

UNIVERSIDADE FEDERAL DE ALFENAS

RENÍSIO BRAULIO BALDINI

**APLICAÇÃO DO MODELO AMMI NA AVALIAÇÃO DA INTERAÇÃO GENÓTIPO ×
AMBIENTE DE ARROZ DE TERRAS ALTAS EM MINAS GERAIS**

ALFENAS/MG

2026

RENÍSIO BRAULIO BALDINI

**APLICAÇÃO DO MODELO AMMI NA AVALIAÇÃO DA INTERAÇÃO GENÓTIPO ×
AMBIENTE DE ARROZ DE TERRAS ALTAS EM MINAS GERAIS**

Dissertação apresentada ao Programa de Pós-Graduação em Estatística Aplicada e Biometria, Área de concentração em Estatística Aplicada e Biometria da Universidade Federal de Alfenas, MG, como parte dos requisitos para a obtenção do título de Mestre.

Orientador: Prof. Dr. Fabrício Goecking Avelar

Coorientadora: Prof. Dr. Carlos Pereira da Silva

ALFENAS/MG

2026

Sistema de Bibliotecas da Universidade Federal de Alfenas
Biblioteca Central

Baldini, Renísio Braulio.

Aplicação do modelo AMMI na avaliação da interação genótipo × ambiente de arroz de terras altas em Minas Gerais / Renísio Braulio Baldini. - Alfenas, MG, 2026.

62 f. : il. -

Orientador(a): Fabrício Goecking Avelar.

Dissertação (Mestrado em Estatística Aplicada e Biometria) - Universidade Federal de Alfenas, Alfenas, MG, 2026.

Bibliografia.

1. Interação G×A. 2. Modelo AMMI. 3. Algoritmo EM. 4. Dados Ausentes. 5. Estabilidade Genotípica. I. Avelar, Fabrício Goecking, orient.
II. Título.

Ficha gerada automaticamente com dados fornecidos pelo autor.

RENÍSIO BRÁULIO**AJUSTE DO MODELO AMMI COM DADOS FALTANTES UTILIZANDO O ALGORITMO EM: APLICAÇÃO A GENÓTIPOS DE ARROZ DE TERRAS ALTAS**

A Banca examinadora abaixo-assinada aprova a Dissertação apresentada como parte dos requisitos para a obtenção do título de Mestre em Estatística Aplicada e Biometria pela Universidade Federal de Alfenas. Área de concentração: Estatística Aplicada e Biometria.

Aprovada em: 15 de dezembro de 2025.

Prof. Dr. Carlos Pereira da Silva
Instituição: Universidade Federal de Lavras - UFLA

Prof. Dr. Luciano Antonio de Oliveira
Instituição: Universidade Federal da Grande Dourados - UFGD

Profa. Dra. Natalia da Silva Martins Fonseca
Instituição: Universidade Federal de Alfenas - UNIFAL-MG



Documento assinado eletronicamente por **Carlos Pereira da Silva, Usuário Externo**, em 15/12/2025, às 18:15, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://sei.unifal-mg.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1691315** e o código CRC **B05EECF7**.

Dedico este trabalho, primeiramente, a Deus, pela força e sabedoria concedidas ao longo desta caminhada; à minha esposa, Mislene, pelo amor, apoio e incentivo constantes.

AGRADECIMENTOS

Agradeço, primeiramente, a Deus, por me conceder força, sabedoria e perseverança ao longo desta caminhada, permitindo que eu superasse os desafios encontrados durante a realização deste trabalho. À minha esposa, Mislene, expresso minha profunda gratidão pelo amor, companheirismo, apoio e incentivo constantes, fundamentais para que eu pudesse seguir firme nesta trajetória. Agradeço à minha família, em especial ao meu pai, Sr. Anísio Braulio, e à minha mãe, Vanda Maria Braulio, por todos os ensinamentos, dedicação, apoio e valores transmitidos ao longo da minha vida. Sem vocês, esta conquista não seria possível. Estendo também meus agradecimentos aos meus amigos que, mesmo não sendo possível mencionar todos individualmente, contribuíram de diferentes formas com palavras de incentivo, apoio e amizade durante esta jornada.

Aos professores do Programa de Pós-Graduação em Estatística Aplicada e Biometria, agradeço pelos ensinamentos, experiências e conhecimentos compartilhados ao longo do curso. De forma especial, agradeço ao meu coorientador, Prof. Dr. Carlos Pereira da Silva, pelas valiosas contribuições, orientações e apoio acadêmico fundamentais para o desenvolvimento deste trabalho. Ao meu orientador, Prof. Dr. Fabrício Goecking Avelar, registro minha sincera gratidão pela dedicação, empenho, paciência e confiança depositada em mim ao longo desta trajetória, tornando possível a realização deste trabalho. Agradeço ao Programa de Pós-Graduação em Estatística Aplicada e Biometria e à sua equipe gestora pela oportunidade de crescimento acadêmico e profissional. Estendo também meus agradecimentos à Universidade Federal de Alfenas (UNIFAL-MG), por proporcionar ensino público, gratuito e de qualidade. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

A tua palavra é lâmpada que ilumina os meus passos e luz que clareia o meu caminho.

(Salmos 119:105)

RESUMO

O aumento da eficiência na produção agrícola depende fundamentalmente do melhoramento de plantas, estratégia que busca elevar a produtividade de forma sustentável. No entanto, o desempenho dos genótipos é frequentemente influenciado pelas variações ambientais, fenômeno conhecido como interação genótipo \times ambiente ($G \times A$). A compreensão dessa interação é crucial para a identificação de cultivares que apresentem adaptabilidade e estabilidade em diferentes cenários. Entre as metodologias estatísticas utilizadas, destaca-se o modelo AMMI (Additive Main Effects and Multiplicative Interaction), que integra a análise de variância (ANOVA) para os efeitos principais e a análise de componentes principais (ACP) para a estrutura multiplicativa da interação. Contudo, o modelo AMMI apresenta limitações significativas na presença de dados ausentes, comuns em experimentos de campo devido a falhas operacionais ou adversidades climáticas. Este trabalho propõe a utilização do algoritmo EM (Expectation-Maximization) como uma alternativa robusta para lidar com o desbalanceamento de dados. O algoritmo opera de forma iterativa nas etapas de Expectativa (E) e Maximização (M), permitindo a estimação consistente dos parâmetros mesmo com lacunas na matriz de dados. O objetivo principal é conduzir a análise AMMI-EM para identificar genótipos resilientes e fornecer recomendações precisas de cultivares. Espera-se que a integração do algoritmo EM ao modelo AMMI amplie a robustez analítica, permitindo interpretações mais confiáveis da interação $G \times A$ e fortalecendo a eficiência dos programas de melhoramento genético.

Palavras-chave: Interação $G \times A$, Modelo AMMI, Algoritmo EM, Dados Ausentes, Estabilidade Genotípica.

ABSTRACT

Increasing efficiency in agricultural production fundamentally depends on plant breeding, a strategy aimed at enhancing productivity sustainably. However, genotype performance is often influenced by environmental variations, a phenomenon known as genotype \times environment (G \times E) interaction. Understanding this interaction is crucial for identifying cultivars that exhibit adaptability and stability across different scenarios. Among the statistical methodologies used, the AMMI (Additive Main Effects and Multiplicative Interaction) model stands out, integrating analysis of variance (ANOVA) for main effects and principal component analysis (PCA) for the multiplicative structure of the interaction. Nevertheless, the AMMI model faces significant limitations in the presence of missing data, which are common in field experiments due to operational failures or climatic adversities. This work proposes the use of the EM (Expectation-Maximization) algorithm as a robust alternative for handling unbalanced data. The algorithm operates iteratively through Expectation (E) and Maximization (M) steps, allowing for consistent parameter estimation even with gaps in the data matrix. The primary objective is to conduct AMMI-EM analysis to identify resilient genotypes and provide accurate cultivar recommendations. It is expected that integrating the EM algorithm into the AMMI model will enhance analytical robustness, enabling more reliable interpretations of G \times E interaction and strengthening the efficiency of plant breeding programs.

Keywords: G \times E Interaction, AMMI Model, EM Algorithm, Missing Data, Genotypic Stability.

LISTA DE FIGURAS

Figura 1 – Folhas e seus rótulos prod x gen.....	41
Figura 2 – Folhas e seus rótulos prod x amb.....	42
Figura 3 – Análise dos resíduos do modelo.....	43
Figura 4 – Produtividade em cada ambiente.....	43
Figura 5 – Resíduos vs valores ajustados.	45
Figura 6 – Biplot AMMI1 mostrando a relação entre produtividade média e estabilidade dos genótipos.	47
Figura 7 – Biplot AMMI2 representando os dois primeiros componentes principais da interação (PC1 e PC2).....	48
Figura 8 – Biplot AMMI1 a 10% de perdas aleatórias.....	52
Figura 9 – Biplot AMMI2 a 10% de perdas aleatórias.....	53

LISTA DE TABELAS

Tabela 1 – Representação dos dados médios de (g) genótipos avaliados em (a) ambientes para um caráter genérico Y	18
Tabela 2 – ANOVA Genótipo x Ambiente ($G \times A$).	19
Tabela 3 – Parcelas da soma de quadrados da interação (SQ_{GA}).	24
Tabela 4 – Análise de variância pelo sistema de Gollob baseada em médias.....	26
Tabela 5 – Ambientes (E) compostos por locais agrícolas e anos.	32
Tabela 6 – Estrutura do planejamento experimental para o conjunto de dados de arroz de terras altas.....	32
Tabela 7 – Análise de variância para grupos de experimentos.	45
Tabela 8 – Teste F de Cornelius (1988) para determinação do número de componentes principais significativos da interação.	46
Tabela 9 – Média das métricas de desempenho (1000 cenários simulados de perdas aleatórias).....	49
Tabela 10 – Análise AMMI para a variável produtividade (dados ausentes imputados).....	50
Tabela 11 – Teste F de Cornelius (1988) para determinar o número de eixos responsáveis pela $G \times A$ com os dados incompletos com 10% de perdas aleatórias.	51

SUMÁRIO

1	INTRODUÇÃO	11
1.1	OBJETIVO GERAL	13
1.1.1	Objetivos específicos	13
2	REVISÃO DE LITERATURA	14
2.1	AS INTERAÇÕES GENÓTIPOS POR AMBIENTES (GXA)	14
2.2	MÉTODOS DE ANÁLISE DA INTERAÇÃO (GXA)	15
2.3	ANÁLISE AMMI	16
2.3.1	Primeiro estágio do ajuste do modelo AMMI	21
2.3.2	Segundo estágio do ajuste do modelo AMMI	22
2.3.3	Seleção do modelo AMMI	25
2.4	REPRESENTAÇÃO GRÁFICA EM BILOT	26
2.5	LIMITAÇÕES DA ANÁLISE AMMI	28
2.5.1	Algoritmo EM - Método para a imputação de dados faltantes	29
3	Materiais e Métodos	32
3.1	ÁREA EXPERIMENTAL	32
3.2	AMMI MODELO LINEAR	33
3.2.1	Função de Verossimilhança	34
3.2.2	Log-Verossimilhança	34
3.2.3	Derivadas Parciais	34
3.2.4	Sistema de Equações Normais na Forma Matricial	35
3.2.5	Detalhamento das Equações	36
3.3	EXPECTATION-MAXIMIZATION (EM)	37
3.3.1	Algoritmo EM para o Modelo AMMI	37
3.3.2	Etapa E (Expectation)	38
3.3.3	Etapa M (Maximization)	38
3.3.4	Critério de Convergência	38
3.3.5	Expressão de cada métrica da avaliação da imputação	39
4	RESULTADOS	41
4.1	ANÁLISE DESCRITIVA DOS DADOS	41
4.2	VERIFICAÇÃO DOS PRESSUPOSTOS DO MODELO	42
4.2.1	Homocedasticidade (igualdade de variâncias)	44

4.2.2	Análise AMMI	45
4.2.3	Determinação do número de eixos significativos pelo teste F de Cornelius . .	46
4.2.4	Análise gráfica: biplots AMMI1 e AMMI2	46
4.3	RESULTADOS DA IMPUTAÇÃO DE DADOS AUSENTES UTILIZANDO O MÉTODO EMAMMI	49
5	Conclusão	54
	REFERÊNCIAS	56

1 INTRODUÇÃO

A chave para a produção agrícola é aumentar a eficiência na utilização de recursos (aumento de produtividade por hectare). Tecnologias como sistemas de irrigação, uso de fertilizantes e pesticidas, uso de sementes geneticamente modificadas, melhoria de métodos de cultivo, diversificação de cultivos, técnicas de controle de qualidade, melhorias nos processos de armazenamento e transporte de alimentos, são alternativas que podem contribuir para o aumento de produtividade (TOTVS, 2021). Entre as alternativas, o melhoramento de plantas é uma das mais valiosas estratégias para o aumento da produtividade de forma sustentável e ecologicamente equilibrada, de acordo com Rohrig (2023). Com o avanço do conhecimento científico, os melhoristas, profissionais dedicados ao melhoramento genético, passaram a conduzir experimentos com o objetivo de compreender de forma mais aprofundada o comportamento e as variações das cultivares em diferentes ambientes.

Segundo Allard (1971), os genótipos podem apresentar desempenho diferenciado em função das variações ambientais. Esse comportamento variável é conhecido como interação genótipo \times ambiente ($G \times A$) e é extremamente relevante para o melhoramento de plantas. A interação ($G \times A$) pode ser usada para aprimorar as características dos cultivares, além disso, pode ajudar a prever como as características dos cultivares se comportarão em diferentes ambientes, a meta principal é verificar se há diferenças significativas entre os genótipos e entre os ambientes. Diversas metodologias estatísticas têm sido empregadas para avaliar a ($G \times A$). Entre elas, destaca-se a análise de variância (ANOVA), inicialmente proposta por Fisher & Mackenzie (1923), além de abordagens mais recentes, como modelos mistos, regressões e técnicas multivariadas. Contudo, a complexidade inerente aos dados provenientes de experimentos multiambiais exige métodos capazes de capturar adequadamente a estrutura da interação, motivando o desenvolvimento de abordagens mais robustas. Essas metodologias serão discutidas na seção de revisão de literatura.

Embora essas metodologias ofereçam informações valiosas, os desafios permanecem na modelagem precisa de interações complexas, particularmente em diversos contextos ambientais. Assim, mais pesquisas são necessárias para refinar essas abordagens e ampliar sua aplicabilidade no melhoramento de plantas. Nesse contexto, a análise AMMI (*Additive Main Effects and Multiplicative Interaction*) consolidou-se como uma das principais abordagens estatísticas para o estudo da interação. O modelo integra a análise de variância (ANOVA), responsável pela decomposição dos efeitos principais de genótipos e ambientes, à análise de componentes prin-

cipais (ACP), aplicada à estrutura multiplicativa da interação. Essa combinação metodológica permite explorar, de forma detalhada, os padrões de adaptação e estabilidade, possibilitando a identificação de genótipos resilientes em condições adversas, bem como daqueles que apresentam desempenho superior e consistente em múltiplos ambientes (Duarte; Vencovsky, 1999). Assim, a análise AMMI constitui ferramenta estratégica na recomendação de cultivares, fornecendo subsídios técnicos para decisões mais precisas em diferentes contextos agrícolas (Dias *et al.*, 2017).

Apesar de suas contribuições, o modelo AMMI apresenta limitações quando aplicado a conjuntos de dados com valores ausentes. A presença de informações incompletas pode comprometer a estimação dos efeitos principais e da interação, introduzir vieses e reduzir a precisão das inferências estatísticas, afetando a confiabilidade das recomendações obtidas (Cummings, 2013). Diante disso, torna-se necessária a adoção de métodos adequados para o tratamento de dados faltantes.

Nesse cenário, o algoritmo EM (*Expectation-Maximization*) destaca-se como uma alternativa metodológica robusta para estimação em modelos com dados incompletos. Fundamentado na maximização da função de verossimilhança, o algoritmo opera de forma iterativa em duas etapas: na etapa E (*Expectation*), calcula-se o valor esperado dos dados ausentes com base nos parâmetros estimados na iteração anterior; na etapa M (*Maximization*), atualizam-se os parâmetros por meio da maximização da verossimilhança. O processo é repetido até que um critério de convergência seja atendido, produzindo estimativas consistentes mesmo na presença de lacunas na matriz de dados (Paderewski; Rodrigues, 2014).

Considerando que experimentos agrícolas conduzidos em condições de campo frequentemente resultam em conjuntos de dados desbalanceados e sujeitos a perdas de informação, decorrentes de falhas experimentais, adversidades climáticas ou restrições operacionais, a incorporação do algoritmo EM ao ajuste do modelo AMMI representa um avanço metodológico relevante. Tal integração amplia a robustez analítica do modelo, viabiliza análises mais realistas e contribui para maior precisão na identificação de genótipos estáveis e adaptáveis, fortalecendo a eficiência dos programas de melhoramento genético.

1.1 OBJETIVO GERAL

O objetivo principal é conduzir uma análise AMMI, utilizando o algoritmo EM para lidar com dados ausentes. A finalidade é identificar genótipos que apresentem estabilidade e proporcionar recomendações específicas de cultivares para ambientes particulares, destacando a adaptabilidade.

1.1.1 Objetivos específicos

- a) Realizar a análise AMMI para avaliar a interação entre genótipos e ambientes;
- b) Tratar a questão dos dados ausentes por meio da aplicação do algoritmo EM;
- c) Verificar a consistência entre a análise AMMI com dados completos e a análise AMMI-EM para dados com perdas de parcelas;
- d) Comparar os padrões na análise biplot do modelo AMMI com os dados completos e do modelo com desbalanceamento via AMMI-EM.

2 REVISÃO DE LITERATURA

De acordo com Tortora, Funke & Case (2016), a relação dos genótipos com os ambientes abrange todos os tipos de organismos vivos como plantas, animais, fungos e até mesmo bactérias, que possuem variações genéticas (genótipos) e são influenciados pelo ambiente onde vivem. A interação entre fatores genéticos e ambientais é fundamental para compreender por que indivíduos ou organismos respondem de maneira diferente às condições do ambiente. Ela influencia tanto a predisposição a doenças em animais quanto a capacidade de plantas e bactérias resistirem a situações adversas, como secas ou a ação de antibióticos.

Na agricultura, os avanços na genômica e na engenharia genética permitiram o desenvolvimento de culturas que podem resistir ao estresse abiótico e, ao mesmo tempo, melhorar seus perfis nutricionais. Essa síntese de técnicas genéticas e métodos tradicionais de melhoramento genético é essencial para a agricultura sustentável e a segurança alimentar (Gangadhara *et al.*, 2024; Umesh *et al.*, 2024). Por isso, os estudos de melhoramento de plantas são favoráveis à humanidade no sentido de contribuir na melhoria do padrão de nutrição da população, isso engloba diversas técnicas de plantios incorporados nas espécies vegetais (Alessandro; Nazzareno; Teodoro, 2021).

2.1 AS INTERAÇÕES GENÓTIPOS POR AMBIENTES ($G \times A$)

Para compreender a interação ($G \times A$), é fundamental definir os conceitos de genótipo e fenótipo. Esses termos foram introduzidos na literatura científica em 1911 por Wilhelm Ludwig Johannsen, no artigo intitulado *The genotype conception of heredity*, publicado na revista *The American Naturalist*. Segundo Johannsen, o genótipo corresponde à constituição genética de um organismo, isto é, ao conjunto de genes herdados de seus ancestrais e responsáveis por seu potencial biológico. O fenótipo engloba as características observáveis influenciadas por fatores genéticos e ambientais. A interação ($G \times A$), inicialmente descrita por Fisher & Mackenzie (1923), refere-se ao desempenho diferencial dos genótipos quando avaliados em diferentes ambientes. Em seus estudos baseados na análise de variância (ANOVA), Fisher destacou a importância de compreender a $G \times A$ para aprimorar o rendimento das culturas e orientar estratégias de melhoramento. Na época, sugeriu que parte das interações poderia ser minimizada por meio do ajuste da escala ambiental, interpretando as variações genéticas como respostas diferencia-

das à sensibilidade ao ambiente, não atribuindo à interação um papel central na determinação das características, mas considerando-a, em certa medida, um obstáculo analítico.

Posteriormente, a interação $G \times A$ passou a ser reconhecida como uma importante fonte de variação fenotípica, ocorrendo quando genótipos distintos respondem de forma diferenciada às variações ambientais. Esse fenômeno não implica alterações no genótipo dos cultivares, mas reflete modificações na expressão fenotípica resultantes da influência do ambiente sobre o potencial genético existente. Assim, a variação observada decorre da combinação entre efeitos genéticos, efeitos ambientais e da própria interação entre esses fatores, sendo frequentemente difícil dissociar completamente as contribuições relativas da variância genética e ambiental (Falconer, 1987). Nesse contexto, a compreensão da interação torna-se essencial para a avaliação do desempenho dos genótipos, especialmente sob a perspectiva da adaptabilidade e da estabilidade.

De acordo com Cruz *et al.* (2004), a adaptabilidade e a estabilidade constituem aspectos fundamentais na determinação do comportamento dos genótipos em diferentes condições ambientais. A adaptabilidade refere-se à capacidade de um genótipo ajustar-se às variações do ambiente, aproveitando de forma eficiente os recursos disponíveis e respondendo positivamente tanto em condições favoráveis quanto adversas. Em termos práticos, trata-se da habilidade do genótipo em expressar bom desempenho produtivo diante de diferentes cenários ambientais. Por outro lado, a estabilidade diz respeito à habilidade do genótipo em manter seu desempenho consistente, independentemente das flutuações ambientais. A estabilidade de um genótipo é influenciada pela sua habilidade em enfrentar e responder aos desafios impostos pelo ambiente. Genótipos mais estáveis demonstram uma capacidade resiliente para lidar com diferentes condições ambientais, mantendo um desempenho consistente ao longo do tempo, essa estabilidade é essencial para garantir resultados previsíveis e confiáveis, mesmo em situações ambientais adversas.

2.2 MÉTODOS DE ANÁLISE DA INTERAÇÃO ($G \times A$)

Compreender como os genótipos respondem a diferentes condições ambientais é essencial para otimizar programas de melhoramento de plantas. A análise da ($G \times A$), inicialmente formalizada por Fisher e Mackenzie (1923), fundamenta-se na avaliação conjunta de múltiplas variáveis ambientais e produtivas, permitindo identificar padrões diferenciais de resposta dos genótipos. A incorporação de dados ambientais multivariados possibilita distinguir efeitos ge-

néticos dos efeitos do ambiente e da própria interação, fornecendo subsídios para estudos de estabilidade e adaptabilidade (Cruz *et al.*, 2004).

No contexto do melhoramento de plantas, compreender o comportamento dos genótipos em diferentes condições ambientais é essencial para estimar estabilidade e adaptabilidade. Metodologias estatísticas têm sido propostas ao longo dos anos para avaliar as respostas de cultivares em diferentes ambientes e analisar as variações genéticas em distintos contextos. Entre as abordagens mais tradicionais, destaca-se a análise de variância (ANOVA), uma das primeiras metodologias aplicadas na detecção de diferenças significativas entre genótipos e ambientes (Fisher; Mackenzie, 1923; Dias; Santos; Mexia, 2023). Com o avanço das técnicas estatísticas, passaram a ser utilizados modelos mistos, que possibilitam a consideração simultânea de efeitos fixos e aleatórios, conferindo maior flexibilidade às análises (Faveri, 2022; Wu *et al.*, 2014). Nesse contexto, métodos baseados em REML/BLUP também ganharam destaque por proporcionarem estimativas mais precisas dos parâmetros genéticos, enquanto modelos de regressão linear, tanto simples quanto múltipla, têm sido empregados para investigar a relação entre variáveis genéticas e ambientais (Li; Gutierrez, 2023). Além disso, técnicas multivariadas, como a análise de componentes principais (PCA), contribuem para a redução da dimensionalidade dos dados e para a interpretação dos padrões de variação (Sahu, 2013), enquanto o modelo GGE Biplot se destaca pela representação gráfica do desempenho relativo dos genótipos em diferentes ambientes (Dias; Santos; Mexia, 2023).

2.3 ANÁLISE AMMI

De acordo com Zobel *et al.* (1988), a interação $G \times A$ é uma forma importante de variabilidade na análise de variância, pois pode afetar drasticamente os resultados da análise. A incompreensão dessa interação pode impossibilitar previsões precisas de ganhos genéticos e prejudicar as recomendações de genótipos (cultivares), comprometendo o sucesso dos programas de melhoramento genético. Por este motivo, torna-se extremamente importante testar essa interação, para que se possa conhecer seus efeitos na análise e tomar decisões baseadas nisso. Segundo Kang & Magari (1996), a verdadeira consequência estatística de uma interação significativa entre fatores é que não é possível interpretar os efeitos principais de forma independente. Em outras palavras, os efeitos principais não podem ser combinados para prever os efeitos gerais, pois a interação interfere nessa previsão. De acordo com Crossa, Crossa, Gauch & Zobel

(1990) o uso de técnicas multivariadas — em particular o modelo AMMI (*Additive Main Effects and Multiplicative Interaction*) — tem se mostrado eficaz para explorar padrões complexos em interações $G \times A$ e melhorar a interpretação dos dados de ensaios multiambiente.

Os fundamentos da análise AMMI remontam a contribuições clássicas relacionadas à decomposição da interação em modelos lineares, destacando-se os trabalhos de Fisher & Mackenzie (1923), Gilbert (1963), Gollob (1968) e Mandel (1971), conforme discutido por Gauch (1992) e Cauwer & Ortiz (1998). A aplicação formal do método na agricultura foi iniciada por Kempton (1984), em estudos com trigo, sendo posteriormente consolidada em experimentos de interação conduzidos com soja por Gauch (1985) e Zobel *et al.* (1988).

A partir desses estudos, o modelo AMMI consolidou-se como uma abordagem alternativa e robusta para a análise da interação. Sua principal contribuição consiste na decomposição da estrutura da interação em componentes interpretáveis, permitindo identificar padrões de estabilidade e adaptabilidade do rendimento e, conseqüentemente, subsidiar estratégias mais eficientes de melhoramento genético Zobel *et al.* (1988). Formalmente, o modelo combina a análise de variância (ANOVA) para os efeitos principais com a análise de componentes principais (*Principal Component Analysis* — *PCA*) aplicada à matriz de interação, sendo amplamente utilizado na modelagem de dados provenientes de ensaios multiambientais. No modelo AMMI, a análise de variância (ANOVA) é utilizada para ajustar os efeitos principais de genótipos e ambientes. Em seguida, a matriz de resíduos da interação é submetida à decomposição por valores singulares (DVS), procedimento matemático que permite expressar essa matriz. A interpretação estatística desses componentes é realizada por meio da análise de componentes principais (PCA), que descreve a estrutura da interação $G \times A$, reduzindo sua complexidade e destacando as fontes estruturadas de variação. A partir dessa decomposição, torna-se possível determinar o número de componentes principais que explicam a maior parte da variação associada à interação.

Para exemplificar o conceito e as conseqüências da interação $G \times A$, considere uma variedade composta por genótipos (g) submetidos a testes experimentais em diferentes ambientes (locais, anos, entre outros). Nesse contexto, as médias observadas para cada combinação de genótipo e ambiente (tratamentos), obtidas a partir de repetições (r) em cada experimento — formando um conjunto de dados balanceado — podem ser organizadas em uma tabela ou matriz de dados representativa, denotada por Y_{ij} , conforme apresentado na Tabela 1.

Tabela 1 – Representação dos dados médios de (g) genótipos avaliados em (a) ambientes para um caráter genérico Y .

Y_{ij} de g genótipos em a ambientes	Ambientes j			
	1	2	...	a
Genótipo i				
1	Y_{11}	Y_{12}	...	Y_{1a}
2	Y_{21}	Y_{22}	...	Y_{2a}
\vdots	\vdots	\vdots	\ddots	\vdots
g	Y_{g1}	Y_{g2}	...	Y_{ga}

Fonte: Elaborado pelo autor.

O modelo mais simples e tradicional para analisar dados multiambientais organizados em uma tabela de dupla entrada, como apresentado na Tabela 1, é a análise de variância conjunta (ANOVA). Esse modelo permite avaliar os efeitos genotípicos, ambientais e a interação, sendo descrito pela seguinte expressão:

$$Y_{ij} = \mu + g_i + a_j + (ga)_{ij} + \varepsilon_{ij} \quad (1)$$

- Y_{ij} : é a resposta média do i -ésimo genótipo no j -ésimo ambiente;
- μ : é uma constante comum às respostas (normalmente a média geral);
- g_i : é o efeito do i -ésimo genótipo ($i = 1, 2, \dots, g$);
- a_j : é o efeito do j -ésimo ambiente ($j = 1, 2, \dots, a$);
- $(ga)_{ij}$: é o efeito da interação do i -ésimo genótipo com o j -ésimo ambiente;
- ε_{ij} : é o erro experimental médio, assumido independente e $\varepsilon_{ij} \sim N(0, \sigma^2)$.

A Tabela 2 apresenta a distribuição dos graus de liberdade associados à análise de variância utilizada para decompor os efeitos de genótipo, ambiente e da interação entre esses fatores. Essa organização segue o modelo clássico adotado em experimentos conduzidos em múltiplos ambientes, no qual os efeitos principais de genótipo e ambiente, bem como a interação GA , são dispostos nas respectivas fontes de variação da ANOVA. Ressalta-se que a tabela apresenta apenas a partição dos graus de liberdade, não configurando uma tabela de ANOVA completa. Para que a tabela fosse considerada uma ANOVA propriamente dita, seriam necessárias informações adicionais, como somas de quadrados, quadrados médios, valores da estatística F e seus

respectivos níveis de significância. Nesse sentido, conforme destacado por Zobel *et al.* (1988), embora a ANOVA permita identificar a presença da interação, ela trata essa interação como um único componente residual, o que limita sua capacidade de descrever a complexidade e as nuances existentes na relação entre genótipos e ambientes.

Tabela 2 – ANOVA Genótipo x Ambiente ($G \times A$).

Fonte de variação	Graus de Liberdade (GL)
Genótipo (G)	$g - 1$
Ambiente (A)	$a - 1$
Interação $G \times A$	$(g - 1)(a - 1)$
Erro médio	$ga(r - 1)$
Total	$gar - 1$

Fonte: Elaborado pelo autor.

- $g - 1$: graus de liberdade, onde “ g ” é o número de níveis ou tratamentos do fator genótipo;
- $a - 1$: graus de liberdade, onde “ a ” é o número de níveis ou tratamentos do fator ambiente;
- $(g - 1)(a - 1)$: graus de liberdade, representando a variação conjunta e específica de cada combinação de genótipo e ambiente;
- $ga(r - 1)$: Graus de Liberdade do erro médio.

Quando o efeito da interação é significativo, os efeitos principais deixam de ser interpretáveis de forma aditiva, pois o desempenho dos genótipos passa a depender das condições ambientais específicas. Nessa situação, torna-se inadequado generalizar os resultados para todos os ambientes, sendo necessário analisar o comportamento dos genótipos dentro de cada ambiente individualmente. Entretanto, esse desdobramento da interação, embora útil para detectar diferenças pontuais, não fornece informações estruturadas sobre a natureza da interação nem permite identificar, de forma clara, quais genótipos apresentam maior estabilidade ou responsividade às variações ambientais. Assim, embora a ANOVA possibilite detectar a presença da interação significativa, ela não descreve sua estrutura. Essa limitação fundamenta a adoção do modelo AMMI, que integra a análise de variância dos efeitos principais à análise de componentes principais aplicada à matriz de interação, permitindo decompor e interpretar a estrutura da GA de maneira mais detalhada.. Essa abordagem possibilita decompor a interação GA em

eixos principais, permitindo interpretar sua estrutura multiplicativa, identificar padrões de resposta diferencial dos genótipos e reduzir a complexidade dos dados, destacando os componentes mais relevantes. Além disso, o modelo AMMI proporciona maior precisão na estimação dos efeitos e melhor capacidade preditiva em experimentos conduzidos em múltiplos ambientes.

Como anteriormente mencionado, o modelo AMMI combina os efeitos principais aditivos de genótipos e ambientes com uma estrutura multiplicativa para descrever a interação. A forma geral do modelo AMMI pode ser representada como:

$$Y_{ij} = \mu + g_i + a_j + \sum_{k=1}^p \lambda_k u_k v_k^\top + \varepsilon_{ij} \quad \text{com } p = \min\{(g-1), (a-1)\} \quad (2)$$

onde:

- Y_{ij} : média observada do genótipo i no ambiente j ;
- μ : média geral;
- g_i : efeito principal do genótipo i ;
- a_j : efeito principal do ambiente j ;
- λ_k : valor singular associado ao componente k ;
- u_{ik} : escore (autovetor) do genótipo i no componente k ;
- v_{jk} : escore (autovetor) do ambiente j no componente k ;
- ε_{ij} : erro experimental, usualmente $\varepsilon_{ij} \sim N(0, \sigma^2)$.

O ajuste do modelo é realizado em dois estágios: no primeiro estágio, estima-se a componente aditiva da equação; no segundo estágio, modela-se a componente multiplicativa por meio da decomposição por valores singulares aplicada aos resíduos da interação, originando os termos. Segundo Segundo Dias & Krzanowski (2006), esta formulação destaca a flexibilidade do modelo AMMI em capturar a estrutura da interação de forma detalhada, ao mesmo tempo em que separa a parte aditiva dos dados.

2.3.1 Primeiro estágio do ajuste do modelo AMMI

A análise de variância efetuada conforme o modelo (2), utilizando os dados da Tabela (2), é de natureza simplificada. Os efeitos principais são determinados através das marginais, as marginais são as médias de cada genótipo em todos os ambientes. As estimativas dos efeitos principais são obtidas pelo método dos mínimos quadrados. A média geral do experimento é dada por $\bar{\mu} = \bar{Y}$. O efeito do genótipo i é definido como $\bar{g}_i = \bar{Y}_i - \bar{Y}$, enquanto o efeito do ambiente j é definido como $\bar{a}_j = \bar{Y}_j - \bar{Y}$, em que \bar{Y} representa a média geral de todas as observações, \bar{Y}_i corresponde à média do genótipo i considerando todos os ambientes e \bar{Y}_j corresponde à média do ambiente j considerando todos os genótipos.

$$\bar{\mu} = \bar{Y} \quad (3)$$

$$\bar{g}_i = \bar{Y}_i - \bar{Y} \quad (4)$$

$$\bar{a}_j = \bar{Y}_j - \bar{Y} \quad (5)$$

Substituindo os parâmetros pelas respectivas estimativas tem-se:

$$Y_{ij} = \bar{Y} + \bar{Y}_i - \bar{Y} + \bar{Y}_j - \bar{Y} + \varepsilon_{ij} \quad (6)$$

$$Y_{ij} = \bar{Y}_i + \bar{Y}_j - \bar{Y} + \varepsilon_{ij} \quad (7)$$

$$\varepsilon_{ij} = Y_{ij} - \bar{Y}_i - \bar{Y}_j + \bar{Y} \quad (8)$$

$$\varepsilon_{ij} = Y_{ij} - \bar{Y}_i - \bar{Y}_j + \bar{Y} = \bar{y}_{ga} = \bar{y}_{ij} \quad (9)$$

A matriz GA é utilizada na segunda etapa do ajuste do modelo AMMI para a obtenção dos parâmetros bilineares que descrevem o efeito da interação.

$$GA_{(ga)} = \begin{bmatrix} \widehat{(ga)}_{11} & \widehat{(ga)}_{12} & \cdots & \widehat{(ga)}_{1i} \\ \widehat{(ga)}_{21} & \widehat{(ga)}_{22} & \cdots & \widehat{(ga)}_{2i} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{(ga)}_{j1} & \widehat{(ga)}_{j2} & \cdots & \widehat{(ga)}_{ji} \end{bmatrix} \quad (10)$$

A matriz GA é utilizada na segunda etapa do ajuste do modelo AMMI para a obtenção dos parâmetros bilineares que descrevem o efeito da interação.

2.3.2 Segundo estágio do ajuste do modelo AMMI

No segundo estágio do modelo AMMI, os parâmetros bilineares são estimados por meio da decomposição por valores singulares (DVS), aplicada à matriz de resíduos da interação após o ajuste dos efeitos principais. A DVS permite decompor uma matriz qualquer (no caso a GA), em um produto de três matrizes.

$$GA_{(ga)} = U\Lambda V^T \quad (11)$$

Em que U : matriz de vetores próprios (autovetores) relacionados aos genótipos. S : matriz diagonal com os valores singulares (raiz quadrada dos autovalores) ordenados. V^T : matriz transposta dos vetores próprios relacionados aos ambientes. De forma equivalente, essa decomposição pode ser expressa como uma soma de componentes bilineares:

$$GA_{(ga)} = U\Lambda V^T = \sum_{k=1}^p \lambda_k u_k v_k^T \quad (12)$$

Onde cada termo $\lambda_k u_k v_k^T$, sendo λ_k o valor singular e u_k e v_k os vetores singulares associados aos efeitos genotípico e ambiental, respectivamente, para o k -ésimo componente principal. O número máximo de componentes extraíveis é dado por $p = \min\{g - 1, a - 1\}$. Essa representação possibilita quantificar a contribuição de cada componente principal para a estrutura da interação, permitindo interpretar sua natureza multiplicativa e facilitando a análise de estabilidade e adaptabilidade em experimentos multiambientais.

A aplicação da DVS, com as propriedades da ACP pode ser utilizada para construir uma representação aproximada de qualquer matriz real, independentemente de sua complexidade, usando um conjunto de parcelas (matrizes) ortogonais entre si e de posto unitário, totalizando p parcelas (Souza *et al.*, 2018).

$$(\mathbf{GA})_{ij} = \lambda_1 \mathbf{u}_1 \mathbf{v}_1^\top + \lambda_2 \mathbf{u}_2 \mathbf{v}_2^\top + \cdots + \lambda_p \mathbf{u}_p \mathbf{v}_p^\top \quad (13)$$

Essa decomposição facilita a interpretação da estrutura da interação, possibilitando identificar padrões de adaptabilidade e estabilidade observados em experimentos multiambientais. A seguir, apresentaremos a partição da soma de quadrados da interação genótipo-ambiente ($GA_{(ga)}$) na análise AMMI. Para tanto, enfatiza-se que:

$$(ga)_{ij} = \sum_{k=1}^p \lambda_k u_{ik} v_{jk}^\top. \quad (14)$$

A soma de quadrados da interação genótipo \times ambiente é definida como a soma dos quadrados de todos os elementos da matriz GA , isto é:

$$(SQ)_{ga} = \sum_{i=1}^g \sum_{j=1}^a (ga)_{ij}^2. \quad (15)$$

Substituindo a expressão de $(ga)_{ij}$ obtida pela DVS, verifica-se que a soma de quadrados da interação corresponde à soma das contribuições associadas a cada eixo singular. Como os vetores singulares são ortogonais, os termos cruzados se anulam, e a soma de quadrados da interação pode ser decomposta como:

$$(SQ)_{ga} = \lambda_1^2 + \lambda_2^2 + \cdots + \lambda_p^2 = \sum_{k=1}^p \lambda_k^2. \quad (16)$$

Essa expressão mostra que cada valor singular λ_k^2 representa a parcela da soma de quadrados da interação associada ao k -ésimo eixo da decomposição, o qual corresponde a um componente principal da interação (IPCA). Assim, a soma de quadrados total da interação pode ser entendida como a soma das contribuições individuais de todos os eixos da DVS. No contexto da análise AMMI, nem todos esses eixos são retidos no modelo final. Os primeiros componentes, associados aos maiores valores singulares, descrevem a parte sistemática da interação, denominada padrão, enquanto os componentes restantes estão associados a variações aleatórias,

interpretadas como ruído. Dessa forma, a soma de quadrados da interação pode ser particionada em SQ_{ga} [padrão] que orresponde à soma dos quadrados dos componentes principais retidos no modelo AMMI e SQ_{ga} [ruídos] que representa a contribuição dos componentes não utilizados.

$$SQ_{ga} = SQ_{ga} \text{ [padrão]} + SQ_{ga} \text{ [ruídos]} \quad (17)$$

Com base no número de termos (eixos singulares ou componentes principais) utilizados para descrever o padrão da interação, o modelo recebe diferentes denominações, como AMMI0, AMMI1, ..., AMMIF. Retendo todos os eixos dessa decomposição, o modelo pode consumir uma quantidade significativa de graus de liberdade, apesar de ajustar perfeitamente a matriz de médias. Desta forma, o objetivo na é condensar a maior parte da interação $GA_{(ga)}$ em apenas alguns eixos SQ_{ga} [padrão], resultando na escolha de um modelo AMMI reduzido, que elimina um resíduo adicional SQ_{ga} [ruídos].

Tabela 3 – Parcelas da soma de quadrados da interação (SQ_{GA}).

Componente	SQ_{IPCA}	AMMI
$\lambda_k u_k v_k^\top$	SQ_{IPCA_k}	AMMI1
$\lambda_k u_k v_k^\top + \lambda_k u_k v_k^\top$	SQ_{IPCA_k}	AMMI2
$\lambda_k u_k v_k^\top + \lambda_k u_k v_k^\top + \lambda_k u_k v_k^\top$	SQ_{IPCA_3}	AMMI3
\vdots	\vdots	\vdots
$\sum_{k=1}^p \lambda_k u_k v_k^\top$	SQ_{IPCA_p}	AMMI p

Fonte: Elaborado pelo autor.

- SQ_{IPCA_p} : representa a soma de quadrados associada ao k -ésimo componente principal da interação $G \times A$, sendo dada por λ_k^2 . Dessa forma, a soma de quadrados acumulada até o p -ésimo componente é expressa por $\sum_{k=1}^p \lambda_k^2$. A proporção da variabilidade da interação explicada pelo modelo AMMI k é obtida pela razão entre SQ_{AMMI_k} e SQ_{ga} , indicando a fração da soma de quadrados total da interação capturada pelos k primeiros eixos principais.
- AMMI p : modelo AMMI completo, obtido pela incorporação dos p componentes principais da interação, explicando a totalidade da soma de quadrados da interação SQ_{ga} .

Para resumir a interação $(ga)_{ij}$, é necessário realizar testes específicos para identificar quais componentes principais (IPCA_{*p*}) são estatisticamente significativos. A IPCA₁ captura a maior parte das variações presentes na interação, enquanto os componentes seguintes apresentam importância decrescente. Com os componentes significativos identificados, é possível selecionar o modelo AMMI mais adequado, reduzindo a dimensionalidade sem perder informações relevantes. A dimensão do modelo escolhido, ou seja, o número de componentes IPCA incluídos, representa quantos padrões principais da interação estão sendo considerados na análise (Gauch; Zobel, 1996).

Os primeiros componentes principais (PCs) capturam uma parte significativa da variância presente na interação (SQ_{GA}), de forma que, à medida que a ordem dos componentes aumenta, a variância capturada diminui. Esse comportamento decrescente em relação aos componentes k indica que componentes de ordens superiores representam proporções menores da variação da interação. Assim, os primeiros PCs podem ser suficientes para capturar grande parte da soma de quadrados da interação, permitindo uma escolha parcimoniosa do modelo AMMI, que consome poucos graus de liberdade e é capaz de descrever a variabilidade da interação.

2.3.3 Seleção do modelo AMMI

Gauch (1992) apresenta critérios usados para o diagnóstico de modelos na literatura AMMI, destacando a significância estatística e a precisão preditiva. Esses critérios desempenham papéis distintos na avaliação e seleção dos modelos mais adequados para análises agrícolas. Embora a significância estatística seja amplamente utilizada, é crucial também considerar a precisão preditiva e a interpretabilidade, especialmente em contextos onde esses aspectos são essenciais para a tomada de decisões. Na análise AMMI, a significância estatística dos componentes principais é geralmente avaliada por meio de testes F aproximados, já que o modelo não é estritamente linear. Nesse contexto, destaca-se inicialmente a proposta de Gollob (1968), amplamente adotada devido à sua simplicidade de aplicação e ao uso de operações aritméticas básicas para atribuição dos graus de liberdade. Assim, o teste F aproximado com base em Gollob tornou-se o mais utilizado para diagnosticar componentes significativos no modelo AMMI. Como alternativa metodológica mais refinada, Cornelius (1993) propôs um teste F com fundamentação estatística mais rigorosa, permitindo uma avaliação mais precisa da significância dos

componentes principais. Essa abordagem é especialmente útil em situações de maior complexidade de interação, contribuindo para a escolha de modelos mais parcimoniosos e robustos.

A seleção do modelo mais adequado, como AMMI0, AMMI1, até AMMIP, é determinada com base na significância dos testes F para os termos sucessivos da interação. O resíduo AMMI, composto pelos termos remanescentes da SQ_{ga} (que representam os ruídos), também pode ser testado para confirmar sua insignificância. Gauch & Zobel (1996) discutem várias abordagens para atribuir graus de liberdade a um modelo AMMI, incluindo o sistema de Gollob (1968). No entanto, eles apontam a falta de consenso entre esses sistemas, deixando a escolha entre eles aberta a considerações práticas e teóricas. O sistema aproximado de Gollob é reconhecido por sua simplicidade, pois requer apenas operações aritméticas básicas para calcular os graus de liberdade, que podem ser expressos como $GL_{IPCA_n} = g + a - 2n$. Na tabela (4) ilustra a estrutura da análise conjunta completa utilizando o sistema de Gollob.

Tabela 4 – Análise de variância pelo sistema de Gollob baseada em médias.

Fontes de variação	Graus de liberdade	Soma de quadrados
Genótipo (G)	$g - 1$	SQ_g
Ambiente (A)	$a - 1$	SQ_a
Interação GA	$(g - 1)(a - 1)$	SQ_{ga}
IPCA 1	$g + a - 1 - 2 \cdot 1$	λ_1^2
IPCA 2	$g + a - 1 - 2 \cdot 2$	λ_2^2
IPCA 3	$g + a - 1 - 2 \cdot 3$	λ_3^2
...
IPCA p	$g + a - 1 - 2 \cdot p$	λ_p^2
Erro médio	$g \cdot (a - 1) \cdot (r - 1)$	SQ_r
Total	$g \cdot (a - 1) \cdot (r - 1)$	SQ_{TOTAL}

Fonte: Elaborado pelo autor com base em Gauch (2006).

2.4 REPRESENTAÇÃO GRÁFICA EM BILOT

O biplot é uma ferramenta gráfica proposta por Gabriel (1971) que permite representar simultaneamente linhas e colunas de uma matriz de dados em um mesmo plano, facilitando a interpretação de relações multivariadas. No contexto da análise AMMI, o biplot é utilizado para visualizar e interpretar a interação entre genótipos e ambientes a partir dos Componentes Principais de Interação (IPCA). Essa representação gráfica auxilia na identificação de padrões de

adaptação e estabilidade, evidenciando como os genótipos respondem aos diferentes ambientes e vice-versa.

Existem dois tipos principais de biplots derivados do modelo AMMI: o biplot AMMI1 e o biplot AMMI2. No biplot AMMI1, o eixo horizontal representa as médias ajustadas dos genótipos e ambientes, enquanto o eixo vertical apresenta os escores do primeiro componente principal da interação (IPCA1). Esse tipo de gráfico combina informação de desempenho médio e estabilidade: genótipos e ambientes com escores de IPCA1 próximos de zero são considerados mais estáveis, enquanto aqueles com valores mais extremos indicam maior interação específica. Assim, o AMMI1 é particularmente útil para identificar genótipos de desempenho médio elevado e comportamento previsível em diferentes ambientes.

O biplot AMMI2, por sua vez, é construído com base nos dois primeiros componentes principais da interação, IPCA1 e IPCA2, representados nos eixos horizontal e vertical, respectivamente. Nesse gráfico, tanto genótipos quanto ambientes são posicionados como pontos em um plano bidimensional, e a proximidade entre eles indica associação positiva em termos de desempenho. O AMMI2 permite uma interpretação mais detalhada da estrutura da interação, possibilitando a identificação de agrupamentos e padrões complexos de adaptação específica.

A interpretação do biplot AMMI2 segue alguns princípios básicos:

- Proximidade entre pontos: Indica similaridade nos padrões de resposta entre genótipos ou ambientes.
- Ângulo entre vetores: Expressa a correlação entre os componentes principais; ângulos pequenos sugerem correlação positiva, perpendiculares indicam ausência de correlação e ângulos obtusos, correlação negativa.
- Comprimento dos vetores: Representa a contribuição relativa do componente principal para a interação; vetores mais longos indicam maior influência.
- Distância em relação à origem: Pontos próximos ao centro indicam estabilidade e baixa interação, enquanto pontos mais distantes refletem maior variabilidade e adaptação específica.

De modo geral, o biplot AMMI1 oferece uma visão mais simples e direta da relação entre média e estabilidade, enquanto o biplot AMMI2 fornece uma perspectiva mais abrangente sobre a estrutura da interação, permitindo interpretações mais refinadas e a identificação de padrões multivariados de adaptação e desempenho.

2.5 LIMITAÇÕES DA ANÁLISE AMMI

A ANOVA permite comparar as médias de diferentes grupos e determinar se as diferenças observadas são devidas ao acaso ou se são estatisticamente significativas. No entanto, como qualquer método estatístico, ela se baseia em algumas premissas importantes. Segundo Schmider *et al.* (2010) as principais premissas da ANOVA incluem a normalidade dos erros e a homogeneidade das variâncias, quando essas premissas não são atendidas, os resultados da ANOVA podem ser incorretos. A não conformidade com as suposições da ANOVA inviabiliza sua aplicação adequada de várias maneiras. Por exemplo, a violação da normalidade dos erros pode levar a resultados enganosos nos testes de significância, enquanto a violação da homogeneidade das variâncias pode distorcer os resultados dos testes F, isso impossibilita a avaliação da significância das interações entre fatores, pois os testes F para as interações assumem que as distribuições dos quadrados médios seguem uma distribuição F sob a hipótese nula. Quando a premissa de homogeneidade das variâncias é violada, o valor do teste F para a interação pode ser inflacionado ou deflacionado, levando a conclusões incorretas. Para lidar com essas situações, são usados testes F aproximados, que ajustam a análise para considerar as violações das premissas.

Para contornar essas limitações, diversas abordagens têm sido propostas por Figueiredo (2017), Malik, Forkman & Piepho (2019). Eles apresentaram um algoritmo de reamostragem baseado em bootstrap não paramétrico e permutação, visando selecionar o número apropriado de termos multiplicativos significativos na decomposição da interação.

Outra alternativa eficaz envolve o uso de modelos mistos, conforme descrito por Piepho (1997). Esses modelos oferecem maior flexibilidade para tratar dados desbalanceados e com variâncias heterogêneas, além de possibilitarem uma modelagem mais precisa da interação. Tais modelos tratam os efeitos de genótipos e ambientes como fixos ou aleatórios, de acordo com a estrutura do experimento, sendo particularmente úteis na presença de dados ausentes ou desbalanceados. Nesse contexto, os modelos mistos têm sido amplamente empregados na análise da interação, pois permitem maior flexibilidade na modelagem da variabilidade e melhor ajuste em situações que violam pressupostos clássicos da análise de variância Silva (2023). Além dos modelos mistos, destacam-se também as abordagens bayesianas, que vêm ganhando espaço na análise da interação $G \times A$ por possibilitarem a incorporação explícita de incertezas por meio da especificação de distribuições a priori para

os parâmetros do modelo. Essas abordagens têm se mostrado promissoras especialmente em cenários com elevada complexidade estrutural, dados ausentes e heterogeneidade de variâncias, ampliando a capacidade inferencial e preditiva dos modelos utilizados Oliveira, Silva & Silva (1954). No contexto da análise robusta, o modelo R-AMMI, proposto por Rodrigues, Monteiro & Lourenço (2016), constitui uma extensão da abordagem tradicional, utilizando métodos robustos em substituição à ANOVA e à DVS, com bons resultados evidenciados por meio de simulações de Monte Carlo.

Outra limitação da análise AMMI é a ocorrência de dados ausentes, que podem surgir por diversas razões, como erros na coleta, falhas experimentais ou problemas de registro. A ausência de observações resulta em perda direta de informações valiosas, comprometendo a precisão e a confiabilidade das conclusões obtidas pela análise AMMI (Crossa; Gauch; Zobel, 1990).

2.5.1 Algoritmo EM - Método para a imputação de dados faltantes

A análise AMMI tradicional apresenta limitações quando aplicada a conjuntos de dados com valores ausentes. A presença de lacunas nos dados experimentais pode comprometer a representatividade das amostras e alterar a distribuição dos valores observados, resultando em estimativas viesadas dos parâmetros do modelo e, conseqüentemente, em interpretações imprecisas da interação $G \times A$. Uma estratégia para contornar esse problema consiste em aplicar modelos estatísticos capazes de estimar as funções ou parâmetros relevantes com base apenas nas observações disponíveis. Trabalhos clássicos, como os de Birkes & Dodge (1976), Dodge (1985) e Dodge & Zoppe (2004), apresentam revisões abrangentes sobre métodos que buscam ajustar modelos mesmo na presença de dados faltantes, sem necessariamente realizar substituições diretas. Por outro lado, uma abordagem distinta foi proposta por Rubin (1996) e Schafer & Graham (2002), que recomendam o uso de técnicas de imputação múltipla paramétrica ou de métodos baseados em verossimilhança, como o algoritmo EM (*Expectation-Maximization*). Essas abordagens procuram estimar os valores ausentes de maneira probabilística, preservando a estrutura e a variabilidade dos dados originais.

O algoritmo EM é um procedimento iterativo amplamente utilizado na estimação de parâmetros em modelos com dados incompletos. Ele opera em duas etapas principais: (i) a etapa de *Expectation* (E), em que se calcula o valor esperado da função de verossimilhança

completa, condicionada aos dados observados e aos parâmetros estimados na iteração anterior; e (ii) a etapa de *Maximization* (M), na qual os parâmetros são atualizados de modo a maximizar essa função esperada. Formalmente, considere um conjunto de dados observados $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$, que dependem de variáveis ausentes $\mathbf{Z} = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N\}$. O objetivo do algoritmo é estimar os parâmetros θ que maximizam a função de verossimilhança marginal dos dados observados, dada por:

1. Etapa E (Expectation): Calcula a expectativa da log-verossimilhança completa, condicionada aos dados observados \mathbf{X} e à estimativa atual dos parâmetros $\theta^{(t)}$. Esta etapa envolve o cálculo do valor esperado da log-verossimilhança com relação à distribuição posterior das variáveis latentes \mathbf{Z} :

$$Q(\theta|\theta^{(t)}) = \mathbb{E}_{\mathbf{Z}|\mathbf{X},\theta^{(t)}}[\log p(\mathbf{X}, \mathbf{Z}|\theta)] \quad (18)$$

- $Q(\theta|\theta^{(t)})$: Função de expectativa, que depende dos novos parâmetros θ e dos parâmetros atuais $\theta^{(t)}$.
- $\mathbb{E}_{\mathbf{Z}|\mathbf{X},\theta^{(t)}}$: Expectativa com relação à distribuição posterior das variáveis latentes \mathbf{Z} , dado os dados observados \mathbf{X} e os parâmetros atuais $\theta^{(t)}$.
- $\log p(\mathbf{X}, \mathbf{Z}|\theta)$: Logaritmo da probabilidade conjunta dos dados observados e das variáveis latentes, dado os novos parâmetros θ .

2. Etapa M (Maximization): Atualiza os parâmetros θ para maximizar a expectativa calculada na etapa E:

$$\theta^{(t+1)} = \arg \max_{\theta} Q(\theta|\theta^{(t)}) \quad (19)$$

- $\theta^{(t+1)}$: Novos parâmetros estimados após a iteração t .
- $\arg \max_{\theta}$: Operação que encontra os parâmetros θ que maximizam a função de expectativa $Q(\theta|\theta^{(t)})$.

O algoritmo alterna entre essas duas etapas até que a mudança nos parâmetros estimados seja menor que um determinado limiar ou até que o número máximo de iterações seja alcançado.

A convergência é garantida pelo fato de que, ao longo das iterações, a função de verossimilhança $L(\theta; \mathbf{X})$ não diminuirá e, frequentemente, aumentará até que a convergência seja alcançada (Dempster; Laird; Rubin, 1977).

Segundo Gauch & Zobel (1990), o algoritmo Expectation-Maximization (EM) é implementado para ajustar o modelo AMMI apesar dos dados ausentes. Esta versão do AMMI que lida com dados ausentes é denominada "EM-AMMI". O EM-AMMI é utilizado para quantificar as informações diretas e indiretas em um ensaio de rendimento, fornecendo insights teóricos sobre o ganho em precisão observado e sobre o processo de imputação de dados ausentes.

3 MATERIAIS E MÉTODOS

3.1 ÁREA EXPERIMENTAL

Treze genótipos de arroz de terras altas foram avaliados em nove ambientes distintos, totalizando 351 observações. Os experimentos foram conduzidos em três localidades do estado de Minas Gerais — Lavras, Patos de Minas e Lambari — ao longo de três safras agrícolas (2013/14, 2014/15 e 2015/16), conforme detalhado na Tabela 5.

Tabela 5 – Ambientes (E) compostos por locais agrícolas e anos.

Ambiente (E)	Local	Ano agrícola
E1	UFLA/Lavras	2013/14
E2	Patos de Minas	2013/14
E3	UFLA/Lavras	2014/15
E4	Patos de Minas	2014/15
E5	Lambari	2014/15
E6	Muquém/Lavras	2014/15
E7	Patos de Minas	2015/16
E8	Lambari	2015/16
E9	UFLA/Lavras	2015/16

Fonte: Próprio autor.

Os experimentos foram conduzidos em delineamento em blocos casualizados (DBC), com três repetições em cada ambiente. Cada parcela experimental foi composta por quatro linhas de 5,0 m de comprimento, espaçadas em 0,35 m, sendo consideradas para a colheita apenas as duas linhas centrais. A adubação e o manejo fitossanitário seguiram as recomendações técnicas para o cultivo de arroz de terras altas, de modo a minimizar efeitos não controlados de ambiente. A variável avaliada foi a **produtividade de grãos** ($\text{kg}\cdot\text{ha}^{-1}$), ajustada para umidade de 13%. Cada genótipo foi avaliado em todas as combinações de ambientes e repetições, compondo um delineamento balanceado (13 genótipos \times 9 ambientes \times 3 repetições = 351 observações). A estrutura geral do planejamento experimental está apresentada na Tabela 6.

Tabela 6 – Estrutura do planejamento experimental para o conjunto de dados de arroz de terras altas.

Fator	Níveis	Observações por nível	Total
Genótipos	13	27	351
Ambientes	9	39	351
Repetições	3	117	351

Fonte: Elaborado pelo autor a partir do delineamento experimental.

De acordo com Inácio *et al.* (2018), as localidades utilizadas apresentam variações edafoclimáticas expressivas, com predomínio de clima tropical semiúmido em Patos de Minas e Lambari e tropical de altitude em Lavras, o que contribui para a expressão diferencial dos genótipos avaliados.

3.2 AMMI MODELO LINEAR

O modelo AMMI combina a análise de variância (ANOVA) com a análise de componentes principais (PCA) para descrever a (G×A). Contudo, o modelo AMMI não é ajustado diretamente aos dados brutos; inicialmente, estima-se o modelo fatorial aditivo, e somente após essa etapa realiza-se a decomposição da interação por meio da PCA. O modelo linear fatorial utilizado como base para o processo de estimação e imputação é expresso como:

$$\mathbf{Y} = \mathbf{J}\mu + \mathbf{Z}\mathbf{g} + \mathbf{X}_1\mathbf{a} + \mathbf{W}\mathbf{ga} + \mathbf{X}_2\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \quad (20)$$

onde:

Y : vetor ($n \times 1$) de respostas observadas;

J : vetor ($n \times 1$) de 1's associado à média geral μ ;

Z : matriz ($n \times g$) de delineamento para os efeitos dos genótipos (**g**);

X₁ : matriz ($n \times a$) de delineamento para os efeitos dos ambientes (**a**);

W : matriz ($n \times ga$) de delineamento para os efeitos da interação (**ga**);

X₂ : matriz ($n \times b$) de delineamento para os efeitos dos blocos dentro dos ambientes (**γ**);

$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_n)$: vetor ($n \times 1$) de erros aleatórios independentes e homocedásticos.

(21)

Nesse contexto, a imputação de dados ausentes foi realizada sobre o efeito da interação ga , uma vez que ele constitui o principal componente de variação de interesse no modelo AMMI. Após essa estimação inicial via algoritmo EM, a interação é decomposta por análise de componentes principais, obtendo-se o modelo AMMI convencional, expresso como:

3.2.1 Função de Verossimilhança

Assumindo que os erros seguem uma distribuição normal multivariada com média zero e variância constante, tem-se:

$$\varepsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{I}) \quad (22)$$

A função de verossimilhança é dada por:

$$L(\mu, \mathbf{g}, \mathbf{a}, \mathbf{ga}, \gamma, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left(-\frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{J}\mu - \mathbf{Z}\mathbf{g} - \mathbf{X}_1\mathbf{a} - \mathbf{W}\mathbf{ga} - \mathbf{X}_2\gamma)^\top (\mathbf{Y} - \mathbf{J}\mu - \mathbf{Z}\mathbf{g} - \mathbf{X}_1\mathbf{a} - \mathbf{W}\mathbf{ga} - \mathbf{X}_2\gamma) \right) \quad (23)$$

3.2.2 Log-Verossimilhança

Tomando o logaritmo da função de verossimilhança, obtemos:

$$\log L(\mu, \mathbf{g}, \mathbf{a}, \mathbf{ga}, \gamma, \sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{J}\mu - \mathbf{Z}\mathbf{g} - \mathbf{X}_1\mathbf{a} - \mathbf{W}\mathbf{ga} - \mathbf{X}_2\gamma)^\top (\mathbf{Y} - \mathbf{J}\mu - \mathbf{Z}\mathbf{g} - \mathbf{X}_1\mathbf{a} - \mathbf{W}\mathbf{ga} - \mathbf{X}_2\gamma) \quad (24)$$

3.2.3 Derivadas Parciais

Para obter as equações normais, derivamos a log-verossimilhança em relação a cada parâmetro:

1. Para μ :

$$\frac{\partial \log L}{\partial \mu} = \frac{1}{\sigma^2} \mathbf{J}^\top (\mathbf{Y} - \mathbf{J}\mu - \mathbf{Z}\mathbf{g} - \mathbf{X}_1\mathbf{a} - \mathbf{W}\mathbf{ga} - \mathbf{X}_2\gamma) = \mathbf{0}$$

2. Para \mathbf{g} :

$$\frac{\partial \log L}{\partial \mathbf{g}} = \frac{1}{\sigma^2} \mathbf{Z}^\top (\mathbf{Y} - \mathbf{J}\mu - \mathbf{Z}\mathbf{g} - \mathbf{X}_1\mathbf{a} - \mathbf{W}\mathbf{g}\mathbf{a} - \mathbf{X}_2\gamma) = \mathbf{0}$$

3. Para \mathbf{a} :

$$\frac{\partial \log L}{\partial \mathbf{a}} = \frac{1}{\sigma^2} \mathbf{X}_1^\top (\mathbf{Y} - \mathbf{J}\mu - \mathbf{Z}\mathbf{g} - \mathbf{X}_1\mathbf{a} - \mathbf{W}\mathbf{g}\mathbf{a} - \mathbf{X}_2\gamma) = \mathbf{0}$$

4. Para $\mathbf{g}\mathbf{a}$:

$$\frac{\partial \log L}{\partial \mathbf{g}\mathbf{a}} = \frac{1}{\sigma^2} \mathbf{W}^\top (\mathbf{Y} - \mathbf{J}\mu - \mathbf{Z}\mathbf{g} - \mathbf{X}_1\mathbf{a} - \mathbf{W}\mathbf{g}\mathbf{a} - \mathbf{X}_2\gamma) = \mathbf{0}$$

5. Para γ :

$$\frac{\partial \log L}{\partial \gamma} = \frac{1}{\sigma^2} \mathbf{X}_2^\top (\mathbf{Y} - \mathbf{J}\mu - \mathbf{Z}\mathbf{g} - \mathbf{X}_1\mathbf{a} - \mathbf{W}\mathbf{g}\mathbf{a} - \mathbf{X}_2\gamma) = \mathbf{0}$$

6. Para σ^2 :

$$\frac{\partial \log L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} (\mathbf{Y} - \mathbf{J}\mu - \mathbf{Z}\mathbf{g} - \mathbf{X}_1\mathbf{a} - \mathbf{W}\mathbf{g}\mathbf{a} - \mathbf{X}_2\gamma)^\top (\mathbf{Y} - \mathbf{J}\mu - \mathbf{Z}\mathbf{g} - \mathbf{X}_1\mathbf{a} - \mathbf{W}\mathbf{g}\mathbf{a} - \mathbf{X}_2\gamma) = 0$$

Resolvendo essa equação para σ^2 , obtemos o estimador de máxima verossimilhança:

$$\hat{\sigma}^2 = \frac{1}{n} (\mathbf{Y} - \mathbf{J}\mu - \mathbf{Z}\mathbf{g} - \mathbf{X}_1\mathbf{a} - \mathbf{W}\mathbf{g}\mathbf{a} - \mathbf{X}_2\gamma)^\top (\mathbf{Y} - \mathbf{J}\mu - \mathbf{Z}\mathbf{g} - \mathbf{X}_1\mathbf{a} - \mathbf{W}\mathbf{g}\mathbf{a} - \mathbf{X}_2\gamma)$$

3.2.4 Sistema de Equações Normais na Forma Matricial

Estimamos os parâmetros por máxima verossimilhança, assumindo erros normais, resultando no seguinte sistema:

$$\begin{bmatrix} \mathbf{J}^\top \mathbf{J} & \mathbf{J}^\top \mathbf{Z} & \mathbf{J}^\top \mathbf{X}_1 & \mathbf{J}^\top \mathbf{W} & \mathbf{J}^\top \mathbf{X}_2 \\ \mathbf{Z}^\top \mathbf{J} & \mathbf{Z}^\top \mathbf{Z} & \mathbf{Z}^\top \mathbf{X}_1 & \mathbf{Z}^\top \mathbf{W} & \mathbf{Z}^\top \mathbf{X}_2 \\ \mathbf{X}_1^\top \mathbf{J} & \mathbf{X}_1^\top \mathbf{Z} & \mathbf{X}_1^\top \mathbf{X}_1 & \mathbf{X}_1^\top \mathbf{W} & \mathbf{X}_1^\top \mathbf{X}_2 \\ \mathbf{W}^\top \mathbf{J} & \mathbf{W}^\top \mathbf{Z} & \mathbf{W}^\top \mathbf{X}_1 & \mathbf{W}^\top \mathbf{W} & \mathbf{W}^\top \mathbf{X}_2 \\ \mathbf{X}_2^\top \mathbf{J} & \mathbf{X}_2^\top \mathbf{Z} & \mathbf{X}_2^\top \mathbf{X}_1 & \mathbf{X}_2^\top \mathbf{W} & \mathbf{X}_2^\top \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha \\ \tau \\ \alpha\tau \\ \gamma \end{bmatrix} = \begin{bmatrix} \mathbf{J}^\top \mathbf{Y} \\ \mathbf{Z}^\top \mathbf{Y} \\ \mathbf{X}_1^\top \mathbf{Y} \\ \mathbf{W}^\top \mathbf{Y} \\ \mathbf{X}_2^\top \mathbf{Y} \end{bmatrix}$$

Esse sistema pode ser escrito de forma mais compacta como:

$$\mathbf{M}\theta = \mathbf{b},$$

onde:

- \mathbf{M} é a matriz dos coeficientes (matriz de covariância dos efeitos);
- $\theta = [\mu, \alpha, \tau, \alpha\tau, \gamma]^\top$ é o vetor dos parâmetros desconhecidos;
- \mathbf{b} é o vetor dos produtos internos das variáveis explicativas com \mathbf{Y} .

3.2.5 Detalhamento das Equações

Cada equação dentro do sistema representa uma condição de **ortogonalidade** entre os resíduos e os componentes do modelo:

1. **Estimador para μ** (média geral):

$$\hat{\mu} = (J^\top J)^{-1} J^\top (Y - Z\alpha - X_1\tau - W\alpha\tau - X_2\gamma)$$

2. **Estimador para α** (efeito dos genótipos):

$$\hat{\alpha} = (Z^\top Z)^{-1} Z^\top (Y - J\mu - X_1\tau - W\alpha\tau - X_2\gamma)$$

3. **Estimador para τ** (efeito dos ambientes):

$$\hat{\tau} = (X_1^\top X_1)^{-1} X_1^\top (Y - J\mu - Z\alpha - W\alpha\tau - X_2\gamma)$$

4. **Estimador para $\alpha\tau$** (efeito da interação):

$$\widehat{\alpha\tau} = (W^\top W)^{-1} W^\top (Y - J\mu - Z\alpha - X_1\tau - X_2\gamma)$$

5. **Estimador para γ** (efeito dos blocos dentro dos ambientes):

$$\hat{\gamma} = (X_2^\top X_2)^{-1} X_2^\top (Y - J\mu - Z\alpha - X_1\tau - W\alpha\tau)$$

3.3 EXPECTATION-MAXIMIZATION (EM)

O Algoritmo EM é uma técnica iterativa utilizada para estimar parâmetros em modelos estatísticos quando há dados faltantes. No contexto da análise GA, o algoritmo EM não é aplicado diretamente ao modelo AMMI, mas sim ao modelo linear com interação, cujo preditor inclui os efeitos principais e o termo de interação.

O objetivo do EM, nesse caso, é estimar os parâmetros do modelo linear e, principalmente, obter uma estimativa completa da matriz de interação \mathbf{GA} , que representa o efeito combinado dos genótipos nos diferentes ambientes. Após essa etapa, a decomposição dessa matriz DVS é o que conduz ao modelo AMMI propriamente dito, permitindo a separação dos efeitos multiplicativos em componentes bilineares.

3.3.1 Algoritmo EM para o Modelo AMMI

Definimos o preditor linear:

$$\mathbf{H} = \mathbf{J}\mu + \mathbf{Z}\alpha + \mathbf{X}_1\tau + \mathbf{W}\alpha\tau + \mathbf{X}_2\gamma \quad (25)$$

3.3.2 Etapa E (Expectation)

Calculamos a expectativa da log-verossimilhança completa dos dados, condicional à observação \mathbf{Y}_0 e aos parâmetros estimados $\theta^{(t)}$:

$$\log L(\mathbf{H} | \mathbf{Y}_0, \mathbf{Y}_f) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{H})^T (\mathbf{Y} - \mathbf{H}) \quad (26)$$

A função Q do algoritmo EM será:

$$Q(\theta | \theta^{(t)}) = \mathbb{E}_{\mathbf{Y}_f | \mathbf{Y}_0, \theta^{(t)}} [\log L(\mathbf{H} | \mathbf{Y}_0, \mathbf{Y}_f)] \quad (27)$$

$$Q(\theta | \theta^{(t)}) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \mathbb{E}_{\mathbf{Y}_f | \mathbf{Y}_0, \theta^{(t)}} [(\mathbf{Y} - \mathbf{H})^T (\mathbf{Y} - \mathbf{H})] \quad (28)$$

3.3.3 Etapa M (Maximization)

Maximizamos a função $Q(\theta | \theta^{(t)})$ em relação a θ , isto é:

$$\theta^{(t+1)} = \arg \min_{\theta} \mathbb{E}_{\mathbf{Y}_f | \mathbf{Y}_0, \theta^{(t)}} [(\mathbf{Y} - \mathbf{H})^T (\mathbf{Y} - \mathbf{H})] \quad (29)$$

3.3.4 Critério de Convergência

A convergência é considerada atingida quando o critério adotado for satisfeito. Esse critério pode ser:

(i) Norma Euclidiana dos parâmetros:

$$\|\theta^{(t+1)} - \theta^{(t)}\| = \sqrt{(\mu^{(t+1)} - \mu^{(t)})^2 + \|\beta^{(t+1)} - \beta^{(t)}\|^2 + \dots} \leq \varepsilon \quad (30)$$

(ii) Diferença na log-verossimilhança:

$$\left| \log L(\theta^{(t+1)}) - \log L(\theta^{(t)}) \right| \leq \varepsilon \quad (31)$$

ou o atingimento de um número máximo de iterações.

3.3.5 Expressão de cada métrica da avaliação da imputação

A qualidade do ajuste após a imputação dos dados ausentes foi avaliada por meio de métricas de erro e de associação entre os valores reais e os valores imputados. As métricas utilizadas foram:

RMSE – Root Mean Square Error (Erro Quadrático Médio):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (32)$$

Em que y_i representa o valor observado da variável resposta na i -ésima observação, \hat{y}_i corresponde ao valor estimado ou predito pelo modelo para essa mesma observação, e n denota o número total de observações consideradas na avaliação. Essa métrica quantifica a magnitude média do erro de predição, sendo expressa na mesma unidade da variável resposta, de modo que valores menores de RMSE indicam melhor desempenho preditivo do modelo.

MAE – Mean Absolute Error (Erro Absoluto Médio):

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (33)$$

Mede o erro médio absoluto entre valores observados e imputados.

MAPE – Mean Absolute Percentage Error (Erro Percentual Médio Absoluto):

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (34)$$

Expressa o erro médio em termos percentuais.

SMAPE – Symmetric Mean Absolute Percentage Error (Erro Percentual Médio Simétrico):

$$SMAPE = \frac{100}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{\left(\frac{|y_i| + |\hat{y}_i|}{2} \right)} \quad (35)$$

Versão simétrica do MAPE, mais estável quando os valores reais são pequenos.

MDAE – Median Absolute Error (Erro Absoluto Mediano):

$$MDAE = \text{mediana}(|y_i - \hat{y}_i|) \quad (36)$$

Essa métrica é robusta à presença de valores discrepantes (outliers).

Correlação de Pearson entre valores reais e imputados:

$$r = \frac{\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2}} \quad (37)$$

Em que y_i representa o valor observado da variável resposta na i -ésima observação, \hat{y}_i corresponde ao valor imputado ou estimado para essa mesma observação, \bar{y} e $\bar{\hat{y}}$ denotam as médias amostrais dos valores observados e imputados, respectivamente, e n é o número total de observações consideradas. O coeficiente r assume valores no intervalo $[-1, 1]$, sendo que valores próximos de 1 indicam forte associação linear positiva entre os valores reais e imputados.

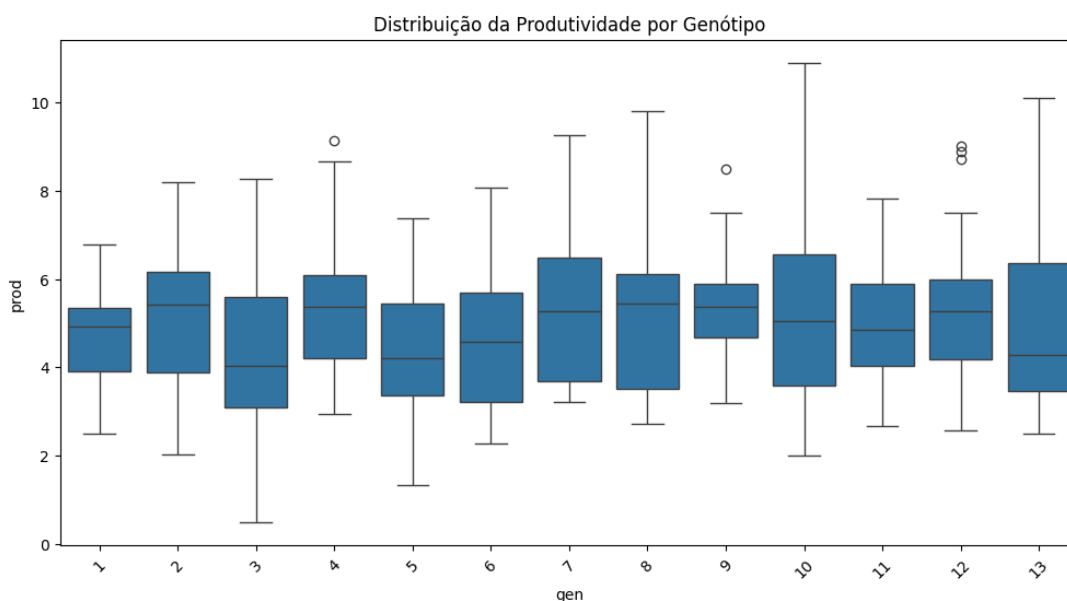
4 RESULTADOS

4.1 ANÁLISE DESCRITIVA DOS DADOS

Conforme discutido anteriormente, os dados usados neste trabalho são referentes a 13 genótipos de arroz avaliados em 9 ambientes distintos entre nos anos agrícolas e os experimentos em cada um dos locais foram avaliados segundo um delineamento em blocos balanceados causalizados com três repetições cada bloco. A variável de interesse neste estudo é a produtividade de grãos, expressa em toneladas por hectare (t/ha). Antes do ajuste do modelo AMMI, foram calculadas medidas descritivas básicas dessa variável, com o objetivo de caracterizar a variação geral observada entre os genótipos e os ambientes.

Na Figura 1, observa-se uma variabilidade na produtividade entre os 13 genótipos avaliados. Alguns materiais, como os genótipos 7, 10 e 13, destacam-se por apresentarem maiores medianas e maior amplitude de variação, indicando elevado potencial produtivo, porém com sensibilidade às condições ambientais. EM contraste, outros genótipos apresentam distribuições mais concentradas, sugerindo maior estabilidade. A presença de valores extremos (outliers) em diversos genótipos evidencia comportamento discrepante em determinados ambientes, reforçando a importância da análise da interação $G \times A$.

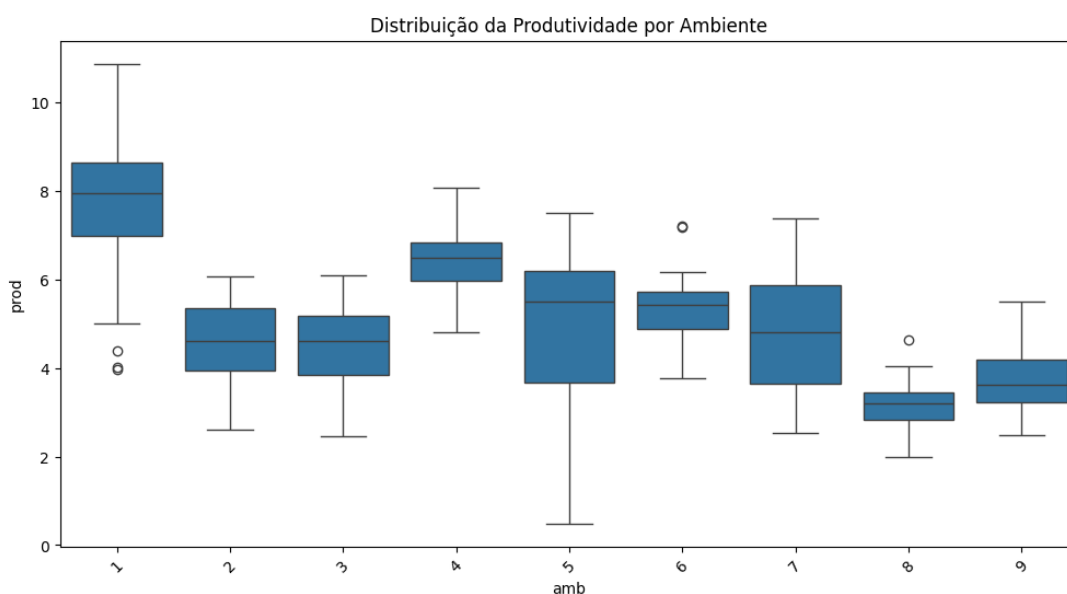
Figura 1 – Folhas e seus rótulos prod x gen.



Fonte: Elaborado pelo autor a partir dos dados experimentais.

Já na Figura 2, evidencia-se uma variação expressiva na produtividade entre os nove ambientes avaliados, entendidos como combinações específicas de locais e anos de avaliação. O ambiente 1, por exemplo, apresenta as maiores produtividades, com distribuição assimétrica e maior dispersão, possivelmente associada a condições favoráveis de solo e clima observadas naquele local e ano específicos. Por outro lado, os ambientes 8 e 9 registraram medianas mais baixas, com menor variabilidade, indicando possíveis restrições ambientais decorrentes da combinação local–ano considerada. Essa heterogeneidade entre os ambientes justifica a aplicação de modelos que consideram explicitamente a interação genótipo \times ambiente, como o modelo AMMI.

Figura 2 – Folhas e seus rótulos prod x amb.



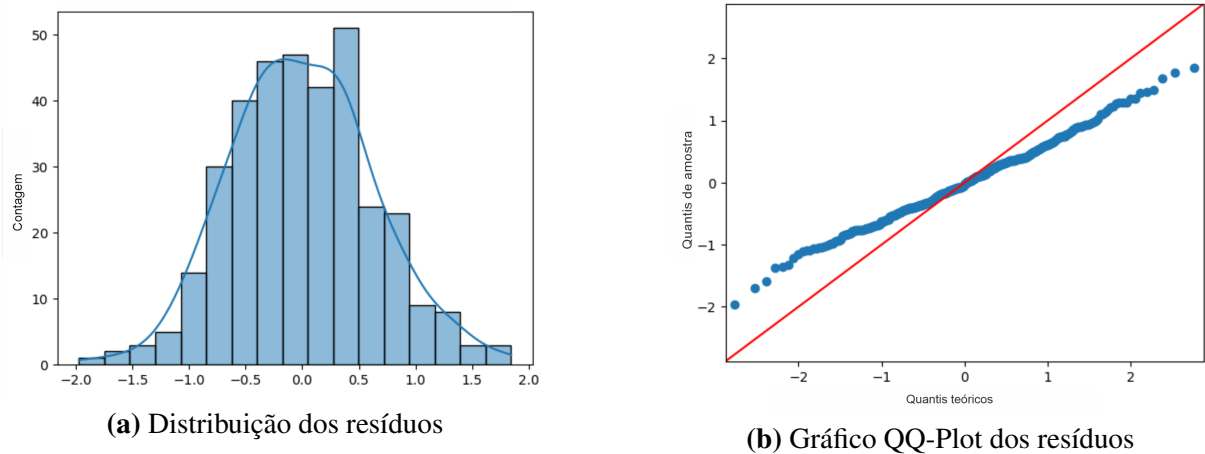
Fonte: Elaborado pelo autor a partir dos dados experimentais.

4.2 VERIFICAÇÃO DOS PRESSUPOSTOS DO MODELO

Antes da aplicação da análise AMMI, foram avaliados os pressupostos herdados da análise de variância (ANOVA), uma vez que o modelo AMMI é extensão desse modelo. Os pressupostos são: normalidade dos resíduos, homocedasticidade e independência dos resíduos. Na Figura 3 apresenta, respectivamente, a distribuição dos resíduos do modelo com histograma com curva de densidade estimada e o gráfico QQ-Plot. Na distribuição dos resíduos, observa-se que os resíduos apresentam uma distribuição aproximadamente simétrica, com formato aproximadamente normal. Já o gráfico QQ-Plot mostra que os pontos seguem de forma razoavelmente

próxima a linha de referência de 45° , indicando boas concordâncias entre os quantis observados e os quantis teóricos de uma distribuição normal. Esses resultados visuais sustentam a suposição de normalidade dos resíduos, posteriormente confirmado pelo teste de Shapiro-Wilk.

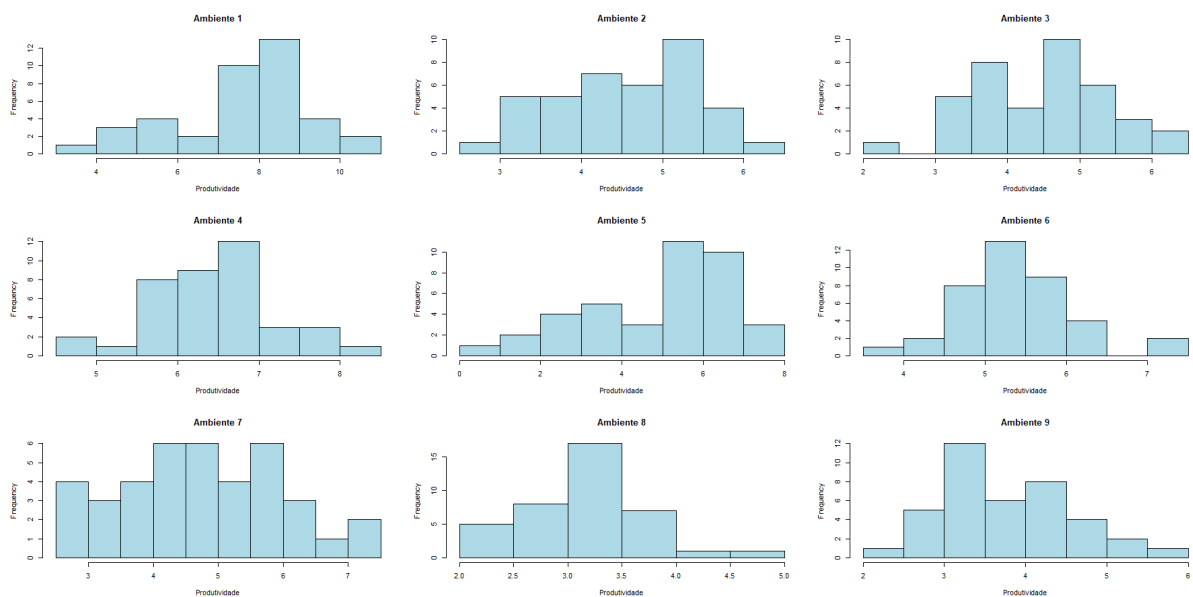
Figura 3 – Análise dos resíduos do modelo.



Fonte: Elaborado pelo autor.

A distribuição da produtividade foi avaliada separadamente por ambiente, evidenciando diferenças na forma, dispersão e tendência central entre os ambientes, reforçando a importância da análise estratificada.

Figura 4 – Produtividade em cada ambiente.



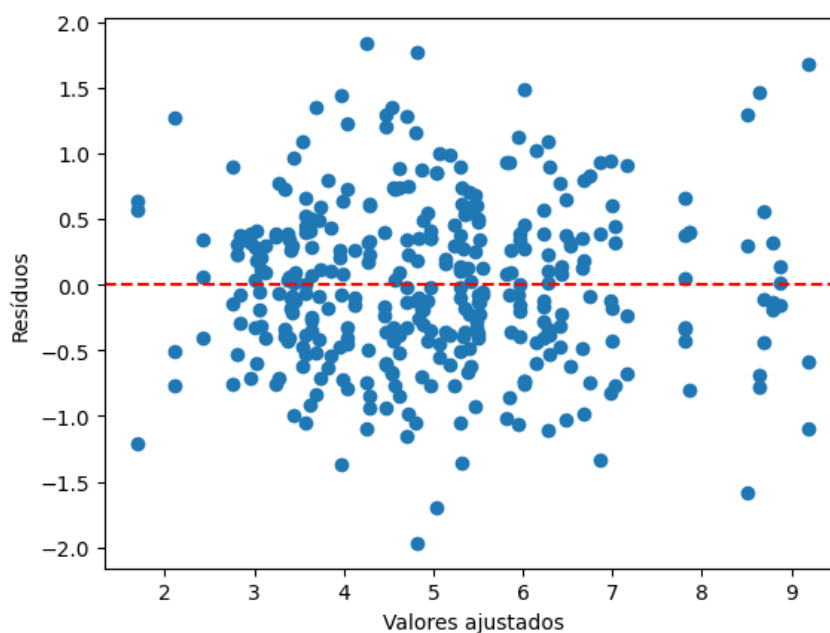
Fonte: Elaborado pelo autor.

4.2.1 Homocedasticidade (igualdade de variâncias)

A homocedasticidade foi avaliada por meio de diferentes testes estatísticos, incluindo os testes de *Levene*, *Brown–Forsythe*, *Bartlett*, *Fligner–Killeen* e *Breusch–Pagan*. Os resultados indicaram heterocedasticidade significativa entre ambientes (por exemplo, teste de Levene: estatística = 9,41; $p < 0,001$), enquanto entre genótipos as variâncias mostraram-se homogêneas ($p = 0,08$). Esse padrão é comum em experimentos multiambientais, uma vez que condições ambientais distintas, como solo, clima e manejo, tendem a aumentar a variabilidade residual entre locais.

A presença de heterocedasticidade também pode ser visualizada no gráfico de resíduos versus valores ajustados (Figura 5), no qual a dispersão dos pontos não se mantém constante ao longo do eixo horizontal. Essa variação sugere que a amplitude do erro depende do nível de produtividade, caracterizando uma violação da suposição de variâncias homogêneas. Embora o modelo AMMI clássico pressuponha homocedasticidade, essa violação não inviabiliza sua aplicação quando o objetivo principal é a análise exploratória dos padrões de interação, especialmente em conjuntos de dados balanceados. Conforme discutido no contexto do GGE Biplot, métodos baseados em decomposição de componentes principais tendem a apresentar robustez na identificação dos padrões dominantes de interação, mesmo na presença de desvios moderados dos pressupostos clássicos (Yan; Kang, 2003). No presente estudo, não foram observados indícios de que a heterogeneidade de variâncias tenha alterado substancialmente a estrutura dos padrões de interação identificados. Ainda assim, recomenda-se cautela na interpretação dos resultados e reconhece-se que, em cenários com heterogeneidade mais pronunciada ou dados desbalanceados, a utilização de versões ponderadas do modelo AMMI ou de abordagens baseadas em modelos mistos, que permitem considerar variâncias específicas por ambiente, constitui uma alternativa metodológica (Piepho, 1997; Crossa; Cornelius; Vargas, 2006).

Figura 5 – Resíduos vs valores ajustados.



Fonte: Elaborado pelo autor.

4.2.2 Análise AMMI

A análise de variância conjunta realizada no primeiro estágio do modelo AMMI para a variável produtividade revelou efeito significativo de ambiente ($ENV, p < 0,001$) e de genótipo ($GEN, p < 0,001$), assim como a interação genótipo x ambiente ($valor - p < 0,001$), são significativos. Isso indica que o desempenho dos genótipos varia substancialmente entre os ambientes. Resultado que confirma a coerência do uso do modelo AMMI para decompor e interpretar adequadamente essa interação.

Tabela 7 – Análise de variância para grupos de experimentos.

Fonte de variação	GL	Soma de Q.	Q. Médio	Valor F	Pr.(f)
Ambiente	8	573,0928	71,6366	35,1960	1,90E-09
Bloco (Ambiente)	18	36,6365	2,0354	4,3353	6,66E-08
Genótipo	12	47,5292	3,9608	8,4363	4,81E-13
Interação (G×E)	96	236,0786	2,4592	5,2379	3,50E-24
Resíduo	216	101,4100	0,4695	–	–
Total	446	1.230,8256	2,7597	–	–

Fonte: Elaborado pelo autor.

4.2.3 Determinação do número de eixos significativos pelo teste F de Cornelius

De acordo com o teste *F* de *Cornelius* (1988), apenas os dois primeiros componentes principais (PC1 e PC2) apresentaram valores de *p* inferiores a 0,05, sendo considerados estatisticamente significativos. A partir do terceiro componente (PC3), os valores de *p* tornaram-se não significativos, indicando que esses eixos estão associados predominantemente a ruído. Assim, recomenda-se a retenção do modelo AMMI2, que proporciona melhor equilíbrio entre capacidade explicativa e parcimônia, conforme mostrado na Tabela 8.

Tabela 8 – Teste *F* de *Cornelius* (1988) para determinação do número de componentes principais significativos da interação.

Fonte	GL	Soma de Q.	Q. Médio	Valor F	Pr.(f)	Proporção (%)	Acumulado (%)
PC1	19	153,4152	8,0745	7,4647	0,0000	64,98	64,98
PC2	17	39,0232	2,2955	2,1221	0,0063	16,53	81,51
PC3	15	14,3231	0,9549	0,8828	0,5841	6,07	87,58
PC4	13	11,1556	0,8581	0,7933	0,6673	4,73	92,31
PC5	11	9,0582	0,8235	0,7613	0,6787	3,84	96,14
PC6	9	4,7438	0,5271	0,4873	0,8829	2,01	98,15
PC7	7	2,5811	0,3687	0,3409	0,9347	1,09	99,25
PC8	5	1,7783	0,3557	0,3288	0,8954	0,75	100,00

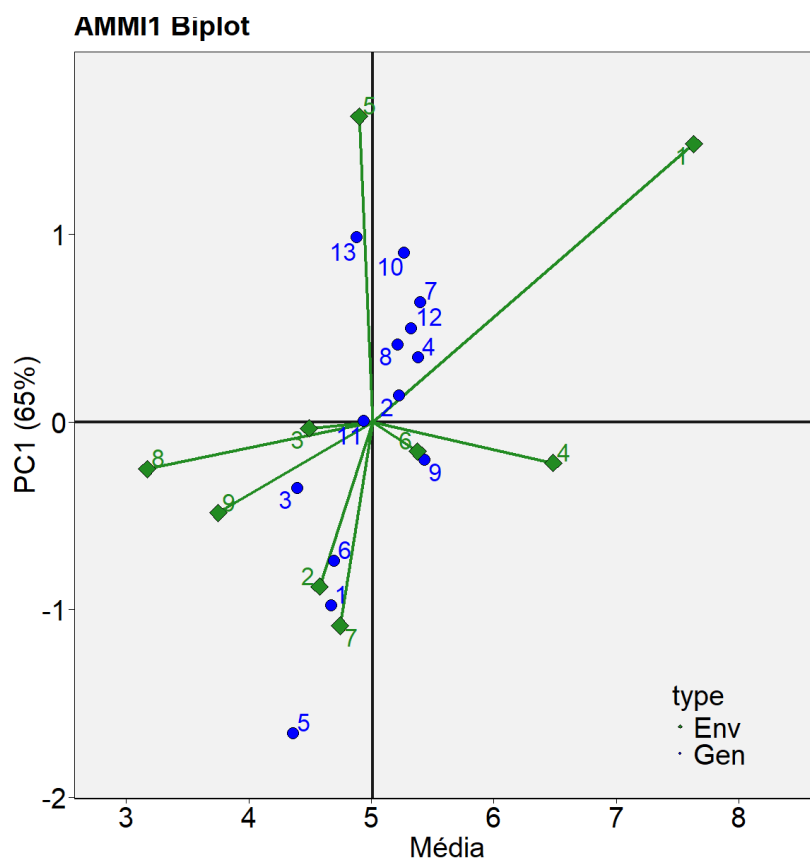
Fonte: Elaborado pelo autor.

4.2.4 Análise gráfica: biplots AMMI1 e AMMI2

O biplot AMMI1 apresentado na Figura 6 combina a produtividade média no eixo horizontal com os escores do primeiro componente principal da interação (PC1) no eixo vertical. Nesse gráfico, a estabilidade está diretamente relacionada à proximidade de PC1 = 0: genótipos e ambientes com escores próximos de zero apresentam menor contribuição para a interação e, portanto, comportamento mais estável, enquanto aqueles com valores mais afastados de zero (em valor absoluto) indicam maior interação e maior instabilidade. Considerando os genótipos, observa-se que os genótipos 1 e 5 apresentam os escores mais afastados de zero no eixo vertical, indicando maior contribuição para a interação e, conseqüentemente, maior instabilidade. Em termos de desempenho médio, o genótipo 1 apresenta a maior média, seguido pelo genótipo 4. Entretanto, enquanto o genótipo 1 possui maior evidência de interação (maior afastamento no eixo vertical), o genótipo 4 combina média elevada com escore de PC1 próximo de zero, sendo, portanto, um dos genótipos mais estáveis entre os de melhor desempenho.

Em relação aos ambientes, verifica-se que o ambiente 1 é o mais afastado da origem no eixo vertical, sendo o que mais contribui para a interação (maior evidência de resposta diferencial dos genótipos). O ambiente 4 também apresenta contribuição elevada, configurando-se como o segundo mais influente na interação, enquanto ambientes mais próximos de PC1 = 0 apresentam menor participação na interação. Além disso, o sinal do PC1 auxilia na interpretação: ambientes com PC1 positivo tendem a favorecer genótipos com PC1 positivo, ao passo que ambientes com PC1 negativo tendem a favorecer genótipos com PC1 negativo, refletindo padrões de adaptação diferencial. Assim, no AMMI1 a leitura principal está na avaliação conjunta de desempenho médio (eixo x) e estabilidade (proximidade de PC1 = 0), permitindo identificar materiais com alto desempenho e baixa interação. Nesse sentido, destaca-se o genótipo 4 como material com desempenho médio elevado e maior estabilidade, enquanto genótipos como 1 e 5 apresentam maior instabilidade devido à maior contribuição para a interação.

Figura 6 – Biplot AMMI1 mostrando a relação entre produtividade média e estabilidade dos genótipos.

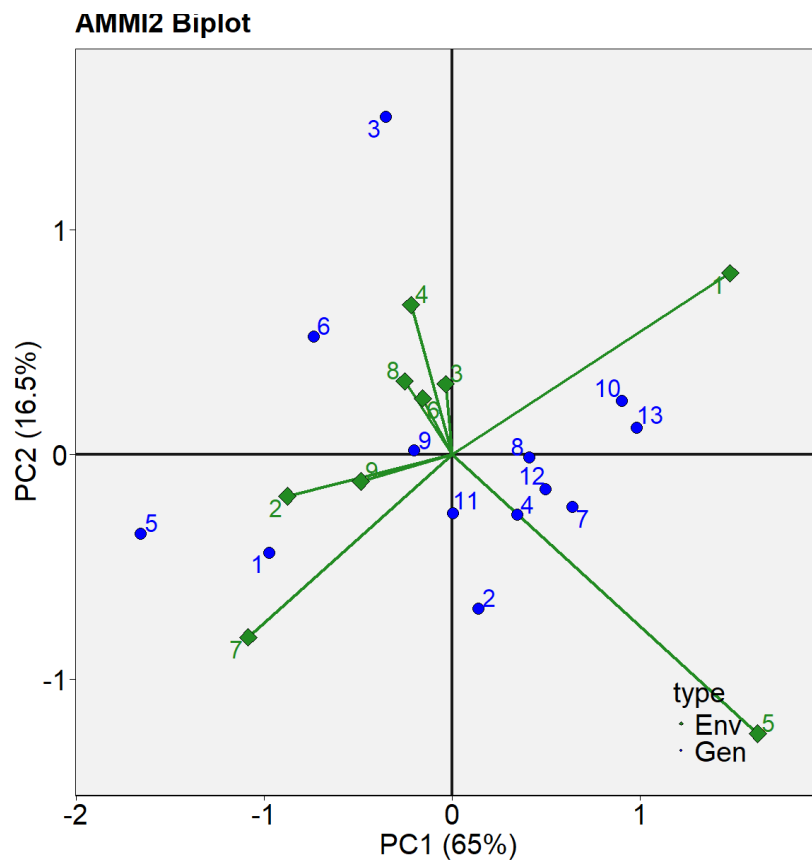


O biplot AMMI2 (Figura 7) representa graficamente os dois primeiros componentes principais da interação (PC1 e PC2), responsáveis por explicar 65% e 16,5% da variação da

interação, respectivamente, totalizando 81,5% da soma de quadrados da interação. Essa representação permite identificar padrões complexos da interação, bem como grupos de genótipos com comportamento semelhantes e ambientes que mais discriminam.

Os genótipos próximos à origem demonstram menor interação e, portanto, maior estabilidade, enquanto genótipos e ambientes distantes da origem refletem interações específicas. O agrupamento de genótipos próximos a determinados ambientes indica afinidade ou adaptação específica, sendo possível distinguir grupos de genótipos com comportamentos semelhantes. Essa representação permite compreender a estrutura da interação G×E de forma mais intuitiva e facilita recomendações, conforme observado na Figura 7.

Figura 7 – Biplot AMMI2 representando os dois primeiros componentes principais da interação (PC1 e PC2).



Fonte: Elaborado pelo autor.

4.3 RESULTADOS DA IMPUTAÇÃO DE DADOS AUSENTES UTILIZANDO O MÉTODO EM-AMMI

O objetivo desta etapa foi usar o algoritmo Expectation–Maximization (EM) na imputação de dados ausentes em experimentos multiambientais, bem como analisar o impacto desse procedimento na modelagem da interação por meio do modelo AMMI. Para tanto, utilizou-se como referência o conjunto de dados completo de produtividade de arroz, previamente analisado, o qual serviu como cenário de comparação (baseline) para a avaliação da qualidade da imputação. Os dados originais foram artificialmente desbalanceados pela introdução aleatória de ausências nos valores de produtividade, considerando dois níveis de perda: 10% e 20%. A remoção foi realizada de forma completamente aleatória (MCAR — Missing Completely At Random), assegurando que a ausência de valores não estivesse relacionada a qualquer estrutura do experimento. Para a condução das análises, serão considerados conjuntos de dados com 10% de dados ausentes.

Em ambos os cenários, assegurou-se que não houvesse exclusão completa de genótipos em ambientes específicos, preservando a estrutura mínima necessária para o ajuste do modelo conjunto. Para cada nível de perda, foram realizadas 1000 simulações independentes, permitindo uma avaliação robusta do desempenho médio do método. O desempenho do algoritmo EM foi avaliado por meio das métricas RMSE, MAE, MAPE, SMAPE, MDAE e coeficiente de determinação (R^2), calculadas a partir da comparação entre os valores imputados e os valores observados no conjunto de dados completo, conforme a tabela abaixo.

Tabela 9 – Média das métricas de desempenho (1000 cenários simulados de perdas aleatórias).

Perc. de perda	RMSE	MAE	MAPE	SMAPE	MDAE	R²
10%	0,9036	0,7346	17,5232	0,1609	0,6210	0,7198
20%	0,9578	0,7574	18,2113	0,1666	0,6277	0,6914

Fonte: Elaborado pelo autor.

Os resultados médios (Tabela 9) das simulações indicaram que, mesmo com o aumento do percentual de dados ausentes, as métricas de erro permaneceram em níveis consistentes. Para 10% de perda, o RMSE médio foi de 0,9036 e o MAE de 0,7346, enquanto no cenário de 20% esses valores aumentaram moderadamente para 0,9578 e 0,7574, respectivamente. Esse incremento relativamente pequeno sugere que o algoritmo EM foi capaz de explorar de forma eficiente a estrutura conjunta dos dados, minimizando os efeitos do desbalanceamento. Os

valores médios de R^2 foram de 0,7198 para 10% de perda e 0,6914 para 20%, indicando que aproximadamente 72% e 69% da variabilidade original da produtividade foram explicadas pelos valores imputados. Esses resultados evidenciam uma boa capacidade do algoritmo EM em recuperar a informação perdida, preservando as relações estruturais entre genótipos, ambientes e blocos experimentais

A análise de variância realizada com o conjunto de dados após a imputação das observações ausentes indicou que os efeitos de genótipo, ambiente e da interação foram altamente significativos. Esses resultados indicam que o desempenho produtivo dos genótipos continuou variando de forma expressiva entre os diferentes ambientes, mesmo após a imputação dos valores ausentes. A significância dos efeitos principais e da interação confirma a coerência dos dados imputados e evidencia que o procedimento de imputação via algoritmo EM preservou adequadamente a estrutura experimental original. Dessa forma, os resultados reforçam a adequação do uso do modelo AMMI para decompor, particionar e interpretar a interação no conjunto de dados após a imputação, permitindo análises consistentes de adaptabilidade e estabilidade dos genótipos avaliados. Tabela : Biplot AMMI2 representando os dois primeiros componentes principais da interação (PC1 e PC2) com 10% de perdas aleatórias.

Tabela 10 – Análise AMMI para a variável produtividade (dados ausentes imputados).

Fonte de variação	GL	Soma de Q.	Q. Médio	Valor F	Pr.(f)
Ambiente	8	599,3320	74,9170	48,4500	$1,30 \times 10^{-10}$
Bloco (Ambiente)	18	27,8330	1,5460	3,5600	$4,13 \times 10^{-6}$
Genótipo	12	46,9520	3,9130	9,0000	$6,05 \times 10^{-14}$
Interação (G×E)	96	244,0130	2,5420	5,8500	$3,24 \times 10^{-27}$
Resíduo	216	93,8830	0,4350	–	–
Total	446	1.256,0260	2,8160	–	–

Fonte: Elaborado pelo autor.

Após análise de variância os dados imputados, foram submetidos à análise AMMI, de forma análoga àquela realizada com os dados completos. O teste F de Cornelius foi aplicado para determinar o número de componentes principais (IPCA) significativos da interação (Tabela 11).

Tabela 11 – Teste F de Cornelius (1988) para determinar o número de eixos responsáveis pela GxA com os dados incompletos com 10% de perdas aleatórias.

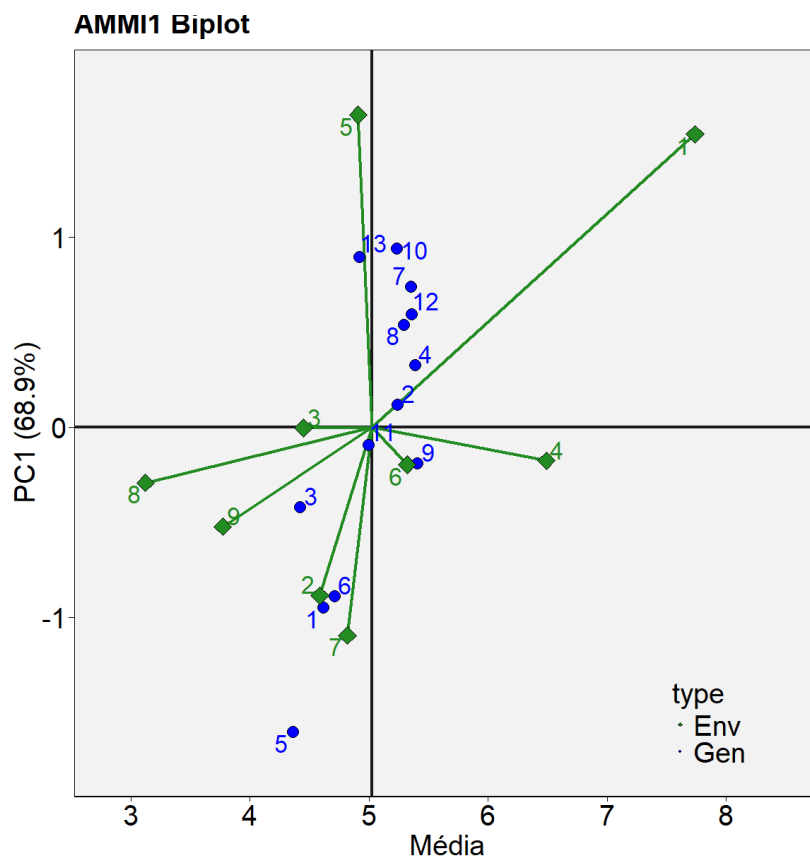
Fonte	GL	Soma de Q.	Q. Médio	Valor F	Pr.(f)	Proporção	Pro. Acumulado
PC1	19	168,0043	8,8423	8,1647	0,00E+00	68,85%	68,85%
PC2	17	41,5138	2,4420	2,2548	0,0034	17,01%	85,86%
PC3	15	13,3246	0,8883	0,8202	0,6549	5,46%	91,32%
PC4	13	8,4588	0,6507	0,6008	0,8533	3,47%	94,79%
PC5	11	5,3485	0,4862	0,4490	0,9326	2,19%	96,98%
PC6	9	4,4827	0,4981	0,4599	0,9007	1,84%	98,82%
PC7	7	2,0117	0,2874	0,2654	0,9668	0,82%	99,64%
PC8	5	0,8688	0,1738	0,1604	0,9767	0,36%	100,00%
PC9	3	0,0000	0,0000	0,0000	1,0000	0,00%	100,00%

Fonte: Elaborado pelo autor.

Os resultados indicam que, tanto para o conjunto de dados completos quanto para o conjunto de dados incompletos, apenas os dois primeiros componentes principais apresentaram valores-p inferiores a 0,05, sendo, portanto, considerados estatisticamente significativos de acordo com o teste *F* de Cornelius (1988). A partir do terceiro componente, os valores-p tornaram-se não significativos, sugerindo que esses eixos adicionais estão associados predominantemente a ruído e não contribuem de forma relevante para a explicação da interação. Dessa forma, em ambos os cenários, recomenda-se a retenção do modelo AMMI2, evidenciando que o processo de imputação não alterou o número de eixos relevantes da interação. Esse resultado é coerente com a natureza conservadora do teste F de Cornelius, que tende a selecionar modelos mais parcimoniosos, evitando a retenção excessiva de componentes associados à variabilidade aleatória.

Observa-se ainda que os dois primeiros componentes principais (PC1 e PC2) explicam a maior parte da variabilidade genótipo \times ambiente, tanto nos dados completos quanto nos dados incompletos, conforme apresentado nas Tabelas correspondentes. Tal comportamento indica que a estrutura principal da interação foi preservada mesmo após a imputação dos valores ausentes. Os biplots AMMI1 e AMMI2 obtidos a partir dos dados imputados no cenário com 10% de perdas aleatórias permitiram avaliar visualmente o efeito da imputação sobre a interpretação da interação. Na Figura 8, que relaciona a produtividade média com os escores do primeiro componente principal, observou-se que a ordenação relativa dos genótipos foi amplamente preservada em relação aos dados completos, indicando que o procedimento de imputação não alterou substancialmente os padrões de adaptabilidade e estabilidade nesse cenário.

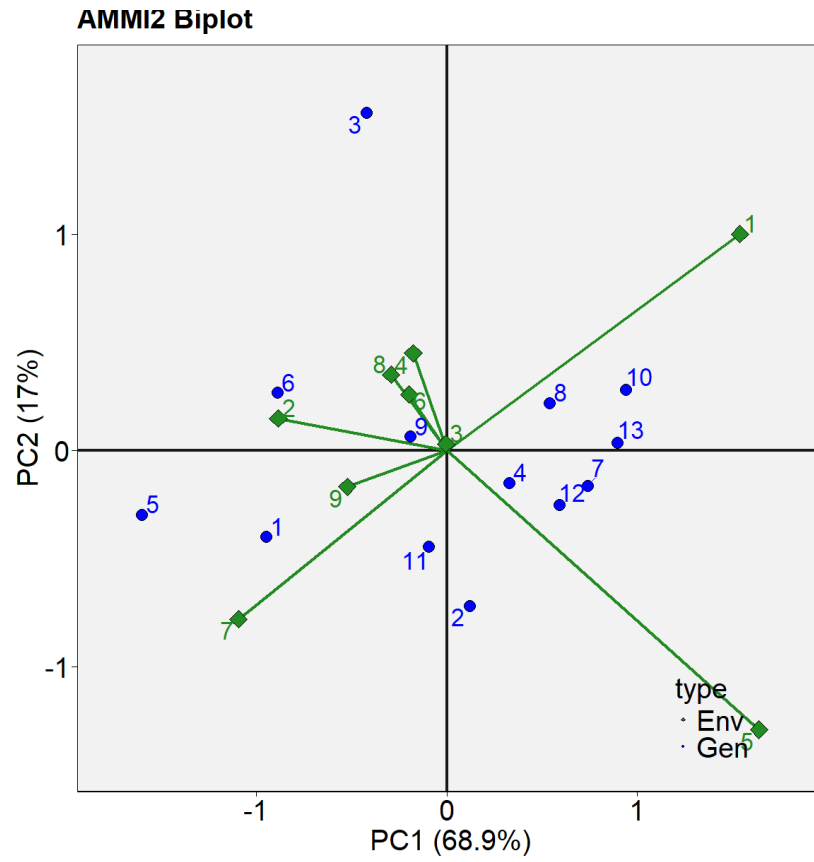
Figura 8 – Biplot AMMI1 a 10% de perdas aleatórias.



Fonte: Elaborado pelo autor.

Genótipos previamente identificados como produtivos nos dados completos, como G4, G7, G10 e G12, mantiveram-se posicionados à direita do gráfico, mesmo após a imputação, indicando estabilidade na classificação produtiva. O genótipo G11 permaneceu próximo à linha zero do eixo vertical, caracterizando-se como altamente estável e com baixa contribuição para a interação, enquanto o genótipo G5 continuou associado a baixa produtividade e alta instabilidade. O biplot AMMI2 (Figura 9), que projeta simultaneamente os escores de PC1 e PC2, confirmou esses padrões. Genótipos como G9, G11 e G2 permaneceram próximos à origem, reforçando sua ampla adaptação, enquanto G3 e G5 apresentaram maiores distâncias ao centro, refletindo comportamento instável e resposta específica a determinados ambientes. A geometria e a orientação dos vetores ambientais, em especial do Ambiente 1, mostraram-se consistentes em relação aos dados completos, indicando que a imputação não distorceu as correlações entre ambientes nem sua capacidade discriminante.

Figura 9 – Biplot AMMI2 a 10% de perdas aleatórias.



Fonte: Elaborado pelo autor.

5 CONCLUSÃO

A análise dos treze genótipos de arroz de terras altas avaliados em nove ambientes demonstrou a existência de diferenças significativas para os efeitos de genótipos, ambientes e interação, evidenciando que o desempenho produtivo dos materiais não foi uniforme entre os locais avaliados. Os resultados confirmam a importância da análise da interação em experimentos multiambientais, uma vez que determinados genótipos apresentaram comportamento distinto conforme as condições ambientais. Os ambientes avaliados apresentaram elevada contribuição para a variabilidade total observada, indicando forte influência das condições ambientais sobre a produtividade dos genótipos. A decomposição da interação por meio do modelo AMMI mostrou que os dois primeiros componentes principais explicaram grande parte da variação associada à interação genótipo \times ambiente, permitindo interpretação consistente dos padrões de estabilidade e adaptabilidade.

Os biplots AMMI1 e AMMI2 possibilitaram identificar materiais com elevado potencial produtivo e ampla adaptação. Os genótipos G4, G7, G10 e G12 destacaram-se por apresentarem maiores médias de produtividade, mantendo posicionamento favorável nos gráficos mesmo após a imputação dos dados ausentes, indicando estabilidade na classificação produtiva. Entre eles, os genótipos G7 e G10 demonstraram bom equilíbrio entre produtividade e resposta à interação, sendo considerados materiais promissores para recomendação em diferentes ambientes. Os genótipos G9, G11 e G2 apresentaram escores próximos à origem nos biplots AMMI2, caracterizando comportamento mais estável e ampla adaptabilidade, com menor contribuição para a interação. O genótipo G11 destacou-se especialmente pela elevada estabilidade, apresentando reduzida sensibilidade às variações ambientais. Em contrapartida, os genótipos G3 e G5 apresentaram maiores distâncias em relação à origem dos componentes principais, evidenciando comportamento instável e adaptação específica a determinados ambientes. Além disso, o genótipo G5 associou baixa produtividade à elevada instabilidade, indicando menor potencial de recomendação.

Em relação aos ambientes, observou-se que alguns apresentaram maior capacidade discriminante, contribuindo de forma mais intensa para a interação. O Ambiente 1 destacou-se pela forte influência sobre a dispersão dos genótipos nos biplots, evidenciando elevada capacidade de diferenciação do desempenho dos materiais avaliados. A manutenção da geometria e da orientação dos vetores ambientais após a imputação dos dados ausentes demonstrou que a estrutura das correlações entre ambientes foi preservada. A aplicação do algoritmo EM para im-

putação de dados ausentes apresentou resultados satisfatórios, mesmo nos cenários com perdas aleatórias. O procedimento manteve a estrutura principal da interação, preservando a ordenação relativa dos genótipos e os padrões de adaptabilidade e estabilidade observados nos dados completos. Os resultados obtidos indicaram forte concordância entre as análises realizadas com os dados originais e os dados imputados, demonstrando a eficiência do algoritmo na recuperação das informações perdidas. De modo geral, a associação entre o modelo AMMI e o algoritmo EM mostrou-se eficiente para análise de experimentos multiambientais desbalanceados, permitindo interpretar adequadamente a interação e reduzir os impactos causados pela ausência de dados. Os resultados obtidos fornecem informações relevantes para programas de melhoramento genético, especialmente na seleção de genótipos mais produtivos, estáveis e adaptados às diferentes condições de cultivo de arroz de terras altas.

REFERÊNCIAS

- ALESSANDRO, N.; NAZZARENO, A.; TEODORO, C. Breeding for nutritional and organoleptic quality in vegetable crops: The case of tomato and cauliflower. **Agriculture**, 2021.
- ALLARD, R. **Princípios do melhoramento de plantas**. [S.l.]: Edgard Blücher, 1971. 381 p.
- BIRKES, D.; DODGE, Y. **Alternative methods of regression**. New York: Wiley, 1976.
- CAUWER, B. D.; ORTIZ, R. Genotype \times environment interaction and adaptation in rice breeding: A simple method for analyzing interaction in a factorial experiment. **Euphytica**, Springer, v. 100, n. 1-3, p. 143–148, 1998.
- CORNELIUS, P. L. Statistical tests and retention of terms in the additive main effects and multiplicative interaction model for cultivar trials. **Crop Science**, v. 33, n. 5, p. 1186–1193, 1993.
- CROSSA, J.; CORNELIUS, P. L.; VARGAS, G. Theoretical and statistical considerations for the analysis of genotype \times environment interaction. **Crop Science**, v. 46, n. 2, p. 561–571, 2006.
- CROSSA, J.; GAUCH, H. G.; ZOBEL, R. W. Additive main effects and multiplicative interaction analysis of two international maize cultivar trials. **Crop Science**, v. 30, n. 3, p. 493–500, 1990.
- CRUZ, C. *et al.* **Modelos Biométricos Aplicados ao Melhoramento Genético**. [S.l.]: Editora Viçosa, 2004.
- CUMMINGS, P. Missing data and multiple imputation. **JAMA Pediatrics**, v. 167, n. 7, p. 656–661, 2013.
- DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the em algorithm. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 39, n. 1, p. 1–22, 1977.
- DIAS, C.; SANTOS, C. A.; MEXIA, J. T. Genotype \times environment interaction: A comparison between joint regression analysis and weighted biplot models. **WSEAS Transactions on Environment and Development**, v. 19, p. 134, 2023. Disponível em: <https://doi.org/10.37394/232015.2023.19.134>.
- DIAS, C. T. dos S. *et al.* **A Metodologia AMMI: com aplicação ao melhoramento genético**. Universidade de São Paulo. Escola Superior de Agricultura Luiz de Queiroz, 2017. ISBN 9788586481369. Disponível em: <https://www.livrosabertos.sibi.usp.br/portaldelivrosUSP/catalog/book/50>.
- DIAS, S.; KRZANOWSKI, W. J. **Comparison of Discrimination Methods for the Classification of Tumors Using Gene Expression Data**. [S.l.]: Oxford University Press, 2006.
- DODGE, Y. **Statistical data analysis based on the L1-norm and related methods**. Amsterdam: North-Holland, 1985.

DODGE, Y.; ZOPPE, M. J. **Regression analysis: methods and applications**. New York: Springer-Verlag, 2004.

DUARTE, J.; VENCOVSKY, R. **Interação genótipos x ambientes: Uma introdução à análise "AMMI"**. [S.l.]: Sociedade Brasileira de Genética, 1999.

FALCONER, D. S. **Introduction to quantitative genetics**. [S.l.]: Longman Scientific & Technical, 1987. 293–298 p.

FAVERI, J. D. Random regression models for multi-environment, multi-time data from crop breeding selection trials. **Crop & Pasture Science**, 2022. Disponível em: <https://doi.org/10.1071/cp21732>.

FIGUEIREDO, A. Bootstrap and permutation tests in anova for directional data. **Computational Statistics**, v. 32, p. 1213–1240, 2017.

FISHER, R. A.; MACKENZIE, W. A. Studies in crop variation: I. an examination of the yield of dressed grain from broadbalk. **Journal of Agricultural Science**, Cambridge University Press, v. 13, n. 3, p. 311–320, 1923.

GABRIEL, K. The biplot graphic display of matrices with application to principal component analysis. **Biometrika**, v. 58, p. 453–467, 1971.

GANGADHARA, D. *et al.* Transformative gene editing methods: Precision in genetically modified crops through trait modification. **Journal of Scientific Research and Reports**, v. 30, 2024.

GAUCH, H.; ZOBEL, R. Imputing missing yield trial data. **Theoret. Appl. Genetics**, Springer, v. 79, n. 6, p. 753–761, 1990.

GAUCH, H. G. Continued fractions in analysis of variance. **Biometrics**, Wiley, v. 41, n. 4, p. 1085–1092, 1985.

GAUCH, H. G. **Statistical Analysis of Regional Yield Trials: AMMI Analysis**. Amsterdam: Elsevier, 1992.

GAUCH, H. G. Statistical analysis of yield trials by ammi and gge. **Crop Science**, v. 46, p. 1488–1500, 2006.

GAUCH, H. G.; ZOBEL, R. W. Ammi analysis of yield trials. In: KANG, M. S.; GAUCH, H. G. (Ed.). **Genotype-by-Environment Interaction**. [S.l.]: CRC Press, 1996. p. 85–122.

GILBERT, G. A. Statistical methods for geochemical prospecting. **Journal of Geochemical Exploration**, Elsevier, v. 2, n. 3, p. 265–275, 1963.

GOLLOB, H. F. A statistical model which combines features of factorial analysis of variance and analysis of covariance. **Psychometrika**, Springer, v. 33, n. 1, p. 73–115, 1968.

INÁCIO, H. *et al.* Phenotypic plasticity of upland rice lines cultivated in minas gerais state. **Aust J Crop Sci**, v. 12, n. 5, p. 763–769, 2018.

KANG, M.; MAGARI, R. New developments in selecting for phenotypic stability in crop breeding. In: KANG, M.; GAUCH, H. (Ed.). **Genotype by Environment Interaction**. [S.l.]: CRC Press, 1996. cap. 1, p. 1–14.

KEMPTON, R. A. The use of a randomized block design in the analysis of covariance. **Biometrics**, Wiley, v. 40, n. 2, p. 547–554, 1984.

LI, Z.; GUTIERREZ, L. Editorial: Statistical methods for analyzing multiple environmental quantitative genomic data. **Frontiers in Genetics**, 2023. Disponível em: <https://doi.org/10.3389/fgene.2023.1212804>.

MALIK, W. A.; FORKMAN, J.; PIEPHO, H. P. Testing multiplicative terms in ammi and gge models for multienvironment trials with replicates. **Theoretical and Applied Genetics**, v. 132, n. 7, p. 2087–2096, jul 2019.

MANDEL, J. An example of multivariate analysis of variance. **Journal of the American Statistical Association**, Taylor & Francis, v. 66, n. 336, p. 875–879, 1971.

OLIVEIRA, G. A.; SILVA, K. N.; SILVA, A. S. S. Melhoria genética de plantas: conceitos e técnicas. **Revista de Agricultura Neotropical**, Institute of Mathematical Statistics, v. 25, n. 2, p. 290–302, 1954.

PADEREWSKI, J.; RODRIGUES, P. C. The usefulness of em-ammi to study the influence of missing data pattern and application to polish post-registration winter wheat data. **Australian Journal of Crop Science**, v. 8, n. 4, p. 640–645, 2014.

PIEPHO, H.-P. An algorithm for mixed model analysis of incomplete data in factorial designs. **Journal of Agronomy and Crop Science**, v. 178, n. 1, p. 49–55, 1997.

RODRIGUES, P.; MONTEIRO, A.; LOURENÇO, V. M. A robust ammi model for the analysis of genotype-by-environment data. **Bioinformatics**, v. 32, n. 1, p. 58–66, jan 2016.

ROHRIG, B. **Entenda como o melhoramento genético de plantas é um aliado de quem produz**. 2023. Disponível em: <https://blog.aegro.com.br/melhoramento-genetico-de-plantas/>.

RUBIN, D. B. **Multiple imputation for nonresponse in surveys**. New York: Wiley-Interscience, 1996.

SAHU, P. K. Analysis related to breeding researches. In: _____. **Statistical Methods for Agricultural and Biological Sciences**. Springer, 2013. p. 225–255. Disponível em: https://doi.org/10.1007/978-81-322-1020-7_11.

SCHAFFER, J. L.; GRAHAM, J. W. Missing data: our view of the state of the art. **Psychological Methods**, v. 7, n. 2, p. 147–177, 2002.

SCHMIDER, E. *et al.* Is it really robust? **Methodology**, v. 6, n. 4, p. 147–151, 2010.

SILVA, M. F. d. **Robustez na capacidade preditiva dos modelos AMMI e Fatoriais Analíticos no estudo de dados multi-ambientais desbalanceados**. Tese (Doutorado) — Universidade Federal de Lavras, 2023. Disponível em: http://repositorio.ufla.br/jspui/bitstream/1/12572/2/DISSERTA%C3%87%C3%83O_Robustez%20na%20capacidade%20preditiva%20dos%20modelos%20AMMI%20e%20Fatoriais%20Anal%C3%ADticos%20no%20estudo%20de%20dados%20multi-ambientais%20desbalanceados.pdf.

SOUZA, T. L. *et al.* A influência da interação genótipo por ambiente em estudos de melhoramento genético. **Revista Brasileira de Agricultura Irrigada**, v. 12, n. 4, p. 1991–2001, 2018.

TORTORA, G. J.; FUNKE, B. R.; CASE, C. L. **Microbiologia**. 12. ed. Porto Alegre: Artmed Editora, 2016.

TOTVS, E. **Produtividade Agrícola: saiba como otimizar**. 2021. Acessado em: 26 de Setembro, 2023. Disponível em: <https://www.totvs.com/blog/gestao-agricola/produtividade-agricola/>.

UMESH, S. *et al.* Genomics for crop improvement: From genes to fields. 2024.

WU, C. *et al.* Integrative analysis of gene–environment interactions under a multi-response partially linear varying coefficient model. **Statistics in Medicine**, 2014. Disponível em: <https://doi.org/10.1002/SIM.6287>.

YAN, W.; KANG, M. S. **GGE Biplot Analysis: A Graphical Tool for Breeders, Geneticists, and Agronomists**. [S.l.]: CRC Press, 2003.

ZOBEL, R. W. *et al.* Statistical analysis of a yield trial. **Agronomy Journal**, v. 80, p. 388–393, 1988.