(list=ls())

# Codigo das funções de amostragem
# para cada parâmetro

#Amostra

beta <- function(delta_beta){
  var_beta <- XXi
  mu_beta  <- XXi\%*\%t(X)\%*\%delta_beta
  beta <- c(rmvnorm(1,mu_beta,var_beta))
  return(beta=beta)
}
```

```
# AG aleatorio

Amostra_g <- function(delta_g){

  M     <- solve(t(Z)\%*\%Z + vE*Gi)

  meanr <- M\%*\%t(Z)\%*\%delta_g
  g     <- c(rmvnorm(1,meanr,M))
  return(g=g)
}

AG_VarComp = Amostra_g(delta_g)

# Problema! não está funcionando com delta_beta

Amostra VCg <- function(g,Gi){

  nu_g      <- nu_0_g + q

  s2_post_g <- as.double(t(g)\%*\%Gi\%*\%g + nu_0_g*s2_0_g)

  VCg       <- s2_post_g/rchisq(1,nu_g)
  return(VCg=VCg)
}

Amostra_VCe <- function(res){

  nu_e      <- nu_0 + n
  s2_post   <- as.double(t(res)\%*\%res + nu_0*s2_0)

  vE        <- s2_post/rchisq(1,nu_e)
  return(vE=vE)
```

```
}

# Leitura de dados e carga das funções importantes

load("Ef_Gen_Sim.rda")
load("dados.rda")
attach(dados)

X <- model.matrix(~ -1+Bloc)

Z <- model.matrix(~ -1+Gen)

source("Leitura.R")
source("AmostradoresGibbs.R")



# coluna <- 3
# OK

vies <- NULL; eqm <- NULL; cgp <- NULL; cgs <- NULL

for(coluna in 3:7){
  y <- dados[,coluna]
  n <- length(y);
  p <- dim(X)[2];
  q <- dim(Z)[2]
  vG <- var(y); vE <- var(y)

  nu_0   <- 0.5; nu_0_g <- 0.5; s2_0   <- 0.5; s2_0_g <- 0.5

  G  <- diag(q)*vG # Genótipos
  Gi <- solve(G) # Inverso de Genótipos
```

```
  # Valores iniciais de efeitos fixos

  XXi <- solve(t(X)\%*\%X)
  beta <- XXi\%*\%t(X)\%*\%y
  g    <- rnorm(q)

  # Valores iniciais de efeitos b:
  delta_beta <- y-Z\%*\%g
  delta_g    <- y-X\%*\%beta

  var_g <- t(Z)\%*\%Z+Gi
  mu_g  <- g

  aceita <- 0

###################### Controla MCCM ################################

  NEF <- 5000

  BI  <- 100

  JMP <- 10

  NMI <- BI+JMP*NEF

  count <- 0

  while(count<NMI){
    count <- count+1
    print(c(coluna,round(100*count/NMI,1)))
  # Atualizar Comp Var
```

```
    res <- y-X\%*\%beta-Z\%*\%g
    vE <- Amostra_VCe(res)
    vG <- Amostra_VCg(g,Gi)
    Gi <- (1/vG)*diag(q)

  # Amostrar médias de bloco (ef. fixos)
    delta_beta <- y-Z\%*\%g
    beta <- Amostra_beta(delta_beta)

  # Amostrar ef. aleatórios de genótipos

    delta_g <- y-X\%*\%beta
    g <- Amostra_g(delta_g)

    vies <- sum(g-Ef_Gen_Sim[,coluna-2])

    eqm <- sum((g-Ef_Gen_Sim[,coluna-2])^2)

    cgp <- cor(g,Ef_Gen_Sim[,coluna-2])

    cgs <- cor(g,Ef_Gen_Sim[,coluna-2],method="spearman")

  # gravar saida a cada passo util:

    saida <- c(names(dados)[coluna],vE,vG,beta,g,vies,eqm,cgp,cgs)

 nomes <- c("Caracter","Var_E","Var_G","Bl1","Bl2","Bl3","Bl4"

"g1","g2","g3","g4","g5","g6","g7","g8","g9","g10"

"g11","g12","g13","g14","g15","g16","g17","g18","g19","g20"
```

```
"Viés","EQM","rho_p","rho_s")

length(nomes)==length(saida)
    if(coluna==3&count==1){
      write(t(nomes),file="saida.txt",append=TRUE,ncol=length(nomes))
    }
    if(count>BI & count\%\%JMP==0){
      write(t(saida),file="saida.txt",append=TRUE,ncol=length(saida))
    }
  }
}

library(coda)

cad <- as.matrix(read.table("saida.txt",head=T))

va<-cad[,c(1,2:3)]

bloco <- as.matrix(read.table("saida_b.txt",head=T))

va <- data.frame(carac = factor(va[,1],levels = unique(va[,1]))

apply(va[,-1],2,as.double))

ci_va <- HPDinterval(as.mcmc(apply(va[,-1],2,as.double)), prob = .95)

M_va  <- colMeans(apply(va[,-1],2,as.double), na.rm = TRUE)

herd  <- va\$Var_G / (va\$Var_G + va\$Var_E)

# Herdabilidade média (ignorando NAs)
```

```
mean_h2 <- mean(herd, na.rm = TRUE)

# Intervalo de confiança

(IC 95\%) ignorando NAs
ic_h2 <- quantile(herd, probs = c(0.025, 0.975), na.rm = TRUE)

# Exibir resultados
cat("Herdabilidade média:", mean_h2, "\n")
cat("IC 95\% da herdabilidade: <", ic_h2[1], ":", ic_h2[2],">","\n")

# Definir o número de repetições

r <- 3

# Calcular a herdabilidade para seleção entre médias de tratamentos

herd_m <-  va\$Var_G / (va\$Var_G + va\$Var_E/r)

# Herdabilidade média (ignorando NAs)
mean_h2_medias <- mean(herd_m, na.rm = TRUE)

# Intervalo de confiança (IC 95\%) ignorando NAs
ic_h2_medias <- quantile(herd_m, probs = c(0.025, 0.975), na.rm = TRUE)

# Exibir resultados
cat("Herdabilidade média para seleção entre médias de tratamentos:",
    mean_h2_medias, "\n")

cat("IC 95\% da herdabilidade (médias de tratamentos):

<", ic_h2_medias[1], ";", ic h2medias[2],">", "\n")
```

```
Carregar os dados MCMC

corps <- as.matrix(read.table("saida.txt", head = TRUE))

co<-corps[,c(1,30:31),na.rm = TRUE]
co\$rho_p<as.numeric(co\$rho_p)

co\$rho_s<-as.numeric(co\$rho_s)

# Extrair colunas de interesse e converter para numérico

#rho_p <- as.numeric(co[, "rho_p"],as.double)

#rho_s <- as.numeric(co[,
"rho_s"],as.double)

# Filtrar valores para garantir que estão dentro do intervalo (-1, 1)

rho_p <- rho_p[rho_p > -1 & rho_p < 1]

rho_s <- rho_s[rho_s > -1 & rho_s < 1]

# Função para calcular o intervalo de confiança

usando a transformação de Fisher

calc_ic <- function(rho, n) {

  fisher_z <- atanh(rho)           # Transformação de Fisher

  se <- 1 / sqrt(n - 3)
```

```
  # Erro padrão

  z_ci <- qnorm(0.975) * se


  Quantil da distribuição normal para 95%

  lower <- tanh(fisher_z - z_ci)

  # Limite inferior

  upper <- tanh(fisher_z + z_ci)

  # Limite superior

  return(c(lower, upper))
}
# Determinar o tamanho da amostra (número de iterações MCMC)

n <- length(rho_p)
plot_radar
# Calcular ICs para Pearson e Spearman

pearson_ic <- t(sapply(rho_p, calc_ic, n))

spearman_ic <- t(sapply(rho_s, calc_ic, n))

# Exibir resultados médios dos ICs

cat("IC médio de Pearson (95\%):", colMeans(pearson_ic), "\n")
```

```
# Carregar os dados MCMC

co <- as.matrix(read.table("saida_co.txt", head = TRUE))

# Extrair colunas de interesse e converter para numérico

rho_p <- as.numeric(co[, "rho_p"])

rho_s <- as.numeric(co[, "rho_s"])

# Filtrar valores para garantir que estão dentro do intervalo (-1, 1)

rho_p <- rho_p[rho_p > -1 & rho_p < 1]

rho_s <- rho_s[rho_s > -1 & rho_s < 1]

# Função para calcular o intervalo de confiança

usando a transformação de Fisher

calc_ic <- function(rho, n) {

  fisher_z <- atanh(rho)

  # Transformação de Fisher

  se <- 1 / sqrt(n - 3)    # Erro padrao

  z_ci <- qnorm(0.975) * se   # Quantil da distribuição normal para 95%

  lower <- tanh(fisher_z - z_ci)   # Limite inferior
```

```
  upper <- tanh(fisher_z + z_ci) # Limite superior
  return(c(lower, upper))
}

# Determinar o tamanho da amostra (número de iterações MCMC)

n <- length(rho_p)

# Calcular ICs para Pearson e Spearman

pearson_ic <- t(sapply(rho_p, calc_ic, n))

spearman_ic <- t(sapply(rho_s, calc_ic, n))

# Exibir resultados médios dos ICs

cat("IC médio de Pearson (95\%):", colMeans(pearson_ic), "\n")
cat("IC médio de Spearman (95\%):", colMeans(spearman_ic), "\n")
```

APENDICE C – Códigos R para Dados reais de um experimento: Dados trigo sorgo.

```
(dados<-read.csv("Dados trigo sorgo.csv",h=TRUE,dec=","))

dados<-dados[, -c(2,3,4,5,7, 9, 10, 15, 16)]

is.na(dados)

# Identificando os NaN

# dados[30,13:16]

# dados[77,13:16]

# fora <- c(30,77)

head(dados)

str(dados)

attach(dados)

B <- factor(IBLOCK)

G <- GENOTYPE

X <- model.matrix(~ -1+B)

# modelo de médias de bloco

Z <- model.matrix(~ -1+G)#
```

```
modelo de médias de Genotipos

# Criar boxplots para todas as variáveis numéricas
boxplot(dados[, sapply(dados, is.numeric)],
        main = "Boxplot de dados de Sorgo  Sacarino",
        las = 2, col = "lightblue")

# Criar uma matriz de gráficos de dispersão para todas

as variáveis numéricas
pairs(dados[, sapply(dados, is.numeric)],
      main = "Matriz de Dispersão de dados de Sorgo  Sacarino",
      pch = 19, col = rgb(0, 0, 1, 0.5))

############################## DEFINICOES ################

y <- dados[,8] # usando FLOR para começar

n <- length(y);

p <- dim(X)[2]; q <- dim(Z)[2]

vG <- var(y);

vE <- var(y)

nu_0   <- 0.5;

nu_0_g <- 0.5;

s2_0   <- 0.5;
```

```
s2_0_g <- 0.5

G  <- diag(q)*vG # Genótipos

Gi <- solve(G) # Inverso de Genótipos

# Valores iniciais de efeitos fixos

beta <- solve(t(X)\%*\%X)\%*\%t(X)\%*\%y

g <- rnorm(q)

library(MASS)

# Valores iniciais de efeitos b:

delta_beta <- y-Z\%*\%g

delta_g    <- y-X\%*\%beta

var_g <- t(Z)\%*\%Z+Gi

mu_g  <- g

aceita <- 0

###################### ControlaMCCM ################################

NEF <- 4000

BI  <- 0
```

```
JMP <- 1

NMI <- BI+JMP*NEF

count <- 0
while(count<NMI){

  count <- count+1

  # Atualizar Comp Var

  res <- y-X\%*\%beta-Z\%*\%g

  vE <- Amostra_VCe(res)

  vG <- Amostra_VCg(g,Gi)

  Gi <- (1/vG)*diag(q)

  # Amostrar médias de bloco (ef. fixos)

  delta_beta <- y-Z\%*\%g

  beta <- Amostra_beta(delta_beta)

  # Amostrar ef. aleatórios de genótipos

  delta_g <- y-X\%*\%beta

  g <- Amostra_g(delta_g)

  # gravar saida a cada passo util:
```

```
  for(coluna in 8){
    print(names(dados)[coluna])

    y <- dados[,coluna]
  if(count>BI & count \%*\%JMP==0){
    write(t(c(names(dados)[coluna],beta)),file="beta.txt",
    append=TRUE,ncol=(length(beta)+1))
    write(t(c(names(dados)[coluna],g)),file="saida_gR.txt",
    append=TRUE,ncol=(length(g)+1))
    write(t(c(c(names(dados)[coluna],vE,vG))),file="vE_e_vG.txt",
    append=TRUE,ncol=3) }}}

source("Leitura.R") # leitura dos dados

y <- dados[,7] # usando FLOR para começar
source("DefineDelineamento.R")
source("FuncAmostragem.R")

names(dados)
for(coluna in 7:12){
  print(names(dados)[coluna])

  y <- dados[,coluna]
  source("DefineDelineamento.R")
  source("ControlaMCMC.R")
}

coluna <- 13

dentro <- which(is.na(POL)==FALSE)

X <- X[dentro,]
```

```
Z <- Z[dentro,]

for(coluna in 13:16){
  print(names(dados)[coluna])

  y <- dados[dentro,coluna]
  source("DefineDelineamento.R")
  source("ControlaMCMC.R")
}
```

APENDICE D – Códigos R para executar a Construção do Grafico.

```
# Carrega bibliotecas

library(sf)

library(ggplot2)

library(coda)

library(tidyverse)


#c1 <- as.matrix(read.table("g1.txt"))

#c2 <- as.matrix(read.table("g2.txt"))

#c3 <- as.matrix(read.table("g3.txt"))

#c4 <- as.matrix(read.table("g4.txt"))

#c5 <- as.matrix(read.table("g5.txt"))

#cc <- rbind((c1),(c2),(c3),(c4),(c5))

#cad <- data.frame(caract = factor(rep(1:5,each=nrow(c1))),cc)


# Carrega os dados sem considerar a primeira linha

cad1 <- as.matrix(read.table("saida.txt", header = FALSE))
```

```
cad<-cad1[,c(1,8:27)]

# Ajusta para 25000 linhas e 20 colunas (se necessário)

cad <- cad[2:25001, 1:21] # Remover a primeira linha e
                          # ajustar o número de colunas

# Cria o data frame com a coluna "carac" como fator

cad <- data.frame(
  carac = factor(cad[, 1],

levels = unique(cad[, 1])), # Converte a primeiracoluna em fator
  apply(cad[, -1], 2,
        as.double)  # Converte as outras colunas para numérico
)

cad <- as.matrix(read.table("saida_g.txt"))
cad <- data.frame(carac = factor(cad[,1],levels = unique(cad[,1])),
apply(cad[,-1],2,as.double))
summary(cad)

# funcões

cria_ eixos <- function(matriz,centro){

  buraco <- centro

  n <- length(matriz)

  r <- c(rep(1, n))
```

```
  r <- c(r, r[length(r)])
  angle_shift <- pi / 2

  angles <- sort(seq(0, 2 * pi, length.out = n + 1) + angle_shift,
                 decreasing = TRUE)

  x <- r * cos(angles)

  y <- r * sin(angles)

  v_df <- data.frame(x, y)

  v <- as.matrix(round(v_df, 3))
  return(v = v)
}

# Função para criar eixos de origem

cria_eixos_origem <- function(matriz,centro){
  buraco <- centro

  n <- length(matriz)

  r <- c(rep(buraco, n))

  r <- c(r, r[length(r)])
  angle_shift <- pi / 2
  angles <- sort(seq(0, 2 * pi, length.out = n + 1) + angle_shift,
                 decreasing = TRUE)

  x <- r * cos(angles)
```

```
  y <- r * sin(angles)

  v_df <- data.frame(x, y)

  v <- as.matrix(round(v_df, 3))
  return(v = v)
}

# Função que converte um vetor de valores em um polC-gono do radar

line2pol <- function(vetor){

  r <- vetor

  angle_shift <- pi / 2

  angles <- sort(seq(0, 2 * pi, length.out = length(r)) + angle_shift,
                 decreasing = TRUE)

  x <- r * cos(angles)

  y <- r * sin(angles)

  v_df <- data.frame(x, y)

  v <- as.matrix(round(v_df, 3))
  return(v = v)
}

plot_radar <- function(dados, Gen, Prob, col1='dodgerblue',

col2 = 'violetred4',
```

```
col3,alpha = 0.25,centro = 0.1,cadeia = F, lin_hpd_gen = F,
                        hpd_med=F, lin_hpd_med = F){

# entrada de dados

# Caract <- Caract

Gen <- Gen

Prob <- Prob

centro <- centro

cad <- dados

buraco <- centro

alpha <- alpha

#cad <- data.frame(caract = factor(rep(1:5,each=nrow(c1))),cc)

cad  <- dados
#cad <- data.frame(caract = factor(rep(1:5,each=nrow(c1))),cc)

# Numero de caracteres, genótipos e tamanho das cadeias
gen_x
Car <- factor(cad[,1],levels = unique(cad[,1]))
p <- nlevels(Car)
N <- dim(cad)[1]/p
v <- dim(cad)[2]-1
```

```
Valor <- t(cad[,2:(v+1)])
Valor <-  apply(Valor,2,as.double)

# Define vetores para caracter, genotipo e valor

carac <- rep(Car,v)

Genotipo <- rep(1:v, p*N)

#Valor    <- t(cad[,2:(v+1)])

#Valor   <-  apply(Valor,2,as.double)# as.double(Valor)

# Calcula medias, minimos e maximos

vecValor <- as.double(c(Valor))

minc <- as.double(tapply(vecValor,carac,min))

maxc <- as.double(tapply(vecValor,carac,max))

medias_gerais <- as.double(tapply(vecValor,carac,mean))

medg <- t(matrix(rep(medias_gerais, N), p, N))

ming <- t(matrix(rep(minc, N), p, N))

maxg <- t(matrix(rep(maxc, N), p, N))

# Define o tamanho do centro do radar
#buraco <- 0.1
```

```
#for(j in 1:v){
#  for(col in 1:N){

    # Selecioando o genotipo

    gen_x <- (Valor[Gen,])

    carat  <- data.frame(caract = cad\$carac, gen= gen_x)

gen_j1 <- carat \%>\% group_by(caract) \%>\%

      mutate(n = row_number()) \%>\%
      spread(key = caract, value = gen)

gen_j1 <- gen_j1[,-1]

    gen_j <- as.data.frame(gen_j1)

# ajuste por genotipo
  mj <- gen_j - ming

  dif <- abs(maxg) + abs(ming)

  # min e maximo feito para o genotipo medio

  m_pre <- mj / dif

  m <- m_pre*(1 - buraco) + buraco

# media do efeito medio dos genotipos

  mg_rel_pre <- (medias_gerais - minc)/(abs(minc) + abs(maxc))
```

```
  medias_gerais_relativas <- mg_rel_pre * (1 - buraco) + buraco

# transformando a escala da resposta do genotipo

  r_inicial <- m

  r_inicial <- as.data.frame(r_inicial)

  rn <- c(r_inicial, r_inicial[1])

  rn <- as.data.frame(rn)
  angle_shift <- pi / 2

  angles <- sort(seq(0, 2 * pi, length.out = p + 1) + angle_shift,
                 decreasing = TRUE)

  anglesn <- matrix((rep(angles, nrow(rn))), byrow = TRUE, ncol = p + 1,
                    nrow = nrow(rn))

  xn <- rn * cos(anglesn)

  yn <- rn * sin(anglesn)

  definindo os valores para os eixos

  eixo <- cria_eixos(r_inicial,buraco)

  eixo_orig <- cria_eixos_origem(r_inicial,buraco)

# genotipo medio
```

```
  medias <- line2pol(c(medias_gerais_relativas,
  medias_gerais_relativas[1]))

  #lines(medias, col = 'black')

  # media da cadeia por genotipo

  medias_g <- cbind(apply(xn,2,mean),apply(yn,2,mean))
  # line2pol(c(ap, medias_gerais_g[1]))

  #lines(medias_g, col = 'green4',lwd=2)

# Agrupando os dados para fazer o HPD do feito medio

  gen_d <- data.frame(caract = carac, gen= vecValor)

  gen <- gen_d \%>\% group_by(caract) \%>\%
    mutate(n = row_number()) \%>\%
    spread(key = caract, value = gen)

  gen <- gen[,-1]
  length(gen)

  ci <- HPDinterval(as.mcmc(gen), prob = Prob)

  ci_inf_pre1 <- ci[, 1] - minc #apply(gen_j1,2,min)

  ci_inf_pre2 <- ci_inf_pre1 / dif[1,]
  ci_inf <- ci_inf_pre2 * (1 - buraco) + buraco
  ci_sup_pre1 <- ci[, 2] - minc#min_g

  ci_sup_pre2 <- ci_sup_pre1 / dif[1,]
```

```
  ci_sup <- ci_sup_pre2 * (1 - buraco) + buraco

  ci_inf_pol <- line2pol(c(ci_inf, ci_inf[1]))

  ci_sup_pol <- line2pol(c(ci_sup, ci_sup[1]))

# HPD do genotipo indice 'col'

  xn <- rn * cos(anglesn)

  yn <- rn * sin(anglesn)

  rnn <- HPDinterval(as.mcmc(rn), prob = Prob)

  xi <- rnn*cos(angles)

  yi <- rnn*sin(angles)

# Exibir

  inf_gen <- cbind((xi)[,1],(yi)[,1])

  sup_gen <- cbind((xi)[,2],(yi)[,2])

# Plot

  #par(mar = c(.2, .2, .2, .2))

  plot(eixo, type = 'l', xlim = c(-1, 1), ylim = c(-1, 1),

  frame.plot = FALSE, axes = FALSE, xlab = '',
```

```
ylab = '', lty = 2, lwd = 1.5,
             col = 'gray')

lines(eixo_orig, lty = 2, lwd = 1.5, col = 'gray')

mtext(side = 3, adj = 0.05, paste('Genotipo', Gen), line = -1.5)

for (i in 1:(nrow(eixo) - 1)) {
  lines(c(eixo_orig[i, 1], eixo[i, 1]),

  c(eixo_orig[i, 2], eixo[i, 2]), lwd = 1.5, col = 'gray', lty = 2)
}

#text(paste('y', 1:p, sep = ''), x = eixo[, 1], y = eixo[, 2])
text(paste(unique(Car), sep = ''), x = eixo[, 1], y = eixo[, 2])

if(cadeia == T){
  for (i in 1:nrow(xn)) {
    lines(xn[i, ], yn[i, ], col = alpha(rgb(0, 0, 1), 0.02), lwd = 0.5)
  }
}

polygon(xi,yi,
        col=adjustcolor(col1, alpha=alpha), border=NA)
lines(medias_g, col = col1,lwd=2) # media do genotipo

if(hpd_med==T){
polygon(cbind(ci_inf_pol[,1],ci_sup_pol[,1]),
        cbind(ci_inf_pol[,2],ci_sup_pol[,2]),
        col=adjustcolor(col2, alpha=alpha), border=NA)
}
```

```
  lines(medias, col = col2) # media geral

  if(lin_hpd_med==T){
  # HPD do genotipo medio
  lines(ci_inf_pol, col = col1, lty = 2)

  lines(ci_sup_pol, col = col1, lty = 2)
  }
  if(lin_hpd_gen==T){

# HPD do genotipo selecionado

  lines(inf_gen, col = col2, lty = 2)
  lines(sup_gen, col = col2, lty = 2)
  }

  comp <- medias >= medias_g

  comp1<- cbind(comp,medias,medias_g)
  print(comp1)

}

# plot

jpeg(filename = "genotipo_20.jpeg",width = 2440, height = 2440,res=300)
par(mar = c(.2, .2, .2, .2))

plot_radar(cad,9,Prob=0.95,col1='dodgerblue', col2 = 'violet', col3,
           alpha =0.25, centro = 0.1,

cadeia = F, lin_hpd_gen = F,hpd_med=F, lin_hpd_med = F)
```

```
dev.off()

par(mar = c(.2, .2, .2, .2))
for(i in 1:20){

plot_radar(cad,i,Prob=0.95,col1='dodgerblue', col2 = 'violet', col3,

alpha =0.25,

centro = 0.1,cadeia = F, lin_hpd_gen = F, hpd_med=F, lin_hpd_med = F)

}

par(mar = c(.2, .2, .2, .2))
#plot_radar(cad,9,Prob=0.95,col1='dodgerblue', col2 = 'white', col3,
            alpha =0.45, centro = 0.1,

#cadeia = F, lin_hpd_gen = F, hpd_med=F, lin_hpd_med = F)
#par(new=T)

plot_radar(cad,15,Prob=0.95,col1='orange', col2 = 'black', col3,
           alpha =0.45, centro = 0.1,
           cadeia = F, lin_hpd_gen = F, hpd_med=F, lin_hpd_med = F)
par(new=T)

plot_radar(cad,16,Prob=0.95,col1='blue', col2 = 'violet', col3,
           alpha =0.45, centro = 0.1,
           cadeia = F, lin_hpd_gen = F, hpd_med=F, lin_hpd_med = F)

plot_radar(cad,1,Prob=0.95,col1='orange', col2 = 'black', col3,
           alpha =0.45, centro = 0.1,
```

```
cadeia = F, lin_hpd_gen = F, hpd_med=F, lin_hpd_med = F)
par(new=T)

plot_radar(cad,20,Prob=0.95,col1='blue', col2 = 'violet', col3,
           alpha =0.45, centro = 0.1,
           cadeia = F, lin_hpd_gen = F, hpd_med=F, lin_hpd_med = F)

cad1<-subset(cad,cad[,1]=="Y2")
dim(cad1)

ci <- HPDinterval(as.mcmc(cad1[,-1]), prob = .95)

Med<-apply(cad1[,-1],2,mean)

Desv<-apply(cad1[,-1],2,sd)

Result<-cbind(Med,Desv)

ResultFLOR<-cbind(Result,ci)
#plot(mcmc(cad1[,-1]))

par(mfrow=c(1,2))
plot((cad1[,2]),type="l")
plot(density(cad1[,2]))

cad1<-subset(cad,cad[,1]=="1")
dim(cad1)

ci <- HPDinterval(as.mcmc(cad1[,-1]), prob = .95)
#plot(mcmc(cad1[,-1]))

par(mfrow=c(1,2))
```

```
plot((cad1[,2]),type="l")
plot(density(cad1[,2]))
#Medidas de variacao

Med<-apply(cad1[,-1],2,mean)
Desv<-apply(cad1[,-1],2,sd)

Result<-cbind(Med,Desv)

ResultFLOR<-cbind(Result,ci)

dim(Result)
dim(ci)
```

APENDICE E – Códigos R para executar a Analises Univariadas.

```
# Analises Univariadas por Local ( Retirar de dentro do loop antes
# de rodar essa função)

library(lme4)

library(lmerTest)

library(data.table)

#i=7

for (i in 7:ncol(dados))
{
  cat("Anovas Individuais (Local) - Variável:", colnames(dados)[i], "\n")

  dados.i <- dados[,c(1:8,i)]
  dados.i <- na.omit(dados.i)

  nomloc <- levels(droplevels(dados.i)\$local)

  statsloc <- matrix(0, nlevels(droplevels(dados.i)\$local),7)

  rownames(statsloc) <- levels(droplevels(dados.i)\$local)

  colnames(statsloc) <- c("Vg", "Ve", "P-valor - Ho:Vg=0", "Media",
                          "H2-Cullis","Acuracia", "CVe")

  #j=1
```

```
  for (j in 1:nlevels(droplevels(dados.i)\$local))
  {
    cat("Anova - Variável:", nomloc[j], colnames(dados)[i], "\n")

    dados.ij <- droplevels(dados.i[dados.i\$local==nomloc[j],])

    g.ran <-lmer(dados.ij[,ncol(dados.i)]-
    ~ rep + (1|rep:block) + (1|trat),
                data=dados.ij)

# to obtain var-cov-matrix for BLUPs of gen effect.

    vc   <- as.data.table(VarCorr(g.ran))

# extract estimated variance components (vc)

# R = varcov-matrix for error term

n    <- length(summary(g.ran)\$residuals)

# numer of observations

    vc.e <- vc[grp=="Residual", vcov]      # error vc

    R    <- diag(n)*vc.e                   # R matrix = I * vc.e

    n.g  <- summary(g.ran)\$ ngrps["trat"]    #number of genotypes

    vc.g <- vc[grp=="trat", vcov]
    # genotypic vc

    G.g <- diag(n.g)*vc.g              # gen part of G matrix = I * vc.g
```

```
# varcov-matrix for block effect
    n.b  <- summary(g.ran)\$ngrps["rep:block"]
    # number of incomplete blocks

    vc.b <- vc[grp=="rep:block", vcov] # incomplete block vc
    G.b  <- diag(n.b)*vc.b
    # incomplete block part of G matrix = I * vc.b

    G <- bdiag(G.g, G.b) # G is blockdiagonal with G.g and G.b

    # Design Matrices
    X <- as.matrix(getME(g.ran, "X")) # Design matrix fixed effects

    Z <- as.matrix(getME(g.ran, "Z")) # Design matrix random effects

    # Mixed Model Equation (HENDERSON 1986; SEARLE et al. 2006)

    C11 <- t(X) \%*\%solve(R) \%*\% X

    C12 <- t(X)\%*\%solve(R) \%*\% Z

    C21 <- t(Z) \%*\%solve(R) \%*\% X

    # C22 <- t(Z) \%*\%solve(R) \%*\% Z + solve(G)

    C22 <- t(Z) \%*\%solve(R) \%*\%Z + MASS::ginv(as.matrix(G))

    C <- as.matrix(rbind(cbind(C11, C12),
    # Combine components into one matrix C
                         cbind(C21, C22)))
```

```
# Mixed Model Equation Solutions
# C.inv <- solve(C)                                  # Inverse of C
C.inv <- MASS::ginv(C)

# Inverse of C
rownames(C.inv) <- colnames(C.inv) <- rownames(C)
Teste LRT
C22.g <- C.inv[levels(droplevels(dados.ij\$trat)),
         levels(droplevels(dados.ij\$trat))]

# subset of C.inv that refers to genotypic BLUPs
#dim(C22.g)

# Mean variance of BLUP-difference from C22 matrix of genotypic BLUPs
one        <- t(t(rep(1, n.g)))                # vector of 1s
P.mu       <- diag(n.g, n.g) - one \%*\% t(one)
# P.mu = matrix that centers for overall-mean
#vdBLUP.sum <- psych::tr(P.mu \%*\% C22.g)
vdBLUP.sum <- sum(diag(P.mu \%*\% C22.g))
# sum of all variance of differences = trace of P.mu*C22.g
vdBLUP.avg <- vdBLUP.sum * (2/(n.g*(n.g-1)))

# mean variance of BLUP-difference =
=divide sum by number of genotype pairs

# H2 Cullis
H2Cullis <- 1 - (vdBLUP.avg / 2 / vc.g)

# Acuracia
PEV <- sum(diag(C22.g))/n.g # PEV
rgg <- sqrt(1-(PEV/vc.g))
```

```
# Media
Media <- mean(dados.i[dados.i\$local==nomloc[j],ncol(dados.i)])

# CVe
CVe <- sqrt(vc.e)/Media*100

# Teste LRT das Comp. Var

ran.anova <- ranova(g.ran)

P.valor <- ran.anova["(1 | trat)","Pr(>Chisq)"]

statsloc[nomloc[j],1] <- vc.g

statsloc[nomloc[j],2] <- vc.e

statsloc[nomloc[j],3] <- P.valor

statsloc[nomloc[j],4] <- Media

statsloc[nomloc[j],5] <- H2Cullis

statsloc[nomloc[j],6] <- rgg

statsloc[nomloc[j],7] <- CVe

#resid <- residuals(lme.i,type=c("deviance"))

#print (shapiro.test(resid))

#qqnorm(resid, main = paste("Normal Q-Q Plot of" , nomloc[j],
colnames(dados)[i]))
```

```
    #identify(qqnorm(resid, main = paste("Normal Q-Q Plot of" , nomloc[j],
    colnames(dados)[i])))
    #dados.ij <- dados.i[dados.i\$LOCAL==nomloc[j],]

    #dados.ij[c(24,71,103),]
  }

  cat("Anovas Individuais (Local) - Variável:", colnames(dados)[i], "\n")
  print(statsloc)
}
```