

UNIVERSIDADE FEDERAL DE VIÇOSA

**Otimização da seleção em programa de melhoramento do feijoeiro comum com
ênfase em ensaios sequenciais e multiambientes**

Deurimar Herênio Gonçalves Júnior
Doctor Scientiae

**VIÇOSA - MINAS GERAIS
2025**

DEURIMAR HERÊNIO GONÇALVES JÚNIOR

Otimização da seleção em programa de melhoramento do feijoeiro comum com ênfase em ensaios sequenciais e multiambientes

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Doctor Scientiae*.

Orientador: Pedro C. Souza Carneiro

Coorientador: Jose E. de Souza Carneiro

**VIÇOSA - MINAS GERAIS
2025**

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade
Federal de Viçosa - Campus Viçosa**

T

G635o
2025

Gonçalves Júnior, Deurimar Herênio, 1992-
Otimização da seleção em programa de melhoramento do
feijoeiro comum com ênfase em ensaios sequenciais e
multiambientes / Deurimar Herênio Gonçalves Júnior. – Viçosa,
MG, 2025.

1 tese eletrônica (108 f.): il. (algumas color.).

Orientador: Pedro Crescêncio Souza Carneiro.
Tese (doutorado) - Universidade Federal de Viçosa,
Departamento de Biologia Geral, 2025.

Inclui bibliografia.

DOI: <https://doi.org/10.47328/ufvbbt.2025.275>

Modo de acesso: World Wide Web.

1. Feijão - Melhoramento genético - Métodos estatísticos.
2. Modelos multiníveis (Estatísticas). 3. Análise fatorial.
4. Análise de variância. I. Carneiro, Pedro Crescêncio Souza,
1966-. II. Universidade Federal de Viçosa. Departamento de
Biologia Geral. Programa de Pós-Graduação em Genética e
Melhoramento. III. Título.

CDD 22. ed. 635.6522

DEURIMAR HERÊNIO GONÇALVES JÚNIOR

Otimização da seleção em programa de melhoramento do feijoeiro comum com ênfase em ensaios sequenciais e multiambientes

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Doctor Scientiae*.

APROVADA: 9 de abril de 2025.

Assentimento:

Deurimar Herênio Gonçalves Júnior
Autor

Pedro Crescencio Souza Carneiro
Orientador

Essa tese foi assinada digitalmente pelo autor em 09/05/2025 às 20:27:07 e pelo orientador em 12/05/2025 às 07:36:15. As assinaturas têm validade legal, conforme o disposto na Medida Provisória 2.200-2/2001 e na Resolução nº 37/2012 do CONARQ. Para conferir a autenticidade, acesse <https://siadoc.ufv.br/validar-documento>. No campo 'Código de registro', informe o código **BYN5.KN7K.HSVG** e clique no botão 'Validar documento'.

AGRADECIMENTOS

À minha família, pelo apoio incondicional em todas as etapas da minha vida.

A todos os professores que contribuíram para a minha formação, desde a infância até este momento.

À Universidade Federal de Viçosa (UFV), pela oportunidade de realizar a pós-graduação.

Ao Programa de Pós-Graduação em Genética e Melhoramento (PPGGM), por toda a estrutura, corpo docente e pela chance de realizar o doutorado.

Ao meu orientador, Professor Pedro Crescêncio Souza Carneiro, pela excelente orientação, disposição, ensinamentos, apoio e confiança.

Ao meu coorientador, Professor José Eustáquio de Souza Carneiro, pela orientação e apoio.

Ao Professor Kaio Olímpio das Graças Dias, pela disposição, suporte, ensinamentos e valiosa contribuição para a realização deste trabalho.

Ao Laboratório de Biometria, pelo acolhimento, suporte e amizade.

Aos colegas do Programa Feijão, pela ajuda na condução dos experimentos, coleta de dados, elaboração de croquis das áreas experimentais, além da amizade, convivência e oportunidade de aprender com todos.

Aos membros das bancas de qualificação e defesa de tese, pelas contribuições ao trabalho e disponibilidade.

Aos funcionários do laboratório do Programa Feijão e da Estação Experimental de Coimbra, pelo auxílio na realização dos experimentos.

Aos funcionários da secretaria do PPGGM, pela disponibilidade e assistência.

Aos meus amigos Larissa, Ney, Guilherme, José Tiago, Jana e tantos outros, pela amizade, incentivo e apoio constante durante toda a jornada.

À CAPES, pela concessão da bolsa de estudos.

"Este trabalho foi realizado com o apoio das seguintes agências de pesquisa brasileiras: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001, Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) e Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

A todos que, direta ou indiretamente, contribuíram para a realização deste trabalho.

Muito obrigado!

"Na vida real, as descobertas são fruto da cooperação entre cientistas, de ligar os pontos entre uma descoberta e outra. A ciência é coletiva, não é individual!"
Bonassa, Freitas & Araújo (2021, p. 8)
(Super-Heróis da Ciência: 52 cientistas e suas pesquisas transformadoras)

RESUMO

JÚNIOR, Deurimar Herênio Gonçalves, D.Sc., Universidade Federal de Viçosa, abril de 2025. **Otimização da seleção em programa de melhoramento do feijoeiro comum com ênfase em ensaios sequenciais e multiambientes.** Orientador: Pedro Crescencio Souza Carneiro. Coorientador: Jose Eustaquio de Souza Carneiro.

Neste estudo, abordou-se o melhoramento do feijoeiro comum (*Phaseolus vulgaris* L.) por meio de análises de dados oriundos de ensaios sequenciais realizados em diferentes gerações e ambientes, enfrentando os desafios decorrentes do desbalanceamento genético e estatístico. No Capítulo I, 400 progênies F4:6 foram avaliadas inicialmente sob delineamento de blocos casualizados completos, e, a partir delas, 95 progênies selecionadas foram submetidas a avaliações em três safras subsequentes (F4:7, F4:8 e F4:9) utilizando um delineamento látice triplo. Foram testados cinco modelos com distintas estruturas de variância-covariância para os efeitos residuais, e a metodologia REML/BLUP foi empregada para estimar os componentes de variâncias genéticos e residuais, permitindo a identificação do modelo que melhor se ajustou aos dados, o qual adotou a estrutura fator analítico de primeira ordem para o efeito de progênie dentro de cada geração, associada a uma variância residual diagonal heterogênea. Esse modelo apresentou um ganho genético médio 68% superior aos demais e incrementou significativamente a herdabilidade, além de modificar a classificação das progênies com base nos valores genotípicos preditos ao longo das safras. No Capítulo II, a investigação foi ampliada para avaliar desempenho e estabilidade de progênies utilizando dados de múltiplas características e ambientes, por meio da aplicação de *Factor Analytic Selection Tools* (FAST), que integraram dados quantitativos e categóricos para fornecer uma análise multivariada robusta. A aplicação da FAST possibilitou a identificação precisa de genótipos com alto desempenho e estabilidade, demonstrando a eficácia dessa abordagem na seleção de progênies superiores. Em síntese, a integração de modelos mistos com estruturas de variância-covariância ajustadas e as *Factor Analytic Selection Tools* mostrou-se a estratégia mais eficaz para lidar com a complexidade dos dados desbalanceados em ensaios multiambientes, proporcionando maior acurácia na predição dos valores genotípicos e fundamentando a recomendação de progênies para recombinação ou extração de linhagens. Assim, o presente trabalho contribui significativamente para o aprimoramento dos programas de melhoramento do feijoeiro, ao demonstrar que metodologias estatísticas inovadoras podem potencializar os ganhos genéticos e aumentar a confiabilidade das análises de seleção em ambientes experimentais

complexos.

Palavras-chave: melhoramento do feijão comum; modelos mistos; análise fator-analítica; ensaios multiambientais; estruturas de variância-covariância

ABSTRACT

JÚNIOR, Deurimar Herênio Gonçalves, D.Sc., Universidade Federal de Viçosa, April, 2025. **Optimization of selection in common bean breeding program with emphasis on sequential and multi-environment trials.** Adviser: Pedro Crescencio Souza Carneiro. Co-adviser: Jose Eustaquio de Souza Carneiro.

In this study, we addressed common bean (*Phaseolus vulgaris* L.) breeding through advanced analyses of sequential trials across generations and environments, targeting challenges posed by unbalanced data in genetic and statistical analyses. In Chapter I, 400 F4:6 progenies were initially evaluated under a randomized complete block design; from these, 95 selected progenies were subsequently assessed over three following seasons (F4:7, F4:8, and F4:9) using a triple lattice design. Five models with distinct variance–covariance structures for the residual effects were tested, and the REML/BLUP methodology was employed to estimate genetic and residual components. This approach allowed the identification of the best-fitting model, which adopted a first-order factor analytic structure for the progeny effect within each generation combined with a heterogeneous diagonal residual variance. This model produced a 68% higher average genetic gain compared to the alternative models, significantly increased heritability, and altered the classification of progenies based on predicted genotypic values throughout the seasons. In Chapter II, the investigation was extended to evaluate progeny performance and stability using data on multiple traits and environments through the application of Factor Analytic Selection Tools (FAST), which integrated quantitative and categorical data to provide a robust multivariate analysis. The use of FAST enabled the precise identification of high-performing, stable genotypes, demonstrating the effectiveness of this approach in selecting superior progenies. In summary, the integration of mixed models with tailored variance–covariance structures and factor analytic tools proved to be the most effective strategy for managing the complexity of unbalanced data in multi-environment trials, delivering enhanced accuracy in predicting genotypic values and supporting the recommendation of progenies for recombination. Thus, this work significantly contributes to the advancement of common bean improvement programs by demonstrating that innovative statistical methodologies can potentiate genetic gains and increase the reliability of selection analyses in complex experimental settings.

Keywords: common bean breeding; mixed models; factor-analytic analysis; multi-environment trials; variance-covariance structures

BIOGRAFIA

DEURIMAR HERÊNIO GONÇALVES JUNIOR, filho de Naurinete Silva Gonçalves e Deurimar Herênio Gonçalves, nasceu em 03 de abril de 1992, na cidade de Imperatriz, no estado do Maranhão, Brasil.

Graduou-se em Agronomia pela Universidade Estadual do Maranhão (UEMA) em 2017, onde desenvolveu pesquisas em melhoramento de feijão-caupi, avaliando a adaptabilidade de genótipos de porte ereto. Durante a graduação, atuou como bolsista de iniciação científica e monitor voluntário da disciplina de Genética sob orientação da professora Dra. Anatércia Ferreira, além de realizar estágios em áreas como produção animal, reflorestamento e melhoramento genético de feijão.

Em 2019, concluiu o mestrado em Genética e Melhoramento de Plantas pela Universidade Estadual do Norte Fluminense Darcy Ribeiro (UENF), sob orientação do Prof. Dr. Alexandre Pio Viana. Sua dissertação focou no desenvolvimento de cultivares de maracujazeiro-azedo resistentes ao vírus do endurecimento dos frutos (CABMV), utilizando técnicas de retrocruzamento e seleção recorrente.

Paralelamente ao doutorado, em 2025, especializou-se em MBA em Inteligência Artificial e Big Data pela Universidade de São Paulo (USP), com ênfase em abordagens de aprendizado de máquina aplicadas ao melhoramento de plantas. Conjuntamente, acumulou formações complementares em programação, análise estatística de dados, ciência de dados, reforçando sua expertise em metodologias computacionais aplicadas à agricultura.

Em 2025, concluiu seu Doutorado no Programa de Genética e Melhoramento de Plantas da Universidade Federal de Viçosa (UFV), com a tese "Otimização da seleção em programa de melhoramento do feijoeiro comum com ênfase em ensaios sequenciais e multiambientes", orientado pelo Prof. Pedro Crescêncio Souza Carneiro. Durante o doutorado, integrou projetos de pesquisa em biometria e análise de dados, contribuindo para o melhoramento do feijoeiro comum.

LISTA DE FIGURAS

- Figura 1.** Concordância (correlação de Spearman) entre a classificação das progênies com base nos valores dos BLUPs para cada modelo em cada geração (A) e nas médias dos BLUPs das progênies considerando todas as gerações (B).44
- Figura 2.** Correlações genéticas entre pares gerações.45
- Figura 3.** Concordância (correlação de *Spearman*) na classificação de todos os genótipos (A) e dos 30 melhores (B) com base nos BLUPs em cada geração, utilizando o modelo selecionado (Modelo 5).46
- Figura 4.** Esquema da análise em estágios.....76
- Figura 5.** Porcentagem de variação explicada acumulada por cada modelo (entre parênteses) e por cada fator em cada combinação característica-ambiente.88
- Figura 6.** Gráfico de barras horizontais comparando os valores de Herdabilidade (barras azuis) e Coeficiente de Variação (CV) (barras laranjas) para diferentes combinações de características e safras. No eixo X, são apresentados os valores numéricos de Herdabilidade e CV, enquanto no eixo Y são listadas as combinações de Característica-Safra89
- Figura 7.** Heatmap das correlações genéticas entre combinações característica-ambiente. ARQ: arquitetura de planta; ASP: aspecto comercial de grão; PROD: produtividade de grãos.....90
- Figura 8.** Performance Geral (OP, eixo y) e estabilidade (RMSD, eixo x) das 95 progênies de feijão carioca, com destaque para as 20 selecionadas. A barra de cores indica a confiabilidade da seleção.91

LISTA DE TABELAS

Tabela 1: Descrição dos quatro ensaios utilizados no presente trabalho.	34
Tabela 2: Estruturas de (co)variância ajustadas aos efeitos aleatórios descritos na Equação [1].	39
Tabela 3: Número de parâmetros (NP), logaritmo do ponto de máximo da função de máxima verossimilhança (LogL), valores dos critérios de informação Akaike (AIC) e bayesiano (BIC), acurácia média (r), herdabilidade média (H^2), ganhos genéticos esperados (GS%).	43
Tabela 4: Componentes de variância e parâmetros genéticos estimados com base no Modelo 5 (FA1) para o conjunto de dados.	45
Tabela 5: Confiabilidade (r^2) acurácia (r) e erro padrão de predição (SEP) das melhores predições lineares não viesadas (BLUPs) para valores genotípicos individuais das 30 melhores progênies em cada geração, com base no modelo selecionado (FA1).	47
Tabela 6: Descrição dos cinco ensaios utilizados no presente trabalho.	70
Tabela 7. Escala de avaliação de arquitetura de planta.	72
Tabela 8: Descrição da escala de notas de aspecto de grãos do tipo carioca, adaptada de Ramalho et al. (1998).	73
Tabela 9: Resumo das características agronômicas no conjunto de dados multiambientes e multicaracterísticas.	74
Tabela 10: Logaritmo da função de máxima verossimilhança (LogLik), Critério de Informação de Akaike (AIC), porcentagem de variância explicada por todos os k fatores ($\%v$) e razão média das semivariâncias (ASVR) para os 10 modelos fator-analíticos ajustados.	87

LISTA DE ABREVIações E ACRÔNIMOS

AIC	Critério de informação de Akaike (<i>Akaike Information Criterion</i>)
AMMI	Modelo de efeitos principais aditivos e interação multiplicativa (<i>Additive Main Effects and Multiplicative Interaction</i>)
ANOVA	Análise de variância (<i>Analysis of variance</i>)
ARQ	Arquitetura de planta
ASVR	Razão média das semivariâncias (<i>Average Semivariance Ratio</i>)
ASP	Aspecto comercial de grão
BIC	Critério de informação bayesiano (<i>Bayesian Information Criterion</i>)
BLUE	Melhor estimador linear não viesado (<i>Best Linear Unbiased Estimator</i>)
BLUP	Melhor preditor linear não viesado (<i>Best Linear Unbiased Prediction</i>)
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CS	Estrutura de (co)variância simetria composta (<i>Compound Symmetry</i>)
CV	Coeficiente de variação experimental
CVres	Coeficiente de variação residual
DBA	Delineamento de blocos aumentados
DBC	Delineamento de blocos completos
DIAG	Estrutura de (co)variância diagonal heterogênea
DP	Desvio padrão
e-BLUEs	Melhores estimativas lineares não viesadas empíricas (<i>Empirical Best Linear Unbiased Estimators</i>)
e-BLUPs	Melhores predições lineares não viesadas empíricas (<i>Empirical Best Linear Unbiased Predictions</i>)
EM	Ensaio multiambientes
FA	Modelos fator-analíticos (<i>Factor Analytic</i>)
FA1	Estrutura de (co)variância fator analítico de 1ª ordem (<i>First-Order Factor Analytic</i>)
FA8	Modelo fator-analítico de oitava ordem (<i>Eighth-Order Factor Analytic</i>)
FAI-BLUP	Índice FAI-BLUP (ideotype-design index)
FA	<i>Factor Analytic Mixed Model</i>
FAST	<i>Factor Analytic Selection Tools</i>
FOP	<i>Fusarium oxysporum</i> f. sp. <i>phaseoli</i>

G×A	Interação genótipo-ambiente
GEI	Interação genótipo-ambiente (<i>Genotype-by-Environment Interaction</i>)
GETI	Interação genótipo-ambiente-característica (<i>Genotype-Environment-Trait Interaction</i>)
GG	Ganho genético
GGE	Genótipo + interação genótipo-ambiente (<i>Genotype + Genotype-by-Environment</i>)
ID	Índice de seleção
kg ha	Quilogramas por hectare
LogLik	Logaritmo da função de máxima verossimilhança
MGIDI	<i>Multitrait Genotype-Ideotype Distance Index</i>
MME	Equações de modelo misto (<i>Mixed Model Equations</i>)
MNAR	Dados ausentes não aleatório (<i>Missing Not at Random</i>)
MRS	Seleção recorrente multipopulacional (<i>Multipopulation Recurrent Selection</i>)
OLS	Mínimos quadrados ordinários (<i>Ordinary Least Squares</i>)
OP	Desempenho geral (<i>Overall Performance</i>)
PEV	Variância do erro de predição (<i>Prediction Error Variance</i>)
PPGGM	Programa de Pós-Graduação em Genética e Melhoramento
PROD	Produtividade de grãos
RE	Responsividade (<i>Responsiveness</i>)
REML	Máxima verossimilhança restrita (<i>Restricted Maximum Likelihood</i>)
RMSD	Raiz dos quadrados médios dos desvios (<i>Root Mean Square Deviation</i>)
RR	Repetibilidade e reprodutibilidade (<i>Repeatability and Reproducibility</i>)
SEP	Erro padrão de predição (<i>Standard Error of Prediction</i>)
SG	Ganho de seleção (<i>Selection Gain</i>)
SVD	Decomposição em valores singulares (<i>Singular Value Decomposition</i>)
UFV	Universidade Federal de Viçosa
UN	Estrutura de (co)variância não estruturada com correlação geral (<i>Unstructured</i>)

SUMÁRIO

INTRODUÇÃO GERAL	15
REFERÊNCIAS	19
CAPÍTULO I	29
ANÁLISE MULTIGERAÇÃO EM PROGRAMA DE MELHORAMENTO DE FEIJOEIRO COMUM: ABORDAGENS PARA LIDAR COM DADOS DESBALANCEADOS	29
RESUMO	30
ABSTRACT	31
1 INTRODUÇÃO	32
2 MATERIAL E MÉTODOS	34
2.2 Material vegetal e delineamentos estatísticos	34
2.3 Metodologia REML/BLUP e modelagem de estruturas de covariância	35
2.4 Seleção do modelo	39
2.5 Estimativas de parâmetros genéticos e não genéticos	40
3 RESULTADOS	43
3.1 Seleção do modelo	43
3.2 Parâmetros genéticos e não genéticos	44
4 DISCUSSÃO	49
4.1 Seleção do modelo	49
4.2 Parâmetros genéticos e não genéticos	51
5 CONCLUSÕES	55
REFERÊNCIAS	56
CAPÍTULO II	65
DESEMPENHO E ESTABILIDADE DE PROGÊNIES DE FEIJOEIRO COM DADOS DE MÚLTIPLAS CARACTERÍSTICAS E AMBIENTES: APLICAÇÃO DAS <i>FACTOR</i> <i>ANALYTIC SELECTION TOOLS</i> - FAST	65
RESUMO	66
ABSTRACT	67
1 INTRODUÇÃO	68
2.MATERIAL E MÉTODOS	70
2.1 Material	70
2.2 Características avaliadas	71
2.3 Análises estatísticas	74
2.4 Análise dos dados em estágios	75

2.5 Análises individuais para cada ensaio – Estágio 1.....	76
2.6 Cálculo dos pesos para o estágio 2	78
2.7 Estágio 2 – Análise de múltiplas características e ambientes	79
2.8 Ajuste dos modelos FA para a interação genótipo-ambiente-característica (GETI)	80
2.9 Seleção do modelo FA de melhor ajuste	82
2.10 Estimativas dos parâmetros genéticos e não genéticos	83
2. 11 <i>Factor Analytic Selection Tools</i> - FAST.....	85
3 RESULTADOS	87
3.1 Seleção do Modelo Fator-Analítico (FA) de Melhor Ajuste	87
3.2 Estimativas dos parâmetros genéticos e não genéticos	88
3.3 <i>Factor Analytic Selection Tools</i>	91
4 DISCUSSÃO	92
4.1 Análise de dois estágios dos dados	92
4.2 Ajuste dos modelos FA para a interação genótipo x ambiente x característica (GETI)	93
4.3 Seleção do modelo FA de melhor ajuste	94
4.4 Estimativas dos parâmetros genéticos e não genéticos	95
4.5 <i>Factor Analytic Selection Tools</i>	97
5 CONCLUSÕES	99
REFERÊNCIAS.....	100

INTRODUÇÃO GERAL

Os programas de melhoramento do feijoeiro comum (*Phaseolus vulgaris* L.) visam desenvolver linhagens superiores com características desejáveis, como alta produtividade, grãos de aceitação comercial, arquitetura ereta e resistência a estresses bióticos e abióticos (RAMALHO; ABREU, 2015, p. 96-103). Esse processo é realizado de forma sequencial, em que o sucesso de cada etapa depende de decisões acertadas nas fases anteriores (RAMALHO; CARVALHO; NUNES, 2013; ROCHA et al., 2019). Uma metodologia eficaz para características quantitativas é a seleção recorrente, que busca aumentar a frequência de alelos favoráveis por meio de ciclos contínuos de avaliação, seleção e recombinação. Essa abordagem promove progresso genético e preserva variabilidade genética (BERNARDO, 2020, pp. 299–318).

Para espécies autógamas, como o feijoeiro, que apresentam limitações nos cruzamentos entre múltiplos genitores, destaca-se a seleção recorrente multipopulacional (MRS) (FOUILLLOUX; BANNEROT, 1988, p. 503-542). Nessa estratégia, os melhoristas realizam cruzamentos entre numerosos pares de genitores, originando diversas populações. A MRS promove o avanço das populações por meio de gerações endogâmicas, selecionando as melhores progênies em cada ciclo para recombinação subsequente, com cruzamentos intrapopulacionais e pareados (MAICH et al., 2017; MORAIS JÚNIOR et al., 2017; POSADAS et al., 2014; RAMALHO; ABREU; DOS SANTOS, 2005). Esse método facilita a exploração da variabilidade genética e permite a introdução gradual de novos pais ao longo dos ciclos (ALVES et al., 2015; MOURA et al., 2018).

No entanto, mesmo com estratégias como a MRS, o melhoramento do feijoeiro enfrenta desafios adicionais, como a escassez de sementes nas etapas iniciais, limitações financeiras, espaço experimental restrito e a grande quantidade de genótipos a serem avaliados (ZYSTRO; COLLEY; DAWSON, 2018). Para superar essas dificuldades, são empregados diferentes delineamentos experimentais ao longo do melhoramento do feijoeiro (FEDERER, 1956; WILLIAMS; PIEPHO; WHITAKER, 2011), permitindo a seleção e descarte de genótipos de acordo com os objetivos dos programas em cada safra, ano e local (PIEPHO; MÖHRING, 2006; SANTANA et al., 2023).

Na etapa final, os genótipos mais promissores são avaliados em múltiplos ambientes, com delineamentos que garantem maior precisão na estimativa do erro experimental, aumentando a confiabilidade das recomendações (CASLER, 2015).

Todo esse procedimento gera heterogeneidade nos dados, decorrente de fatores como a seleção contínua de genótipos, perdas em experimentos de campo, delineamentos não ortogonais e diferentes delineamentos experimentais (PIEPHO; MÖHRING, 2006; VIANA et al., 2012). Essa complexidade é agravada quando a seleção se concentra apenas nos ensaios mais recentes, sem aproveitar os dados históricos acumulados, o que pode reduzir a confiabilidade dos resultados das análises (ARIEF et al., 2019; RESENDE, 2002). Além disso, a interação genótipos x ambientes (GxA) e a correlação de dados entre parcelas em diferentes ambientes tornam a avaliação genotípica ainda mais desafiadora (ARIEF et al., 2019; RESENDE, 2002). Para agravar ainda mais a situação, dados ausentes não aleatórios (MNAR), decorrentes de exclusões baseadas em características específicas, podem introduzir viés nas estimativas de variância, dificultando análises e recomendações confiáveis (HARTUNG; PIEPHO, 2021; PIEPHO; MÖHRING, 2006).

Nesse contexto, a interação GxA surge como um dos principais desafios para a recomendação de cultivares (PRIYADARSHAN, 2019, p. 457-472). Nesse sentido, os melhoristas recorrem a ensaios multiambientes (EM), que possibilitam identificar padrões de interação GxA e selecionar genótipos com maior adaptabilidade e estabilidade (KELLY et al., 2007; SMITH; CULLIS; GILMOUR, 2001). Inicialmente, a análise desses ensaios se baseava no método de mínimos quadrados ordinários (OLS) para avaliar essas interações (MALOSETTI; RIBAUT; VAN EEUWIJK, 2013).

A análise de variância (ANOVA) é amplamente utilizada no melhoramento de plantas, embora suas pressuposições nem sempre são atendidas, especialmente em casos de desbalanceamento, perda de parcelas ou variação nos delineamentos experimentais (COCHRAN, 1947; PIMENTEL-GOMES, 2009; SOKAL; ROHLF, 1995). Além disso, a ANOVA tradicional não considera o parentesco entre indivíduos nem a covariância entre efeitos genéticos e não genéticos em ensaios multiambientes (HENDERSON, 1975).

Para superar essas limitações, os modelos mistos via REML/BLUP oferecem uma abordagem mais robusta e flexível, permitindo estruturas complexas de variância-covariância e resultados mais confiáveis em condições experimentais desafiadoras (HILL; ROSENBERGER, 1985; PIEPHO et al., 2008).

Com o tempo, novas abordagens surgiram para melhorar a análise da interação GxA, como a análise de regressão conjunta, proposta por Yates e Cochran (1938) e aprimorada por Finlay e Wilkinson (1963) e Eberhart e Russell (1966), modela a interação GxA com base na regressão do desempenho genotípico na média ambiental. Apesar de sua relevância, essa abordagem pode introduzir vieses devido aos erros compartilhados entre o regressor e a variável resposta (CROSSA, 1990; MALOSETTI; RIBAUT; VAN EEUWIJK, 2013; SMITH; CULLIS; THOMPSON, 2005). Métodos como AMMI (GAUCH, 1988) e biplots GGE (YAN et al., 2000) surgiram para explorar a interação GxA usando termos multiplicativos e representações gráficas (GABRIEL, 1971), no entanto, esses métodos podem ser difíceis de interpretar em conjuntos de dados grandes (DIAS et al., 2022; SMITH; CULLIS, 2018) e dependem de condições específicas, como dados balanceados e efeitos genotípicos fixos, o que acaba limitando sua aplicação em situações mais complexas (CROSSA, 2012; SMITH; CULLIS; THOMPSON, 2005).

Nesse contexto, metodologias baseadas em modelos mistos, como o Modelo Misto Fator-Analítico (FA), oferecem uma abordagem mais robusta para interpretar a interação GxA, superando limitações de métodos tradicionais. Desenvolvido por Piepho (1998) e refinado em estudos subsequentes, o FA considera a heterogeneidade da covariância de GEI, a correlação entre esses efeitos e utiliza estruturas de variância espacial adequadas, sem exigir equilíbrio genético ou estatístico, aproveitando os benefícios do efeito genotípico aleatório (SMITH; CULLIS; THOMPSON, 2005, 2001). Além disso, identificar genótipos que se destacam em múltiplas características simultaneamente é outro desafio (OLIVOTO; NARDINO, 2021), já que a análise de características únicas pode introduzir vieses na seleção, especialmente em contextos multivariados. A correlação entre características, causada pelo pleiotropismo, desequilíbrio na fase gamética, e/ou influência comum do ambiente, exige metodologias que integrem essas complexidades para uma seleção mais precisa e eficiente (POLLAK; VAN DER WERF; QUAAS, 1984).

Com o avanço dos estudos, diversas abordagens foram desenvolvidas para a avaliação simultânea de múltiplas características, destacando-se o uso amplamente disseminado dos índices de seleção. Introduzidos por Smith (1936) no melhoramento de plantas e Hazel (1943) no melhoramento animal, esses índices apresentam desafios como a definição de pesos econômicos adequados (CERÓN-ROJAS et al., 2006; STEPHENS et al., 2012), frequentemente baseados em parâmetros genéticos (BHERING et al., 2012) ou atribuídos arbitrariamente (CRUZ; CARNEIRO; REGAZZI, 2014, p. 221-294), o que pode limitar os ganhos genéticos e desconsiderar a interação GxA (OLIVOTO; NARDINO, 2021). Métodos mais avançados, como o índice FAI-BLUP (ROCHA; MACHADO; CARNEIRO, 2018), que incorpora análise de fatores e distância genótipo-ideótipo, e o índice MGIDI, baseado em modelos mistos, têm superado essas limitações ao reduzir problemas de multicolinearidade e eliminar a necessidade de atribuição arbitrária de pesos, oferecendo maior eficiência na seleção multivariada (OLIVOTO; NARDINO, 2021).

Em síntese, o melhoramento do feijoeiro comum enfrenta desafios complexos, desde a seleção de genótipos superiores até a interpretação da interação genótipos x ambientes (GxA) e a avaliação de múltiplas características em diferentes ambientes. A utilização de metodologias avançadas, como modelos mistos via REML/BLUP e abordagens fator-analíticas, tem se mostrado essencial para superar essas dificuldades, especialmente em cenários de desbalanceamento genético e estatístico. Esta tese busca contribuir para o avanço do melhoramento do feijoeiro, explorando estratégias que integram análises robustas de dados sequenciais e multiambientais. No primeiro capítulo, a conexão de experimentos sequenciais permitiu identificar o modelo de melhor ajuste, promovendo ganhos genéticos superiores. No segundo capítulo, a aplicação de *Factor Analytic Selection Tools* (FAST) destacou-se na identificação de progênies com alto desempenho e estabilidade em múltiplas características e ambientes, oferecendo uma estrutura robusta para a seleção de genótipos superiores. Dessa forma, esta tese visa aprimorar a eficiência dos programas de melhoramento de plantas, garantindo a recomendação de genótipos mais produtivos, estáveis e adaptados a diferentes condições de cultivo.

REFERÊNCIAS

- ACHARYA, J. P. et al. Breeding Alfalfa (*Medicago sativa* L.) Adapted to Subtropical Agroecosystems. **Agronomy**, v. 10, n. 5, p. 742, 2020.
- AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, v. 19, n. 6, p. 716–723, 1974.
- ALVES, A. F. et al. Genetic progress and potential of common bean families obtained by recurrent selection. **Crop Breeding and Applied Biotechnology**, v. 15, n. 4, p. 218–226, 2015.
- ARAÚJO, M. DOS S. et al. Modeling covariance structures for genetic and non-genetic effects in cowpea multi-environment trials. **Agronomy Journal**, v. 115, n. 3, p. 1248–1256, 2023.
- ARAÚJO, M. S. et al. GIS-FA: an approach to integrating thematic maps, factor-analytic, and envirotyping for cultivar targeting. **Theoretical and Applied Genetics**, v. 137, n. 4, p. 80, 2024.
- ARIEF, V. N. et al. Utilization of Multiyear Plant Breeding Data to Better Predict Genotype Performance. **Crop Science**, v. 59, n. 2, p. 480–490, 2019.
- BAKARE, M. A. et al. Parsimonious genotype by environment interaction covariance models for cassava (*Manihot esculenta*). **Frontiers in Plant Science**, v. 13, 2022.
- BANČIČ, J. et al. Genomic selection for genotype performance and stability using information on multiple traits and multiple environments. **Theoretical and Applied Genetics**, v. 136, n. 5, p. 104, 2023.
- BARILI, L. D. **Evolução dos cultivares de feijão carioca recomendados no Brasil**. Tese (Doutorado em Genética e Melhoramento)—Viçosa: Universidade Federal de Viçosa, 2015.
- BERNARDO, R. Recurrent Selection. Em: BERNARDO, R. (Ed.). **Breeding for Quantitative Traits in Plants**. 3. ed. Woodbury: Stemma Press, 2020a. p. 299–318.
- BERNARDO, R. Reinventing quantitative genetics for plant breeding: something old, something new, something borrowed, something BLUE. **Heredity**, v. 125, n. 6, p. 375–385, 2020b.
- BHERING, L. L. et al. Genetic gains in physic nut using selection indexes. **Pesquisa Agropecuária Brasileira**, v. 47, n. 3, p. 402–408, 2012.
- BRITO, G. M. DE; CARNEIRO, P. C. S.; SILVA, M. B. P. DA. **Avaliação e seleção de progênies visando o melhoramento do feijão carioca por seleção recorrente**. Simpósio de Integração Acadêmica - SIA 2021. **Anais...**Viçosa: Editora UFV, 2021.

BURNHAM, K. P.; ANDERSON, D. R. Practical Use of the Information-Theoretic Approach. Em: **Model Selection and Inference**. New York, NY: Springer New York, 1998. p. 75–117.

BUTLER, D. G. et al. **ASReml-R Reference Manual Version 4.2**. Hemel Hempstead VSN International Ltd., , 2023.

CARNEIRO, J. E.; DE PAULA JÚNIOR, T.; BORÉM, A. **Feijão: do plantio à colheita**. 1. ed. Viçosa: Editora UFV, 2015.

CASLER, M. D. Fundamentals of Experimental Design: Guidelines for Designing Successful Experiments. **Agronomy Journal**, v. 107, n. 2, p. 692–705, 2015.

CERÓN-ROJAS, J. J. et al. A Selection Index Method Based on Eigenanalysis. **Crop Science**, v. 46, n. 4, p. 1711–1721, 2006.

CHAVES, S. F. S. et al. Employing factor analytic tools for selecting high-performance and stable tropical maize hybrids. **Crop Science**, v. 63, n. 3, p. 1114–1125, 2023a.

CHAVES, S. F. S. et al. Analysis of repeated measures data through mixed models: An application in *Theobroma grandiflorum* breeding. **Crop Science**, v. 63, n. 4, p. 2131–2144, 2023b.

CHRISTENSEN, W. **Agreeing to disagree: Using SAS to make reasoned decisions when information criteria select different models**. SAS Conference Proceedings: Western Users of SAS Software 2018. **Anais...** Sacramento: Paper 099–2018, 2018.

COCHRAN, W. G. Some Consequences When the Assumptions for the Analysis of Variance are not Satisfied. **Biometrics**, v. 3, n. 1, p. 22, 1947.

COLLICCHIO, E.; RAMALHO, M. A. P.; ABREU, A. F. B. Association between plant architecture of the bean plant and seed size. (In Portuguese, with English abstract.). **Pesquisa Agropecuária Brasileira**, v. 32, n. 3, p. 297–304, 1997.

COWLING, W. A. et al. Optimal Contribution Selection Improves the Rate of Genetic Gain in Grain Yield and Yield Stability in Spring Canola in Australia and Canada. **Plants**, v. 12, n. 2, p. 383, 2023.

CROSSA, J. Statistical Analyses of Multilocation Trials. Em: BRADY, N. C. (Ed.). **Advances in Agronomy**. Cambridge: Academic Press, 1990. v. 44p. 55–85.

CROSSA, J. From Genotype × Environment Interaction to Gene × Environment Interaction. **Current Genomics**, v. 13, n. 3, p. 225–244, 2012.

- CRUZ, C. D.; CARNEIRO, P. C. S.; REGAZZI, A. J. Seleção simultânea de caracteres. Em: CRUZ, C. D.; CARNEIRO, P. C. S.; REGAZZI, A. J. (Eds.). **Modelos biométricos aplicados ao melhoramento genético**. 3. ed. Viçosa: Editora UFV, 2014. v. 2p. 221–294.
- CULLIS, B. R. et al. Analysis of yield and oil from a series of canola breeding trials. Part II. Exploring variety by environment interaction using factor analysis. **Genome**, v. 53, n. 11, p. 1002–1016, 2010.
- CULLIS, B. R. et al. Factor analytic and reduced animal models for the investigation of additive genotype-by-environment interaction in outcrossing plant species with application to a *Pinus radiata* breeding programme. **Theoretical and Applied Genetics**, v. 127, n. 10, p. 2193–2210, 2014.
- CULLIS, B. R.; SMITH, A. B.; COOMBES, N. E. On the design of early generation variety trials with correlated data. **Journal of Agricultural, Biological, and Environmental Statistics**, v. 11, n. 4, p. 381–393, 2006.
- DIAS, K. O. G. et al. Leveraging probability concepts for cultivar recommendation in multi-environment trials. **Theoretical and Applied Genetics**, v. 135, n. 4, p. 1385–1399, 2022.
- DOS ANJOS, R. S. R. et al. New Proposals to Estimate Unbiased Selection Gain and Coefficient of Variation in Traits Evaluated Using Score Scales. **Crop Science**, v. 59, n. 3, p. 937–944, 2019.
- DWIVEDI, S. L. et al. Epistasis and pleiotropy-induced variation for plant breeding. **Plant Biotechnology Journal**, v. 22, n. 10, p. 2788–2807, 2024.
- EBERHART, S. A.; RUSSELL, W. A. Stability Parameters for Comparing Varieties. **Crop Science**, v. 6, n. 1, p. 36–40, 1966.
- EVANGELISTA, J. S. P. C. et al. Modeling covariance structures and optimizing *Jatropha curcas* breeding. **Tree Genetics & Genomes**, v. 19, n. 2, p. 21, 2023.
- FEDERER, W. T. Augmented (or hoonuiaku) designs. **Biometrics Unit Technical Reports**, p. 1–33, 1956.
- FELDMANN, M. J. et al. Average semivariance yields accurate estimates of the fraction of marker-associated genetic variance and heritability in complex trait analyses. **PLOS Genetics**, v. 17, n. 8, p. e1009762, 2021.
- FINLAY, K. W.; WILKINSON, G. N. The analysis of adaptation in a plant-breeding programme. **Australian Journal of Agricultural Research**, v. 14, n. 6, p. 742–754, 1963.
- FOUILLLOUX, G.; BANNEROT, H. Selection methods in the common bean (*Phaseolus vulgaris*). Em: GEPTS, P. (Ed.). **Genetic Resources of Phaseolus Beans. Current Plant Science and Biotechnology in Agriculture**. 1. ed. Dordrecht: Springer, 1988. v. 6p. 503–542.

GABRIEL, K. R. The Biplot Graphic Display of Matrices with Application to Principal Component Analysis. **Biometrika**, v. 58, n. 3, p. 453–467, 1971.

GAUCH, H. G. Model Selection and Validation for Yield Trials with Interaction. **Biometrics**, v. 44, n. 3, p. 705–715, 1988.

GU, Z.; EILS, R.; SCHLESNER, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. **Bioinformatics**, v. 32, n. 18, p. 2847–2849, 2016.

HARTUNG, J.; PIEPHO, H. Effect of missing values in multi-environment trials on variance component estimates. **Crop Science**, v. 61, n. 6, p. 4087–4097, 2021.

HAZEL, L. N. The genetic basis for constructing selection indexes. **Genetics**, v. 28, n. 6, p. 476–490, 1943.

HENDERSON, C. R. Best Linear Unbiased Estimation and Prediction under a Selection Model. **Biometrics**, v. 31, n. 2, p. 423, jun. 1975.

HERNÁNDEZ, M. V. et al. Modeling Genotype × Environment Interaction Using a Factor Analytic Model of On-Farm Wheat Trials in the Yaqui Valley of Mexico. **Agronomy Journal**, v. 111, n. 6, p. 2647–2657, 2019.

HILL, R. R.; ROSENBERGER, J. L. Methods for Combining Data from Gemplasm Evaluation Trials ¹. **Crop Science**, v. 25, n. 3, p. 467–470, 1985.

HUNT, C. H. et al. Multi-environment analysis of sorghum breeding trials using additive and dominance genomic relationships. **Theoretical and Applied Genetics**, v. 133, n. 3, p. 1009–1018, 2020.

ISIK, F.; HOLLAND, J.; MALTECCA, C. Multi Environmental Trials. Em: ISIK, F.; HOLLAND, J.; MALTECCA, C. (Eds.). **Genetic Data Analysis for Plant and Animal Breeding**. Cham: Springer International Publishing, 2017. p. 227–262.

KELLY, A. M. et al. The Accuracy of Varietal Selection Using Factor Analytic Models for Multi-Environment Plant Breeding Trials. **Crop Science**, v. 47, n. 3, p. 1063–1070, 2007.

KEMPTON, R. A. The use of biplots in interpreting variety by environment interactions. **The Journal of Agricultural Science**, v. 103, n. 1, p. 123–135, 1984.

LITTLE, R. J. A.; RUBIN, D. B. Missing Data in Experiments. Em: BALDING, D. J. et al. (Eds.). **Statistical Analysis with Missing Data**. 2. ed. New York: John Wiley & Sons, 2002. p. 24–40.

MAICH, R. H. et al. Association between grain yield, grain quality and morpho-physiological traits along ten cycles of recurrent selection in bread wheat (*Triticum aestivum* L.). **Cereal Research Communications**, v. 45, n. 1, p. 146–153, 2017.

MALOSETTI, M.; RIBAUT, J.-M.; VAN EEUWIJK, F. A. The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. **Frontiers in Physiology**, v. 4, 2013.

MASOOD, M. A. et al. Improvement in Precision of Agricultural field Experiments through Design and analysis. **Pakistan Journal of Life and Social Sciences**, v. 6, n. 2, p. 89–91, 2008.

MELO, V. L. DE et al. Modeling (co)variance structures for genetic and non-genetic effects in the selection of common bean progenies. **Euphytica**, v. 216, n. 5, p. 77, 18 maio 2020.

MEYER, K. Factor-analytic models for genotype $\times$ environment type problems and structured covariance matrices. **Genetics Selection Evolution**, v. 41, n. 1, p. 21, 2009.

MÖHRING, J.; PIEPHO, H. -P. Comparison of Weighting in Two-Stage Analysis of Plant Breeding Trials. **Crop Science**, v. 49, n. 6, p. 1977–1988, 2009.

MORAIS JÚNIOR, O. P. et al. Effectiveness of Recurrent Selection in Irrigated Rice Breeding. **Crop Science**, v. 57, n. 6, p. 3043–3058, 2017.

MOURA, L. M. et al. Combining ability of common bean parents in different seasons, locations and generations. **Euphytica**, v. 214, n. 10, p. 181, 25 out. 2018.

MRODE, R. A. Univariate Models with One Random Effect. Em: **Linear Models for the Prediction of Animal Breeding Values**. 3. ed. Wallingford: Cabi, 2014a. p. 34–60.

MRODE, R. A. **Linear models for the prediction of animal breeding values**. 3. ed. Wallingford: CABI Publishing, 2014b.

OGUT, F. et al. Genetic Analysis of Diallel Progeny Test Data Using Factor Analytic Linear Mixed Models. **Forest Science**, v. 60, n. 1, p. 119–127, 2014.

OLIVEIRA, R. L. DE et al. Precisão experimental em ensaios com a cultura do feijão. **Ciência e Agrotecnologia**, v. 33, n. 1, p. 113–119, 2009.

OLIVEIRA, I. C. M. et al. Genotype-by-environment interaction and yield stability analysis of biomass sorghum hybrids using factor analytic models and environmental covariates. **Field Crops Research**, v. 257, p. 107929, 2020.

OLIVOTO, T.; NARDINO, M. MGIDI: toward an effective multivariate selection in biological experiments. **Bioinformatics**, v. 37, n. 10, p. 1383–1389, 2021.

PATTERSON, H. D.; THOMPSON, R. Recovery of Inter-Block Information when Block Sizes are Unequal. **Biometrika**, v. 58, n. 3, p. 545, 1971.

PEEL, M. C.; FINLAYSON, B. L.; MCMAHON, T. A. Updated world map of the Köppen-Geiger climate classification. **Hydrology and Earth System Sciences**, v. 11, n. 5, p. 1633–1644, 2007a.

PEEL, M. C.; FINLAYSON, B. L.; MCMAHON, T. A. Updated world map of the Köppen-Geiger climate classification. **Hydrology and Earth System Sciences**, v. 11, n. 5, p. 1633–1644, 11 out. 2007b.

PEIXOTO, M. A. et al. Selection of maize hybrids: an approach with multi-trait, multi-environment, and ideotype-design. **Crop Breeding and Applied Biotechnology**, v. 21, n. 2, 2021.

PIEPHO, H. et al. A stage-wise approach for the analysis of multi-environment trials. **Biometrical Journal**, v. 54, n. 6, p. 844–860, 2012.

PIEPHO, H. A coefficient of determination R^2 for generalized linear mixed models. **Biometrical Journal**, v. 61, n. 4, p. 860–872, 2019.

PIEPHO, H.; MÖHRING, J. Selection in Cultivar Trials—Is It Ignorable? **Crop Science**, v. 46, n. 1, p. 192–201, 2006.

PIEPHO, H. P. et al. BLUP for phenotypic selection in plant breeding and variety testing. **Euphytica**, v. 161, n. 1–2, p. 209–228, 2008.

PIEPHO, H.-P. Analyzing Genotype-Environment Data by Mixed Models with Multiplicative Terms. **Biometrics**, v. 53, n. 2, p. 761, 1997.

PIEPHO, H.-P. Empirical best linear unbiased prediction in cultivar trials using factor-analytic variance-covariance structures. **Theoretical and Applied Genetics**, v. 97, n. 1–2, p. 195–201, 1998a.

PIEPHO, H.-P. Empirical best linear unbiased prediction in cultivar trials using factor-analytic variance-covariance structures. **Theoretical and Applied Genetics**, v. 97, n. 1–2, p. 195–201, 1998b.

PIMENTEL-GOMES, F. **Curso de Estatística Experimental**. 15. ed. Piracicaba: FEALQ, 2009.

POLLAK, E. J.; VAN DER WERF, J.; QUAAS, R. L. Selection Bias and Multiple Trait Evaluation. **Journal of Dairy Science**, v. 67, n. 7, p. 1590–1595, 1984.

POSADAS, L. G. et al. Elite Performance for Grain Yield from Unadapted Exotic Soybean Germplasm in Three Cycles of a Recurrent Selection Experiment. **Crop Science**, v. 54, n. 6, p. 2536–2546, 2014.

PRIYADARSHAN, P. M. Genotype-by-Environment Interactions. Em: **PLANT BREEDING: Classical to Modern**. Singapore: Springer Singapore, 2019. p. 457–472.

R CORE TEAM. **R: A language and environment for statistical computing**. Vienna, Austria R Foundation for Statistical Computing, , 2023.

RAMALHO, M. A. P. et al. **Aplicações da Genética Quantitativa no Melhoramento de Plantas Autógamas**. 1. ed. Lavras: Editora UFLA, 2012.

RAMALHO, M. A. P.; ABREU, A. DE F. B. Obtenção de cultivares. Em: CARNEIRO, J. E.; PAULA JÚNIOR, T. J. DE; BORÉM, A. (Eds.). **Feijão: do plantio a colheita**. Viçosa: Editora UFV, 2015. p. 96–114.

RAMALHO, M. A. P.; ABREU, Â. DE F. B.; DOS SANTOS, J. B. Genetic progress after four cycles of recurrent selection for yield and grain traits in common bean. **Euphytica**, v. 144, n. 1–2, p. 23–29, 2005.

RAMALHO, M. A. P.; CARVALHO, B. L.; NUNES, J. A. R. Perspectives for the Use of Quantitative Genetics in Breeding of Autogamous Plants. **ISRN Genetics**, v. 2013, p. 1–6, 2013.

RAMALHO, M. A. P.; PIROLA, L. H.; ABREU, A. F. B. Alternativas na seleção de plantas de feijoeiro com porte ereto e grão tipo carioca. **Pesquisa Agropecuária Brasileira**, v. 33, n. 12, p. 1989–1994, 1998.

RESENDE, M. D. V. DE; ALVES, R. S. Statistical significance, selection accuracy, and experimental precision in plant breeding. **Crop Breeding and Applied Biotechnology**, v. 22, n. 3, 2022.

RESENDE, M. D. V. DE. **Genética biométrica e estatística no melhoramento de plantas perenes**. [s.l.] Emprapa Informação Tecnológica, 2002.

RESENDE, M. D. V.; DUARTE, J. B. Precisão e controle de qualidade em experimentos de avaliação de cultivares. **Pesquisa Agropecuária Tropical**, v. 37, n. 3, p. 182–194, 16 out. 2007.

ROCHA, J. R. DO A. S. DE C. et al. Selection of Superior Inbred Progenies toward the Common Bean Ideotype. **Agronomy Journal**, v. 111, n. 3, p. 1181–1189, 2019.

ROCHA, J. R. DO A. S. DE C.; MACHADO, J. C.; CARNEIRO, P. C. S. Multitrait index based on factor analysis and ideotype-design: proposal and application on elephant grass breeding for bioenergy. **GCB Bioenergy**, v. 10, n. 1, p. 52–60, 2018.

RODRÍGUEZ-ÁLVAREZ, M. X. et al. Correcting for spatial heterogeneity in plant breeding experiments with P-splines. **Spatial Statistics**, v. 23, p. 52–71, mar. 2018.

SAE-LIM, P. et al. Identifying environmental variables explaining genotype-by-environment interaction for body weight of rainbow trout (*Onchorynchus mykiss*): reaction norm and factor analytic models. **Genetics Selection Evolution**, v. 46, n. 1, p. 16, 2014.

SALVADOR, F. V. et al. Correcting experimental data for spatial trends in a common bean breeding program. **Crop Science**, v. 62, n. 2, p. 825–838, 14 mar. 2022.

SALVADOR, F. V. et al. Accounting for heterogeneous genotypic variance of progenies within populations in autogamous plant breeding: An evaluation in common bean. **Crop Science**, v. 61, n. 6, p. 3947–3961, 2021.

SANTANA, A. S. et al. Analysis of advanced generation multistage field trials data in autogamous plant breeding: An evaluation in common Bean. **Crop Science**, v. 63, n. 3, p. 1221–1233, 2023.

SCHMIDT, P. et al. Estimating Broad-Sense Heritability with Unbalanced Data from Agricultural Cultivar Trials. **Crop Science**, v. 59, n. 2, p. 525–536, 2019.

SCHWARZ, G. Estimating the Dimension of a Model. **The Annals of Statistics**, v. 6, n. 2, 1978.

SILVA, A. C. et al. Selection of common bean parents and segregating populations targeting fusarium wilt resistance and grain yield. **Agronomy Science and Biotechnology**, v. 9, p. 1–20, 2023.

SMITH, A. et al. Plant Variety Selection Using Interaction Classes Derived From Factor Analytic Linear Mixed Models: Models With Independent Variety Effects. **Frontiers in Plant Science**, v. 12, p. 1–17, 2021.

SMITH, A. B. et al. Varietal selection for perennial crops where data relate to multiple harvests from a series of field trials. **Euphytica**, v. 157, n. 1–2, p. 253–266, 2007.

SMITH, A. B. et al. Factor analytic mixed models for the provision of grower information from national crop variety testing programs. **Theoretical and Applied Genetics**, v. 128, n. 1, p. 55–72, 2015.

SMITH, A. B.; CULLIS, B. R. Plant breeding selection tools built on factor analytic mixed models for multi-environment trial data. **Euphytica**, v. 214, n. 8, p. 143, 2018.

SMITH, A. B.; CULLIS, B. R.; THOMPSON, R. The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. **The Journal of Agricultural Science**, v. 143, n. 6, p. 449–462, 2005.

SMITH, A.; CULLIS, B.; GILMOUR, A. Applications: The Analysis of Crop Variety Evaluation Data in Australia. **Australian & New Zealand Journal of Statistics**, v. 43, n. 2, p. 129–145, 2001.

SMITH, A.; CULLIS, B.; THOMPSON, R. Analyzing Variety by Environment Data Using Multiplicative Mixed Models and Adjustments for Spatial Field Trend. **Biometrics**, v. 57, n. 4, p. 1138–1147, dez. 2001.

SMITH, H. F. A discriminant function for plant selection. **Annals of Eugenics**, v. 7, n. 3, p. 240–250, 1936.

SOKAL, R. R.; ROHLF, F. J. **The principles and practice of statistics in biological research**. 3. ed. New York: W. H. Freeman, 1995.

STEPHENS, M. J. et al. Genetic Parameters and Development of a Selection Index for Breeding Red Raspberries for Processing. **Journal of the American Society for Horticultural Science**, v. 137, n. 4, p. 236–242, 2012.

STUDNICKI, M. et al. Yield response of winter wheat cultivars to environments modeled by different variance-covariance structures in linear mixed models. **Spanish Journal of Agricultural Research**, v. 14, n. 2, p. e0703, 2016.

TOLHURST, D. J. et al. Genomic selection in multi-environment plant breeding trials using a factor analytic linear mixed model. **Journal of Animal Breeding and Genetics**, v. 136, n. 4, p. 279–300, 2019.

VERBYLA, A. P. A note on model selection using information criteria for general linear models estimated using REML. **Australian & New Zealand Journal of Statistics**, v. 61, n. 1, p. 39–50, 9 mar. 2019.

VIANA, J. M. S. et al. Relevance of Pedigree, Historical Data, Dominance, and Data Unbalance for Selection Efficiency. **Agronomy Journal**, v. 104, n. 3, p. 722–728, 2012.

WICKHAM, H. **ggplot2: Elegant Graphics for Data Analysis**. New York, NY: Springer-Verlag, 2016.

WILLIAMS, E.; PIEPHO, H.-P.; WHITAKER, D. Augmented p-rep designs. **Biometrical Journal**, v. 53, n. 1, p. 19–27, 2011.

YAN, W. et al. Cultivar Evaluation and Mega-Environment Investigation Based on the GGE Biplot. **Crop Science**, v. 40, n. 3, p. 597–605, 2000.

YATES, F.; COCHRAN, W. G. The analysis of groups of experiments. **The Journal of Agricultural Science**, v. 28, n. 4, p. 556–580, 1938.

ZHANG, R.; HAN, D.; HU, X. Analyzing the performance of corn in China using a factor-analytic variance-covariance structure with multiple factors. **Crop Science**, v. 60, n. 1, p. 190–201, 2020.

ZYSTRO, J.; COLLEY, M.; DAWSON, J. Alternative Experimental Designs for Plant Breeding. Em: **Plant Breeding Reviews**. [s.l.] Wiley, 2018. p. 87–117.

CAPÍTULO I

ANÁLISE MULTIGERAÇÃO EM PROGRAMA DE MELHORAMENTO DE FEIJOEIRO COMUM: ABORDAGENS PARA LIDAR COM DADOS DESBALANCEADOS

RESUMO

No melhoramento do feijoeiro comum (*Phaseolus vulgaris* L.), são conduzidos ensaios sequenciais ao longo de sucessivas gerações, enfrentando desafios como desbalanceamento genético e/ou estatístico. Nesse contexto, diferentes delineamentos experimentais são utilizados, e, à medida que o programa avança, os ensaios são realizados em diversos ambientes. Para lidar com dados desbalanceados, a metodologia de modelos mistos é amplamente empregada, permitindo modelar as matrizes de variância-covariância dos efeitos aleatórios. Assim, o objetivo do presente estudo foi conectar experimentos sequenciais de avaliação de progênies de feijoeiro, considerando diferentes delineamentos experimentais com desbalanceamento genético e estatístico, com a finalidade de selecionar o modelo de melhor ajuste e avaliar o impacto da utilização dos dados de rendimento de grãos (kg ha^{-1}) disponíveis nos ensaios sequenciais no processo de seleção. Inicialmente, um grupo de 20 experimentos, totalizando 400 progênies $F_{4:6}$, foi avaliado durante a safra seca de 2019, utilizando delineamento de blocos completos casualizados. Posteriormente, 95 progênies selecionadas foram avaliadas em três safras subsequentes: águas/2019, inverno/2020 e águas/2020 ($F_{4:7}$, $F_{4:8}$ e $F_{4:9}$, respectivamente), conduzidas em delineamento látice triplo 10x10. Foram testados cinco modelos, considerando diferentes estruturas de variância-covariância para os efeitos residuais de progênies dentro de cada geração. O modelo que assumiu estrutura Fator Analítico de 1ª Ordem para o efeito de progênie dentro de geração e variância residual diagonal heterogênea foi o que apresentou melhor ajuste aos dados. O ganho genético médio predito obtido a partir do modelo de melhor ajuste foi 68% superior ao dos demais modelos. Da mesma forma, houve um incremento significativo na herdabilidade média. Observou-se também alteração na classificação das progênies com base nos valores genotípicos preditos ao longo das safras. A utilização de modelos mistos, ajustando as matrizes de variância-covariância, mostrou-se a melhor opção em comparação ao modelo tradicional (simetria composta), especialmente em cenários com desbalanceamento estatístico e/ou genético.

Palavras-chave: feijão carioca; modelos mistos; variância-covariância.

ABSTRACT

In our efforts to improve common bean (*Phaseolus vulgaris* L.), we conducted sequential trials over successive generations, facing challenges from unbalanced genetic and statistical data. We employed different experimental designs as the breeding program progressed, carrying out trials in multiple environments. To handle unbalanced data, we applied mixed models methodology, which enabled us to model variance-covariance matrices of random effects effectively. This study aimed to connect sequential experiments for evaluating bean progenies—accounting for different experimental designs with unbalanced genetic and statistical data—to select the best-fitting model and assess how using available grain yield data (kg ha^{-1}) from sequential trials impacted the selection process. We initially evaluated 400 $F_{4:6}$ progenies across 20 experiments during the 2019 dry season using a randomized complete block design. We then assessed 95 selected progenies in three subsequent seasons (wet season 2019, winter 2020, and wet season 2020; $F_{4:7}$, $F_{4:8}$, and $F_{4:9}$ generations) through a triple lattice design (10×10). We tested five models with different variance-covariance structures for progeny residual effects within each generation. The model combining a first-order factor analytic structure for progeny effects with heterogeneous diagonal residual variance provided the best fit. This optimal model delivered 68% higher predicted average genetic gain than alternative models and significantly increased average heritability. We also observed changes in progeny rankings based on predicted genotypic values across seasons. Our results demonstrated that mixed models with adjusted variance-covariance matrices outperformed traditional compound symmetry models, particularly in scenarios with unbalanced genetic or statistical data.

Keywords: carioca bean; mixed models; variance-covariance.

1 INTRODUÇÃO

Os programas de melhoramento do feijão comum (*Phaseolus vulgaris* L.) visam desenvolver genótipos superiores com características como alta produtividade, qualidade de grãos, resistência a estresses bióticos e abióticos, e adaptação a vários locais (RAMALHO; ABREU, 2015, p. 96-103). O melhoramento do feijão comum envolve várias etapas sequenciais e requer decisões bem-sucedidas em cada uma delas (RAMALHO; CARVALHO; NUNES, 2013; ROCHA et al., 2019). A seleção recorrente é uma metodologia chave, a qual envolve ciclos contínuos de avaliação, seleção e recombinação para incrementar as frequências dos alelos desejados e preservar a variabilidade genética (BERNARDO, 2020a, p. 299–318).

Tal processo enfrenta vários desafios, incluindo escassez de sementes nas fases iniciais, restrições financeiras e de recursos, e a necessidade de avaliar muitos genótipos (ZYSTRO; COLLEY; DAWSON, 2018, p. 87-117). Para lidar com esses desafios, são empregados diferentes delineamentos experimentais (FEDERER, 1956; WILLIAMS; PIEPHO; WHITAKER, 2011), com a seleção e descarte de genótipos menos promissores após testes sob diversas condições (CASLER, 2015). No entanto, a redução deliberada de dados cria padrões informativos de dados ausentes não aleatórios (MNAR) (LITTLE; RUBIN, 2002, p. 24-40), que podem introduzir viés nas estimativas dos componentes de variância, especialmente se a seleção dos genótipos depender da característica em questão (HARTUNG; PIEPHO, 2021; PIEPHO; MÖHRING, 2006).

O método tradicional de Análise de Variância (ANOVA), comumente usados no melhoramento de plantas (COCHRAN, 1947; PIMENTEL-GOMES, 2009; SOKAL; ROHLF, 1995), enfrenta limitações como suposições nem sempre atendidas, desbalanceamento estatístico e/ou genético e a incapacidade de contabilizar o parentesco entre indivíduos ou a covariância entre efeitos genéticos e residuais em diferentes ambientes. Estes desafios exigem abordagens alternativas no melhoramento de plantas (Resende, 2002; Piepho e Möhring, 2006; Viana et al., 2012).

A metodologia de modelo misto, particularmente por meio de máxima verossimilhança restrita/melhor preditor linear imparcial (REML/BLUP), oferece uma solução mais flexível e robusta (HENDERSON, 1975). Ao incorporar

estruturas complexas de covariância e explorar interações, o REML/BLUP melhora a precisão na estimativa de valores genotípicos, aumentando a eficiência do programa de melhoramento de plantas. Esta metodologia permite a inclusão de matrizes de parentesco e interação genótipo-ambiente, fornecendo representações mais fiéis da natureza experimental e permitindo decisões de seleção e recomendação mais informadas (HILL; ROSENBERGER, 1985; PIEPHO et al., 2008).

Aumentar a acurácia das inferências sobre os valores genotípicos e ordená-los para fins de seleção e/ou recomendação pode tornar os programas de melhoramento de plantas mais eficientes (RESENDE; DUARTE, 2007). Conseguir tal feito envolve o emprego de modelos estatísticos que representem mais fielmente a natureza dos experimentos, onde seja possível a incorporação de estruturas de (co)variância que expliquem as relações entre os indivíduos, facilitadas pela matriz de parentesco, bem como a interação $G \times A$, entre outras possíveis interações (MELO et al., 2020; RODRÍGUEZ-ÁLVAREZ et al., 2018; SANTANA et al., 2023).

Deste modo, os objetivos deste estudo consistiram em: i) conectar experimentos sequenciais de avaliação de progênies de feijoeiro comum com delineamentos distintos, desbalanceamento genético e estatístico através da metodologia de modelos mistos; ii) investigar diferentes estruturas (co)variância para as matrizes de efeitos aleatórios considerando os ensaios multiambientes; iii) selecionar o modelo de melhor ajuste aos dados; iv) avaliar o impacto da utilização dos dados disponíveis nos ensaios sequenciais e; v) identificar e selecionar progênies para recombinação.

2 MATERIAL E MÉTODOS

2.2 Material vegetal e delineamentos estatísticos

Foram utilizados dados provenientes de cinco ensaios do Programa de Seleção Recorrente (Ciclo C0) de Feijão Carioca, conduzidos pela Universidade Federal de Viçosa (UFV) entre 2018 e 2020, com foco na resistência à murcha-de-fusário (*Fusarium oxysporum* f. sp. *phaseoli* – FOP), cuja avaliação deve ser conduzida em linhagens homozigotas, e não em famílias (progênies) ou indivíduos segregantes, pois a replicação do genótipo é necessária para garantir a confiabilidade dos resultados. Por essa razão, a seleção para resistência à murcha-de-fusário não é realizada durante os estágios iniciais do programa de melhoramento do feijoeiro-comum, mas sim em um grupo reduzido de linhagens previamente selecionadas com base em características agrônômicas desejáveis. A Tabela 1 apresenta os detalhes dos experimentos.

Tabela 1: Descrição dos quatro ensaios utilizados no presente trabalho.

Ensaio	Ano	Safr	Geração	Tratamento*	Delineamento	Repetição	Parcela**
1	2019	Seca	F _{4:6}	20 (5)	20 DBC	3	1L x 2m
2	2019	Águas	F _{4:7}	95 (5)	Látice	3	2L x 2m
3	2020	Inverno	F _{4:8}	95 (5)	Látice	3	2L x 2m
4	2020	Águas	F _{4:9}	95 (5)	Látice	3	2L x 2m

*Número de testemunhas comuns entre parênteses.

** L: linhas, m: metros

O programa de seleção recorrente para resistência à murcha-de-fusário iniciou-se em 2014, com a avaliação do potencial de 12 genótipos de feijoeiro em esquema de dialelo parcial. Foram selecionadas cultivares comerciais de elite e linhagens dos principais programas de melhoramento de feijão do Brasil, abrangendo genes de resistência às principais doenças que afetam a cultura. A partir dos cruzamentos dialélicos, obtiveram-se 36 populações F₁, que foram multiplicadas, resultando em 34 populações F₂. Estas foram conduzidas em *bulk* até a geração F₄, das quais foram selecionadas as 20 melhores populações para continuidade do programa. A partir dessas 20 populações, iniciaram-se as avaliações das progênies, com 800 progênies F_{4:5} avaliadas em delineamento de blocos aumentados na safra águas/2018 (SILVA et al., 2023).

Da geração F_{4:5}, foram selecionadas as 20 melhores progênies de cada uma das 20 populações, resultando em 400 progênies F_{4:6}. A geração

$F_{4:6}$ foi avaliada em 20 experimentos (um experimento para cada população), em delineamento de blocos casualizados, com três repetições, 20 tratamentos (progênes) e cinco testemunhas (BRITO; CARNEIRO; SILVA, 2021). A partir das 400 progênes $F_{4:6}$, selecionaram-se as 95 melhores progênes, representantes de todas as 20 populações, que foram avaliadas em delineamento látice triplo em três safras (Tabela 1). As testemunhas utilizadas em todos os experimentos foram: Manteigão Fosco (resistente a FOP), Meia Noite (suscetível a FOP), Pérola (resistente a FOP e com padrão de grãos comerciais desejável), BRS Estilo (suscetível a FOP e alto potencial de rendimento de grãos) e VC25 (linhagem elite resistente a FOP).

Todos os ensaios foram realizados na Unidade de Ensino, Pesquisa e Extensão de Coimbra da UFV, localizada no município de Coimbra, Minas Gerais. Segundo a classificação climática de Köppen, o clima da região é do tipo Cwa, subtropical de inverno seco (PEEL; FINLAYSON; MCMAHON, 2007b).

A escolha do delineamento experimental para cada ensaio foi baseada no número de tratamentos, viabilidade da área experimental e disponibilidade de sementes para repetição dos tratamentos. Embora o tamanho das parcelas tenha variado entre os experimentos (conforme mostrado na Tabela 1), os sulcos foram preparados de forma consistente, com espaçamento de 0,5 metros entre si e densidade de 15 sementes por metro em todos os experimentos. O manejo e os tratos culturais, como adubação e controle de pragas e doenças, seguiram as recomendações técnicas para a cultura do feijoeiro na região (CARNEIRO; DE PAULA JÚNIOR; BORÉM, 2015).

2.3 Metodologia REML/BLUP e modelagem de estruturas de covariância

Para abordar o desbalanceamento experimental e genético observado nos experimentos, foram ajustados cinco modelos, considerando diferentes estruturas de (co)variância para efeitos genéticos e residuais. A metodologia da máxima verossimilhança residual/melhor predição linear não viesada (REML/BLUP) (HENDERSON, 1975; PATTERSON; THOMPSON, 1971) foi utilizada para estimar os componentes de (co)variância e prever os valores genotípicos. O modelo é descrito pela Equação [1]:

$$y = X\beta + Z_1g + Z_2b + e \quad [1]$$

em que:

- y é o vetor de observações fenotípicas;
- β é o vetor de efeitos fixos de gerações, repetições dentro de gerações e blocos dentro da geração $F_{4:6}$, somado à média;
- g é o vetor de efeitos aleatórios de progênies dentro de geração, com $g \sim N(0, G)$;
- b é o vetor de efeitos aleatórios de blocos dentro das gerações onde os experimentos foram instalados em blocos incompletos, com $b \sim N(0, B)$;
- e é o vetor de resíduos, com $e \sim N(0, R)$;
- X , Z_1 e Z_2 são as matrizes de incidência associadas aos efeitos β , g e b , respectivamente.

O vetor de soluções representando os efeitos fixos e aleatórios do modelo descrito na Equação [1] pode ser estimado usando a Equação [2], conforme descrito por Henderson *et al.* (1959):

$$\begin{bmatrix} X^T R^{-1} X & X^T R^{-1} Z_g & X^T R^{-1} Z_b \\ Z_g^T R^{-1} X & Z_g^T R^{-1} Z_g + G_g^{-1} & Z_g^T R^{-1} Z_b \\ Z_b^T R^{-1} X & Z_b^T R^{-1} Z_g & Z_b^T R^{-1} + G_b^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{g} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} X^T R^{-1} y \\ Z_g^T R^{-1} y \\ Z_b^T R^{-1} y \end{bmatrix} \quad [2]$$

Diferentes estruturas de (co)variância para G , B e R foram modeladas, resultando em cinco modelos (Tabela 2). No Modelo 1, R assumiu homogeneidade de variâncias, denominada componente de variância (CV), $\sigma_e^2 I_e$, onde σ_e^2 é o componente de variância associado aos resíduos e I_e é uma matriz identidade associada aos resíduos. Do Modelo 2 ao 5, R assumiu estrutura de (co)variância diagonal heterogênea (DIAG), dada por

$$R = \bigoplus_{j=1}^t \sigma_{ej}^2 I_{ej}$$

assumindo variâncias residuais para cada safra e ausência de covariância entre pares de safras. Para o efeito de blocos nas gerações onde os experimentos foram instalados em blocos incompletos, foi empregada a

estrutura de (co)variância diagonal heterogênea (DIAG) para os cinco modelos testados:

$$B = \bigoplus_{j=1}^t \sigma_{bj}^2 I_{bj}$$

Para o efeito de progênes dentro de geração (G), foi inicialmente testada a estrutura de (co)variância simetria composta (CS), onde:

$$G = \sigma_g^2 J_t + J_t \otimes I_g$$

Essa estrutura postula variância comum e independência dos efeitos genéticos entre as gerações (Modelos 1 e 2). No Modelo 3, a estrutura de (co)variância empregada para G foi a diagonal heterogênea (DIAG):

$$G = \bigotimes_{j=1}^t \sigma_{gj}^2$$

assumindo variâncias genéticas para cada geração e ausência de covariância entre pares de gerações.

Para avaliar a heterogeneidade das variâncias genéticas, foi aplicado um modelo que avalia correlações entre pares de gerações, permitindo heterogeneidade na variância genética e covariância entre diferentes gerações (UN – não estruturada com correlação geral) (Modelo 4). Sua estrutura de (co)variância segue a forma:

$$G = UN = \begin{bmatrix} \sigma_{ge1}^2 & \rho_{12}\sigma_{12}^2 & \rho_{13}\sigma_{13}^2 & \rho_{14}\sigma_{14}^2 \\ \rho_{12}\sigma_{12}^2 & \sigma_{ge2}^2 & \rho_{23}\sigma_{23}^2 & \rho_{24}\sigma_{24}^2 \\ \rho_{13}\sigma_{13}^2 & \rho_{23}\sigma_{23}^2 & \sigma_{ge3}^2 & \rho_{34}\sigma_{34}^2 \\ \rho_{14}\sigma_{14}^2 & \rho_{24}\sigma_{24}^2 & \rho_{34}\sigma_{34}^2 & \sigma_{ge4}^2 \end{bmatrix}$$

No Modelo 5, foi utilizada a estrutura de (co)variância Fator Analítico (FA) para G. O modelo FA apresenta uma estrutura mais parcimoniosa e flexível do que o modelo não estruturado (KELLY et al., 2007), exigindo menos parâmetros. Neste modelo, o efeito aleatório da i -ésima progênie na j -ésima geração, denotado como g_{ij} , é expresso da seguinte forma:

$$g_{ij} = \sum_{k=1}^t \lambda_{jk} f_{ik} + \delta_{ij} \quad [3]$$

onde

- λ_{ij} é a carga fatorial k na j -ésima geração;
- f_{ik} é o *score* da i -ésima progênie para o fator latente k relacionado à j -ésima geração em λ_{jk} ;
- δ_{ij} é o termo residual que representa a falta de ajuste ao modelo.

Assim, o modelo FA expressa o efeito aleatório da i -ésima geração na j -ésima progênie como uma função linear dos fatores latentes λ_{jk} com sensibilidade aleatória f_{ik} para $k = 1, 2, \dots, t$ mais um termo de erro δ_{ij} . A especificação do modelo FA em forma de variância é:

$$G = \Lambda \Lambda^T + \psi \quad [4]$$

em que:

$$\Lambda \Lambda^T = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1k} \\ \lambda_{21} & \lambda_{22} & \dots & \lambda_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{s1} & \lambda_{s2} & \dots & \lambda_{sk} \end{bmatrix} \begin{bmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{s1} \\ \lambda_{21} & \lambda_{22} & \dots & \lambda_{s2} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{1k} & \lambda_{2k} & \dots & \lambda_{sk} \end{bmatrix} + \begin{bmatrix} \psi_1 & 0 & \dots & 0 \\ 0 & \psi_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \psi_s \end{bmatrix} \quad [5]$$

onde Λ é uma matriz $s \times t$ de cargas, com a k -ésima coluna contendo as cargas ambientais (geração) para o k -ésimo fator latente ($k = 1, 2, \dots, t$), e ψ é uma matriz diagonal $s \times s$ com uma variância específica para cada geração. O número de fatores latentes é chamado de ordem do modelo, e, no presente estudo, FA k representa um modelo FA de ordem k . Foi utilizado o modelo FA1.

Tabela 2: Estruturas de (co)variância ajustadas aos efeitos aleatórios descritos na Equação [1].

Matrizes de (co)variância*	Modelos**				
	1	2	3	4	5
G	CS	CS	DIAG	UN	FA1
R	CV	DIAG	DIAG	DIAG	DIAG

*G: Matriz de (co)variância dos efeitos genéticos; R: Matriz de (co)variância residual.

**DIAG: Diagonal Heterogênea; CS: Simetria Composta; UN: Não Estruturada com Correlação Geral; FA1: Fator Analítico de 1ª Ordem; CV: Componente de Variância.

2.4 Seleção do modelo

Segundo Verbyla (2019), o log-verossimilhança residual não pode ser utilizado para comparar modelos lineares gerais com diferentes efeitos fixos. Como os modelos ajustados no presente trabalho possuem os mesmos efeitos fixos, os critérios de informação baseados nas razões do log-verossimilhança podem ser aplicados. A qualidade do ajuste dos modelos aos dados foi avaliada utilizando o critério de informação de Akaike (AIC) (AKAIKE, 1974) e o critério de informação bayesiano (BIC) (SCHWARZ, 1978):

$$AIC = -2\text{Log}(\text{likelihood}) + 2p \quad [6]$$

$$BIC = -2\text{Log}(\text{likelihood}) + p \text{Log } n \quad [7]$$

onde

- p é o número de parâmetros estimados;
- n é o número de observações.

Os modelos com menores valores de AIC e BIC são considerados os de melhor ajuste aos dados.

Os ganhos de seleção esperados (GS%) para cada modelo foram calculados considerando as 30 melhores progênies (aproximadamente 7,5% do conjunto de dados):

$$GS\% = \left(\frac{\bar{g}_s - \bar{g}_0}{\bar{y}} \right) * 100 \quad [8]$$

onde:

- $\bar{g}_s$ é a média dos valores BLUP das progênies selecionadas;
- $\bar{g}_0$ é a média dos valores BLUP da população avaliada;

- $\bar{y}$ é a média fenotípica.

No modelo FA1, foi calculada a porcentagem de variância acumulada explicada pelo fator k ($\% \bar{V}$) (SMITH et al., 2015):

$$\% \bar{V} = \left(\frac{tr(\Lambda \Lambda')}{tr(\Sigma g)} \right) \quad [9]$$

onde Λ^* é a matriz $h \times k$ de rotação de cargas.

2.5 Estimativas de parâmetros genéticos e não genéticos

A variância fenotípica ($\hat{\sigma}_p^2$), a herdabilidade generalizada (H^2) para dados desbalanceados, conforme proposto por Cullis *et al.* (2006), e a acurácia (r) (MRODE, 2014, p. 45) foram calculadas utilizando as seguintes equações:

$$\hat{\sigma}_p^2 = \hat{\sigma}_g^2 + \hat{\sigma}_b^2 + \hat{\sigma}_e^2 \quad [10]$$

$$H^2 = 1 - \frac{\bar{V}(\Delta)}{2\hat{\sigma}_g^2} \quad [11]$$

$$r = \sqrt{1 - \frac{PEV}{\hat{\sigma}_g^2}} \quad [12]$$

onde:

- $\bar{V}(\Delta)$ é a variância média do erro de predição entre progênies par a par;
- $\hat{\sigma}_g^2$ é variância de genótipos;
- PEV é a variância do erro de predição *per se*, obtida a partir da diagonal da inversa generalizada da matriz de coeficientes das equações de modelo misto (MME).

Os valores de variância fenotípica ($\hat{\sigma}_p^2$), variância de bloco ($\hat{\sigma}_b^2$) nos experimentos em delineamento de blocos incompletos e variância residual ($\hat{\sigma}_e^2$) também foram estimados. O coeficiente de variação experimental (CV_{e_i}) para cada geração foi calculado por meio da seguinte equação:

$$CV_{e_i}(\%) = 100 \times \frac{\sqrt{\hat{\sigma}_{e_i}^2}}{\mu_i} \quad [13]$$

onde $\hat{\sigma}_e^2$ representa a variância residual na safra i e μ é a média fenotípica na safra i .

As correlações genéticas entre pares de gerações (ρ) no modelo Fator Analítico foram estimadas utilizando as cargas fatoriais (λ_{kj}) (CULLIS et al., 2010):

$$\rho = \frac{\sum_1^k \lambda_{ks/j}^* \lambda_{ks'/j'}^*}{\sqrt{\sigma_{g_{s/j}}^2 \sigma_{g_{s'/j'}}^2}} \quad [14]$$

onde $\lambda_{ks/j}^*$ é a carga fatorial rotacionada do k -ésimo fator na s -ésima geração/safra. No modelo não estruturado, as correlações genéticas entre as gerações foram estimadas utilizando as covariâncias da matriz de (co)variância G :

$$\rho = \frac{\hat{\sigma}_{g_{ij}}}{\sqrt{\hat{\sigma}_{g_i}^2 \times \hat{\sigma}_{g_j}^2}} \quad [15]$$

onde:

- $\hat{\sigma}_{g_{ij}}$ é a covariância genética das progênies entre os pares de gerações ij ;
- $\hat{\sigma}_{g_i}^2$ é a variância genética entre as progênies na geração i ;
- $\hat{\sigma}_{g_j}^2$ é a variância genética entre as progênies na geração j .

Dado o desbalanceamento do conjunto de dados, avaliar a confiabilidade na seleção de cada progênie torna-se essencial para estimar a incerteza associada à escolha de uma progênie específica. Para isso, a confiabilidade de seleção de cada progênie em cada geração foi calculada utilizando a seguinte equação, conforme descrito por Mrode (2014, p. 45):

$$r_i^2 = 1 - \frac{PEV_i}{\hat{\sigma}_g^2} \quad [16]$$

onde:

- PEV_{gi} é a variância do erro de predição do i -ésimo genótipo, obtida pelo quadrado do erro padrão de predição (SEP) para cada progênie em cada safra;
- $\hat{\sigma}_g^2$ é variância genotípica em cada geração.

A correlação de Spearman ($r_{MM'}$) foi calculada para avaliar a concordância entre as classificações das 30 melhores progênies com base nos valores genotípicos preditos (BLUPs) em cada geração para os cinco modelos testados. Além disso, também foi calculada a correlação na classificação das 30 melhores progênies a partir da média dos BLUPs considerando todas as gerações para os cinco modelos. Para o modelo selecionado (FA1), esse procedimento foi repetido, avaliando a concordância na classificação de todas as progênies e das 30 melhores, considerando todas as gerações.

As análises estatísticas foram conduzidas no ambiente de programação R (versão 4.2.3), utilizando o pacote ASReml-R (BUTLER et al., 2023) na versão 4.1.0.176. Os gráficos apresentados no presente trabalho foram construídos com os pacotes ggplot2 (versão 3.4.2) (WICKHAM, 2016) e ComplexHeatmap (versão 2.14.0) (GU; EILS; SCHLESNER, 2016).

3 RESULTADOS

3.1 Seleção do modelo

De acordo com os critérios apresentados na Tabela 3, o modelo que obteve melhor ajuste, apresentando os menores valores de AIC e BIC, foi o modelo com estrutura FA1 para a matriz G e diagonal heterogênea para a matriz R (Modelo 5). A acurácia média variou entre 0,55 e 0,91 nos modelos testados (Tabela 3), representando um incremento de aproximadamente 25% entre o modelo selecionado e o modelo base (Modelo 1).

Tabela 3: Número de parâmetros (NP), logaritmo do ponto de máximo da função de máxima verossimilhança (LogL), valores dos critérios de informação Akaike (AIC) e bayesiano (BIC), acurácia média ($\bar{r}$), herdabilidade média ($\bar{H}^2$), ganhos genéticos esperados (GS%).

Modelo	Matriz de (co)variância*		NP	log(L)	AIC	BIC	$\bar{r}$	$\bar{H}^2$	GS%	% $\bar{V}$
	G	R								
1	CS	CV	5	-636.90	1283.80	1312.65	0.55	0.30	6.48	-
2	CS	DIAG	8	-634.69	1285.39	1331.55	0.91	0.30	6.71	-
3	DIAG	DIAG	11	-630.17	1282.35	1345.82	0.91	0.30	6.76	-
4	UN	DIAG	17	-575.04	1184.09	1282.17	0.78	0.64	10.82	-
5	FA1	DIAG	15	-575.21	1180.43	1266.98	0.69	0.50	10.84	66.09

*CV: Componente de variância; DIAG: Diagonal heterogênea; UN: Não estruturada com correlação geral; FA1: Fator analítico de 1ª ordem.

A herdabilidade média entre as gerações para os modelos testados variou de 0,30 a 0,64, representando um aumento de mais de 100%. Enquanto o ganho de seleção médio apresentou um incremento de aproximadamente 67% entre o modelo selecionado e o modelo base (Tabela 3), os Modelos 4 e 5 tiveram os maiores ganhos de seleção, com valores de 10,82% e 10,84%, respectivamente. Por outro lado, os Modelos 1, 2 e 3 tiveram ganhos de seleção mais baixos, variando de 6,48% a 6,76%. No Modelo 5, cuja estrutura de (co)variância utilizada para G foi a FA1, a porcentagem de variância acumulada foi de 66,09%.

A Figura 1 apresenta os valores de correlação entre os cinco modelos testados na classificação das progênies avaliadas nas quatro safras, com base no valor genotípico predito das progênies em cada geração e ao longo das gerações. Os resultados mostram que a correlação entre os rankings dos modelos é alta, variando de 0,89 a 1,00 na classificação individual das progênies em cada geração (Figura 1A) e de 0,97 a 1,00 na classificação média das progênies ao longo das gerações (Figura 1B).

Tanto na classificação das progênies em cada safra quanto ao longo das safras, nota-se que a mudança no posto das progênies, com base nos BLUPs, altera-se apenas quando se comparam os Modelos 1, 2 e 3 com os Modelos 4 e 5, que possuem estruturas de (co)variância mais complexas, exceto na comparação entre os Modelos 5 e 4, cuja classificação foi idêntica.

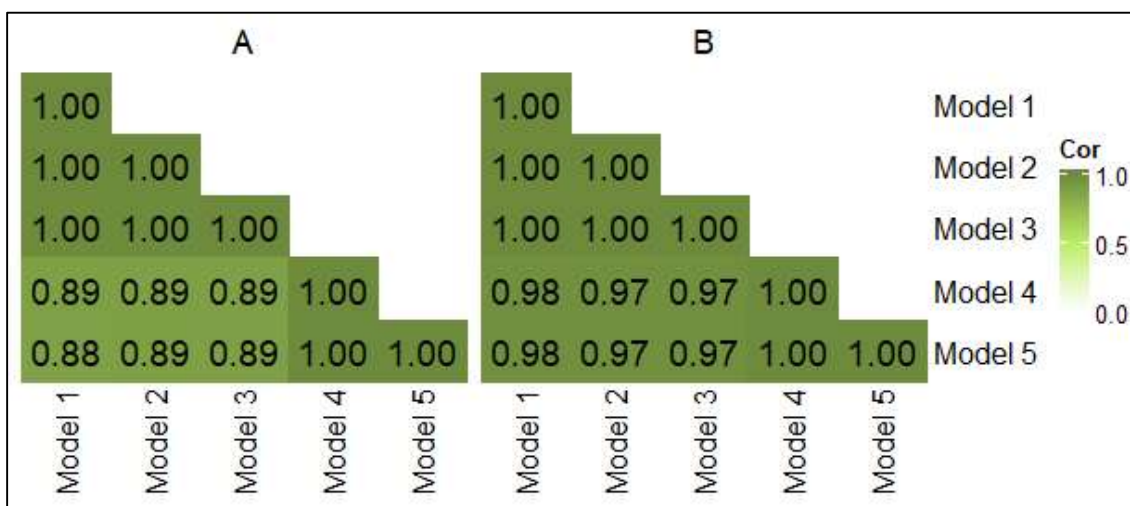


Figura 1. Concordância (correlação de Spearman) entre a classificação das progênies com base nos valores dos BLUPs para cada modelo em cada geração (A) e nas médias dos BLUPs das progênies considerando todas as gerações (B).

3.2 Parâmetros genéticos e não genéticos

Considerando o modelo de melhor ajuste (Modelo 5), a variância fenotípica apresentou valores entre 700.682,05 ($F_{4:9}$) e 914.200,97 ($F_{4:7}$) entre as gerações, enquanto a variância genotípica, representando a contribuição genética para a variabilidade total, variou de 273.605,0 ($F_{4:8}$) a 516.042,5 ($F_{4:7}$) (Tabela 4). Em relação às herdabilidades, houve variação de 0,37 ($F_{4:9}$) a 0,70 ($F_{4:6}$), com um declínio significativo ao longo das gerações, de aproximadamente 47%. A acurácia seletiva variou de 0,59 ($F_{4:8}$) a 0,84 ($F_{4:6}$) (Tabela 4).

Tabela 4: Componentes de variância e parâmetros genéticos estimados com base no Modelo 5 (FA1) para o conjunto de dados.

Geração	Componentes de variância*				Parâmetros genéticos**		CV_{e_i} (%)
	$\hat{\sigma}_p^2$	$\hat{\sigma}_g^2$	$\hat{\sigma}_b^2$	$\hat{\sigma}_e^2$	H^2	r	
F4:6	737649.33	317444.0	-	420205.33	0.70	0.84	16.21
F4:7	914200.9	516042.5	44012.28	354146.12	0.52	0.72	17.98
F4:8	788123.81	273605.0	146313.48	368205.33	0.38	0.59	17.24
F4:9	700682.05	360003.1	15216.91	325462.04	0.37	0.60	20.30

* $\hat{\sigma}_p^2$: Variância fenotípica; $\hat{\sigma}_g^2$: variância genotípica; $\hat{\sigma}_b^2$: variância do bloco; $\hat{\sigma}_e^2$: variância residual.

** H^2 : herdabilidade; r : acurácia. CV_{e_i} (%): coeficiente de variação ambiental.

Os valores do coeficiente de variação ambiental (CV_{e_i}) variaram de 16,21% (F4:6) a 20,30% (F4:9) (Tabela 4). As correlações genéticas entre as gerações foram moderadas, variando de 0,49 a 0,79.

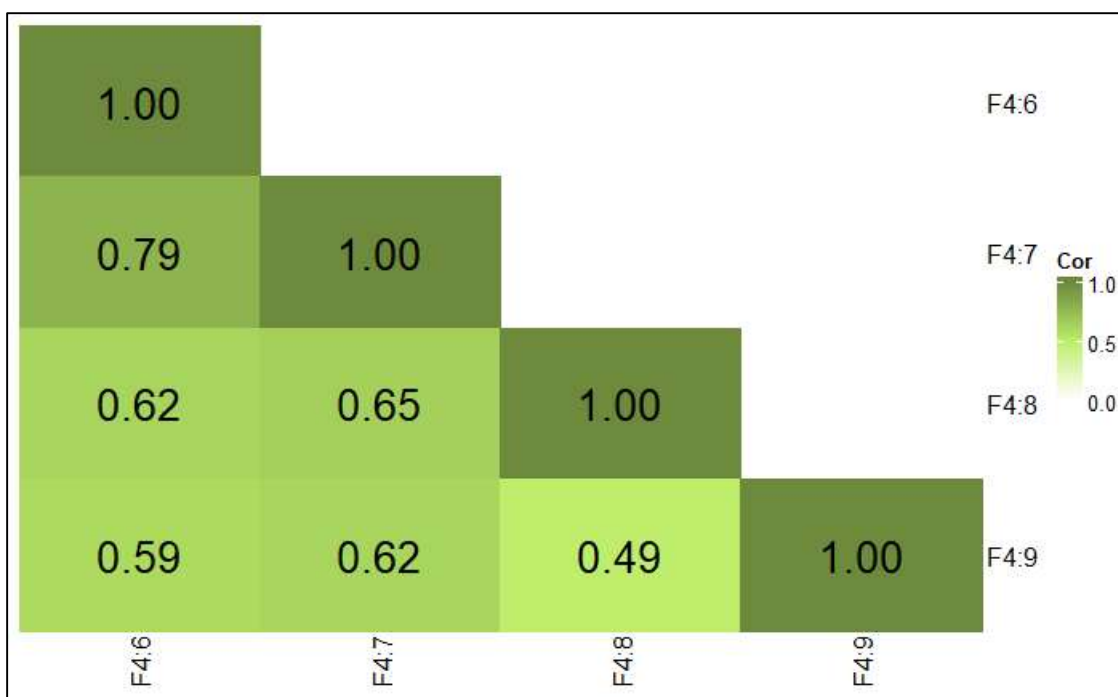


Figura 2. Correlações genéticas entre pares gerações.

A correlação entre a classificação considerando todas as progênies avaliadas e as 30 progênies de melhor desempenho, com base nos valores dos BLUPs em cada uma das gerações utilizando o Modelo 5, está ilustrada na Figura 3A e 3B, respectivamente. É fundamental destacar que a classificação das progênies de melhor desempenho em uma geração não necessariamente se alinha com a classificação em outra, embora as correlações entre os ranks estejam próximas de 1.

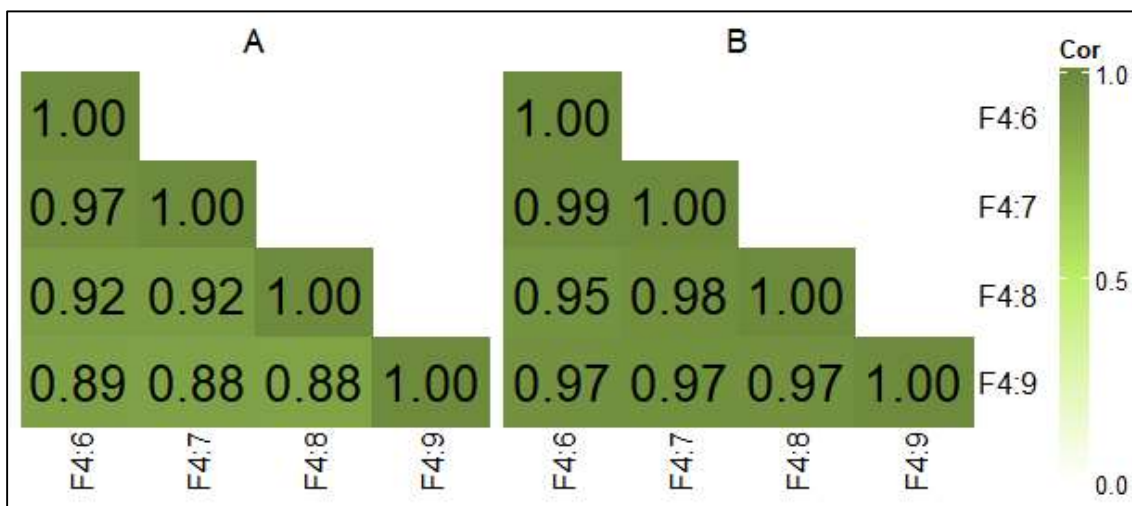


Figura 3. Concordância (correlação de *Spearman*) na classificação de todos os genótipos (A) e dos 30 melhores (B) com base nos BLUPs em cada geração, utilizando o modelo selecionado (Modelo 5).

Em geral, a confiabilidade das melhores predições lineares não viesadas (BLUPs) dos valores genotípicos individuais das 30 melhores progênies em cada safra apresentou valores bem próximos (Tabela 5). Na geração $F_{4:6}$, a variação foi de 4,30%; na $F_{4:7}$, de 0,0095%; na $F_{4:8}$, de 0,0063%; e na $F_{4:9}$, de 0,0024%. É possível observar que 80% das progênies estão presentes entre as 30 melhores em todas as gerações. A ausência das progênies 256, 679, 1, 510, 772 e 680 em todas as gerações deve-se à mudança no posto em função do valor da confiabilidade.

As progênies 256 e 680 foram avaliadas apenas na geração $F_{4:6}$. No entanto, a progênie 680 apresentou valor de 0,8397637, ocupando a 31ª colocação na geração $F_{4:7}$, mas na geração $F_{4:9}$, o valor de confiabilidade decaiu drasticamente. Já a progênie 256 apresentou um declínio significativo a partir da geração $F_{4:7}$, de aproximadamente 66,2% no total.

Pode-se constatar que cada geração tem seu próprio padrão de valores de confiabilidade. A geração $F_{4:7}$ apresenta progênies com confiabilidades consistentemente altas, enquanto na $F_{4:6}$, há maior variação nas confiabilidades das progênies. Além disso, considerando todas as progênies avaliadas, a variação entre os valores de confiabilidade é expressivamente maior do que entre as 30 melhores

Tabela 5: Confiabilidade (r^2) acurácia (r) e erro padrão de predição (SEP) das melhores predições lineares não viesadas (BLUPs) para valores genotípicos individuais das 30 melhores progênie em cada geração, com base no modelo selecionado (FA1).

Classificação	Progênie F4:6 (Seca/2019)	r^2	r	SEP	Progênie F4:7 (Águas/2019)	r^2	r	SEP	Progênie F4:8 (Inverno/2020)	r^2	r	SEP	Progênie F4:9 (Águas/2020)	r^2	r	SEP
1	668	0.7731	0.8793	268.3705	668	0.8398	0.9164	287.4833	668	0.7107	0.8430	281.3413	30	0.7885	0.8880	275.9229
2	738	0.7731	0.8793	268.3813	30	0.8398	0.9164	287.4973	30	0.7107	0.8430	281.3449	224	0.7885	0.8880	275.9232
3	191	0.7731	0.8793	268.3824	224	0.8398	0.9164	287.5006	740	0.7107	0.8430	281.3474	191	0.7885	0.8880	275.9238
4	224	0.7731	0.8793	268.3860	191	0.8398	0.9164	287.5040	224	0.7107	0.8430	281.3476	738	0.7885	0.8880	275.9239
5	733	0.7731	0.8793	268.3869	740	0.8398	0.9164	287.5049	733	0.7107	0.8430	281.3480	740	0.7885	0.8880	275.9241
6	740	0.7731	0.8793	268.3873	738	0.8398	0.9164	287.5050	191	0.7107	0.8430	281.3483	53	0.7885	0.8880	275.9243
7	30	0.7731	0.8792	268.3935	733	0.8398	0.9164	287.5057	738	0.7107	0.8430	281.3487	364	0.7885	0.8880	275.9254
8	439	0.7731	0.8792	268.3982	53	0.8398	0.9164	287.5078	53	0.7107	0.8430	281.3505	48	0.7885	0.8880	275.9261
9	419	0.7731	0.8792	268.3992	364	0.8398	0.9164	287.5149	48	0.7107	0.8430	281.3521	187	0.7885	0.8880	275.9261
10	214	0.7731	0.8792	268.4038	48	0.8398	0.9164	287.5174	454	0.7107	0.8430	281.3524	454	0.7885	0.8880	275.9262
11	364	0.7731	0.8792	268.4048	187	0.8398	0.9164	287.5180	187	0.7107	0.8430	281.3526	439	0.7885	0.8880	275.9265
12	454	0.7731	0.8792	268.4059	454	0.8398	0.9164	287.5184	679	0.7107	0.8430	281.3528	419	0.7885	0.8880	275.9265
13	679	0.7731	0.8792	268.4059	679	0.8398	0.9164	287.5199	364	0.7107	0.8430	281.3531	42	0.7885	0.8880	275.9268
14	187	0.7731	0.8792	268.4069	439	0.8398	0.9164	287.5217	214	0.7107	0.8430	281.3544	214	0.7885	0.8880	275.9268
15	48	0.7731	0.8792	268.4072	419	0.8398	0.9164	287.5223	439	0.7107	0.8430	281.3551	668	0.7885	0.8880	275.9270
16	53	0.7730	0.8792	268.4101	214	0.8398	0.9164	287.5226	772	0.7107	0.8430	281.3552	711	0.7885	0.8880	275.9278
17	287	0.7730	0.8792	268.4179	42	0.8398	0.9164	287.5241	419	0.7107	0.8430	281.3558	287	0.7885	0.8880	275.9282
18	10	0.7730	0.8792	268.4224	711	0.8398	0.9164	287.5290	42	0.7107	0.8430	281.3567	10	0.7885	0.8880	275.9288
19	1	0.7730	0.8792	268.4241	287	0.8398	0.9164	287.5324	711	0.7107	0.8430	281.3568	158	0.7885	0.8880	275.9290
20	42	0.7730	0.8792	268.4263	772	0.8398	0.9164	287.5338	503	0.7107	0.8430	281.3582	101	0.7885	0.8880	275.9298
21	711	0.7730	0.8792	268.4269	1	0.8398	0.9164	287.5344	10	0.7107	0.8430	281.3584	510	0.7885	0.8880	275.9298
22	688	0.7730	0.8792	268.4375	10	0.8398	0.9164	287.5350	1	0.7107	0.8430	281.3586	80	0.7885	0.8880	275.9302
23	101	0.7730	0.8792	268.4399	158	0.8398	0.9164	287.5387	287	0.7107	0.8430	281.3589	503	0.7885	0.8880	275.9305

24	510	0.7730	0.8792	268.4401	503	0.8398	0.9164	287.5428	688	0.7107	0.8430	281.3600	733	0.7885	0.8880	275.9307
25	503	0.7730	0.8792	268.4415	101	0.8398	0.9164	287.5430	158	0.7107	0.8430	281.3622	688	0.7885	0.8880	275.9313
26	158	0.7730	0.8792	268.4440	510	0.8398	0.9164	287.5430	551	0.7107	0.8430	281.3623	551	0.7885	0.8880	275.9319
27	80	0.7730	0.8792	268.4447	80	0.8398	0.9164	287.5444	80	0.7107	0.8430	281.3624	571	0.7885	0.8880	275.9322
28	772	0.7730	0.8792	268.4475	688	0.8398	0.9164	287.5481	571	0.7107	0.8430	281.3626	220	0.7885	0.8880	275.9325
29	680	0.7730	0.8792	268.4623	551	0.8398	0.9164	287.5522	680	0.7107	0.8430	281.3628	498	0.7885	0.8880	275.9348
30	220	0.7730	0.8792	268.4627	571	0.8398	0.9164	287.5539	101	0.7107	0.8430	281.3633	350	0.7885	0.8880	275.9351

4 DISCUSSÃO

4.1 Seleção do modelo

A utilização de modelos mistos apresenta-se como uma alternativa viável em cenários marcados por desbalanceamentos de dados, tornando a análise mais gerenciável ao permitir a consideração de grupos de efeitos como fatores aleatórios, em vez de fatores fixos (PIEPHO et al., 2008; SMITH; CULLIS; THOMPSON, 2001). Um argumento antigo contra o emprego de modelos mistos no melhoramento de plantas era a ausência de software adequado ou poder computacional suficiente. No entanto, no cenário atual, recursos computacionais mais robustos estão disponíveis, permitindo a utilização de metodologias estatísticas mais eficientes, juntamente com ferramentas de software apropriadas.

Vários critérios de informação atendem a propósitos distintos, cada um otimizado para objetivos específicos. Por exemplo, o AIC é conhecido por sua eficiência, enquanto o BIC é reconhecido por sua consistência (CHRISTENSEN, 2018). Ambos permitem comparar múltiplos modelos, sejam aninhados ou não, e avaliar efetivamente a incerteza associada à seleção do modelo (BURNHAM; ANDERSON, 1998). No presente estudo, o AIC e o BIC indicaram que as estruturas de covariância DIAG e FA1 são adequadas para modelar os efeitos residuais e genéticos, respectivamente.

O Modelo 1, considerado padrão ao assumir homogeneidade de (co)variância para R , apresentou valores de AIC e BIC menores que o Modelo 2, cuja estrutura de (co)variância para R já assume heterogeneidade entre as diferentes gerações. No entanto, admitir que os erros são não correlacionados e que as variâncias são homogêneas entre as gerações é um pressuposto irreal dentro do contexto dos dados analisados e, de modo geral, no campo do melhoramento de plantas. Tal simplificação pode enviesar as previsões dos valores genotípicos e as estimativas dos demais parâmetros de interesse (MEYER, 2009).

Além disso, espera-se que qualquer estrutura de (co)variância proposta para a matriz R , em contraste com a abordagem convencional $I\sigma^2$, melhore significativamente o ajuste do modelo aos dados. Isso ocorre porque se alinha à noção lógica de contabilização de correlações entre gerações (ACHARYA et al., 2020; PEIXOTO et al., 2021; SMITH et al., 2007).

A modelagem de (co)variância genética utilizando a estrutura Fator Analítico (FA), que é uma aproximação da estrutura Não Estruturada (UN), ganhou força e

provou ser eficaz, principalmente em análises multiambientes (EM) e multicaracterísticas (ARAÚJO et al., 2023; EVANGELISTA et al., 2023; HERNÁNDEZ et al., 2019).

O modelo não estruturado assume uma estrutura ampla, onde G é totalmente não estruturado e envolve $J(J + 1)/2$ parâmetros para variações genéticas distintas em cada geração, bem como covariâncias genéticas separadas entre pares de gerações. No entanto, ao tratar de um número substancial de gerações avaliadas, estimar essa matriz torna-se ineficiente ou inviável, mesmo quando o número de gerações é moderado. Portanto, uma estrutura mais parcimoniosa, como a FA, é preferível (SMITH; CULLIS; THOMPSON, 2001).

Em relação à herdabilidade, observa-se que não houve variação entre os Modelos 1, 2 e 3, cujas estruturas de (co)variância eram mais simples. No entanto, houve um aumento de mais de 100% no Modelo 4. Segundo Schmidt *et al.* (2019), a utilização de modelos inadequados pode levar à superestimação ou subestimação da herdabilidade.

Quanto aos ganhos de seleção esperados, não houve uma variação percentual considerável entre os modelos com estruturas de (co)variância mais simples. No entanto, entre o Modelo 1, considerado o mais simples, e o modelo selecionado (Modelo 5), houve um incremento de aproximadamente 68%. Ramalho *et al.* (1993) observaram que o fator específico da safra exerce uma influência substancial na interação genótipo $\times$ ambiente ($G \times A$), concluindo que a realização de avaliações em várias safras é essencial para garantir recomendações mais robustas e confiáveis em programas de melhoramento de feijão.

De fato, a expressão de um genótipo pode ser influenciada por condições ambientais variáveis, conforme observado por Meyer (2009). Esse fenômeno pode levar a diferenças na classificação das progênies entre as diferentes safras, como observado na Figura 3. Consequentemente, quando as avaliações são conduzidas exclusivamente em um único local ou quando se assume independência e homogeneidade de variância, a interação $G \times A$ pode inflacionar a estimativa da variância genética e levar a estimativas tendenciosas dos ganhos genéticos de seleção, resultando em ganhos reais inferiores aos previstos.

A seleção e o descarte de genótipos realizados a partir da geração $F_{4:6}$ influenciam diretamente a classificação das progênies nas gerações subsequentes, uma vez que nem todas as progênies avaliadas inicialmente foram reavaliadas nos

ciclos seguintes. Esse processo pode introduzir vieses de seleção, já que a perda de diversidade genética ou a eliminação de genótipos promissores em ambientes específicos não é trivialmente mensurada.

Nesse contexto, para selecionar progênies de alto desempenho em ambientes heterogêneos – especialmente para características de herança complexa –, a adoção de modelos estatísticos que incorporam estruturas de (co)variância para efeitos genéticos, ambientais e interações G×A torna-se indispensável. Esses modelos permitem estimar valores genéticos com maior precisão, ajustando a variabilidade ambiental e a heterogeneidade dos dados (OGUT et al., 2014; SMITH et al., 2015; SMITH; CULLIS; THOMPSON, 2005).

4.2 Parâmetros genéticos e não genéticos

Os resultados das herdabilidades demonstraram variações notáveis ao longo das gerações. Na safra da seca, foi observada a maior estimativa de herdabilidade. Melo *et al.* (2020) relataram comportamento semelhante ao avaliar progênies de feijão vermelho. Por outro lado, nas safras subsequentes, as estimativas de herdabilidade apresentaram tendência de redução, o que pode ser explicado pela baixa variabilidade genética observada.

Essa diminuição nas estimativas de herdabilidade ao longo das safras deve-se a uma série de fatores. Um deles é a seleção realizada nas safras anteriores, que reduz a variabilidade genética disponível para seleção. Além disso, condições ambientais específicas em cada safra influenciam a expressão genética da característica. De modo geral, é comum encontrar valores de baixa magnitude para herdabilidade em feijoeiro para caracteres quantitativos (MELO et al., 2020; SALVADOR et al., 2022, 2021; SANTANA et al., 2023).

De acordo com Schmidt *et al.* (2019), como a H_{cullis} depende da precisão das comparações pareadas envolvendo os BLUPs genotípicos, suas estimativas tendem a diminuir ao passar de 100% de sobreposição genotípica entre anos/safras/gerações para 0%. Esse declínio ocorre porque um número substancial de comparações indiretas torna-se necessário ao avaliar genótipos que não foram cultivados nos mesmos anos/safras/gerações. De qualquer forma, o uso da herdabilidade padrão, que utiliza a relação entre variâncias em dados desbalanceados, tende a superestimar H^2 , enquanto métodos alternativos proporcionam uma melhor aproximação (SCHMIDT et al., 2019).

Bernardo (2020) comenta que, uma vez que a herdabilidade quantifica a proporção de efeitos genéticos que podem ser transmitidos aos descendentes, os melhoristas de plantas estão normalmente mais interessados em avaliar o desempenho dos próprios candidatos do que o desempenho de sua descendência, como ocorre no melhoramento animal. O autor argumenta ainda que a confiabilidade pode ser uma alternativa à herdabilidade clássica, uma vez que está estritamente ligada à unidade de seleção utilizada. No entanto, os melhoristas de plantas manterão seu interesse em compreender a interação entre a genética e as influências ambientais nas características quantitativas (BERNARDO, 2020b).

Assim como os resultados das herdabilidades ao longo das gerações, a acurácia apresentou um decréscimo significativo entre a geração $F_{4:6}$ e $F_{4:9}$, interferindo diretamente na relação entre esses dois parâmetros (Tabela 4), uma vez que a raiz quadrada da herdabilidade pode ser interpretada como a correlação entre o genótipo e o fenótipo (Ramalho et al., 2012; Resende e Alves, 2022).

Em relação aos valores de herdabilidade e acurácia, os distintos delineamentos utilizados nas diferentes safras também podem explicar o declínio observado. O uso de um delineamento em blocos completos casualizados (DBC) na primeira safra, em contraste com o látice nas três safras subsequentes, pode ter proporcionado uma avaliação inicial mais precisa, ao minimizar o impacto de variações ambientais, conforme destacado por Masood *et al.* (2008). Além disso, os valores de CVres reforçam esse argumento, embora os valores se encontrem na faixa de precisão média em todas as safras, o CVres foi menor na primeira safra (OLIVEIRA et al., 2009).

No entanto, o melhoramento do feijoeiro envolve um processo sequencial de avaliação e seleção até a obtenção de linhagens endogâmicas em gerações avançadas, como F_6 ou posteriores, em que o uso do delineamento em blocos completos casualizados se torna inviável, seja pelo número insuficiente de sementes nos estágios iniciais ou pela grande quantidade de genótipos a serem testados, entre outros fatores. Dessa forma, diferentes delineamentos são utilizados ao longo do melhoramento do feijoeiro para otimizar as etapas.

A Figura 2 indica a importância do efeito específico da geração no comportamento genotípico das progênes avaliadas. A correlação genética moderada observada entre as gerações deve-se principalmente à interação entre o genótipo e as condições ambientais de cada safra em que as progênes foram

testadas. Em termos gerais, as distintas condições climáticas entre as safras levam a padrões variados de expressão gênica entre as progênies, conforme observado por Malosetti *et al.* em 2013. As diferenças nas características ambientais provenientes das diferentes safras e anos em que as progênies foram avaliadas tendem a influenciar diretamente as características. No Brasil, o cultivo do feijão é realizado anualmente em três safras: Águas, Seca e Inverno (CARNEIRO; DE PAULA JÚNIOR; BORÉM, 2015).

As disparidades nas condições ambientais entre a safra de inverno, que depende de irrigação, e a safra das águas, cultivada em condições secas, levam à presença de estresses bióticos e abióticos que prejudicam a produtividade do feijoeiro. Essa heterogeneidade de variância presente no modelo possibilita inferir sobre tais fatores ambientais, que se diluem na interação $G \times A$ (MELO *et al.*, 2020). Por exemplo, as safras colhidas durante a estação chuvosa geralmente produzem grãos de qualidade comercial e industrial inferior, pois o momento da colheita frequentemente coincide com o período chuvoso (SANTANA *et al.*, 2023).

A confiabilidade de um valor genotípico predito depende de sua variância do erro de predição (PEV) em relação à variância genética. Em essência, pode ser vista como uma métrica que encapsula o valor informativo derivado do cálculo do BLUP (MRODE, 2014, p. 44). Além disso, a confiabilidade depende da quantidade de informações empregadas nas análises. Algumas progênies são avaliadas em todas as safras, enquanto outras são descartadas ao longo das avaliações. Essa afirmação pode ser ratificada pelo fato de as progênies 256 e 680, avaliadas apenas na geração $F_{4:6}$, terem seus valores genotípicos preditos drasticamente reduzidos nas demais gerações.

No contexto de experimentos sequenciais em programas de melhoramento de feijoeiro comum, considerando o desbalanceamento genético e/ou estatístico, a metodologia de modelos mistos aliada à modelagem de matrizes de (co)variância dos efeitos genéticos e não genéticos aumenta a eficiência da seleção, uma vez que é possível representar mais fielmente a natureza dos experimentos e a especificidade dos dados. Além disso, a inclusão dos dados de genótipos descartados nas análises não pode ser ignorada, pois tem o potencial de afetar diretamente a tomada de decisões (Piepho e Möhring, 2006; Viana *et al.*, 2012). Embora as pesquisas que adotam essas abordagens sejam escassas, elas revelam-

se promissoras na consecução de seus objetivos, apresentando resultados satisfatórios (MELO et al., 2020; SALVADOR et al., 2022; SANTANA et al., 2023).

5 CONCLUSÕES

O presente estudo demonstrou que a modelagem adequada de estruturas de (co)variância em modelos mistos é fundamental para otimizar a seleção de progênies em programas de melhoramento de feijoeiro comum, especialmente em cenários de desbalanceamento genético e estatístico. A adoção do Modelo Fator Analítico de 1ª Ordem (FA1) destacou-se como a abordagem mais eficiente, proporcionando ganhos genéticos 68% superiores aos modelos convencionais, além de maior precisão nas predições de valores genotípicos. Esses resultados reforçam a importância de estruturas de covariância flexíveis para capturar a heterogeneidade ambiental e genética em experimentos sequenciais

A inclusão de dados históricos (genótipos selecionados e descartados) revelou-se crítica para mitigar viés de seleção, uma vez que a classificação das progênies se alterou significativamente entre gerações, principalmente devido à interação genótipo $\times$ ambiente ($G \times A$) e à redução progressiva da variabilidade genética. Esse achado ressalta a necessidade de abordagens estatísticas integrativas, que considerem não apenas os dados atuais, mas também o histórico de seleção, para garantir decisões mais robustas em programas de melhoramento de plantas.

Os resultados sugerem que a tomada de decisão dos melhoristas em programas de melhoramento de plantas que avaliam numerosas progênies em diferentes locais, anos e safras deve incluir, na análise estatística dos dados, informações prévias das progênies por meio de abordagens que considerem o desbalanceamento de dados. Essa alternativa não acarreta custos adicionais para o programa de melhoramento de plantas e pode melhorar significativamente a precisão das seleções.

REFERÊNCIAS

- ACHARYA, J. P. et al. Breeding Alfalfa (*Medicago sativa* L.) Adapted to Subtropical Agroecosystems. **Agronomy**, v. 10, n. 5, p. 742, 2020.
- AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, v. 19, n. 6, p. 716–723, 1974.
- ALVES, A. F. et al. Genetic progress and potential of common bean families obtained by recurrent selection. **Crop Breeding and Applied Biotechnology**, v. 15, n. 4, p. 218–226, 2015.
- ARAÚJO, M. DOS S. et al. Modeling covariance structures for genetic and non-genetic effects in cowpea multi-environment trials. **Agronomy Journal**, v. 115, n. 3, p. 1248–1256, 2023.
- ARAÚJO, M. S. et al. GIS-FA: an approach to integrating thematic maps, factor-analytic, and envirotyping for cultivar targeting. **Theoretical and Applied Genetics**, v. 137, n. 4, p. 80, 2024.
- ARIEF, V. N. et al. Utilization of Multiyear Plant Breeding Data to Better Predict Genotype Performance. **Crop Science**, v. 59, n. 2, p. 480–490, 2019.
- BAKARE, M. A. et al. Parsimonious genotype by environment interaction covariance models for cassava (*Manihot esculenta*). **Frontiers in Plant Science**, v. 13, 2022.
- BANČIČ, J. et al. Genomic selection for genotype performance and stability using information on multiple traits and multiple environments. **Theoretical and Applied Genetics**, v. 136, n. 5, p. 104, 2023.
- BARILI, L. D. **Evolução dos cultivares de feijão carioca recomendados no Brasil**. Tese (Doutorado em Genética e Melhoramento)—Viçosa: Universidade Federal de Viçosa, 2015.
- BERNARDO, R. Recurrent Selection. Em: BERNARDO, R. (Ed.). **Breeding for Quantitative Traits in Plants**. 3. ed. Woodbury: Stemma Press, 2020a. p. 299–318.
- BERNARDO, R. Reinventing quantitative genetics for plant breeding: something old, something new, something borrowed, something BLUE. **Heredity**, v. 125, n. 6, p. 375–385, 2020b.
- BHERING, L. L. et al. Genetic gains in physic nut using selection indexes. **Pesquisa Agropecuária Brasileira**, v. 47, n. 3, p. 402–408, 2012.
- BRITO, G. M. DE; CARNEIRO, P. C. S.; SILVA, M. B. P. DA. **Avaliação e seleção de progênies visando o melhoramento do feijão carioca por seleção recorrente**. Simpósio de Integração Acadêmica - SIA 2021. **Anais...**Viçosa: Editora UFV, 2021.

BURNHAM, K. P.; ANDERSON, D. R. Practical Use of the Information-Theoretic Approach. Em: **Model Selection and Inference**. New York, NY: Springer New York, 1998. p. 75–117.

BUTLER, D. G. et al. **ASReml-R Reference Manual Version 4.2**. Hemel Hempstead VSN International Ltd., , 2023.

CARNEIRO, J. E.; DE PAULA JÚNIOR, T.; BORÉM, A. **Feijão: do plantio à colheita**. 1. ed. Viçosa: Editora UFV, 2015.

CASLER, M. D. Fundamentals of Experimental Design: Guidelines for Designing Successful Experiments. **Agronomy Journal**, v. 107, n. 2, p. 692–705, 2015.

CERÓN-ROJAS, J. J. et al. A Selection Index Method Based on Eigenanalysis. **Crop Science**, v. 46, n. 4, p. 1711–1721, 2006.

CHAVES, S. F. S. et al. Employing factor analytic tools for selecting high-performance and stable tropical maize hybrids. **Crop Science**, v. 63, n. 3, p. 1114–1125, 2023a.

CHAVES, S. F. S. et al. Analysis of repeated measures data through mixed models: An application in *Theobroma grandiflorum* breeding. **Crop Science**, v. 63, n. 4, p. 2131–2144, 2023b.

CHRISTENSEN, W. **Agreeing to disagree: Using SAS to make reasoned decisions when information criteria select different models**. SAS Conference Proceedings: Western Users of SAS Software 2018. **Anais...**Sacramento: Paper 099–2018, 2018.

COCHRAN, W. G. Some Consequences When the Assumptions for the Analysis of Variance are not Satisfied. **Biometrics**, v. 3, n. 1, p. 22, 1947.

COLLICCHIO, E.; RAMALHO, M. A. P.; ABREU, A. F. B. Association between plant architecture of the bean plant and seed size. (In Portuguese, with English abstract.). **Pesquisa Agropecuária Brasileira**, v. 32, n. 3, p. 297–304, 1997.

COWLING, W. A. et al. Optimal Contribution Selection Improves the Rate of Genetic Gain in Grain Yield and Yield Stability in Spring Canola in Australia and Canada. **Plants**, v. 12, n. 2, p. 383, 2023.

CROSSA, J. Statistical Analyses of Multilocation Trials. Em: BRADY, N. C. (Ed.). **Advances in Agronomy**. Cambridge: Academic Press, 1990. v. 44p. 55–85.

CROSSA, J. From Genotype × Environment Interaction to Gene × Environment Interaction. **Current Genomics**, v. 13, n. 3, p. 225–244, 2012.

CRUZ, C. D.; CARNEIRO, P. C. S.; REGAZZI, A. J. Seleção simultânea de caracteres. Em: CRUZ, C. D.; CARNEIRO, P. C. S.; REGAZZI, A. J. (Eds.). **Modelos biométricos aplicados ao melhoramento genético**. 3. ed. Viçosa: Editora UFV, 2014. v. 2p. 221–294.

CULLIS, B. R. et al. Analysis of yield and oil from a series of canola breeding trials. Part II. Exploring variety by environment interaction using factor analysis. **Genome**, v. 53, n. 11, p. 1002–1016, 2010.

CULLIS, B. R. et al. Factor analytic and reduced animal models for the investigation of additive genotype-by-environment interaction in outcrossing plant species with application to a *Pinus radiata* breeding programme. **Theoretical and Applied Genetics**, v. 127, n. 10, p. 2193–2210, 2014.

CULLIS, B. R.; SMITH, A. B.; COOMBES, N. E. On the design of early generation variety trials with correlated data. **Journal of Agricultural, Biological, and Environmental Statistics**, v. 11, n. 4, p. 381–393, 2006.

DIAS, K. O. G. et al. Leveraging probability concepts for cultivar recommendation in multi-environment trials. **Theoretical and Applied Genetics**, v. 135, n. 4, p. 1385–1399, 2022.

DOS ANJOS, R. S. R. et al. New Proposals to Estimate Unbiased Selection Gain and Coefficient of Variation in Traits Evaluated Using Score Scales. **Crop Science**, v. 59, n. 3, p. 937–944, 2019.

DWIVEDI, S. L. et al. Epistasis and pleiotropy-induced variation for plant breeding. **Plant Biotechnology Journal**, v. 22, n. 10, p. 2788–2807, 2024.

EBERHART, S. A.; RUSSELL, W. A. Stability Parameters for Comparing Varieties. **Crop Science**, v. 6, n. 1, p. 36–40, 1966.

EVANGELISTA, J. S. P. C. et al. Modeling covariance structures and optimizing *Jatropha curcas* breeding. **Tree Genetics & Genomes**, v. 19, n. 2, p. 21, 2023.

FEDERER, W. T. Augmented (or hoynuiaku) designs. **Biometrics Unit Technical Reports**, p. 1–33, 1956.

FELDMANN, M. J. et al. Average semivariance yields accurate estimates of the fraction of marker-associated genetic variance and heritability in complex trait analyses. **PLOS Genetics**, v. 17, n. 8, p. e1009762, 2021.

FINLAY, K. W.; WILKINSON, G. N. The analysis of adaptation in a plant-breeding programme. **Australian Journal of Agricultural Research**, v. 14, n. 6, p. 742–754, 1963.

FOUILLOUX, G.; BANNEROT, H. Selection methods in the common bean (*Phaseolus vulgaris*). Em: GEPTS, P. (Ed.). **Genetic Resources of Phaseolus Beans. Current Plant Science and Biotechnology in Agriculture**. 1. ed. Dordrecht: Springer, 1988. v. 6p. 503–542.

GABRIEL, K. R. The Biplot Graphic Display of Matrices with Application to Principal Component Analysis. **Biometrika**, v. 58, n. 3, p. 453–467, 1971.

GAUCH, H. G. Model Selection and Validation for Yield Trials with Interaction. **Biometrics**, v. 44, n. 3, p. 705–715, 1988.

GU, Z.; EILS, R.; SCHLESNER, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. **Bioinformatics**, v. 32, n. 18, p. 2847–2849, 2016.

HARTUNG, J.; PIEPHO, H. Effect of missing values in multi-environment trials on variance component estimates. **Crop Science**, v. 61, n. 6, p. 4087–4097, 2021.

HAZEL, L. N. The genetic basis for constructing selection indexes. **Genetics**, v. 28, n. 6, p. 476–490, 1943.

HENDERSON, C. R. Best Linear Unbiased Estimation and Prediction under a Selection Model. **Biometrics**, v. 31, n. 2, p. 423, jun. 1975.

HENDERSON, C. R.; KEMPTHORNE, O.; SEARLE, S. R.; VON KROSIGK, C. M. The estimation of environmental and genetic trends from records subject to culling. **Biometrics**, [S. l.], v. 15, n. 2, p. 192, 1959.

HERNÁNDEZ, M. V. et al. Modeling Genotype × Environment Interaction Using a Factor Analytic Model of On-Farm Wheat Trials in the Yaqui Valley of Mexico. **Agronomy Journal**, v. 111, n. 6, p. 2647–2657, 2019.

HILL, R. R.; ROSENBERGER, J. L. Methods for Combining Data from Gemrplasm Evaluation Trials ¹. **Crop Science**, v. 25, n. 3, p. 467–470, 1985.

HUNT, C. H. et al. Multi-environment analysis of sorghum breeding trials using additive and dominance genomic relationships. **Theoretical and Applied Genetics**, v. 133, n. 3, p. 1009–1018, 2020.

ISIK, F.; HOLLAND, J.; MALTECCA, C. Multi Environmental Trials. Em: ISIK, F.; HOLLAND, J.; MALTECCA, C. (Eds.). **Genetic Data Analysis for Plant and Animal Breeding**. Cham: Springer International Publishing, 2017. p. 227–262.

KELLY, A. M. et al. The Accuracy of Varietal Selection Using Factor Analytic Models for Multi-Environment Plant Breeding Trials. **Crop Science**, v. 47, n. 3, p. 1063–1070, 2007.

KEMPTON, R. A. The use of biplots in interpreting variety by environment interactions. **The Journal of Agricultural Science**, v. 103, n. 1, p. 123–135, 1984.

LITTLE, R. J. A.; RUBIN, D. B. Missing Data in Experiments. Em: BALDING, D. J. et al. (Eds.). **Statistical Analysis with Missing Data**. 2. ed. New York: John Wiley & Sons, 2002. p. 24–40.

MAICH, R. H. et al. Association between grain yield, grain quality and morpho-physiological traits along ten cycles of recurrent selection in bread wheat (*Triticum aestivum* L.). **Cereal Research Communications**, v. 45, n. 1, p. 146–153, 2017.

MALOSETTI, M.; RIBAUT, J.-M.; VAN EEUWIJK, F. A. The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. **Frontiers in Physiology**, v. 4, 2013.

MASOOD, M. A. et al. Improvement in Precision of Agricultural field Experiments through Design and analysis. **Pakistan Journal of Life and Social Sciences**, v. 6, n. 2, p. 89–91, 2008.

MELO, V. L. DE et al. Modeling (co)variance structures for genetic and non-genetic effects in the selection of common bean progenies. **Euphytica**, v. 216, n. 5, p. 77, 18 maio 2020.

MEYER, K. Factor-analytic models for genotype $\times$ environment type problems and structured covariance matrices. **Genetics Selection Evolution**, v. 41, n. 1, p. 21, 2009.

MÖHRING, J.; PIEPHO, H. -P. Comparison of Weighting in Two-Stage Analysis of Plant Breeding Trials. **Crop Science**, v. 49, n. 6, p. 1977–1988, 2009.

MORAIS JÚNIOR, O. P. et al. Effectiveness of Recurrent Selection in Irrigated Rice Breeding. **Crop Science**, v. 57, n. 6, p. 3043–3058, 2017.

MOURA, L. M. et al. Combining ability of common bean parents in different seasons, locations and generations. **Euphytica**, v. 214, n. 10, p. 181, 25 out. 2018.

MRODE, R. A. Univariate Models with One Random Effect. Em: **Linear Models for the Prediction of Animal Breeding Values**. 3. ed. Wallingford: Cabi, 2014a. p. 34–60.

MRODE, R. A. **Linear models for the prediction of animal breeding values**. 3. ed. Wallingford: CABI Publishing, 2014b.

OGUT, F. et al. Genetic Analysis of Diallel Progeny Test Data Using Factor Analytic Linear Mixed Models. **Forest Science**, v. 60, n. 1, p. 119–127, 2014.

OLIVEIRA, R. L. DE et al. Precisão experimental em ensaios com a cultura do feijão. **Ciência e Agrotecnologia**, v. 33, n. 1, p. 113–119, 2009.

OLIVEIRA, I. C. M. et al. Genotype-by-environment interaction and yield stability analysis of biomass sorghum hybrids using factor analytic models and environmental covariates. **Field Crops Research**, v. 257, p. 107929, 2020.

OLIVOTO, T.; NARDINO, M. MGIDI: toward an effective multivariate selection in biological experiments. **Bioinformatics**, v. 37, n. 10, p. 1383–1389, 2021.

PATTERSON, H. D.; THOMPSON, R. Recovery of Inter-Block Information when Block Sizes are Unequal. **Biometrika**, v. 58, n. 3, p. 545, 1971.

PEEL, M. C.; FINLAYSON, B. L.; MCMAHON, T. A. Updated world map of the Köppen-Geiger climate classification. **Hydrology and Earth System Sciences**, v. 11, n. 5, p. 1633–1644, 2007a.

PEEL, M. C.; FINLAYSON, B. L.; MCMAHON, T. A. Updated world map of the Köppen-Geiger climate classification. **Hydrology and Earth System Sciences**, v. 11, n. 5, p. 1633–1644, 11 out. 2007b.

PEIXOTO, M. A. et al. Selection of maize hybrids: an approach with multi-trait, multi-environment, and ideotype-design. **Crop Breeding and Applied Biotechnology**, v. 21, n. 2, 2021.

PIEPHO, H. et al. A stage-wise approach for the analysis of multi-environment trials. **Biometrical Journal**, v. 54, n. 6, p. 844–860, 2012.

PIEPHO, H. A coefficient of determination R^2 for generalized linear mixed models. **Biometrical Journal**, v. 61, n. 4, p. 860–872, 2019.

PIEPHO, H.; MÖHRING, J. Selection in Cultivar Trials—Is It Ignorable? **Crop Science**, v. 46, n. 1, p. 192–201, 2006.

PIEPHO, H. P. et al. BLUP for phenotypic selection in plant breeding and variety testing. **Euphytica**, v. 161, n. 1–2, p. 209–228, 2008.

PIEPHO, H.-P. Analyzing Genotype-Environment Data by Mixed Models with Multiplicative Terms. **Biometrics**, v. 53, n. 2, p. 761, 1997.

PIEPHO, H.-P. Empirical best linear unbiased prediction in cultivar trials using factor-analytic variance-covariance structures. **Theoretical and Applied Genetics**, v. 97, n. 1–2, p. 195–201, 1998a.

PIEPHO, H.-P. Empirical best linear unbiased prediction in cultivar trials using factor-analytic variance-covariance structures. **Theoretical and Applied Genetics**, v. 97, n. 1–2, p. 195–201, 1998b.

PIMENTEL-GOMES, F. **Curso de Estatística Experimental**. 15. ed. Piracicapa: FEALQ, 2009.

POLLAK, E. J.; VAN DER WERF, J.; QUAAS, R. L. Selection Bias and Multiple Trait Evaluation. **Journal of Dairy Science**, v. 67, n. 7, p. 1590–1595, 1984.

POSADAS, L. G. et al. Elite Performance for Grain Yield from Unadapted Exotic Soybean Germplasm in Three Cycles of a Recurrent Selection Experiment. **Crop Science**, v. 54, n. 6, p. 2536–2546, 2014.

PRIYADARSHAN, P. M. Genotype-by-Environment Interactions. Em: **PLANT BREEDING: Classical to Modern**. Singapore: Springer Singapore, 2019. p. 457–472.

R CORE TEAM. **R: A language and environment for statistical computing**. Vienna, Austria R Foundation for Statistical Computing, , 2023.

RAMALHO, M. A. P. et al. **Aplicações da Genética Quantitativa no Melhoramento de Plantas Autógamas**. 1. ed. Lavras: Editora UFLA, 2012.

RAMALHO, M. A. P.; ABREU, A. DE F. B. Obtenção de cultivares. Em: CARNEIRO, J. E.; PAULA JÚNIOR, T. J. DE; BORÉM, A. (Eds.). **Feijão: do plantio a colheita**. Viçosa: Editora UFV, 2015. p. 96–114.

RAMALHO, M. A. P.; ABREU, Â. DE F. B.; DOS SANTOS, J. B. Genetic progress after four cycles of recurrent selection for yield and grain traits in common bean. **Euphytica**, v. 144, n. 1–2, p. 23–29, 2005.

RAMALHO, M. A. P.; CARVALHO, B. L.; NUNES, J. A. R. Perspectives for the Use of Quantitative Genetics in Breeding of Autogamous Plants. **ISRN Genetics**, v. 2013, p. 1–6, 2013.

RAMALHO, M. A. P.; PIROLA, L. H.; ABREU, A. F. B. Alternativas na seleção de plantas de feijoeiro com porte ereto e grão tipo carioca. **Pesquisa Agropecuária Brasileira**, v. 33, n. 12, p. 1989–1994, 1998.

RESENDE, M. D. V. DE; ALVES, R. S. Statistical significance, selection accuracy, and experimental precision in plant breeding. **Crop Breeding and Applied Biotechnology**, v. 22, n. 3, 2022.

RESENDE, M. D. V. DE. **Genética biométrica e estatística no melhoramento de plantas perenes**. [s.l.] Emprapa Informação Tecnológica, 2002.

RESENDE, M. D. V.; DUARTE, J. B. Precisão e controle de qualidade em experimentos de avaliação de cultivares. **Pesquisa Agropecuária Tropical**, v. 37, n. 3, p. 182–194, 16 out. 2007.

ROCHA, J. R. DO A. S. DE C. et al. Selection of Superior Inbred Progenies toward the Common Bean Ideotype. **Agronomy Journal**, v. 111, n. 3, p. 1181–1189, 2019.

ROCHA, J. R. DO A. S. DE C.; MACHADO, J. C.; CARNEIRO, P. C. S. Multitrait index based on factor analysis and ideotype-design: proposal and application on elephant grass breeding for bioenergy. **GCB Bioenergy**, v. 10, n. 1, p. 52–60, 2018.

RODRÍGUEZ-ÁLVAREZ, M. X. et al. Correcting for spatial heterogeneity in plant breeding experiments with P-splines. **Spatial Statistics**, v. 23, p. 52–71, mar. 2018.

SAE-LIM, P. et al. Identifying environmental variables explaining genotype-by-environment interaction for body weight of rainbow trout (*Onchorynchus mykiss*): reaction norm and factor analytic models. **Genetics Selection Evolution**, v. 46, n. 1, p. 16, 2014.

SALVADOR, F. V. et al. Correcting experimental data for spatial trends in a common bean breeding program. **Crop Science**, v. 62, n. 2, p. 825–838, 14 mar. 2022.

SALVADOR, F. V. et al. Accounting for heterogeneous genotypic variance of progenies within populations in autogamous plant breeding: An evaluation in common bean. **Crop Science**, v. 61, n. 6, p. 3947–3961, 2021.

SANTANA, A. S. et al. Analysis of advanced generation multistage field trials data in autogamous plant breeding: An evaluation in common Bean. **Crop Science**, v. 63, n. 3, p. 1221–1233, 2023.

SCHMIDT, P. et al. Estimating Broad-Sense Heritability with Unbalanced Data from Agricultural Cultivar Trials. **Crop Science**, v. 59, n. 2, p. 525–536, 2019.

SCHWARZ, G. Estimating the Dimension of a Model. **The Annals of Statistics**, v. 6, n. 2, 1978.

SILVA, A. C. et al. Selection of common bean parents and segregating populations targeting fusarium wilt resistance and grain yield. **Agronomy Science and Biotechnology**, v. 9, p. 1–20, 2023.

SMITH, A. et al. Plant Variety Selection Using Interaction Classes Derived From Factor Analytic Linear Mixed Models: Models With Independent Variety Effects. **Frontiers in Plant Science**, v. 12, p. 1–17, 2021.

SMITH, A. B. et al. Varietal selection for perennial crops where data relate to multiple harvests from a series of field trials. **Euphytica**, v. 157, n. 1–2, p. 253–266, 2007.

SMITH, A. B. et al. Factor analytic mixed models for the provision of grower information from national crop variety testing programs. **Theoretical and Applied Genetics**, v. 128, n. 1, p. 55–72, 2015.

SMITH, A. B.; CULLIS, B. R. Plant breeding selection tools built on factor analytic mixed models for multi-environment trial data. **Euphytica**, v. 214, n. 8, p. 143, 2018.

SMITH, A. B.; CULLIS, B. R.; THOMPSON, R. The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. **The Journal of Agricultural Science**, v. 143, n. 6, p. 449–462, 2005.

SMITH, A.; CULLIS, B.; GILMOUR, A. Applications: The Analysis of Crop Variety Evaluation Data in Australia. **Australian & New Zealand Journal of Statistics**, v. 43, n. 2, p. 129–145, 2001.

SMITH, A.; CULLIS, B.; THOMPSON, R. Analyzing Variety by Environment Data Using Multiplicative Mixed Models and Adjustments for Spatial Field Trend. **Biometrics**, v. 57, n. 4, p. 1138–1147, dez. 2001.

SMITH, H. F. A discriminant function for plant selection. **Annals of Eugenics**, v. 7, n. 3, p. 240–250, 1936.

SOKAL, R. R.; ROHLF, F. J. **The principles and practice of statistics in biological research**. 3. ed. New York: W. H. Freeman, 1995.

STEPHENS, M. J. et al. Genetic Parameters and Development of a Selection Index for Breeding Red Raspberries for Processing. **Journal of the American Society for Horticultural Science**, v. 137, n. 4, p. 236–242, 2012.

STUDNICKI, M. et al. Yield response of winter wheat cultivars to environments modeled by different variance-covariance structures in linear mixed models. **Spanish Journal of Agricultural Research**, v. 14, n. 2, p. e0703, 2016.

TOLHURST, D. J. et al. Genomic selection in multi-environment plant breeding trials using a factor analytic linear mixed model. **Journal of Animal Breeding and Genetics**, v. 136, n. 4, p. 279–300, 2019.

VERBYLA, A. P. A note on model selection using information criteria for general linear models estimated using REML. **Australian & New Zealand Journal of Statistics**, v. 61, n. 1, p. 39–50, 9 mar. 2019.

VIANA, J. M. S. et al. Relevance of Pedigree, Historical Data, Dominance, and Data Unbalance for Selection Efficiency. **Agronomy Journal**, v. 104, n. 3, p. 722–728, 2012.

WICKHAM, H. **ggplot2: Elegant Graphics for Data Analysis**. New York, NY: Springer-Verlag, 2016.

WILLIAMS, E.; PIEPHO, H.-P.; WHITAKER, D. Augmented p-rep designs. **Biometrical Journal**, v. 53, n. 1, p. 19–27, 2011.

YAN, W. et al. Cultivar Evaluation and Mega-Environment Investigation Based on the GGE Biplot. **Crop Science**, v. 40, n. 3, p. 597–605, 2000.

YATES, F.; COCHRAN, W. G. The analysis of groups of experiments. **The Journal of Agricultural Science**, v. 28, n. 4, p. 556–580, 1938.

ZHANG, R.; HAN, D.; HU, X. Analyzing the performance of corn in China using a factor-analytic variance-covariance structure with multiple factors. **Crop Science**, v. 60, n. 1, p. 190–201, 2020.

ZYSTRO, J.; COLLEY, M.; DAWSON, J. Alternative Experimental Designs for Plant Breeding. Em: **Plant Breeding Reviews**. [s.l.] Wiley, 2018. p. 87–117.

CAPÍTULO II

**DESEMPENHO E ESTABILIDADE DE PROGÊNIES DE FEIJOEIRO COM DADOS
DE MÚLTIPLAS CARACTERÍSTICAS E AMBIENTES: APLICAÇÃO DAS *FACTOR*
ANALYTIC SELECTION TOOLS - FAST**

RESUMO

O objetivo deste estudo foi selecionar progênies de feijão comum (*Phaseolus vulgaris* L.) com alto desempenho e estabilidade fenotípica em múltiplos ambientes, utilizando ferramentas estatísticas. Para isso, adotou-se uma abordagem em dois estágios: (1) estimação de médias ajustadas por ensaio e (2) análise de modelo misto com estrutura de variância-covariância fator-analítica (FA) aplicada às médias ajustadas. Os dados utilizados foram obtidos de cinco ensaios em múltiplos ambientes conduzidos entre 2018 e 2020 no âmbito do programa de seleção recorrente da Universidade de Viçosa. O modelo fator-analítico de oitava ordem (FA8) foi selecionado com base no Critério de Informação de Akaike (AIC). Posteriormente, as *Factor Analytic Selection Tools* (FAST) foram aplicadas para identificar progênies superiores. Como resultado, 20 progênies destacaram-se por apresentarem alto desempenho geral (OP, *Overall Performance*), baixo desvio quadrático médio (RMSD, *Root Mean Square Deviation*) e confiabilidade superior a 0,80. Esses achados reforçam a importância da integração de análises multiambientes e multivariadas em programas de melhoramento de plantas. O FAST demonstrou ser uma ferramenta eficaz para sintetizar dados complexos em métricas decisórias, permitindo a identificação de genótipos com desempenho e estabilidade desejáveis. Conclui-se que a estrutura proposta oferece um arcabouço robusto para a seleção de progênies adaptáveis e produtivas, sendo recomendada sua incorporação em programas de melhoramento de feijão comum. Além disso, a adoção dessa abordagem pode aprimorar a análise de dados de múltiplas características e ambientes, garantindo maior eficiência na seleção de genótipos superiores.

Palavras-chave: seleção multiambiente; modelos fator-analíticos; desempenho e estabilidade fenotípica.

ABSTRACT

We aimed to select common bean (*Phaseolus vulgaris* L.) progenies with high performance and phenotypic stability across multiple environments using statistical tools. We adopted a two-stage approach: (1) estimating trial-adjusted means and (2) applying a mixed model analysis with a factor-analytic (FA) variance-covariance structure to the adjusted means. We obtained data from five multi-environment trials conducted between 2018 and 2020 under the recurrent selection program of the Universidade Federal de Viçosa. Based on the Akaike Information Criterion (AIC), we selected the eighth-order factor-analytic model (FA8). We then applied Factor Analytic Selection Tools (FAST) to identify superior progenies. Our results highlighted 20 progenies that demonstrated high overall performance (OP), low root mean square deviation (RMSD), and reliability above 0.80. These findings reinforce how integrating multi-environment and multivariate analyses benefits plant breeding programs. FAST effectively synthesized complex data into decision-making metrics, enabling us to identify genotypes with desirable performance and stability. We conclude that this framework provides a robust structure for selecting adaptable and productive progenies and recommend its adoption in common bean breeding programs. Moreover, this approach could enhance multi-trait and multi-environment data analysis, improving efficiency in selecting superior genotypes.

Keywords: multi-environment selection; factor-analytic models; phenotypic performance and stability.

1 INTRODUÇÃO

A recomendação de cultivares de feijão comum enfrenta um desafio central: a interação genótipos x ambientes (GxA), que influencia diretamente a adaptabilidade e estabilidade dos genótipos (Priyadarshan, 2019, p. 457-472). Para contornar essa complexidade, melhoristas recorrem a ensaios multiambientais (EM), ferramentas essenciais para avaliar e selecionar genótipos superiores (KELLY et al., 2007; SMITH; CULLIS; THOMPSON, 2001). A análise dessa interação evoluiu significativamente nas últimas décadas. Inicialmente, métodos baseados em mínimos quadrados ordinários (OLS) foram utilizados para modelar a interação GxA (MALOSETTI; RIBAUT; VAN EEUWIJK, 2013), seguidos por abordagens como a regressão de Finlay e Wilkinson (1963), que relacionavam o desempenho dos genótipos a índices ambientais. Contudo, a dependência de índices derivados dos próprios dados fenotípicos introduzia viés estatístico, devido à correlação entre regressores e variáveis resposta (CROSSA, 1990; SMITH; CULLIS; THOMPSON, 2005).

Para superar essas limitações, métodos multiplicativos como AMMI (GAUCH, 1988) e biplots GGE (YAN et al., 2000) emergiram, explorando decomposições matriciais e representações gráficas (GABRIEL, 1971). No entanto, sua aplicação prática mostrou-se restrita a conjuntos de dados balanceados e de pequena escala, além de dependerem de pressupostos rígidos, como efeitos genotípicos fixos (DIAS et al., 2022; SMITH; CULLIS, 2018). A virada metodológica ocorreu com a introdução do Modelo Misto Fator-Analítico (FA) por Piepho (1997, 1998a), que incorporou heterogeneidade de covariância entre ambientes e estruturas de variância espacial, além de permitir efeitos genotípicos aleatórios e dados desbalanceados (SMITH; CULLIS; THOMPSON, 2005, 2001).

Apesar desses avanços, a seleção simultânea de múltiplas características em diversos ambientes permanece um desafio, sobretudo devido à correlação entre características quantitativas e ao risco de viés de seleção (POLLAK; VAN DER WERF; QUAAS, 1984). Índices clássicos de seleção, como os propostos por Smith (1936) e Hazel (1943), embora úteis para ponderar características economicamente relevantes, frequentemente negligenciam a complexidade da interação GxA (CERÓN-ROJAS et al., 2006; CRUZ; CARNEIRO; REGAZZI, 2014). Uma inovação promissora nesse contexto foi o *Factor Analytic Selection Tools* (FAST), proposto por Smith e Cullis (2018). Essa ferramenta integra

as estimativas do modelo fator-analítico (FA) em métricas decisórias práticas, como o *Desempenho Geral* (OP, *Overall Performance*) e a *Raiz dos Quadrados Médios dos Desvios* (RMSD, *Root Mean Square Deviation*), que quantifica a estabilidade. Essas métricas permitem identificar genótipos ideais – aqueles com OP elevado, RMSD baixo e RE próximo de zero – e têm sido aplicadas com sucesso em culturas como soja, trigo, milho e algodão (ARAÚJO et al., 2024; BAKARE et al., 2022; CHAVES et al., 2023a; TOLHURST et al., 2019).

No entanto, nenhum estudo reportou a aplicação das FAST no feijoeiro comum, lacuna que pode limitar o avanço de programas de melhoramento dedicados a essa cultura. Diante disso, este trabalho teve como objetivos (i) identificar progênies de feijoeiro comum com alto desempenho agrônômico e estabilidade fenotípica utilizando as FAST, e (ii) validar a eficácia dessa abordagem para a seleção de progênies para extração de linhagens avaliadas para múltiplas características em condições multiambientais.

2. MATERIAL E MÉTODOS

2.1 Material

Foram utilizados dados provenientes de cinco ensaios do Programa de Seleção Recorrente (Ciclo C0) de Feijão Carioca, conduzidos pela Universidade Federal de Viçosa (UFV) entre 2018 e 2020, com foco na resistência à murcha-de-fusário (*Fusarium oxysporum* f. sp. *phaseoli* – FOP), cuja avaliação será conduzida em linhagens homozigotas, e não em famílias (progênes) ou indivíduos segregantes, pois a replicação do genótipo é necessária para garantir a confiabilidade dos resultados. Por essa razão, a seleção para resistência à murcha-de-fusário não é realizada durante os estágios iniciais do programa de melhoramento do feijoeiro-comum, mas sim em um grupo reduzido de linhagens previamente selecionadas com base em características agronômicas desejáveis.

Os experimentos, detalhados na Tabela 6, abrangeram diferentes safras (águas, seca, inverno) e gerações (F_{4:5} a F_{4:9}), com delineamentos variando entre blocos aumentados (DBA), blocos casualizados (DBC) e látice triplo, conforme o número de tratamentos e a disponibilidade de recursos.

Tabela 6: Descrição dos cinco ensaios utilizados no presente trabalho.

Ensaio	Ano	Safr	Geração	Tratamento*	Delineamento	Repetição	Parcela**
1	2018	Águas	F _{4:5}	800 (5)	DBA	-	1L x 2m
2	2019	Seca	F _{4:6}	20 (5)	20 DBC	3	1L x 2m
3	2019	Águas	F _{4:7}	95 (5)	Látice	3	2L x 2m
4	2020	Inverno	F _{4:8}	95 (5)	Látice	3	2L x 2m
5	2020	Águas	F _{4:9}	95 (5)	Látice	3	2L x 2m

*Número de testemunhas comuns entre parênteses.

** L: linha, m: metros

O programa iniciou-se em 2014 com a avaliação de 12 genótipos de feijoeiro em esquema dialelo parcial, selecionando cultivares comerciais de elite e linhagens portadoras de genes de resistência a doenças-chave de programas brasileiros. Dos cruzamentos dialélicos, obtiveram-se 36 populações F₁, das quais 34 progrediram para F₂ (SILVA et al., 2023). Após seleção em *bulk* até a geração F₄, 20 populações superiores foram escolhidas, das quais 800 progênes foram derivadas. Na safra águas/2018, estas 800 progênes F_{4:5} foram avaliadas em delineamento de blocos aumentados (DBA), com parcelas de 1 linha de 2 metros.

Posteriormente, selecionaram-se as 20 melhores progênes F_{4:5} de cada população, totalizando 400 progênes F_{4:6}, avaliadas em um grupo de 20

experimentos (um por população) em delineamento de blocos completos casualizados (DBC), com três repetições, 20 tratamentos e 5 testemunhas comuns (BRITO; CARNEIRO; SILVA, 2021). A partir da avaliação dessas 400 progênies, as 95 progênies mais promissoras foram selecionadas representando as 20 populações originais. Essas 95 progênies selecionadas foram avaliadas em látice triplo em três safras (Tabela 6).

As testemunhas incluíram cultivares com perfis contrastantes: Manteigão Fosco e VC25 (resistência a FOP), Meia Noite e BRS Estilo (suscetibilidade a FOP), e Pérola (com padrão de grãos comerciais desejável e resistência a FOP). A avaliação específica para resistência à FOP, no entanto, será realizada posteriormente em linhagens homozigotas já estabilizadas, por meio de inoculação controlada em condições padronizadas. Essa etapa não foi implementada nas fases iniciais (F_{4:5} a F_{4:9}), nas quais a seleção focou exclusivamente em atributos agronômicos.

Todos os ensaios foram conduzidos na Unidade de Ensino, Pesquisa e Extensão de Coimbra (UFV), localizada em Coimbra-MG, região de clima subtropical de inverno seco (Cwa, segundo Köppen) (Peel et al., 2007). O delineamento experimental foi definido com base no número de tratamentos, viabilidade de área e disponibilidade de sementes. Embora o tamanho das parcelas tenha variado (1L x 2m ou 2L x 2m), o espaçamento entre sulcos (0,5 m) e a densidade de semeadura (15 sementes/m) foram padronizados. O manejo seguiu recomendações técnicas para a cultura, incluindo adubação equilibrada e controle fitossanitário (CARNEIRO; DE PAULA JÚNIOR; BORÉM, 2015).

2.2 Características avaliadas

Foram avaliadas três características agronômicas em diferentes safras: produtividade de grãos (PROD), arquitetura de planta (ARQ) e aspecto comercial de grãos (ASP). A Tabela 9 sumariza os dados coletados, enquanto as metodologias específicas para cada característica são detalhadas abaixo.

2.2.1 Produtividade de grãos (PROD)

A produtividade de grãos foi determinada em todas as safras após colheita, trilha, remoção de impurezas e pesagem dos grãos por parcela. Os valores foram convertidos de gramas para quilogramas por hectare (kg ha^{-1}) utilizando o fator de conversão padrão para a cultura.

2.2.2 Arquitetura de planta (ARQ)

A arquitetura de planta foi avaliada nas safras Águas/2018, Seca/2019 e Águas/2020. Na safra Seca/2019, adotou-se uma escala de notas de 1 a 5, conforme descrito por Ramalho *et al.* (1998). Nas demais safras, utilizou-se uma escala de 1 a 9 proposta por Collicchio *et al.* (1997) (Tabela 7).

Tabela 7. Escala de avaliação de arquitetura de planta.

Nota	Descrição
1	Planta do tipo II: ereta, com uma haste e com inserção alta das primeiras vagens
3	Planta do tipo II: ereta e com algumas ramificações
5	Planta do tipo II ou III: ereta, com muitas ramificações e tendência a prostrar-se
7	Planta do tipo III: semiereta e medianamente prostrada
9	Planta do tipo III: com entrenós longos e muito prostrada

Fonte: Collicchio *et al.* (1997).

Para uniformizar os dados, as notas da escala de 1–5 (Seca/2019) foram convertidas para a escala de 1–9 por meio da equação proposta por Anjos *et al.* (2019).

$$S'_{ui'} = \left[\left(\frac{i' - u'}{i - u} \right) (S_{ui} - u) \right] + u' \quad [17]$$

em que:

- S_{ui} : nota original na escala $u-i$;
- u : nota do fenótipo indesejável na escala original;
- i : nota do ideótipo na escala original;
- $S'_{ui'}$: nota convertida na escala $u' - i'$;
- u' : nota do fenótipo indesejável na escala convertida;
- i' : nota do ideótipo na escala convertida.

2.2.3 Aspecto comercial de grãos (ASP)

O aspecto comercial foi avaliado em todas as safras com base em uma escala adaptada de Ramalho *et al.* (1998) (Tabela 8), onde notas ≤ 2 indicam excelente qualidade comercial (BARILI, 2015).

Tabela 8: Descrição da escala de notas de aspecto de grãos do tipo carioca, adaptada de Ramalho *et al.* (1998).

Nota	Descrição
1	Grão típico carioca: cor creme com estrias marrom-claras, fundo claro, halo creme, peso médio de 100 sementes de 22 a 24 g e não achatado
2	Grão tipo carioca com deficiência em uma das características mencionadas no padrão
3	Grão tipo carioca com deficiência em duas características mencionadas no padrão
4	Grão tipo carioca com deficiência em três características mencionadas no padrão
5	Grão de cor creme com estrias marrom-escuras, fundo escuro, com halo não creme, peso médio de 100 sementes menor que 22g e achatado

Fonte: Ramalho *et al.* (1998).

2.2.4 Padronização das escalas

Como as escalas de ARQ e ASP atribuem maior mérito fenotípico a genótipos mais próximos do ideótipo, aplicou-se a Escala Delta (δ_{ui}) para inverter as notas, facilitando análises comparativas. O mérito fenotípico foi calculado como:

$$\delta_{ui} = |S_{ui} - u| \quad [18]$$

em que:

- S_{ui} : nota original do genótipo;
- u : nota do fenótipo indesejável na escala;
- δ_{ui} : mérito fenotípico na Escala Delta (valores mais altos indicam maior proximidade ao ideótipo).

A Tabela 9 apresenta os valores mínimo, médio e máximo de PROD, ARQ e ASP por ano e safra. Os dados de ARQ não foram coletados nas safras Águas/2019 e Inverno/2020 (indicados por "-"). Todos os valores correspondem aos fenótipos originais, antes da padronização para variância unitária.

Tabela 9: Resumo das características agrônômicas no conjunto de dados multiambientes e multicaracterísticas.

		PROD (Kg ha ⁻¹)			ARQ			ASP		
		Min	Média	Max	Min	Média	Max	Min	Média	Max
Ano	2018	597,60	2600,12	5471,10	2	3,69	7	3,8	2,69	3,8
	2019	275,10	3883,45	7383,50	1	3,28	7	3,5	2,35	3,5
	2020	480,00	3165,83	5900,00	1	3,11	7	3,5	2,54	3,5
Safras	Águas/2018	597,60	2600,12	5471,10	2	3,69	7	3,8	2,69	3,8
	Seca/2019	275,10	3883,45	7383,50	1	3,28	7	3,5	2,35	3,5
	Águas/2019	745,00	3307,90	6870,00	-	-	-	0	2,61	3,5
	Inverno/2020	1020,00	3515,60	5900,00	-	-	-	0	2,61	3,5
	Águas/2020	480,00	2814,88	5460,00	1	3,11	7	0	2,47	3,5

2.3 Análises estatísticas

Para a análise dos dados, adotou-se um procedimento em dois estágios, metodologia amplamente reconhecida por seu equilíbrio entre rigor analítico e viabilidade prática (MÖHRING e PIEPHO, 2009). Nesse esquema, a avaliação inicial é conduzida separadamente para cada ambiente, permitindo a consideração das particularidades experimentais inerentes a cada local, incluindo o delineamento, variáveis avaliadas, número de blocos, repetições e a estrutura de variância residual. Esta abordagem mostrou-se particularmente adequada ao presente estudo, dado: (i) a marcante heterogeneidade entre os diferentes locais de avaliação; (ii) as particularidades dos delineamentos experimentais empregados; e (iii) a significativa economia de recursos computacionais em comparação com análise de estágio único.

Para a análise integrada de múltiplas características e ambientes, os dados fenotípicos foram escalonados para média zero e variância unitária no primeiro estágio, garantindo comparabilidade entre escalas distintas. A padronização foi realizada utilizando o método z-score:

$$Z = \frac{x - \mu}{\sigma} \quad [19]$$

em que:

- x : valor fenotípico observado para cada característica em uma safra;
- μ : média fenotípica da característica na safra;
- σ : desvio padrão fenotípico da característica na safra.

Em ensaios organizados em Delineamento em Blocos Completos Casualizados (DBC), a padronização foi aplicada separadamente para cada bloco, considerando a heterogeneidade genética entre populações.

Os componentes de variância-covariância e os valores genotípicos foram estimados via metodologia REML/BLUP (HENDERSON, 1975; PATTERSON; THOMPSON, 1971). A análise foi conduzida em dois estágios:

1. Primeiro estágio: Ajuste dos modelos mistos para cada ambiente (e nos 20 Delineamentos em Blocos Casualizados - DBC) para obter as médias ajustadas das progênies.
2. Segundo estágio: Integração multivariada dos resultados parciais (médias ajustadas) para estimação de parâmetros genéticos conjuntos.

Todas as análises foram implementadas no ambiente de desenvolvimento integrado (IDE) RStudio, utilizando a linguagem de programação R (R Core Team, 2023, versão 4.2.3) e pacotes especializados para modelagem estatística e visualização de dados. Foram empregados os pacotes ggplot2 (versão 3.4.2) (WICKHAM, 2016) para geração de gráficos, ComplexHeatmap (versão 2.14.0) (GU; EILS; SCHLESNER, 2016) para construção de heatmaps multivariados e ASReml-R versão 4.1.0.176 (BUTLER et al., 2023) para ajuste de modelos. Adicionalmente, utilizou-se a função personalizada fa.outs.R, disponível no repositório GitHub (https://github.com/saulo-chaves/May_b_useful/blob/main/fa.outs.R), que compila as saídas dos modelos fator-analíticos ajustados pelo ASReml-R.

A execução das análises foi realizada em um sistema com processador Intel® Core I5 11400H 2,70GHz, 16GB de RAM DDR4 3200 Mhz, 512GB SSD, placa gráfica NVIDIA® RTX 3050 4GB GDDR6, rodando o sistema operacional Microsoft® Windows 11 Home.

2.4 Análise dos dados em estágios

A análise em dois estágios envolve dois elementos cruciais:

1. **Estimativas imparciais no primeiro estágio:** Para garantir imparcialidade, os efeitos genotípicos foram modelados como fixos em cada ensaio individual, permitindo a obtenção das melhores estimativas lineares não viesadas empíricas (e-BLUEs, do inglês *empirical Best Linear Unbiased*

Estimators). Essas médias ajustadas (e-BLUEs) representam o desempenho fenotípico dos genótipos, livre de variações ambientais e de erro experimental.

2. **Matriz de ponderação no segundo estágio:** As médias ajustadas do primeiro estágio foram analisadas conjuntamente em um modelo misto multivariado, ponderado por uma matriz ω que considera a precisão das estimativas. Essa matriz é definida como $\omega = \text{diag}(V_j^{-1})$, em que V_j^{-1} é a matriz de variância-covariância das médias ajustadas por ambiente (SMITH; CULLIS; GILMOUR, 2001). A Figura 4 ilustra o fluxo analítico, destacando a interdependência entre os estágios.

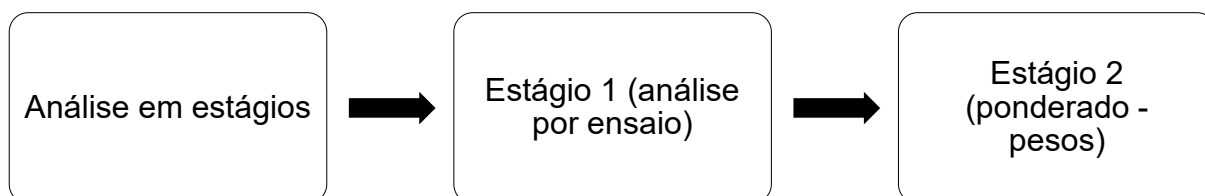


Figura 4. Esquema da análise em estágios.

As subseções seguintes apresentam as análises do estágio 1 e estágio 2 detalhadamente.

2.5 Análises individuais para cada ensaio – Estágio 1

No primeiro estágio da análise, modelos mistos lineares foram ajustados individualmente para cada ambiente/safra (ensaio) e característica, com o objetivo de obter as médias ajustadas das progênies. Os modelos variaram conforme o delineamento experimental utilizado (DBA, DBC e Látice), conforme detalhado a seguir.

Ensaio em Delineamento em Blocos Aumentados (DBA) – Águas/2018

Para o ensaio em DBA, empregou-se o seguinte modelo:

$$y = X_g\mu + Z_r\theta + Z_p\tau + \epsilon \quad [20]$$

em que:

- y : vetor de dados fenotípicos ($n \times 1$), sendo n o número de observações.
- X_g : matriz de incidência do efeito da média geral, de dimensão $n \times 1$;

- μ : constante fixa (média geral das progênes no ensaio);
- Z_r : matriz de incidência dos efeitos de blocos ($n \times 16$);
- θ : vetor de efeitos aleatórios de blocos (16×1);
- Z_p : matriz de incidência dos efeitos genotípicos ($n \times 880$);
- τ : vetor de efeitos fixos genotípicos (880×1);
- ϵ : vetor de erros ($n \times 1$), com $\epsilon \sim N(0, R)$ e $R = I\sigma_\epsilon^2$, em que I é a matriz identidade de dimensão $n \times n$ e σ_ϵ^2 é a variância do erro experimental.

Ensaio em Delineamento em Blocos Casualizados (DBC) – Seca/2019

Foram instalados 20 experimentos em DBC, um para cada população, conforme descrito na subseção "Material". Cada experimento foi analisado individualmente, considerando a heterogeneidade genética entre populações. O modelo utilizado foi:

$$y = X_g\mu + Z_r\theta + Z_p\tau + \epsilon \quad [21]$$

em que:

- y : vetor de dados fenotípicos ($n \times 1$);
- X_g : matriz de incidência do efeito da média geral ($n \times 1$);
- μ : constante fixa (média geral das progênes no ensaio);
- Z_r : matriz de incidência dos efeitos de blocos ($n \times 3$);
- θ : vetor de efeitos fixos de blocos (3×1);
- Z_p : matriz de incidência dos efeitos genotípicos ($n \times 25$) para cada DBC montado;
- τ : vetor de efeitos fixos genotípicos (25×1);
- ϵ : vetor de erros ($n \times 1$), com $\epsilon \sim N(0, R)$ e $R = I\sigma_\epsilon^2$, em que I é a matriz identidade de dimensão $n \times n$ e σ_ϵ^2 é a variância do erro experimental.

Ensaio em Látice Triplo – Águas/2019, Inverno/2020 e Águas/2020

Para os ensaios em látice triplo, utilizou-se o seguinte modelo:

$$y = X_g\mu + Z_r\gamma + Z_b\theta + Z_p\tau + \epsilon \quad [22]$$

em que:

- y : vetor de dados fenotípicos ($n \times 1$);
- X_g : matriz de incidência do efeito da média geral ($n \times 1$);
- μ : constante fixa (média geral das progênies no ensaio);
- Z_r : matriz de incidência dos efeitos de repetições ($n \times 3$);
- γ : vetor dos efeitos fixos de repetições (3×1);
- Z_b : matriz de incidência dos efeitos de blocos ($n \times 30$);
- θ : vetor de efeitos aleatórios de blocos (30×1);
- Z_p : matriz de incidência dos efeitos genotípicos ($n \times 100$);
- τ : vetor de efeitos fixos genotípicos (100×1);
- ϵ : vetor de erros ($n \times 1$), com $\epsilon \sim N(0, R)$ e $R = I\sigma_\epsilon^2$, em que I é a matriz identidade de dimensão $n \times n$ e σ_ϵ^2 é a variância do erro experimental.

2.6 Cálculo dos pesos para o estágio 2

Na análise em dois estágios, a ponderação das estimativas é crítica para garantir robustez estatística. Diversos métodos são propostos na literatura, sendo que os pesos geralmente dependem das variâncias e covariâncias das médias ajustadas obtidas no primeiro estágio (MÖHRING; PIEPHO, 2009). Neste estudo, adotou-se o método de Smith *et al.* (2001), que utiliza a matriz de variância-covariância V_j das médias genotípicas ajustadas (calculadas no Estágio 1) para derivar os pesos do segundo estágio.

A matriz de pesos ω foi definida como:

$$\omega = \text{diag}(V_j^{-1}) \quad [23]$$

em que:

- V_j : matriz de variância-covariância das médias ajustadas para o ambiente j e característica s ;
- $\text{diag}(V_j^{-1})$: matriz diagonal cujos elementos correspondem aos valores da diagonal principal de V_j^{-1} .

Este método aproxima V_j^{-1} por uma matriz diagonal, minimizando os erros quadráticos médios associados às estimativas (SMITH; CULLIS; THOMPSON, 2005; SMITH; CULLIS; GILMOUR, 2001).

2.7 Estágio 2 – Análise de múltiplas características e ambientes

O conjunto de dados analisados compreendeu três características fenotípicas padronizadas ($s = 3$), representadas pelas médias ajustadas obtidas no primeiro estágio da análise: (i) produtividade de grãos (PG), (ii) arquitetura de planta (ARQ) e (iii) aspecto comercial de grão (ASP). Estas foram avaliadas em cinco ambientes ($p = 5$) correspondentes a diferentes safras: Águas/2018, Seca/2019, Águas/2019, Inverno/2020 e Águas/2020, totalizando 805 genótipos ($v = 805$) distribuídos em 3.280 parcelas experimentais ($n = 3.280$).

Para as análises integradas, os dados foram reorganizados hierarquicamente em um vetor y no formato *long*, utilizando a função `pivot_longer()` do pacote *tidyr* no R. Esta transformação converteu a estrutura inicial (*wide*, com uma coluna por característica) para uma disposição longitudinal (uma linha por observação característica-ambiente), conforme representado matricialmente:

$$y = [Y_1^T Y_2^T Y_3^T]^T \quad [24]$$

onde cada subvetor Y_i ($i = 1,2,3$) corresponde à concatenação vertical dos dados da i -ésima característica em todos os ambientes, podendo ser expandido como:

$$Y_i = [Y_{i_1}^T Y_{i_2}^T \dots Y_{i_5}^T]^T \quad [25]$$

Nesta estrutura, Y_{ij} representa um vetor coluna de dimensão $n_j \times 1$ contendo as observações da característica i no ambiente j , sendo n_j o número de parcelas no ambiente j . Das 15 combinações teóricas possíveis (3 características x 5 ambientes), foram obtidas 13 combinações válidas ($h = 1, \dots, 13$), devido à ausência de dados para arquitetura de planta (ARQ) nas safras Águas/2019 e Inverno/2020 - restrição detalhada na Tabela 9. O comprimento total do vetor y foi calculado por:

$$n_s = \sum_{h=1}^{13} n_h = 9.236 \quad [26]$$

onde n_h representa o número de observações na h -ésima combinação característica-ambiente.

Modelo Misto Linear Ponderado

As médias ajustadas ($\bar{y}_{ij}$) das progênies, obtidas no primeiro estágio, foram modeladas por:

$$\bar{y}_{ij} = \mathbf{X}\beta + \mathbf{Z}_g u_g + e \quad [27]$$

Componentes do modelo:

1. Efeitos fixos (β): Médias gerais das combinações característica-ambiente.
2. Efeitos aleatórios (u_g): Interações genótipo $\times$ combinação característica-ambiente, com $u_g \sim N(0, G_g)$.
3. Erro residual (e): $e \sim N(0, V)$, em que V incorpora a variância residual.

Matrizes de variância:

- $G_g = \bigoplus_{h=1}^{13} \sigma_{gi}^2 I_m$: Matriz de variância genética, com σ_{gi}^2 representando a variância genética para cada combinação h .
- I_m : Matriz identidade de ordem $m \times m$, onde m é o número de genótipos.

2.8 Ajuste dos modelos FA para a interação genótipo-ambiente-característica (GETI)

O modelo fator-analítico (FA) foi empregado para modelar covariâncias genéticas entre ambientes por meio de fatores latentes hipotéticos. Para a interação genótipo-ambiente-característica (GETI), o efeito genotípico da progênie i na combinação característica-ambiente h é expresso como:

$$u_{gi} = f_{1i}\lambda_1 + f_{2i}\lambda_2 + \dots + f_{ki}\lambda_k + \delta_i \quad [28]$$

ou, em notação compacta:

$$g_{ih} = \sum_{k=1}^t \lambda_{hk} f_{ik} + \delta_{ih} \quad [29]$$

Componentes do modelo:

- λ_{hk} : carga fatorial do fator latente k na h -ésima combinação característica-ambiente.
- f_{ik} : *score* da progênie i para o fator k , representando sua sensibilidade genética.
- δ_{ih} : termo residual, capturando a variabilidade não explicada pelos fatores.

Decomposição da Variância-Covariância

A matriz de variância-covariância genética (G) é decomposta em:

$$G = (\Lambda\Lambda^T) + \Psi \quad [30]$$

em que:

- Λ : matriz $s \times t$ de cargas fatoriais ($s = 13$ combinações característica-ambiente; $t =$ número de fatores):

$$\Lambda = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1t} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{s1} & \lambda_{s2} & \cdots & \lambda_{st} \end{bmatrix} \quad [31]$$

- Ψ : Matriz diagonal $s \times s$ de variâncias específicas ($\Psi_1, \Psi_2, \dots, \Psi_s$):

$$\Psi = \begin{bmatrix} \Psi_1 & 0 & \cdots & 0 \\ 0 & \Psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Psi_s \end{bmatrix} \quad [32]$$

O número de fatores latentes é chamado de ordem do modelo e no presente estudo FAK representa um modelo FA de ordem k .

2.9 Seleção do modelo FA de melhor ajuste

A seleção do modelo fator-analítico (FA) ótimo baseou-se no equilíbrio entre parcimônia (simplicidade do modelo) e capacidade explicativa (variância capturada pelos fatores latentes). Para isso, adotou-se uma abordagem multicritério, conforme detalhado a seguir:

Critério de Informação de Akaike (AIC)

O AIC, proposto por Akaike (1974), foi calculado como:

$$AIC = -2\ln(L) + 2p \quad [33]$$

em que:

- L : Verossimilhança máxima do modelo;
- p : Número de parâmetros estimados.

Modelos com menor AIC são preferíveis, indicando melhor ajuste relativo. Contudo, em modelos fatoriais (FA), o AIC pode favorecer estruturas excessivamente complexas (ISIK; HOLLAND; MALTECCA, 2017; SMITH et al., 2015). Para contornar limitações do AIC, foram incorporadas métricas adicionais:

1. Variância total explicada pelos fatores ($\% \bar{v}$), segundo Smith et al. (2015):

$$\% \bar{v} = \left(\frac{\text{tr}(\Lambda \Lambda')}{\text{tr}(\Sigma g)} \right) \quad [34]$$

em que:

- Λ : Matriz de cargas fatoriais rotacionadas;
2. Variância explicada por fator em cada combinação (v_{k_h}), segundo Smith et al. (2015):

$$v_{k_h} = \frac{\sum_{k=1}^k \lambda_{k_h}^2}{\sum_{k=1}^k \lambda_{k_h}^2 + \Psi_h} \quad [35]$$

em que:

- $\lambda_{k_h}^*$: Carga fatorial rotacionada do fator k na combinação h ;

- Ψ_h : Variância específica residual da combinação h .

3. Razão Média das Semivariâncias (ASVR)

A Razão Média das Semivariâncias (ASVR) quantifica a proporção da variância estruturada (explicada pelos fatores) em relação à variância total (CHAVES et al., 2023b):

$$ASVR = \frac{Numerador}{Denominador} \times 100 \quad [36]$$

Componentes:

Numerador: Variância estruturada entre combinações h e h' :

$$\frac{2}{h(h-1)} \sum_{h=1}^{H-1} \sum_{h'=h+1}^H \left[\frac{1}{2} \left(\sum_{k=1}^K \hat{\lambda}_{k_h}^{*2} d_k + \sum_{k=1}^K \hat{\lambda}_{k_{h'}}^{*2} d_k \right) - \sum_{k=1}^K \hat{\lambda}_{k_h}^* \hat{\lambda}_{k_{h'}}^* d_k \right] \quad [37]$$

Denominador: Variância total entre h e h' :

$$\frac{2}{h(h-1)} \sum_{h=1}^{H-1} \sum_{h'=h+1}^H \left[\frac{1}{2} \left(\sum_{k=1}^K \hat{\lambda}_{k_h}^{*2} d_k + \hat{\Psi}_h + \sum_{k=1}^K \hat{\lambda}_{k_{h'}}^{*2} d_k + \hat{\Psi}_{h'} \right) - \sum_{k=1}^K \hat{\lambda}_{k_h}^* \hat{\lambda}_{k_{h'}}^* d_k \right] \quad [38]$$

em que:

d_k : Variância do *score* do fator k (elemento diagonal da matriz D);

$\hat{\lambda}_{k_h}^*$: Cargas rotacionadas via decomposição em valores singulares (SVD) (Smith et al., 2021).

As cargas fatoriais (*loadings*) foram rotacionadas empregando o método de decomposição em Valores Singulares (SVD), conforme Smith et al. (2021). A rotação SVD foi adotada para assegurar ortogonalidade entre os fatores e maximizar a variância explicada pelas cargas fatoriais. Essa abordagem simplifica a interpretação biológica dos fatores latentes, além de garantir estabilidade numérica em matrizes multiambientais.

2.10 Estimativas dos parâmetros genéticos e não genéticos

Para calcular a matriz de correlações genéticas entre ambientes (combinação característica-ambiente), a metodologia envolve duas etapas

principais: (i) estimação da matriz de covariâncias genéticas e (ii) normalização para obtenção das correlações. A matriz de covariâncias genéticas entre ambientes (Σ_g) de dimensão 13×13 é definida pela expressão:

$$\Sigma_g = (\hat{\Lambda}^* D \hat{\Lambda}^{*'} + \hat{\Psi}) \otimes I_v \quad [39]$$

em que:

- $\hat{\Lambda}^*$ corresponde à matriz de cargas fatoriais rotacionadas;
- D é a matriz diagonal previamente definida na Seção 2.8;
- $\hat{\Psi}$ representa a matriz diagonal 13×13 de variâncias específicas;
- I_v denota a matriz identidade de dimensão compatível com os ambientes.

A matriz de correlações genéticas (Υ) é derivada da normalização de Σ_g , conforme proposto por Cullis et al., (2010):

$$\Upsilon = \Delta \Sigma_g \Delta \quad [40]$$

em que Δ é uma matriz diagonal cujos elementos são o inverso da raiz quadrada das variâncias genéticas $(diag(\Sigma_g))^{-0,5}$, garantindo a escala unitária das correlações. Para visualização, as correlações foram representadas em um *heatmap* gerado pelo pacote *ComplexHeatmap* v.2.14.0 (GU; EILS; SCHLESNER, 2016), com escalas de cores padronizadas para facilitar a interpretação de padrões ambientais.

Os Melhores Preditores Lineares Imparciais (BLUPs) empíricos condicionais, que levam em conta efeitos ambientais específicos somados ao intercepto, podem ser estimados usando a seguinte equação:

$$g = (\Lambda^* \otimes I_g) \tilde{f}^* + \tilde{\delta} + \mu \quad [41]$$

em que $\tilde{f}^*$ é o vetor de escores rotacionados, I_g é uma matriz identidade cuja dimensão é o número de genótipos, $\tilde{\delta}$ é o vetor de efeito específico (falta de ajuste da regressão múltipla) e μ é o intercepto do modelo. Para a estimação dos e-BLUPs marginais apenas os efeitos comuns são usados na eq. [35], ou seja:

$$g = (\Lambda^* \otimes I_g) \tilde{f}^* \quad [42]$$

A partir do modelo mais bem ajustado, foram estimadas as herdabilidades generalizadas para cada ambiente (H_j^2), conforme descrito por Cullis *et al.* (2006):

$$H_j^2 = 1 - \frac{\bar{V}(\Delta)}{2\hat{\sigma}_{gj}^2} \quad [43]$$

em que $\bar{V}(\Delta)$ é a variância média do erro de predição entre progênies par a par e σ_{gj}^2 é variância genotípica para cada ambiente. Além disso foi calculado o coeficiente de variância experimental (CV_{ej}), dado por:

$$CV_{ej} = \frac{\sqrt{\sigma_{ej}^2}}{\mu_j} \quad [44]$$

onde σ_{ej}^2 representa a variância residual no ambiente j e μ representa a média fenotípica no ambiente j .

2. 11 Factor Analytic Selection Tools - FAST

Para a seleção de progênies superiores, aplicou-se as ferramentas de seleção para desempenho geral e estabilidade FAST (*Factor Analytic Selection Tools*), proposto por Smith e Cullis (2018), em modelos mistos de estrutura fator-analítica. O modelo fator-analítico de ordem reduzida é definido como:

$$g_{ij} = \lambda_1^* f_1^* + \lambda_{2j}^* f_{2i}^* + \lambda_{3j}^* f_{3i}^* + \dots + \lambda_{kj}^* f_{ki}^* \quad [45]$$

Para simplificação computacional, considerou-se a decomposição:

$$g_{ij} = \lambda_1^* f_1^* + \epsilon_{ij} \quad [46]$$

onde $\epsilon_{ij} = \lambda_{2j}^* f_{2i}^* + \lambda_{3j}^* f_{3i}^* + \dots + \lambda_{kj}^* f_{ki}^*$ representa os efeitos residuais associados à estabilidade.

Essa relação facilita o cálculo do desempenho geral (OP_i) e estabilidade ($RMSD_i$), parâmetros cruciais para a seleção das progênies estáveis e de alto desempenho, conforme descrito por Smith e Cullis (2018):

$$OP_i = \frac{1}{m} \sum \lambda_1^* f_1^* \quad [47]$$

$$RMSD_i = \sqrt{\frac{1}{m} \sum_{j=1}^m \epsilon_{ij}^2} \quad [48]$$

Quanto maior o OP_i e menor o $RMSD_i$, superior é o genótipo. Para verificar a estabilidade inerente das progênes selecionadas, foi calculada a capacidade de resposta a todos os fatores, exceto o primeiro. Esta abordagem foi adotada porque o primeiro fator representa performance, enquanto os subsequentes significam estabilidade (CULLIS et al., 2014).

Para determinar as progênes de melhor desempenho, foram empregados OP_i e $RMSD_i$ dentro de um índice de seleção (ID_i) (CHAVES et al., 2023a; COWLING et al., 2023):

$$ID_i = 3 \times \left(\frac{OP_i - \overline{OP}}{\sqrt{V(OP)}} \right) - \left(\frac{RMSD_i - \overline{RMSD}}{\sqrt{V(RMSD)}} \right) \quad [49]$$

Em adicional ao ID_i , foi determinada a confiabilidade de seleção individual de cada progênie (MRODE, 2014, p. 45):

$$r = \sqrt{1 - \frac{PEV_i}{\bar{\sigma}_g^2}} \quad [50]$$

em que PEV_i é a variância do erro de predição do i -ésimo genótipo, obtida a partir da diagonal da inversa generalizada da matriz de coeficientes das equações de modelo misto (MME) e $\bar{\sigma}_g^2$ é a variância genotípica média entre os ambientes.

3 RESULTADOS

3.1 Seleção do Modelo Fator-Analítico (FA) de Melhor Ajuste

Os modelos fator-analíticos (FA) foram ajustados para análises de múltiplas características e múltiplos ambientes, considerando dois critérios principais: (i) a porcentagem de variância explicada por todos os k fatores ($\% \bar{v}$) e (ii) a razão média das semivariâncias (ASVR). Ambos os critérios foram definidos com um limiar mínimo de 70% para garantir robustez explicativa (Tabela 10). Foram testados modelos de ordem 1 a 10, avaliando-se o logaritmo da função de máxima verossimilhança (LogLik), o Critério de Informação de Akaike (AIC), a $\% \bar{v}$ e o ASVR.

Os valores de AIC variaram entre -6083,00 (FA4) e -5925,91 (FA1), indicando que modelos de maior complexidade (ordem 4 a 10) apresentaram melhor ajuste relativo. Contudo, a seleção do modelo final considerou não apenas o AIC, mas também a capacidade explicativa ($\% \bar{v} \geq 70\%$) e o poder preditivo ($ASVR \geq 70\%$). Dessa forma, o modelo FA8 foi escolhido para análises subsequentes, pois atendeu a ambos os critérios, explicando 78,44% da variância total e apresentando um ASVR de 74,49% (Tabela 10).

Tabela 10: Logaritmo da função de máxima verossimilhança (LogLik), Critério de Informação de Akaike (AIC), porcentagem de variância explicada por todos os k fatores ($\% \bar{v}$) e razão média das semivariâncias (ASVR) para os 10 modelos fator-analíticos ajustados.

Modelo	Parâmetro			
	LogLik	AIC	$\% \bar{v}$	ASVR
FA1	2989,00	-5925,91	17,04	11,32
FA2	3045,52	-6017,05	41,08	31,28
FA3	3063,52	-6031,04	48,03	39,18
FA4	3098,50	-6083,00	57,00	48,96
FA5	3104,49	-6079,00	64,26	57,77
FA6	3109,01	-6074,03	69,00	63,13
FA7	3111,32	-6062,64	71,41	66,14
FA8	3112,29	-6058,58	78,44	74,49
FA9	3112,49	-6051,00	81,34	77,95
FA10	3112,61	-6045,22	87,78	85,53

O modelo FA8 demonstrou variabilidade explicativa distinta entre as combinações característica-ambiente. Para a característica ARQ, a porcentagem de variância explicada foi de 41,08% na safra Águas/2018, 48,03% na safra

Águas/2020 e 57,00% na safra Seca/2019. Para ASP, os valores foram 64,26% na safra Águas/2018, 69,00% na safra Águas/2019, 71,41% na safra Águas/2020, 78,44% na safra Inverno/2020 e 81,34% na safra Seca/2019. Já para PROD, observou-se 87,78% de variância explicada na safra Águas/2018, 17,04% na safra Águas/2019, 41,08% na safra Águas/2020, 48,03% na safra Inverno/2020 e 57,00% na safra Seca/2018. Nota-se que algumas combinações característica-ambiente apresentaram baixa variância explicada, como PROD na safra Águas/2019 (17,04%), indicando que fatores não capturados pelo modelo podem influenciar significativamente essas combinações. Essa limitação está associada à porcentagem de variância residual inexplicável, conforme ilustrado na Figura 5.

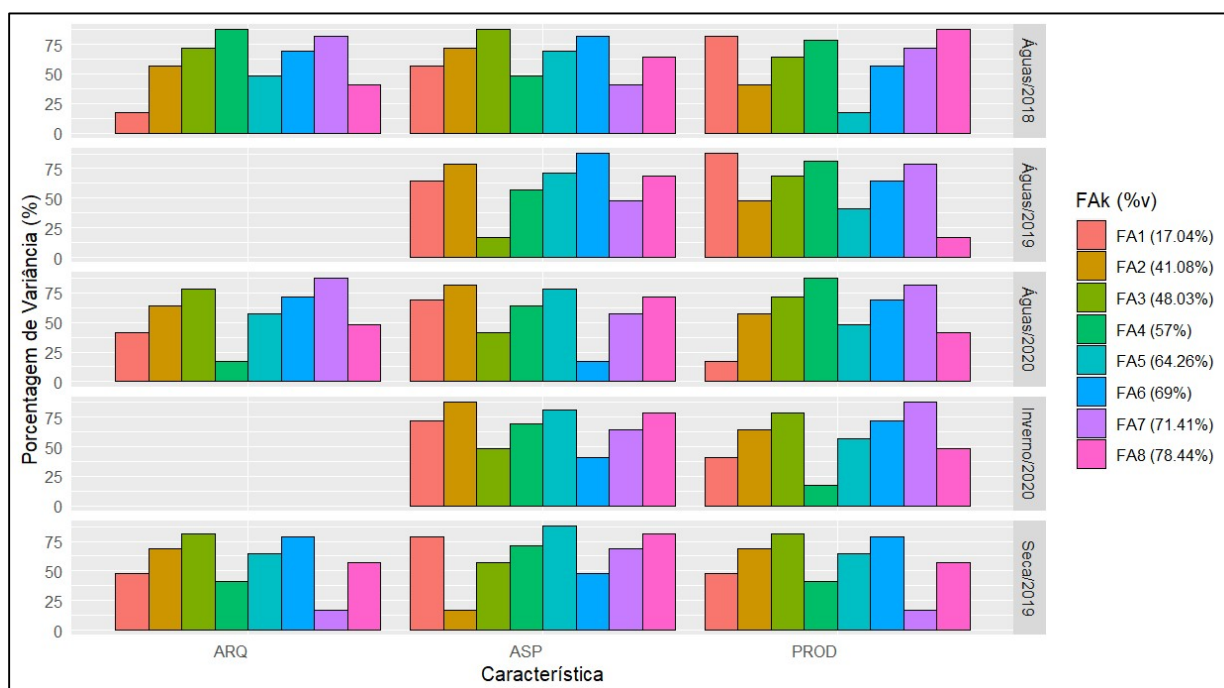


Figura 5. Porcentagem de variação explicada acumulada por cada modelo (entre parênteses) e por cada fator em cada combinação característica-ambiente.

3.2 Estimativas dos parâmetros genéticos e não genéticos

Considerando os resultados do modelo selecionado, as estimativas de herdabilidade generalizada (H^2) e coeficiente de variação experimental (CV) para as combinações característica-safra apresentaram variações significativas (Figura 6). Para a arquitetura de planta (ARQ), os valores de H^2 variaram de 0,11 (águas/2020) a 0,53 (seca/2019), com CV entre 18,79% (águas/2018) e 35,50% (águas/2020), sendo este último o maior CV observado no estudo. Para o aspecto comercial de grão (ASP), a herdabilidade alcançou seu pico na safra Águas/2018 (0,79),

associada ao menor CV (5,99%), enquanto os menores valores de H^2 foram registrados nas safras Inverno/2020 (0,10) e Águas/2020 (0,10), com CV em torno de 20-21%.

Na produtividade de grãos (PROD), as herdabilidades variaram de 0,22 (inverno/2020) a 0,58 (águas/2018), com CV entre 16,08% (seca/2019) e 21,79% (águas/2020). Destaca-se que a combinação aspecto de grão nas águas/2018 apresentou a maior herdabilidade (0,79) e o menor CV (5,99%), indicando alta precisão experimental e forte controle genético para essa característica-ambiente. Em contraste, a arquitetura de planta nas águas/2020 exibiu a menor herdabilidade (0,11) e o maior CV (35,50%), sugerindo maior influência de fatores ambientais ou interações genótipo-ambiente não capturadas pelo modelo. A maioria das combinações característica-safra exibiu herdabilidades moderadas a baixas ($H^2 < 0,60$) e CV experimentais entre 15% e 25%, exceto para os casos extremos mencionados.

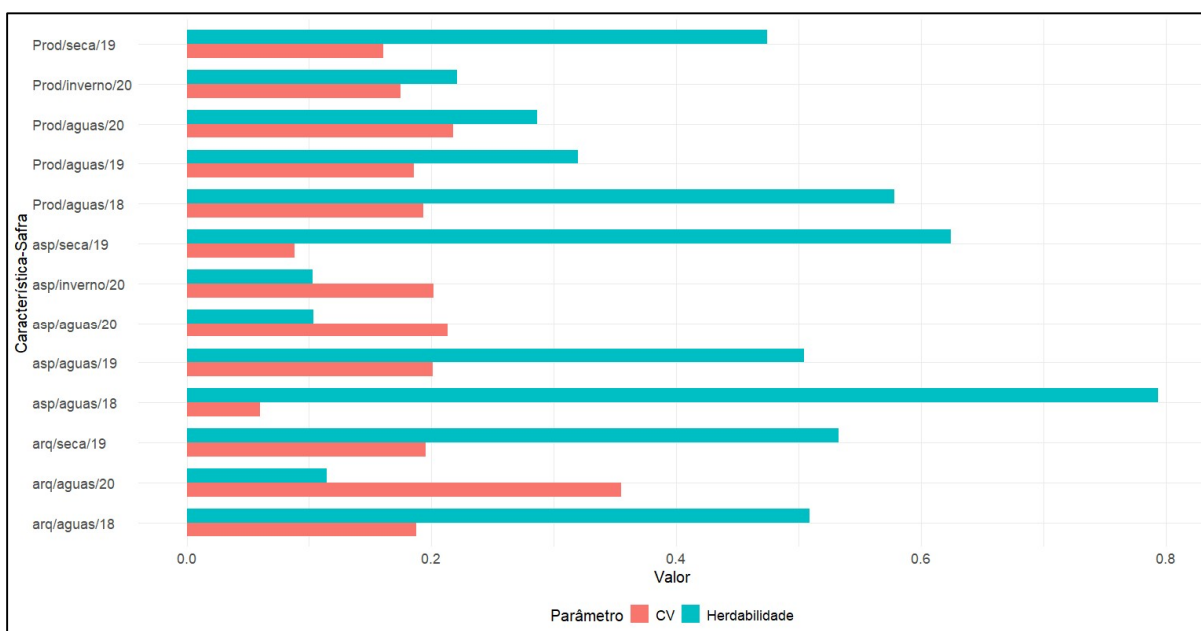


Figura 6. Gráfico de barras horizontais comparando os valores de Herdabilidade (barras azuis) e Coeficiente de Variação (CV) (barras laranjas) para diferentes combinações de características e safras. No eixo X, são apresentados os valores numéricos de Herdabilidade e CV, enquanto no eixo Y são listadas as combinações de Característica-Safra

As correlações genéticas entre as combinações características-ambiente variaram de -0,29 a 0,67, com média geral de 0,14 (Figura 7). A maior correlação positiva (0,67) foi observada entre aspecto comercial de grãos nas safras

Águas/2018 e Águas/2019, enquanto a mais negativa (-0,29) ocorreu entre produtividade de grãos na safra Águas/2020 e aspecto comercial de grãos na safra Inverno/2020. Para a arquitetura de planta (ARQ), a correlação mais elevada (0,67) foi registrada entre as safras Seca/2019 e Águas/2018. Correlações negativas foram observadas entre arquitetura de planta nas safras Águas/2020 e Inverno/2020 (-0,16) e entre arquitetura de planta nas safras Águas/2020 e Seca/2019 (-0,16).

Na produtividade de grãos (PROD), as correlações positivas mais altas ocorreram entre as safras Seca/2019 e Águas/2019 (0,65) e entre as safras Águas/2018 e Águas/2020 (0,52). Uma correlação negativa (-0,19) foi verificada entre produtividade de grãos na safra Águas/2020 e arquitetura de planta na safra Águas/2020. Para o aspecto comercial de grão (ASP), a safra Águas/2018 apresentou correlação positiva com produtividade de grãos na safra Águas/2018 (0,57) e correlação negativa com aspecto comercial de grão na safra Águas/2020 (-0,14). aspecto comercial de grão na safra Seca/2019 correlacionou-se moderadamente com aspecto comercial de grão na safra Águas/2018 (0,59) e ASP/Águas/2019 (0,60).

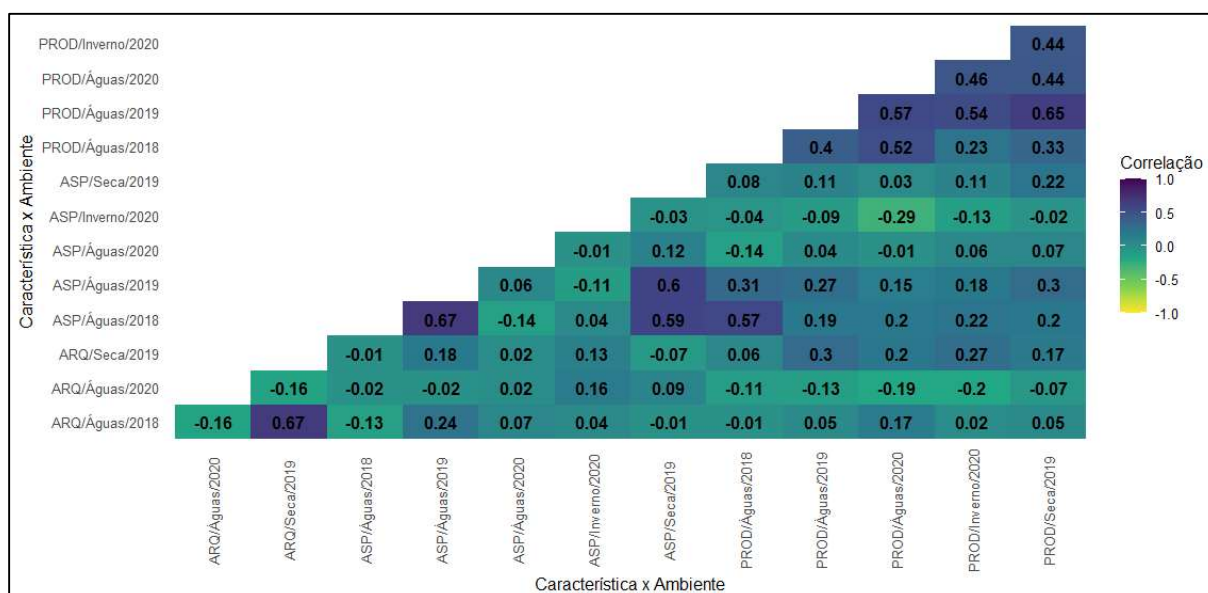


Figura 7. Heatmap das correlações genéticas entre combinações característica-ambiente. ARQ: arquitetura de planta; ASP: aspecto comercial de grão; PROD: produtividade de grãos.

3.3 Factor Analytic Selection Tools

Foram identificadas 20 progênies mais produtivas e estáveis com base nas análises fator-analíticas, utilizando um índice que combinou Performance Geral (OP) e Raiz do Quadrado Médio dos Desvios (RMSD) (Figura 8). As testemunhas foram excluídas do processo de seleção. O índice OP classificou as progênies de acordo com o desempenho em múltiplas características e ambientes, com valores mais altos indicando melhor performance. O RMSD avaliou a estabilidade geral, sendo valores mais baixos associados a maior consistência.

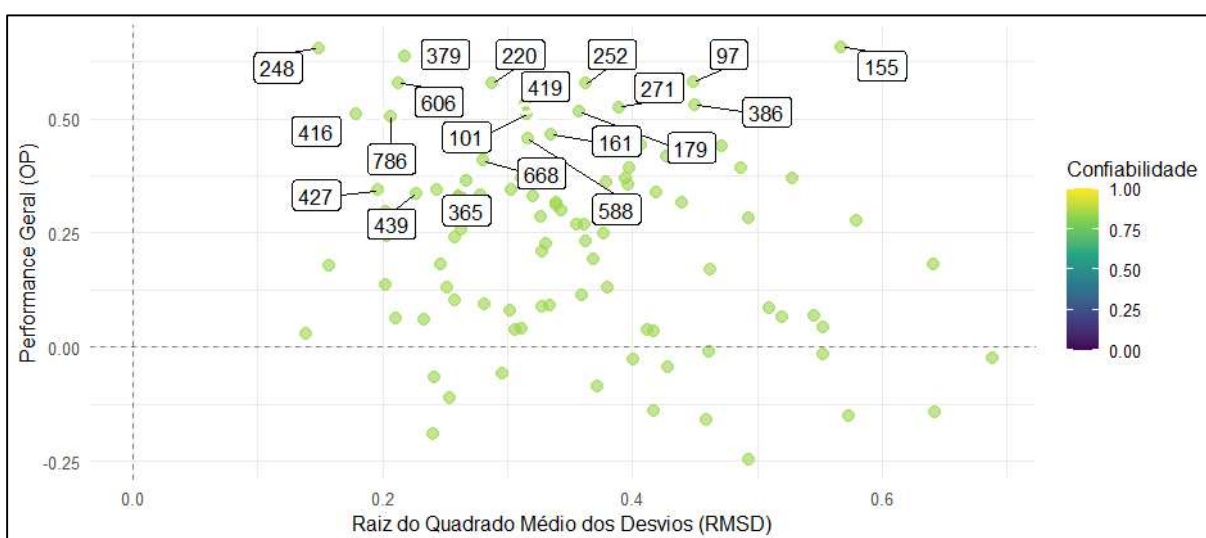


Figura 8. Performance Geral (OP, eixo y) e estabilidade (RMSD, eixo x) das 95 progênies de feijão carioca, com destaque para as 20 selecionadas. A barra de cores indica a confiabilidade da seleção.

A confiabilidade da seleção (Figura 8) está intrinsecamente ligada à quantidade de informações utilizadas nas análises. As 20 progênies selecionadas (248, 379, 606, 416, 220, 786, 252, 419, 101, 97, 155, 179, 271, 588, 161, 386, 668, 427, 365 e 439) estiveram presentes em todas as cinco safras avaliadas, resultando em valores de confiabilidade superiores a 0,80.

4 DISCUSSÃO

Este estudo propõe uma abordagem de avaliação e seleção de progênies de feijão comum em dois estágios, integrando informações de múltiplas características e ambientes no segundo estágio por meio de uma estrutura fator-analítica para a matriz de variância-covariância genética e aplicação da metodologia FAST (*Factor Analytic Selection Tools*). A combinação dessas estratégias permite o uso eficiente da interação genótipo-ambiente-característica (GETI), otimizando predições em características e ambientes correlacionados (BANČIČ et al., 2023). A metodologia em dois estágios mostra-se útil em cenários onde os indivíduos são avaliados para todas as características em um ambiente, mas apenas um subconjunto em outros, ou quando características específicas são mensuradas em grupos distintos de indivíduos entre ambientes. Além disso, a abordagem facilita a análise de conjuntos de dados complexos, compostos por experimentos com delineamentos variados, desbalanceamento genético ou estatístico, e particularidades como *outliers* ou unidades de medida heterogêneas.

A principal vantagem desta abordagem reside em sua flexibilidade operacional, permitindo análises por ambiente e característica de forma isolada no primeiro estágio, seguida de integração via modelo fator-analítico no segundo estágio. Essa estratégia reduz a dimensionalidade dos dados e identifica padrões latentes, simplificando a complexidade multivariada (SMITH et al., 2021). Adicionalmente, ganhos computacionais significativos foram observados, com reduções no tempo de processamento em comparação a análises de estágio único (embora não apresentado), devido à menor quantidade de parâmetros estimados em cada modelo parcial e à convergência mais rápida de algoritmos. Essa eficiência é crucial em programas de melhoramento de plantas de larga escala, onde recursos computacionais são frequentemente limitantes (PIEPHO et al., 2012).

4.1 Análise de dois estágios dos dados

A análise de dois estágios para dados de melhoramento de plantas adota uma abordagem sequencial, na qual os efeitos genotípicos são tratados como fixos no primeiro estágio, conforme proposto por Möhring e Piepho (2009). Essa escolha visa evitar o duplo *shrinkage* (subestimação sistemática de efeitos) que ocorre quando o BLUP (*Best Linear Unbiased Predictor*) é aplicado em ambos os estágios, conforme alertado por Smith et al. (2001). No segundo estágio, a transição para o

uso do BLUP é estratégica, pois este método supera o BLUE (*Best Linear Unbiased Estimator*) em precisão preditiva, especialmente em cenários com desbalanceamento de dados ou interações genótipo-ambiente complexas (PIEPHO et al., 2008).

A recomendação de tratar efeitos genotípicos como fixos em estágios intermediários e reservar o BLUP para o estágio final foi estendida por Piepho *et al.* (2012) a modelos com múltiplos estágios, garantindo que a penalização excessiva de efeitos (*shrinkage*) seja minimizada enquanto se mantém a robustez estatística. Essa abordagem é particularmente vantajosa em programas de melhoramento de plantas, onde a precisão na seleção de genótipos superiores é crítica, e a eficiência computacional é necessária para lidar com grandes volumes de dados de múltiplos ambientes e características. A sinergia entre a fixação de efeitos em estágios preliminares e a predição via BLUP no estágio final pode equilibrar a acurácia estatística com a praticidade operacional, sugerindo como padrão metodológico em análises multivariadas em programas de melhoramento de plantas.

4.2 Ajuste dos modelos FA para a interação genótipo x ambiente x característica (GETI)

No presente estudo, modelos fator-analíticos (FA) com número variável de fatores (FA1 a FA10) foram comparados para analisar a interação genótipo x ambiente x característica (GETI). Embora estruturas não estruturadas de matrizes de covariância genética sejam teoricamente possíveis, sua aplicação prática pode enfrentar desafios computacionais, especialmente em conjuntos de dados extensos, devido ao alto número de parâmetros estimados (KELLY et al., 2007). Nesse contexto, modelos fator-analíticos podem oferecer uma alternativa mais parcimoniosa, reduzindo a complexidade paramétrica para $\frac{t(t+1)}{2}$ quando k fatores são ajustados (SMITH; CULLIS; THOMPSON, 2001). A literatura sugere que essa abordagem pode ser adequada para análises multivariadas, como multiambientes e multicaracterísticas e em vários contextos experimentais (BAKARE et al., 2022; HUNT et al., 2020).

O número ideal de fatores pode variar conforme o número de ambientes (t) e critérios estatísticos. Smith *et al.* (2001b) propõem um limite teórico dado por $k \leq \left\lceil 2t + 1 - \sqrt{(8t + 1)} \right\rceil$, mas estudos empíricos divergem: alguns apontam dois

fatores como suficientes (KELLY et al., 2007; PIEPHO, 1998), enquanto outros defendem até quatro fatores (Studnicki et al., 2016). Essa variação pode estar relacionada a diferenças nos tamanhos das amostras ou na heterogeneidade dos dados. Por exemplo, Studnicki et al. (2016) argumentam que conjuntos de dados maiores podem exigir mais fatores para capturar padrões latentes, ao contrário de estudos com amostras menores. Smith et al. (2015), ao selecionarem o modelo FA5 como ótimo, sugerem a possibilidade de ajustes com mais fatores, mas destacam que limitações computacionais, como problemas de convergência ou conectividade dos dados, podem restringir essa escolha.

Em contraste, Zhang et al. (2020) não relataram dificuldades ao ajustar modelos FA1 a FA17 em dados com 22 ambientes, indicando que a viabilidade pode depender de fatores como algoritmos de otimização e qualidade dos dados. As discrepâncias entre estudos podem decorrer não apenas de diferenças metodológicas, mas também de critérios de avaliação divergentes, como a porcentagem de variância explicada ou a razão média das semivariâncias (CHAVES et al., 2023a; PIEPHO, 2019). Assim, embora o número de fatores não deva ultrapassar o número de ambientes, a definição do modelo ideal pode exigir equilíbrio entre critérios estatísticos, capacidade computacional e objetivos específicos do estudo, como adotado no presente trabalho.

4.3 Seleção do modelo FA de melhor ajuste

A seleção do modelo fator-analítico (FA) de melhor ajuste requer critérios que excedam o uso isolado do AIC (Critério de Informação de Akaike). Conforme destacado por Chaves et al. (2023), priorizar modelos que equilibrem parcimônia, eficiência computacional e capacidade de explicar variação substancial dos dados constitui uma abordagem metodologicamente robusta. Para tanto, recomenda-se integrar ao AIC métricas complementares, como a porcentagem de variância total explicada pelos fatores (ISIK; HOLLAND; MALTECCA, 2017; SMITH et al., 2015; ZHANG; HAN; HU, 2020) e a razão média das semivariâncias (ASVR, *Average Semivariance Ratio*), parâmetro que quantifica o poder explicativo do modelo frente à variabilidade residual.

No presente estudo, embora o AIC indicasse otimalidade para o modelo FA4, optou-se pelo FA8 com base em critérios multidimensionais. Este modelo explicou 78% da variância total (superior ao limiar de 70% recomendado por Isik et

al., 2017) e registrou ASVR de 74%, indicando maior capacidade de capturar padrões estruturais subjacentes aos dados. A preferência por esse arcabouço alinha-se a práticas consolidadas em estudos análogos (ARAÚJO et al., 2024; CHAVES et al., 2023b; FELDMANN et al., 2021), nos quais a combinação de variância explicada e métricas de semivariância demonstra superioridade frente a critérios baseados exclusivamente em ajuste estatístico.

4.4 Estimativas dos parâmetros genéticos e não genéticos

Devido à natureza específica de cada população, as estimativas de herdabilidade podem variar entre estudos, refletindo a diversidade genética e as condições experimentais únicas de cada conjunto de dados. Portanto, não existe uma norma universal para qualificar esse parâmetro, e a herdabilidade não deve ser considerada como o único indicador de um cenário favorável para a seleção. Por exemplo, em um estudo conduzido por Santana et al. (2023), a estimativa de herdabilidade para a característica ASP na safra das Águas/2019 foi de 0,000017, um valor significativamente menor do que o encontrado no presente estudo. Essa disparidade pode ser atribuída a diferenças nas populações avaliadas, bem como em outras características específicas dos experimentos. Ainda neste exemplo, os autores avaliaram feijão-preto em uma população maior e em gerações mais avançadas, entre outras distinções experimentais que podem influenciar as estimativas de herdabilidade.

É possível observar que existe uma relação entre herdabilidade e o coeficiente de variação experimental (Figura 6), todavia, tal relação pode variar dependendo das características específicas estudadas e das condições ambientais nas quais elas são computadas. Em geral, no entanto, tende a haver uma relação inversa entre herdabilidade e o coeficiente de variação experimental. Além disso, a correlação entre herdabilidade e o coeficiente de variação experimental pode variar dependendo da população estudada e dos métodos estatísticos utilizados para estimar esses parâmetros.

A análise das correlações genéticas revela que cerca de 68% delas apresentaram valores positivos, embora nenhuma tenha excedido 0,67. Conforme ilustrado na Figura 7, as estimativas obtidas por meio do Modelo FA8, que avalia as interações entre características e ambientes, confirmam a presença de relações

complexas entre genótipo, ambiente e características. Esses resultados reforçam a ocorrência de interações multifatoriais que influenciam a expressão fenotípica, destacando a importância de considerar esses elementos de forma integrada em estudos de melhoramento de plantas.

A forte correlação positiva (0,67) entre ARQ/Seca2019 e ARQ/Águas2018 indica que a arquitetura de planta apresenta estabilidade genética em safras de estações distintas (seca e águas). Por outro lado, as correlações negativas entre ARQ/Seca2019 e ARQ/Águas2020 (-0,16) sugerem que condições ambientais atípicas durante a safra de Águas/2020.

No caso do ASP (Aspecto Comercial de Grãos), a alta correlação positiva (0,67) entre ASP/Águas2018 e ASP/Águas2019 revela uma estabilidade genética nas condições de cultivo dessas safras, facilitando a identificação e seleção de genótipos com qualidade estável. Contudo, a correlação negativa (-0,14) entre ASP/Águas2020 e ASP/Águas2018 também sugere que condições ambientais adversas em 2020.

De fato, os experimentos da safra Águas/2019, cujas plantas foram cultivadas em outubro e colhidas em janeiro de 2020, enfrentaram condições climáticas desfavoráveis, com chuvas intensas que resultaram em perda de qualidade dos grãos e redução de produtividade devido à perda de várias parcelas. Esses efeitos são evidenciados pelas médias das características avaliadas na Tabela 9. As sementes colhidas em janeiro de 2020 foram utilizadas para plantar os experimentos da safra Águas/2020, o que pode ter contribuído para os resultados observados nessa safra.

Em relação à produtividade, a única característica que se manteve estável em todas as safras, as correlações genéticas observadas ao longo das safras mostram que todas as associações são positivas, variando de 0,23 a 0,65. Isso indica uma estabilidade genética significativa, mesmo diante de variações ambientais. Os genótipos que se destacaram em uma safra também apresentaram bom desempenho nas demais, sugerindo que a interação genótipo-ambiente ($G \times A$) é predominantemente influenciada por sua fração simples. Esse padrão reforça a possibilidade de selecionar genótipos com adaptação ampla, capazes de manter produtividade consistente em diferentes condições.

As correlações negativas recorrentes (por exemplo, PROD/Águas2020 vs. ASP/Inverno2020: -0,29) confirmam que a safra de 2020 foi marcada por condições

ambientais extremas, como precipitação excessiva. Além disso, as baixas correlações genéticas observadas (como ASP/Águas2020 vs. ASP/Águas2018: -0,14) reforçam a instabilidade genética nesse ambiente específico. Por outro lado, as correlações positivas consistentes (por exemplo, PROD/Seca2019 vs. PROD/Águas2019: 0,65) sugerem que esse ambiente é propício para a seleção de genótipos com adaptabilidade a diversas condições, destacando sua relevância para programas de melhoramento de plantas. Adicionalmente, as correlações entre as características podem ser atribuídas à interação entre genes ou ao pleiotropismo (DWIVEDI et al., 2024).

Além disso é possível observar a influência da safra e do ano em que os ensaios foram realizados, por exemplo, ARQ na safra das Águas no ano de 2018 apresentou maior correlação com ARQ na safra da Seca em 2019, enquanto na safra das Águas em 2020 e na safra da Seca em 2019, ARQ apresentou valor negativo. É possível deduzir que as diferenças ambientais foram um ponto chave para tal resultado. Uma análise posterior pode ser realizada incluindo covariáveis ambientais como temperatura, precipitação, radiação solar, humidade, entre outras, para entender melhor as interações entre os ambientes, características e genótipos (OLIVEIRA et al., 2020; SAE-LIM et al., 2014).

4.5 Factor Analytic Selection Tools

A seleção de genótipos para adaptação ampla ou específica representa uma decisão importante no melhoramento de plantas, especialmente na presença da interação GxA. Tradicionalmente, estas decisões de seleção baseiam-se em alguma medida de desempenho geral e estabilidade para cada genótipo. No modelo de regressão conjunta, essas medidas correspondem ao intercepto e à inclinação das respostas dos genótipos individuais, enquanto Kempton (1984) utilizou biplots para visualizar padrões de interação GxA a partir do modelo de efeitos principais aditivos e interação multiplicativa (AMMI). No entanto, estas abordagens apresentam várias deficiências importantes, particularmente a última exigindo dados equilibrados (ou seja, todos os genótipos avaliados para todas as características em todos os ambientes), enquanto ambos os métodos assumem efeitos principais de genótipo fixo e muitas vezes não conseguem fornecer um bom ajuste aos dados (SMITH; CULLIS, 2018).

Cullis (2018) sugeriu empregar as FAST para melhorar a interpretação dos resultados de modelos mistos fator-analíticos (FA). Ao resumir os resultados em três estimativas, os melhoristas obtêm informações valiosas que podem ser utilizadas na construção de índices de seleção ou biplots (figura 8). Esta abordagem proporciona aos melhoristas a flexibilidade de escolher entre diferentes critérios para fins de seleção, oferecendo medidas de performance geral (OP) e estabilidade geral (RMSD), em todos os ambientes amostrados, concisa e significativa.

Em uma analogia ao modelo de simetria composta, onde os efeitos aditivos de GxA são divididos em efeitos principais e de interação GxA, esta estrutura se assemelha a uma forma restrita do modelo FA1 em que todas as cargas são iguais e determinadas pela raiz quadrada da variância do efeito principal de genótipo (SMITH; CULLIS, 2018). Além disso, as variações específicas são uniformes entre os ambientes e ditadas pelo componente de variação da interação GxA. Consequentemente, se o FAST fosse aplicado ao modelo de simetria composta, o OP para um genótipo espelharia a predição do efeito principal aditivo (TOLHURST et al., 2019).

5 CONCLUSÕES

A análise em dois estágios mostrou-se eficaz para diferentes delineamentos experimentais, reduzindo vieses e melhorando a eficiência computacional. Além disso, permite a integração de múltiplos ambientes, facilitando a análise de interações G×E.

O modelo FA8 mostrou-se ideal (78,4% variância explicada), capturando eficientemente as interações G×E. As herdabilidades variaram de 0,10 a 0,79 (maior para aspecto comercial), evidenciando forte influência ambiental. Correlações genéticas (-0,29 a 0,67) indicaram respostas genotípicas distintas entre safras, reforçando a necessidade de avaliação multiambiental.

As ferramentas FAST otimizaram a seleção das 20 melhores progêies (confiabilidade >0,80) através do índice OP×RMSD e biplots, transformando dados complexos em critérios objetivos. A modelagem GETI revelou padrões cruciais nas interações G×E×características, permitindo uma avaliação integrada que identifica genótipos adaptáveis a diferentes condições ambientais.

A integração da análise em dois estágios com modelagem fator-analítica e ferramentas multivariadas (FAST/GETI) cria um sistema eficaz para o melhoramento de plantas, unindo: (1) precisão estatística no ajuste de dados multiambientes e multicaracterísticas e (2) eficiência operacional via critérios objetivos de seleção. Esse framework otimiza desde a avaliação experimental até a recomendação de cultivares superiores e adaptáveis, elevando a eficácia global do programa.

REFERÊNCIAS

- ACHARYA, J. P. et al. Breeding Alfalfa (*Medicago sativa* L.) Adapted to Subtropical Agroecosystems. **Agronomy**, v. 10, n. 5, p. 742, 2020.
- AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, v. 19, n. 6, p. 716–723, 1974.
- ALVES, A. F. et al. Genetic progress and potential of common bean families obtained by recurrent selection. **Crop Breeding and Applied Biotechnology**, v. 15, n. 4, p. 218–226, 2015.
- ARAÚJO, M. DOS S. et al. Modeling covariance structures for genetic and non-genetic effects in cowpea multi-environment trials. **Agronomy Journal**, v. 115, n. 3, p. 1248–1256, 2023.
- ARAÚJO, M. S. et al. GIS-FA: an approach to integrating thematic maps, factor-analytic, and envirotyping for cultivar targeting. **Theoretical and Applied Genetics**, v. 137, n. 4, p. 80, 2024.
- ARIEF, V. N. et al. Utilization of Multiyear Plant Breeding Data to Better Predict Genotype Performance. **Crop Science**, v. 59, n. 2, p. 480–490, 2019.
- BAKARE, M. A. et al. Parsimonious genotype by environment interaction covariance models for cassava (*Manihot esculenta*). **Frontiers in Plant Science**, v. 13, 2022.
- BANČIČ, J. et al. Genomic selection for genotype performance and stability using information on multiple traits and multiple environments. **Theoretical and Applied Genetics**, v. 136, n. 5, p. 104, 2023.
- BARILI, L. D. **Evolução dos cultivares de feijão carioca recomendados no Brasil**. Tese (Doutorado em Genética e Melhoramento)—Viçosa: Universidade Federal de Viçosa, 2015.
- BERNARDO, R. Recurrent Selection. Em: BERNARDO, R. (Ed.). **Breeding for Quantitative Traits in Plants**. 3. ed. Woodbury: Stemma Press, 2020a. p. 299–318.
- BERNARDO, R. Reinventing quantitative genetics for plant breeding: something old, something new, something borrowed, something BLUE. **Heredity**, v. 125, n. 6, p. 375–385, 2020b.
- BHERING, L. L. et al. Genetic gains in physic nut using selection indexes. **Pesquisa Agropecuária Brasileira**, v. 47, n. 3, p. 402–408, 2012.
- BRITO, G. M. DE; CARNEIRO, P. C. S.; SILVA, M. B. P. DA. **Avaliação e seleção de progênies visando o melhoramento do feijão carioca por seleção recorrente**. Simpósio de Integração Acadêmica - SIA 2021. **Anais...**Viçosa: Editora UFV, 2021.

BURNHAM, K. P.; ANDERSON, D. R. Practical Use of the Information-Theoretic Approach. Em: **Model Selection and Inference**. New York, NY: Springer New York, 1998. p. 75–117.

BUTLER, D. G. et al. **ASReml-R Reference Manual Version 4.2**. Hemel HempsteadVSN International Ltd., , 2023.

CARNEIRO, J. E.; DE PAULA JÚNIOR, T.; BORÉM, A. **Feijão: do plantio à colheita**. 1. ed. Viçosa: Editora UFV, 2015.

CASLER, M. D. Fundamentals of Experimental Design: Guidelines for Designing Successful Experiments. **Agronomy Journal**, v. 107, n. 2, p. 692–705, 2015.

CERÓN-ROJAS, J. J. et al. A Selection Index Method Based on Eigenanalysis. **Crop Science**, v. 46, n. 4, p. 1711–1721, 2006.

CHAVES, S. F. S. et al. Employing factor analytic tools for selecting high-performance and stable tropical maize hybrids. **Crop Science**, v. 63, n. 3, p. 1114–1125, 2023a.

CHAVES, S. F. S. et al. Analysis of repeated measures data through mixed models: An application in Theobroma grandiflorum breeding. **Crop Science**, v. 63, n. 4, p. 2131–2144, 2023b.

CHRISTENSEN, W. **Agreeing to disagree: Using SAS to make reasoned decisions when information criteria select different models**. SAS Conference Proceedings: Western Users of SAS Software 2018. **Anais...**Sacramento: Paper 099–2018, 2018.

COCHRAN, W. G. Some Consequences When the Assumptions for the Analysis of Variance are not Satisfied. **Biometrics**, v. 3, n. 1, p. 22, 1947.

COLLICCHIO, E.; RAMALHO, M. A. P.; ABREU, A. F. B. Association between plant architecture of the bean plant and seed size. (In Portuguese, with English abstract.). **Pesqui. Agropecu. Bras.**, v. 32, n. 3, p. 297–304, 1997.

COWLING, W. A. et al. Optimal Contribution Selection Improves the Rate of Genetic Gain in Grain Yield and Yield Stability in Spring Canola in Australia and Canada. **Plants**, v. 12, n. 2, p. 383, 2023.

CROSSA, J. Statistical Analyses of Multilocation Trials. Em: BRADY, N. C. (Ed.). **Advances in Agronomy**. Cambridge: Academic Press, 1990. v. 44p. 55–85.

CROSSA, J. From Genotype × Environment Interaction to Gene × Environment Interaction. **Current Genomics**, v. 13, n. 3, p. 225–244, 2012.

CRUZ, C. D.; CARNEIRO, P. C. S.; REGAZZI, A. J. Seleção simultânea de caracteres. Em: CRUZ, C. D.; CARNEIRO, P. C. S.; REGAZZI, A. J. (Eds.). **Modelos biométricos aplicados ao melhoramento genético**. 3. ed. Viçosa: Editora UFV, 2014. v. 2p. 221–294.

CULLIS, B. R. et al. Analysis of yield and oil from a series of canola breeding trials. Part II. Exploring variety by environment interaction using factor analysis. **Genome**, v. 53, n. 11, p. 1002–1016, 2010.

CULLIS, B. R. et al. Factor analytic and reduced animal models for the investigation of additive genotype-by-environment interaction in outcrossing plant species with application to a *Pinus radiata* breeding programme. **Theoretical and Applied Genetics**, v. 127, n. 10, p. 2193–2210, 2014.

CULLIS, B. R.; SMITH, A. B.; COOMBES, N. E. On the design of early generation variety trials with correlated data. **Journal of Agricultural, Biological, and Environmental Statistics**, v. 11, n. 4, p. 381–393, 2006.

DIAS, K. O. G. et al. Leveraging probability concepts for cultivar recommendation in multi-environment trials. **Theoretical and Applied Genetics**, v. 135, n. 4, p. 1385–1399, 2022.

DOS ANJOS, R. S. R. et al. New Proposals to Estimate Unbiased Selection Gain and Coefficient of Variation in Traits Evaluated Using Score Scales. **Crop Science**, v. 59, n. 3, p. 937–944, 2019.

DWIVEDI, S. L. et al. Epistasis and pleiotropy-induced variation for plant breeding. **Plant Biotechnology Journal**, v. 22, n. 10, p. 2788–2807, 2024.

EBERHART, S. A.; RUSSELL, W. A. Stability Parameters for Comparing Varieties. **Crop Science**, v. 6, n. 1, p. 36–40, 1966.

EVANGELISTA, J. S. P. C. et al. Modeling covariance structures and optimizing *Jatropha curcas* breeding. **Tree Genetics & Genomes**, v. 19, n. 2, p. 21, 2023.

FEDERER, W. T. Augmented (or hoonuiaku) designs. **Biometrics Unit Technical Reports**, p. 1–33, 1956.

FELDMANN, M. J. et al. Average semivariance yields accurate estimates of the fraction of marker-associated genetic variance and heritability in complex trait analyses. **PLOS Genetics**, v. 17, n. 8, p. e1009762, 2021.

FINLAY, K. W.; WILKINSON, G. N. The analysis of adaptation in a plant-breeding programme. **Australian Journal of Agricultural Research**, v. 14, n. 6, p. 742–754, 1963.

FOUILLLOUX, G.; BANNEROT, H. Selection methods in the common bean (*Phaseolus vulgaris*). Em: GEPTS, P. (Ed.). **Genetic Resources of Phaseolus Beans. Current Plant Science and Biotechnology in Agriculture**. 1. ed. Dordrecht: Springer, 1988. v. 6p. 503–542.

GABRIEL, K. R. The Biplot Graphic Display of Matrices with Application to Principal Component Analysis. **Biometrika**, v. 58, n. 3, p. 453–467, 1971.

GAUCH, H. G. Model Selection and Validation for Yield Trials with Interaction. **Biometrics**, v. 44, n. 3, p. 705–715, 1988.

GU, Z.; EILS, R.; SCHLESNER, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. **Bioinformatics**, v. 32, n. 18, p. 2847–2849, 2016.

HARTUNG, J.; PIEPHO, H. Effect of missing values in multi-environment trials on variance component estimates. **Crop Science**, v. 61, n. 6, p. 4087–4097, 2021.

HAZEL, L. N. The genetic basis for constructing selection indexes. **Genetics**, v. 28, n. 6, p. 476–490, 1943.

HENDERSON, C. R. Best Linear Unbiased Estimation and Prediction under a Selection Model. **Biometrics**, v. 31, n. 2, p. 423, jun. 1975.

HERNÁNDEZ, M. V. et al. Modeling Genotype × Environment Interaction Using a Factor Analytic Model of On-Farm Wheat Trials in the Yaqui Valley of Mexico. **Agronomy Journal**, v. 111, n. 6, p. 2647–2657, 2019.

HILL, R. R.; ROSENBERGER, J. L. Methods for Combining Data from Gemrplasm Evaluation Trials ¹. **Crop Science**, v. 25, n. 3, p. 467–470, 1985.

HUNT, C. H. et al. Multi-environment analysis of sorghum breeding trials using additive and dominance genomic relationships. **Theoretical and Applied Genetics**, v. 133, n. 3, p. 1009–1018, 2020.

ISIK, F.; HOLLAND, J.; MALTECCA, C. Multi Environmental Trials. Em: ISIK, F.; HOLLAND, J.; MALTECCA, C. (Eds.). **Genetic Data Analysis for Plant and Animal Breeding**. Cham: Springer International Publishing, 2017. p. 227–262.

KELLY, A. M. et al. The Accuracy of Varietal Selection Using Factor Analytic Models for Multi-Environment Plant Breeding Trials. **Crop Science**, v. 47, n. 3, p. 1063–1070, 2007.

KEMPTON, R. A. The use of biplots in interpreting variety by environment interactions. **The Journal of Agricultural Science**, v. 103, n. 1, p. 123–135, 1984.

LITTLE, R. J. A.; RUBIN, D. B. Missing Data in Experiments. Em: BALDING, D. J. et al. (Eds.). **Statistical Analysis with Missing Data**. 2. ed. New York: John Wiley & Sons, 2002. p. 24–40.

MAICH, R. H. et al. Association between grain yield, grain quality and morpho-physiological traits along ten cycles of recurrent selection in bread wheat (*Triticum aestivum* L.). **Cereal Research Communications**, v. 45, n. 1, p. 146–153, 2017.

MALOSETTI, M.; RIBAUT, J.-M.; VAN EEUWIJK, F. A. The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. **Frontiers in Physiology**, v. 4, 2013.

- MASOOD, M. A. et al. Improvement in Precision of Agricultural field Experiments through Design and analysis. **Pakistan Journal of Life and Social Sciences**, v. 6, n. 2, p. 89–91, 2008.
- MELO, V. L. DE et al. Modeling (co)variance structures for genetic and non-genetic effects in the selection of common bean progenies. **Euphytica**, v. 216, n. 5, p. 77, 18 maio 2020.
- MEYER, K. Factor-analytic models for genotype $\times$ environment type problems and structured covariance matrices. **Genetics Selection Evolution**, v. 41, n. 1, p. 21, 2009.
- MÖHRING, J.; PIEPHO, H. -P. Comparison of Weighting in Two-Stage Analysis of Plant Breeding Trials. **Crop Science**, v. 49, n. 6, p. 1977–1988, 2009.
- MORAIS JÚNIOR, O. P. et al. Effectiveness of Recurrent Selection in Irrigated Rice Breeding. **Crop Science**, v. 57, n. 6, p. 3043–3058, 2017.
- MOURA, L. M. et al. Combining ability of common bean parents in different seasons, locations and generations. **Euphytica**, v. 214, n. 10, p. 181, 25 out. 2018.
- MRODE, R. A. Univariate Models with One Random Effect. Em: **Linear Models for the Prediction of Animal Breeding Values**. 3. ed. Wallingford: Cabi, 2014a. p. 34–60.
- MRODE, R. A. **Linear models for the prediction of animal breeding values**. 3. ed. Wallingford: CABI Publishing, 2014b.
- OGUT, F. et al. Genetic Analysis of Diallel Progeny Test Data Using Factor Analytic Linear Mixed Models. **Forest Science**, v. 60, n. 1, p. 119–127, 2014.
- OLIVEIRA, R. L. DE et al. Precisão experimental em ensaios com a cultura do feijão. **Ciência e Agrotecnologia**, v. 33, n. 1, p. 113–119, 2009.
- OLIVEIRA, I. C. M. et al. Genotype-by-environment interaction and yield stability analysis of biomass sorghum hybrids using factor analytic models and environmental covariates. **Field Crops Research**, v. 257, p. 107929, 2020.
- OLIVOTO, T.; NARDINO, M. MGIDI: toward an effective multivariate selection in biological experiments. **Bioinformatics**, v. 37, n. 10, p. 1383–1389, 2021.
- PATTERSON, H. D.; THOMPSON, R. Recovery of Inter-Block Information when Block Sizes are Unequal. **Biometrika**, v. 58, n. 3, p. 545, 1971.
- PEEL, M. C.; FINLAYSON, B. L.; MCMAHON, T. A. Updated world map of the Köppen-Geiger climate classification. **Hydrology and Earth System Sciences**, v. 11, n. 5, p. 1633–1644, 2007a.

PEEL, M. C.; FINLAYSON, B. L.; MCMAHON, T. A. Updated world map of the Köppen-Geiger climate classification. **Hydrology and Earth System Sciences**, v. 11, n. 5, p. 1633–1644, 11 out. 2007b.

PEIXOTO, M. A. et al. Selection of maize hybrids: an approach with multi-trait, multi-environment, and ideotype-design. **Crop Breeding and Applied Biotechnology**, v. 21, n. 2, 2021.

PIEPHO, H. et al. A stage-wise approach for the analysis of multi-environment trials. **Biometrical Journal**, v. 54, n. 6, p. 844–860, 2012.

PIEPHO, H. A coefficient of determination R^2 for generalized linear mixed models. **Biometrical Journal**, v. 61, n. 4, p. 860–872, 2019.

PIEPHO, H.; MÖHRING, J. Selection in Cultivar Trials—Is It Ignorable? **Crop Science**, v. 46, n. 1, p. 192–201, 2006.

PIEPHO, H. P. et al. BLUP for phenotypic selection in plant breeding and variety testing. **Euphytica**, v. 161, n. 1–2, p. 209–228, 2008.

PIEPHO, H.-P. Analyzing Genotype-Environment Data by Mixed Models with Multiplicative Terms. **Biometrics**, v. 53, n. 2, p. 761, 1997.

PIEPHO, H.-P. Empirical best linear unbiased prediction in cultivar trials using factor-analytic variance-covariance structures. **Theoretical and Applied Genetics**, v. 97, n. 1–2, p. 195–201, 1998a.

PIEPHO, H.-P. Empirical best linear unbiased prediction in cultivar trials using factor-analytic variance-covariance structures. **Theoretical and Applied Genetics**, v. 97, n. 1–2, p. 195–201, 1998b.

PIMENTEL-GOMES, F. **Curso de Estatística Experimental**. 15. ed. Piracicapa: FEALQ, 2009.

POLLAK, E. J.; VAN DER WERF, J.; QUAAS, R. L. Selection Bias and Multiple Trait Evaluation. **Journal of Dairy Science**, v. 67, n. 7, p. 1590–1595, 1984.

POSADAS, L. G. et al. Elite Performance for Grain Yield from Unadapted Exotic Soybean Germplasm in Three Cycles of a Recurrent Selection Experiment. **Crop Science**, v. 54, n. 6, p. 2536–2546, 2014.

PRIYADARSHAN, P. M. Genotype-by-Environment Interactions. Em: **PLANT BREEDING: Classical to Modern**. Singapore: Springer Singapore, 2019. p. 457–472.

R CORE TEAM. **R: A language and environment for statistical computing**. Vienna, Austria R Foundation for Statistical Computing, , 2023.

RAMALHO, M. A. P. et al. **Aplicações da Genética Quantitativa no Melhoramento de Plantas Autógamas**. 1. ed. Lavras: Editora UFLA, 2012.

RAMALHO, M. A. P.; ABREU, A. DE F. B. Obtenção de cultivares. Em: CARNEIRO, J. E.; PAULA JÚNIOR, T. J. DE; BORÉM, A. (Eds.). **Feijão: do plantio a colheita**. Viçosa: Editora UFV, 2015. p. 96–114.

RAMALHO, M. A. P.; ABREU, Â. DE F. B.; DOS SANTOS, J. B. Genetic progress after four cycles of recurrent selection for yield and grain traits in common bean. **Euphytica**, v. 144, n. 1–2, p. 23–29, 2005.

RAMALHO, M. A. P.; CARVALHO, B. L.; NUNES, J. A. R. Perspectives for the Use of Quantitative Genetics in Breeding of Autogamous Plants. **ISRN Genetics**, v. 2013, p. 1–6, 2013.

RAMALHO, M. A. P.; PIROLA, L. H.; ABREU, A. F. B. Alternativas na seleção de plantas de feijoeiro com porte ereto e grão tipo carioca. **Pesquisa Agropecuária Brasileira**, v. 33, n. 12, p. 1989–1994, 1998.

RESENDE, M. D. V. DE; ALVES, R. S. Statistical significance, selection accuracy, and experimental precision in plant breeding. **Crop Breeding and Applied Biotechnology**, v. 22, n. 3, 2022.

RESENDE, M. D. V. DE. **Genética biométrica e estatística no melhoramento de plantas perenes**. [s.l.] Emprapa Informação Tecnológica, 2002.

RESENDE, M. D. V.; DUARTE, J. B. Precisão e controle de qualidade em experimentos de avaliação de cultivares. **Pesquisa Agropecuária Tropical**, v. 37, n. 3, p. 182–194, 16 out. 2007.

ROCHA, J. R. DO A. S. DE C. et al. Selection of Superior Inbred Progenies toward the Common Bean Ideotype. **Agronomy Journal**, v. 111, n. 3, p. 1181–1189, 2019.

ROCHA, J. R. DO A. S. DE C.; MACHADO, J. C.; CARNEIRO, P. C. S. Multitrait index based on factor analysis and ideotype-design: proposal and application on elephant grass breeding for bioenergy. **GCB Bioenergy**, v. 10, n. 1, p. 52–60, 2018.

RODRÍGUEZ-ÁLVAREZ, M. X. et al. Correcting for spatial heterogeneity in plant breeding experiments with P-splines. **Spatial Statistics**, v. 23, p. 52–71, mar. 2018.

SAE-LIM, P. et al. Identifying environmental variables explaining genotype-by-environment interaction for body weight of rainbow trout (*Onchorynchus mykiss*): reaction norm and factor analytic models. **Genetics Selection Evolution**, v. 46, n. 1, p. 16, 2014.

SALVADOR, F. V. et al. Correcting experimental data for spatial trends in a common bean breeding program. **Crop Science**, v. 62, n. 2, p. 825–838, 14 mar. 2022.

SALVADOR, F. V. et al. Accounting for heterogeneous genotypic variance of progenies within populations in autogamous plant breeding: An evaluation in common bean. **Crop Science**, v. 61, n. 6, p. 3947–3961, 2021.

SANTANA, A. S. et al. Analysis of advanced generation multistage field trials data in autogamous plant breeding: An evaluation in common Bean. **Crop Science**, v. 63, n. 3, p. 1221–1233, 2023.

SCHMIDT, P. et al. Estimating Broad-Sense Heritability with Unbalanced Data from Agricultural Cultivar Trials. **Crop Science**, v. 59, n. 2, p. 525–536, 2019.

SCHWARZ, G. Estimating the Dimension of a Model. **The Annals of Statistics**, v. 6, n. 2, 1978.

SILVA, A. C. et al. Selection of common bean parents and segregating populations targeting fusarium wilt resistance and grain yield. **Agronomy Science and Biotechnology**, v. 9, p. 1–20, 2023.

SMITH, A. et al. Plant Variety Selection Using Interaction Classes Derived From Factor Analytic Linear Mixed Models: Models With Independent Variety Effects. **Frontiers in Plant Science**, v. 12, p. 1–17, 2021.

SMITH, A. B. et al. Varietal selection for perennial crops where data relate to multiple harvests from a series of field trials. **Euphytica**, v. 157, n. 1–2, p. 253–266, 2007.

SMITH, A. B. et al. Factor analytic mixed models for the provision of grower information from national crop variety testing programs. **Theoretical and Applied Genetics**, v. 128, n. 1, p. 55–72, 2015.

SMITH, A. B.; CULLIS, B. R. Plant breeding selection tools built on factor analytic mixed models for multi-environment trial data. **Euphytica**, v. 214, n. 8, p. 143, 2018.

SMITH, A. B.; CULLIS, B. R.; THOMPSON, R. The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. **The Journal of Agricultural Science**, v. 143, n. 6, p. 449–462, 2005.

SMITH, A.; CULLIS, B.; GILMOUR, A. Applications: The Analysis of Crop Variety Evaluation Data in Australia. **Australian & New Zealand Journal of Statistics**, v. 43, n. 2, p. 129–145, 2001.

SMITH, A.; CULLIS, B.; THOMPSON, R. Analyzing Variety by Environment Data Using Multiplicative Mixed Models and Adjustments for Spatial Field Trend. **Biometrics**, v. 57, n. 4, p. 1138–1147, dez. 2001.

SMITH, H. F. A discriminant function for plant selection. **Annals of Eugenics**, v. 7, n. 3, p. 240–250, 1936.

SOKAL, R. R.; ROHLF, F. J. **The principles and practice of statistics in biological research**. 3. ed. New York: W. H. Freeman, 1995.

STEPHENS, M. J. et al. Genetic Parameters and Development of a Selection Index for Breeding Red Raspberries for Processing. **Journal of the American Society for Horticultural Science**, v. 137, n. 4, p. 236–242, 2012.

STUDNICKI, M. et al. Yield response of winter wheat cultivars to environments modeled by different variance-covariance structures in linear mixed models. **Spanish Journal of Agricultural Research**, v. 14, n. 2, p. e0703, 2016.

TOLHURST, D. J. et al. Genomic selection in multi-environment plant breeding trials using a factor analytic linear mixed model. **Journal of Animal Breeding and Genetics**, v. 136, n. 4, p. 279–300, 2019.

VERBYLA, A. P. A note on model selection using information criteria for general linear models estimated using REML. **Australian & New Zealand Journal of Statistics**, v. 61, n. 1, p. 39–50, 9 mar. 2019.

VIANA, J. M. S. et al. Relevance of Pedigree, Historical Data, Dominance, and Data Unbalance for Selection Efficiency. **Agronomy Journal**, v. 104, n. 3, p. 722–728, 2012.

WICKHAM, H. **ggplot2: Elegant Graphics for Data Analysis**. New York, NY: Springer-Verlag, 2016.

WILLIAMS, E.; PIEPHO, H.-P.; WHITAKER, D. Augmented p-rep designs. **Biometrical Journal**, v. 53, n. 1, p. 19–27, 2011.

YAN, W. et al. Cultivar Evaluation and Mega-Environment Investigation Based on the GGE Biplot. **Crop Science**, v. 40, n. 3, p. 597–605, 2000.

YATES, F.; COCHRAN, W. G. The analysis of groups of experiments. **The Journal of Agricultural Science**, v. 28, n. 4, p. 556–580, 1938.

ZHANG, R.; HAN, D.; HU, X. Analyzing the performance of corn in China using a factor-analytic variance-covariance structure with multiple factors. **Crop Science**, v. 60, n. 1, p. 190–201, 2020.

ZYSTRO, J.; COLLEY, M.; DAWSON, J. Alternative Experimental Designs for Plant Breeding. Em: **Plant Breeding Reviews**. [s.l.] Wiley, 2018. p. 87–117.