**Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”**

Análise dos modelos AMMI bivariados

Marisol García Peña

Dissertação apresentada para obtenção do título de
Mestre em Agronomia. Área de concentração:
Estatística e Experimentação Agronômica

**Piracicaba
2008**

Marisol García Peña
Estatística

Análise dos modelos AMMI bivariados

Orientador:
Prof. Dr. **CARLOS TADEU DOS SANTOS DIAS**

Dissertação apresentada para obtenção do título de
Mestre em Agronomia. Área de concentração:
Estatística e Experimentação Agronômica

Piracicaba
2008

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - ESALQ/USP**

García Peña, Marisol

Análise dos modelos AMMI bivariados / Marisol García Peña. - - Piracicaba, 2008.
77 p. : il.

Dissertação (Mestrado) - - Escola Superior de Agricultura Luiz de Queiroz, 2008.
Bibliografia.

1. Análise de variância 2. Análise multivariada 3. Correlação genética e ambiental
4. Delineamento experimental 5. Genética estatística 6. Modelo estatístico I. Título

CDD 519.53
G216a

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor"

DEDICATÓRIA

*A minha avó Concepción
in memoriam*

AGRADECIMENTOS

A minha família, por sua constante voz de ânimo e porque, mesmo estando longe, seu amor e carinho estiveram perto de mim em todo momento.

Ao Dr. Carlos Tadeu dos Santos Dias, pelo conhecimento compartilhado, sugestões e apoio ao longo deste trabalho.

Ao Dr. Roland Vencovsky do departamento de Genética e ao Dr. Fred van Eeuwijk da Universidade de Wageningen - Holanda, pela disponibilidade e as orientações.

Aos professores e funcionários do programa de Pós-graduação em Estatística e Experimentação Agronômica da ESALQ/USP, pela atenção, auxílios permanentes e amizade.

Aos meus amigos colombianos pela amizade e as risadas nos muitos momentos de descontração.

Aos colegas do mestrado, pelas horas de trabalho compartilhado.

Ao programa PEC-PG e ao CNPq pela concessão da bolsa de estudos.

SUMÁRIO

RESUMO	6
ABSTRACT	7
LISTA DE FIGURAS	8
LISTA DE TABELAS	10
1 INTRODUÇÃO	10
2 DESENVOLVIMENTO	12
2.1 Interação entre genótipos e ambientes ($G \times E$)	12
2.2 Modelo de Efeitos Aditivos com Interação Multiplicativa AMMI	13
2.2.1 Ajuste dos efeitos principais por ANOVA	15
2.2.2 Desdobramento da Soma de Quadrados da interação ($G \times E$) por DVS	16
2.2.3 Seleção do número de termos no modelo AMMI para descrever a interação	18
2.2.4 Avaliação preditiva por validação cruzada	21
2.2.5 Análise de Procrustes	25
2.2.6 Análise Multivariada da Variância - MANOVA	31
2.3 Materiais e métodos	35
2.3.1 Características dos dados	35
2.3.2 Metodologia	35
2.4 Resultados	36
2.4.1 Análise dos dados originais	36
2.4.2 Representação gráfica - Biplot	43
2.4.3 Análise de procrustes	49
2.4.4 Análise dos dados originais com MANOVA	51
3 CONCLUSÕES	54
REFERÊNCIAS	55
APÊNDICES	58

RESUMO

Análise dos modelos AMMI bivariados

É comum encontrar nos ensaios experimentais a análise de dois fatores, cada um com diferente número de níveis, eles proporcionam uma tabela de dados de dupla entrada. Geralmente a análise destes dados é feita através da análise de variância - ANOVA, cumprindo algumas pressuposições básicas do modelo, mas há outros estudos nos quais é de grande importância a interação, como é o caso dos estudos de melhoramento genético, em que o objetivo é selecionar genótipos com ótimos desempenhos em diferentes ambientes. A pouca eficiência na análise da interação dos genótipos com os ambientes ($G \times E$) da ANOVA pode representar um problema aos melhoristas, que devem tirar proveito dessa interação para os seus estudos. Os modelos aditivos com interação multiplicativa - AMMI, traz vantagens na seleção de genótipos quando comparados com métodos convencionais, pois proporcionam uma melhor análise da interação ($G \times E$), além de permitir combinar componentes aditivos e multiplicativos em um mesmo modelo; estes modelos tem demonstrado ser eficientes na análise quando se tem apenas uma variável resposta, mas quando há mais de uma, ainda não existe um procedimento geral para realizar a análise. O presente trabalho propõe uma metodologia de análise quando se têm modelos AMMI bivariados, realizando análises individuais das variáveis respostas seguidas de uma análise de procrustes, que permite fazer comparações dos resultados obtidos nas análises individuais e finalmente uma confirmação destes resultados através da análise multivariada de variância - MANOVA. Os resultados obtidos permitem concluir que a análises AMMI e procrustes proporcionam uma boa alternativa de análise para os modelos AMMI bivariados.

Palavras-chaves: Modelos AMMI; Interação genótipos $\times$ ambientes; Análise de procrustes; Variáveis respostas

ABSTRACT

Bivariate AMMI models analysis

Is frequently find in the studies the two way factor analysis, each factor with different number of levels, they conform a two way table of data, generally the analysis of the data is made with the analysis of variance - ANOVA, satisfying some assumptions, but there are some studies in which is very important the interaction, like the case of the improvement studies, where the objective is select genotypes with optimum performance in different environments. The poor efficiency in the genotypes and environment interaction ($G \times E$) analysis of the ANOVA can represent a problem for the researchers, that need to take advantage of the interaction. The additive main effects and multiplicative interactions model - AMMI, give advantages in the selection of genotypes when is compare with traditional methods, because give a better interaction ($G \times E$) analysis, also permit combine additive and multiplicative components in the same model, these models have demonstrated be efficient in the analysis with just one response variable but when there is more than one there is not a clear procedure to do the analysis. This work presents a analysis methodology for the bivariate AMMI models, doing individuals analysis in the response variables follow by the procrustes, which permit compare the results of the individuals analysis, and finally a confirmation of these results with the multivariate analysis of variance - MANOVA. From the results can be concluded that the AMMI and the procrustes analysis give a good alternative for the bivariate AMMI models analysis.

Key words: AMMI models; Genotypes $\times$ environments interaction; Procrustes analysis; Response variables

LISTA DE FIGURAS

Figura 1 - Reflexão em uma dimensão como rotação de duas dimensões	28
Figura 2 - Biplot AMMI1 para dados de produtividade de grãos (kg/ha), em feijoeiro, com dezenove genótipos (G) e dezessete ambientes (A). A figura captura 23,6% de variabilidade	45
Figura 3 - Biplot AMMI2 para dados de produtividade de grãos (kg/ha), em feijoeiro, com dezenove genótipos (G) e dezessete ambientes (A). A figura captura 40,1% de variabilidade	46
Figura 4 - Biplot AMMI1 para dados de tempo de cozimento (min.), em feijoeiro, com dezenove genótipos (G) e dezessete ambientes (A). A figura captura 21,9% de variabilidade	48
Figura 5 - Biplot AMMI1 para dados de tempo de cozimento (min.), em feijoeiro, com dezenove genótipos (G) e dezessete ambientes (A). A figura captura 40,7% de variabilidade	49

LISTA DE TABELAS

Tabela 1 - Análise da variância conjunta completa calculada a partir das médias usando os sistemas de Gollob e Cornelius	20
Tabela 2 - Análise da variância conjunta calculada a partir das médias para a variável Produtividade	36
Tabela 3 - Análise da variância conjunta calculada a partir das médias para a variável Tempo de Cozimento	36
Tabela 4 - Médias de produtividade e tempo de cozimento para Genótipos e Ambientes .	37
Tabela 5 - Proporção da $SQ_{G \times E}$ explicada por cada eixo	38
Tabela 6 - Análise da variância conjunta completa calculada a partir das médias usando os sistemas de Gollob e Cornelius para a variável produtividade	39
Tabela 7 - Análise da variância conjunta completa calculada a partir das médias usando os sistemas de Gollob e Cornelius para a variável tempo de cozimento	40
Tabela 8 - Estatísticas PRESS, W e PRECORR para Validação Cruzada de Eastment e Krzanowski segundo as variáveis produtividade e tempo de cozimento	41
Tabela 9 - Coordenadas dos eixos para os gráficos biplot da variável produtividade	43
Tabela 10 - Coordenadas dos eixos para os gráficos biplot da variável tempo de cozimento	46
Tabela 11 - Análise de procrustes (M^2) para os marcadores de genótipos	50
Tabela 12 - Análise de procrustes (M^2) para os marcadores de ambientes	50
Tabela 13 - Análise de variância multivariada para comparar os efeitos e interação	51
Tabela 14 - Testes de hipóteses multivariados para o efeito dos Genótipos	51
Tabela 15 - Testes de hipóteses multivariados para o efeito dos Ambientes	52
Tabela 16 - Testes de hipóteses multivariados para o efeito da interação Genótipo $\times$ Ambiente	52
Tabela 17 - Dados originais para a variável produtividade	59
Tabela 18 - Dados originais para a variável tempo de cozimento	60
Tabela 19 - Estimativa da interação $G \times E$ modelo AMMI4 para a variável produtividade . .	61
Tabela 20 - Estimativa da interação $G \times E$ modelo AMMI6 para a variável tempo de cozimento	62
Tabela 21 - Estimativa $\hat{Y}$ modelo AMMI4 para a variável produtividade	63
Tabela 22 - Estimativa $\hat{Y}$ modelo AMMI6 para a variável tempo de cozimento	64

1 INTRODUÇÃO

Nas atividades de pesquisa, um aspecto que determina a validade dos resultados, as conclusões corretas e inferências é a seleção dos métodos na análise da informação obtida. Estes métodos devem-se ajustar às características dos dados, os quais podem apresentar problemas que devem ser considerados no momento da análise.

Em alguns estudos, é comum encontrar a análise de dois fatores cada um com diferentes níveis, eles proporcionam uma tabela de dados de dupla entrada, geralmente a análise destes dados é feita através da ANOVA, cumprindo algumas pressuposições como a que o número de repetições em cada combinação do fator deve ser maior do que 1, mas tem outros estudos nos quais é importante a interação ainda que as condições não sejam satisfeitas. Por exemplo, no melhoramento, o objetivo é selecionar genótipos com ótimos desempenhos em diferentes ambientes. A baixa eficiência na análise da interação dos genótipos com os ambientes ($G \times E$) pode representar um problema aos melhoristas, que devem tirar proveito dessa interação para os seus estudos.

Segundo Lavoranti (2003) as posições críticas dos estatísticos que atuam em programas de melhoramento genético, referem-se à falta de uma análise criteriosa da estrutura da interação ($G \times E$) como um dos principais problemas para a recomendação de cultivares. Tradicionalmente, a análise dessa estrutura é superficial, não detalhando os efeitos da complexidade da interação.

Os modelos aditivos com interação multiplicativa (AMMI), são uma boa opção para a análise da interação ($G \times E$) pois permitem um detalhamento maior da soma de quadrados da interação e conseqüentemente, traz vantagens na seleção de genótipos, quando comparados a outros métodos tradicionais de análise como a ANOVA. A utilização dessa teoria parece ser uma alternativa eficiente para os programas de melhoramento, já que permite combinar em um único modelo estatístico, componentes aditivos para os efeitos principais, como genótipos e ambientes, e componentes multiplicativos para os efeitos da interação.

Quando um experimento depende de muitas variáveis, não basta conhecer informações estatísticas isoladas para cada variável, mas é necessário também conhecer a totalidade destas informações fornecida pelo conjunto das variáveis. As relações existentes entre as variáveis não são percebidas e assim efeitos antagônicos ou sinérgicos de efeito mútuo entre variáveis complicam a interpretação do fenômeno a partir das variáveis consideradas. A necessidade de entender as relações entre várias variáveis faz com que a análise multivariada seja de grande importância. Com esta análise pode-se reduzir os dados ou simplificar a sua estrutura sem muita perda da informação contida nos dados obtendo assim uma fácil interpretação dos resultados, principalmente

quando as variáveis são correlacionadas.

Os modelos AMMI tem sido muito eficientes na análise de dados quando se tem apenas uma variável resposta, mas quando há mais de uma ainda não existe um procedimento claro de análise, por isso, no presente trabalho propõe-se uma metodologia de análise para este caso, usando algumas técnicas de análise multivariada. Serão usados testes paramétricos para a seleção do modelo levando em conta também a validação cruzada dada a sua importância por ser uma ferramenta de seleção do modelo que não depende de alguma distribuição de probabilidade.

A análise dos modelos AMMI com mais de uma variável resposta será feita através de análises individuais e em seguida comparadas com a análise procrustes. Esta técnica permite fazer a comparação dos resultados obtidos para cada uma das variáveis nas mesmas condições. A comparação de dois conjuntos de coeficientes de componentes principais poderia indicar se existem fontes comuns de variação ou não; no entanto, uma simples comparação dos coeficientes pode ser enganadora. Nesse sentido, a análise de procrustes evita que esta situação, em particular, aconteça.

Com as ferramentas apresentadas, pretende-se oferecer ao pesquisador novas possibilidades na análise dos modelos AMMI, no que diz respeito a análise de duas ou mais variáveis respostas.

2 DESENVOLVIMENTO

2.1 Interação entre genótipos e ambientes ($G \times E$)

Na experimentação agrônômica é comum fazer análise conjunta de grupos de experimentos, por exemplo, quando se trata de resultados de regiões, o experimento é feito em diferentes locais e repetido várias vezes; a interação é a resposta diferente dos genótipos em ambientes (locais) distintos, a interação pode ser devida a fatores físicos, adaptativos entre outros. As variações na resposta dos genótipos ou dos procedimentos agrônômicos nos diferentes ambientes são conhecidas como a interação destes fatores com o ambiente. Nos programas de melhoramento, a interação genótipos por ambientes ($G \times E$) é de extrema importância, pois possibilita a seleção de genótipos, bem como, a determinação do número ideal de ambientes e genótipos a serem avaliados em cada fase da seleção (Fox et al., 1997).

Crossa (1990), revisando alguns métodos de análise estatística para ensaios de produtividade em multiambientes, coloca três principais objetivos agrícolas a serem atingidos nestes ensaios: (a) estimar com precisão e prever a produtividade, baseado em um número reduzido de dados experimentais, (b) determinar a estabilidade dos rendimentos e o padrão de resposta dos genótipos ou dos procedimentos agrônômicos nos diferentes ambientes e (c) permitir uma orientação segura na seleção dos melhores genótipos ou procedimentos agrônômicos. Na prática, para verificar a significância da interação de genótipos com ambientes, é necessário repetir o experimento várias vezes, pois se o experimento for realizado somente em um ambiente, poderá ocorrer uma superestimação dos ganhos genéticos.

Geralmente os dados estão organizados em uma tabela de dupla entrada, com os genótipos nas linhas (g) e os ambientes nas colunas (e), as observações são representadas por Y_{ij} , em que g_i é o efeito do i -ésimo genótipo ($i = 1, 2, \dots, g$) e e_j é o efeito do j -ambiente ($j = 1, 2, \dots, e$). A análise da variância conjunta (ANOVA) permite determinar a magnitude da interação, usando a razão entre o Quadrado Médio da Interação ($QM_{G \times E}$) e o Quadrado Médio do Erro Médio (QM_{EM}). A interação significativa não é muito informativa por si só, então é necessário fazer um estudo mais detalhado sobre este componente. Várias metodologias têm sido propostas no sentido de entender melhor o efeito da interação genótipo $\times$ ambiente. Algumas dessas propostas são: zoneamento ecológico ou estratificação de ambientes, ou seja, identificar regiões ou sub-regiões onde o efeito da interação seja não significativa, o que pode levar a identificação de genótipos que se adaptam a ambientes específicos e ainda identificar genótipos com uma ampla adaptação ou estabilidade (Ramalho et al., 1993).

Uma das metodologias muito usada atualmente é o modelo de efeitos aditivos com interação

multiplicativa, AMMI; que tem como objetivo selecionar modelos que expliquem o padrão de resposta da interação, deixando fora o ruído presente nos dados.

2.2 Modelo de Efeitos Aditivos com Interação Multiplicativa AMMI

O primeiro passo para entender os Modelos Aditivos com Interação Multiplicativa (AMMI) é o tipo de dados para os quais esses modelos são aplicados, geralmente dois tipos são conhecidos:

1. *Fatorial de dupla entrada, com repetições ou não.* Os dados devem estar organizados em uma tabela de dupla entrada, tais como nos delineamentos fatoriais genótipos x ambientes ($G \times E$). Os dois fatores podem ser chamados de linhas e colunas e a combinação entre seus níveis de tratamentos. Os fatoriais de uma e tripla entrada não são permitidos, a razão desta restrição é que embora a parte da Análise da Variância (ANOVA) do AMMI seja flexível, a parte da Análise de Componentes Principais (PCA) precisa de uma estrutura de dupla entrada. No entanto, essa abordagem pode ser feita pelos modelos de Tucker ou Parafac (Bro, 1998). Uma tabela de tripla entrada, por exemplo, genótipos, locais e anos, pode se aproximar através do modelo AMMI como um ou vários subproblemas de dupla entrada. Os dados podem ter repetições ou não quando o AMMI é usado para estimação, seleção, modelamento e outros propósitos. Em resumo, o AMMI pode ser usado para dados com estrutura de dupla entrada com pelo menos três linhas e três colunas, com ou sem repetições.
2. *Tipo de dados.* Os dados para um modelo AMMI devem ser de uma única variável resposta, como rendimento. Não é permitida uma matriz de várias linhas ou colunas que contenham variáveis, como concentrações e temperaturas, pois esta mistura poderia causar parâmetros no modelo para colunas que não têm unidades significativas. Geralmente, se padroniza cada linha de matriz a média zero e variância unitária para eliminar unidades para algumas análises, como o PCA, mas não tem sentido para o modelo AMMI pois isso elimina o significado das linhas ou colunas desse modelo.
3. *Ajuste do modelo.* Com cada análise estatística, o modelo AMMI tem um modelo específico, e só se este modelo é usado para ajustar dados é que então os resultados serão realmente úteis. Em geral, AMMI é o modelo selecionado quando os dados apresentam efeitos principais e interação significativa. (Gauch, 1992).

O modelo AMMI usa dois métodos na sua análise: análise de variância e a decomposição singular; no modelo se unem os termos aditivos dos efeitos principais e os termos multiplicativos

para os efeitos da interação. Na primeira fase a análise de variância é aplicada à matriz de médias ($Y_{(g \times e)}$) composta pelos efeitos principais na parte aditiva (média geral, efeitos genotípicos e ambientais), resultando em um resíduo de não aditividade, isto é, na interação ($G \times E$), dada por $(\widehat{ge})_{ij}$, essa interação constitui a parte multiplicativa do modelo, na segunda fase a interação é analisada pela decomposição por valores singulares (DVS) da matriz de interações ($GE_{(g \times e)} = [(\widehat{ge}_{ij})]$), ou por análise de componentes principais (PCA) como prefere referir um grande número de autores (Duarte e Vencovsky, 1999). O modelo AMMI para dois fatores (G e E) é apresentado como:

$$Y_{ij} = \mu + g_i + e_j + \sum_{k=1}^p \lambda_k \gamma_{ik} \alpha_{jk} + \varepsilon_{ij}; \quad (1)$$

com $(ge)_{ij}$ interação genótipo-ambiente modelada por $\sum_{k=1}^n \lambda_k \gamma_{ik} \alpha_{jk} + \rho_{ij}$.
em que:

- Y_{ij} : resposta média do i -ésimo genótipo no j -ésimo ambiente;
- μ : média geral;
- g_i : efeito do i -ésimo genótipo, ($i = 1, 2, \dots, g$);
- e_j : efeito do j -ésimo ambiente, ($j = 1, 2, \dots, e$);
- λ_k : raiz quadrada do k -ésimo autovalor das matrizes $(GE)(GE)^T$ e $(GE)^T(GE)$ de iguais autovalores não nulos (λ_k^2 é o k -ésimo autovalor); $[GE_{ge} = (\widehat{ge}_{ij})]$ matriz de interações obtida como resíduo do ajuste aos efeitos principais, por ANOVA, aplicada à matriz de médias, ($k = 1, 2, \dots, p$), em que p é o número de raízes características não nulas $p = (1, 2, \dots, \min\{g - 1, e - 1\})$;
- γ_{ik} : i -ésimo elemento (relacionado ao genótipo i) do k -ésimo autovetor de $(GE)(GE)^T$ associado a λ_k^2 ;
- α_{jk} : j -ésimo elemento (relacionado ao ambiente j) do k -ésimo autovetor de $(GE)^T(GE)$ associado a λ_k^2 ;
- ρ_{ij} : ruídos presentes nos dados;
- ε_{ij} : erro experimental médio, $\varepsilon_{ij} \sim N(0, \frac{\sigma^2}{m})$, em que m é o número de repetições;
- n : número de componentes retidos no modelo ($n < p$).

Sob as restrições de identificabilidade $\sum_{i=1}^g g_i = \sum_{j=1}^e e_j = \sum_{i=1}^g (ge)_{ij} = \sum_{j=1}^e (ge)_{ij} = 0$,

A DVS da matriz de interação GE dá origem ao termo $\sum_{k=1}^p \lambda_k \gamma_{ik} \alpha_{jk} + \rho_{ij}$.

O termo $(ge)_{ij}$ é descrito como uma soma de p componentes, cada uma resultante da multiplicação de λ_k , expresso na mesma unidade de Y_{ij} , por um efeito genotípico (γ_{ik}) e um efeito

ambiental (α_{jk}), ambos adimensionais, ou seja, $\sum_{k=1}^p \lambda_k \gamma_{ik} \alpha_{jk}$. O termo λ_k traz uma informação relativa à variação devida à interação $G \times E$, no k -ésimo componente. De forma que a soma dos p componentes recompõe toda a variação ($SQ_{G \times E} = \sum_{k=1}^p \lambda_k^2$). Os efeitos γ_{ik} e α_{jk} representam os pesos do genótipo i e do ambiente j , naquele componente da interação (λ_k^2) (Duarte e Vencovsky, 1999).

Na abordagem AMMI não se busca recuperar toda a $SQ_{G \times E}$, mas apenas a parcela mais fortemente determinada por genótipos e ambientes (linhas e colunas da matriz GE), ou seja, o *padrão* (parte determinística ou sistemática). Assim, a interação do genótipo i com o ambiente j é descrita por: $\sum_{k=1}^n \lambda_k \gamma_{ik} \alpha_{jk}$, descartando-se o resíduo adicional ρ_{ij} dado por: $\sum_{k=n+1}^p \lambda_k \gamma_{ik} \alpha_{jk}$. Da mesma forma que em PCA esses eixos captam, sucessivamente, porções cada vez menores de variação presente na matriz GE ($\lambda_1^2 \geq \lambda_2^2 \geq \dots \geq \lambda_p^2$). Por isso, o método AMMI é visto como um procedimento capaz de separar *padrão e ruído* na análise da $SQ_{G \times E}$: $\sum_{k=1}^n \lambda_k \gamma_{ik} \alpha_{jk}$ e $\sum_{k=n+1}^p \lambda_k \gamma_{ik} \alpha_{jk}$, respectivamente (Weber et al. 1996).

2.2.1 Ajuste dos efeitos principais por ANOVA

Duarte e Vencovsky (1990) expõem que a primeira etapa da análise AMMI consiste numa análise de variância comum aplicada à matriz de médias ($Y_{g \times e}$). O modelo subjacente a essa etapa é:

$$Y_{ij} = \mu + g_i + e_j + \varepsilon_{ij} \quad (2)$$

em que ε_{ij} , agora, representa o resíduo de não aditividade dos efeitos principais, equivalente ao termo $(ge)_{ij}$ do modelo 1. As soluções de *mínimos quadrados ordinários* para o sistema de equações normais correspondente, sob as restrições de identificabilidade $\sum_i g_i = \sum_j e_j = 0$, são dadas por:

$$\begin{aligned} \hat{\mu} &= \bar{Y}_{..}; \quad \hat{g}_i = \bar{Y}_{i.} - \bar{Y}_{..}; \quad \hat{e}_j = \bar{Y}_{.j} - \bar{Y}_{..}; \quad \text{com:} \\ \bar{Y}_{..} &= \sum_{i,j} \frac{Y_{ij}}{ge}; \quad \bar{Y}_{i.} = \sum_j \frac{Y_{ij}}{e}; \quad \text{e} \quad \bar{Y}_{.j} = \sum_i \frac{Y_{ij}}{g} \end{aligned}$$

A aproximação de mínimos quadrados ($\hat{Y}_{ij}$) e seu respectivo resíduo, correspondente ao termo geral de interação $(\hat{ge})_{ij}$, ambos invariantes, são dados por:

$$\begin{aligned}\hat{Y}_{ij} &= \bar{Y}_{i.} + \bar{Y}_{.j} - \bar{Y}_{..}; \quad e \\ \hat{e}_{ij} &= Y_{ij} - \hat{Y}_{ij} \implies (\hat{ge})_{ij} = Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..}\end{aligned}$$

A partir desse último resultado pode-se construir a matriz de interações $GE_{(g \times e)} = [(\hat{ge})_{ij}]$, objeto de estudo da segunda etapa da análise AMMI.

As somas de quadrados para cada fonte de variação podem ser obtidas a partir das médias de cada genótipo em cada ambiente, já ajustadas para a constante μ , pelas seguintes expressões:

$$\begin{aligned}SQ_G &= e \sum_i (\hat{g}_i)^2; \text{ com } (g - 1) \text{ graus de liberdade;} \\ SQ_E &= g \sum_j (\hat{e}_j)^2; \text{ com } (e - 1) \text{ graus de liberdade;} e \\ SQ_{Res} &= \sum_{i,j} \hat{e}_{ij}^2 \implies SQ_{G \times E} = \sum_{i,j} (\hat{ge})_{ij}^2; \text{ com } (g - 1)(e - 1) \text{ graus de liberdade.}\end{aligned}$$

Geralmente a $SQ_{G \times E}$ é de elevada magnitude (pelo menos 20% da $SQ(Tratamentos)$). O objetivo final sempre é identificar algum padrão de resposta agrônomo importante que, uma vez incorporado no modelo, possa trazer uma melhoria na sua capacidade preditiva (Zobel et al., 1988). A análise procura capturar *padrões* presentes na estrutura dos dados, que possam contribuir para melhor explicar a resposta diferencial dos genótipos quando cultivados em diversos ambientes. (Duarte e Vencovsky, 1999).

2.2.2 Desdobramento da Soma de Quadrados da interação $(G \times E)$ por DVS

A decomposição da $SQ_{G \times E}$ é feita através da decomposição por valor singular - DVS, da matriz

$$\underbrace{GE}_B = \underbrace{U}_{(P_{BBT})} \underbrace{S}_{(\Lambda^{1/2})} \underbrace{V^T}_{(P_{B^TB})} = \sum_{k=1}^p \lambda_k \gamma_k \alpha_k^T;$$

com $p = \min\{(g - 1)(e - 1)\}$, γ_k k -ésimo autovetor $(GE)(GE)^T$ e α_k k -ésimo autovetor de $(GE)^T(GE)$.

Duarte e Vencovsky (1999) expõem que a DVS de GE pode ser escrita como uma soma de p matrizes de posto unitário, o que permite ilustrar a propriedade de interesse (partição da soma de quadrados dos elementos da matriz GE):

$$\underbrace{\mathbf{GE}}_{\substack{\mathbf{B} = [b_{ij}] \\ (g \times e)}} = \underbrace{\lambda_1 \gamma_1 \alpha_1^T}_{\substack{\mathbf{B1} = [b1_{ij}] \\ (g \times e)}} + \underbrace{\lambda_2 \gamma_2 \alpha_2^T}_{\substack{\mathbf{B2} = [b2_{ij}] \\ (g \times e)}} + \dots + \underbrace{\lambda_p \gamma_p \alpha_p^T}_{\substack{\mathbf{Bp} = [bp_{ij}] \\ (g \times e)}}$$

$$\begin{aligned} \sum_{i,j} b_{ij}^2 &= \sum_{i,j} (\hat{ge})_{ij}^2 = \sum_{i,j} b1_{ij}^2 + \sum_{i,j} b2_{ij}^2 + \dots + \sum_{i,j} bp_{ij}^2 = SQ_{G \times E} \\ \sum_k \lambda_k^2 &= \sum_{i,j} (\hat{ge})_{ij}^2 = \lambda_1^2 + \lambda_2^2 + \dots + \lambda_p^2 = \sum_k SQ_{IPCA_k} = SQ_{G \times E} \end{aligned}$$

Como $\mathbf{GE}$ é, por construção, uma matriz de desvios centrados e, portanto, com média nula, a soma de quadrados de seus elementos reproduz exatamente a $SQ_{G \times E}$. Assim, essa soma de quadrados é desdobrada em componentes relativos a cada termo da DVS, ou a cada eixo de interação da PCA ($IPCA_k$ conforme notação internacional). Para obter o desdobramento dessa soma de quadrados não é necessário encontrar as p matrizes parciais B_k , todas de posto 1. Mas, é suficiente encontrar os valores singulares de $\mathbf{GE}$ ou, equivalentemente, os autovalores de $(\mathbf{GE})^T(\mathbf{GE})$ ou de $(\mathbf{GE})(\mathbf{GE})^T$. A soma de quadrados relativa ao k -ésimo termo ou eixo de interação equivale ao quadrado do valor singular correspondente (o autovalor λ_k^2): $SQ_{G \times E[IPCA_k]} = \lambda_k^2$. Dada as propriedades desse tipo de composição, as p parcelas obtidas para a $SQ_{G \times E}$, são ortogonais e, portanto, independentes (Duarte e Vencovsky, 1999).

Tomando-se de forma cumulativa os sucessivos termos do desdobramento obtém-se aproximações cada vez melhores para a $SQ_{G \times E}$ original. Porém, sendo $\lambda_1^2 \geq \lambda_2^2 \geq \dots \geq \lambda_p^2$, é possível que alguns poucos primeiros termos já sejam suficientes para descrever uma alta proporção dessa soma de quadrados. Isso pode permitir a seleção de um modelo parsimonioso (apresentando poucos graus de liberdade) e com boa capacidade para descrever a variabilidade devida à interação, bem como para predizer novas observações. Logo, o modelo AMMI procura explicar a $SQ_{G \times E}$ por meio de uma aproximação de posto n para a matriz $\mathbf{GE}$ (com n preferivelmente bem menor do que p). Ou seja, o objetivo da análise é descrever a interação $G \times E$ por um número reduzido de eixos ($n = 1$ a 3), resultando num modelo informativo, mas que tenha poucos graus de liberdade (Duarte e Vencovsky, 1999).

Segundo os autores, com isso espera-se que o modelo seja capaz de : (i) captar o *padrão* subjacente à interação, ou seja, a possível lei geral que governa o fenômeno; e (ii) descartar os *ruídos* presentes nos dados e sem interesse agrônômico, os quais só prejudicam a capacidade preditiva do modelo. Essa separação em padrão e ruído pode também ser expressa em termos de

somas de quadrados, da seguinte forma:

$$SQ_{G \times E} = \sum_{k=1}^p \lambda_k^2 = \left(\sum_{k=1}^n \lambda_k^2 \right) + \left(\sum_{k=n+1}^p \lambda_k^2 \right) = SQ_{G \times E[\text{padrão}]} + SQ_{G \times E[\text{ruídos}]}$$

Dependendo do número n de termos (eixos singulares ou componentes principais) retidos para descrever o padrão da interação, o modelo é denotado: $AMMI0, AMMI1, \dots, AMMIp$, gerando uma família de modelos AMMI. Em $AMMI0$ nenhum eixo é ajustado (modelo aditivo, sem interação), em $AMMI1$ ajusta-se apenas o primeiro eixo de interação e assim por diante até $AMMIp$, o modelo completo de médias de caselas, com $n = p$. Se todos os eixos dessa decomposição (ou PCA) forem retidos, o modelo consumirá muitos graus de liberdade, apesar do ajuste perfeito à matriz de médias. O objetivo, entretanto, é resumir grande parte da interação $G \times E$ em apenas uns poucos eixos ($SQ_{G \times E[\text{padrão}]}$), resultando na seleção de um modelo AMMI reduzido, que descarta um resíduo adicional ($SQ_{G \times E[\text{ruídos}]}$) (Duarte e Vencovsky, 1999).

2.2.3 Seleção do número de termos no modelo AMMI para descrever a interação

O número de eixos que devem ser retidos é o menor possível (dois ou três) para explicar a estrutura da interação. Segundo Duarte e Vencovsky (1999), um dos procedimentos mais usados para a definição do número de eixos a serem retidos consiste em determinar os graus de liberdade associados à parcela da $SQ_{G \times E}$ relacionada a cada membro da família AMMI. Obtém-se, então, o quadrado médio (QM) correspondente a cada parcela (ou modelo). Logo, é obtido um teste F avaliando-se a significância de cada componente em relação ao $QM_{\text{erro médio}}$, este é calculado obtendo a média ponderada dos QM 's do erro, obtidos nas ANOVA's individuais dos e experimentos. O ponto que determina a seleção do modelo ($AMMI0, AMMI1, \dots$, ou $AMMIp$) baseia-se na significância do teste F para os sucessivos eixos da interação. O resíduo $AMMI$, reunindo os eixos descartados da interação, também pode ser testado de maneira a assegurar o seu caráter desprezível (Duarte e Vencovsky, 1999).

Um outro procedimento para determinar o número de eixos é o método de validação cruzada, no qual os dados são divididos aleatoriamente em dois grupos, para a modelagem e para validação dos dados. As respostas previstas obtidas do modelo AMMI, são então comparadas com o conjunto de validação, sendo calculadas as diferenças entre estes valores, logo a soma de quadrados dessas diferenças, dividindo-se o resultado pelo número delas e extraindo a raiz quadrada se obtém a diferença preditiva média (RMSPD).

$$RMSPD = \sqrt{\frac{\sum_{i=1}^g \sum_{j=1}^e (\hat{y}_{ij} - y_{ij})^2}{ge}} \quad (3)$$

em que, y_{ij} é a resposta média do i -ésimo genótipo no j -ésimo ambiente, $\hat{y}_{ij}$ é a resposta predita obtida do modelo AMMI do i -ésimo genótipo no j -ésimo ambiente, g é o número de genótipos e e é o número de ambientes.

Teste de Gollob

O teste de Gollob (1968) distribui graus de liberdade às Somas de Quadrados $SQ_k = m\lambda_k^2$ com $k = 1, 2, \dots, p$ e m o número de repetições, contando o número de parâmetros no k -ésimo termo multiplicativo. Logo, o teste F é calculado como na análise da variância para modelos lineares. O teste F na análise da variância supõe sob a hipótese nula que, o numerador e o denominador da estatística F são distribuídos independentemente como uma variável aleatória qui-quadrado (Cornelius et al 1996).

O teste F de Gollob (1968) não é válido porque os autovalores λ_k^2 são distribuídos como autovalores de uma matriz de Wishart mas não tem distribuição qui-quadrado, além disso, o teste assume que $m\lambda_k^2/\sigma^2$ é distribuído como qui-quadrado e então obviamente não é válido (Dias, 2005).

Na questão dos graus de liberdade, o método de Gollob (1968) é muito popular, este procedimento é muito fácil de aplicar, desde que o número de graus de liberdade para o k -ésimo componente da interação é simplesmente definido como $GL(IPCA_k) = g + e - 1 - 2k$, enquanto muitos outros procedimentos requerem simulações extensivas antes de serem usadas (Duarte e Vencovsky, 1999).

Teste F_R de Cornelius

Este teste foi proposto por Cornelius, alguns estudos realizados por Piepho (1995) mostram que o teste F_R é bem mais robusto que o proposto por Gollob. A estatística do teste é definida como:

$$F_R = \frac{SQ_{G \times E} - \sum_{k=1}^n \lambda_k^2}{f_2 \text{ } QM_{\text{erro médio}}}$$

em que, $f_2 = (g-1-n)(e-1-n)$ com n o número de termos multiplicativos incluídos no modelo. A estatística F_R , sob a hipótese nula de que não haja mais do que n termos determinando a interação, tem uma distribuição F aproximada com f_2 e $GL_{\text{erro médio}}$ graus de liberdade (Cornelius et al. 1996).

Um resultado significativo pelo teste sugere que pelo menos um termo multiplicativo ainda deve ser adicionado aos n já ajustados. Logo F_R pode ser visto como um teste para a significância dos $n+1$ primeiros termos da interação. Observa-se que para $n = 0$, isto é, quando nenhum termo multiplicativo é ajustado, o teste é equivalente ao F para a interação $G \times E$ global, na ANOVA conjunta, que é um teste exato. Nota-se também que os graus de liberdade do numerador de F_R é igual aos graus de liberdade para toda a interação menos os graus de liberdade atribuídos por Gollob (1968) aos n primeiros termos. Então, a aplicação de F_R é equivalente ao teste do resíduo AMMI para $G \times E$ e também ao teste para falta de ajuste da análise de variância da regressão (Duarte e Vencovsky, 1999).

Assim, pelo sistema de Gollob e Cornelius, a análise da variância conjunta completa (computadas a partir das médias) tem a estrutura como mostrada na Tabela 1.

Tabela 1 – Análise da variância conjunta completa calculada a partir das médias usando os sistemas de Gollob e Cornelius

Fontes de variação	GL ¹ Gollob	SQ ² Gollob	GL Cornelius	SQ Cornelius
Genótipo (G)	$g - 1$	SQ_G	-	-
Ambiente (E)	$e - 1$	SQ_E	-	-
Interação (GE)	$(g - 1)(e - 1)$	SQ_{GE}	-	-
$IPCA_1^3$	$g + e - 1 - (2 \times 1)$	λ_1^2	$(g - 1 - 1)(e - 1 - 1)$	$\sum_{k=2}^p \lambda_k^2$
$IPCA_2$	$g + e - 1 - (2 \times 2)$	λ_2^2	$(g - 1 - 2)(e - 1 - 2)$	$\sum_{k=3}^p \lambda_k^2$
$IPCA_3$	$g + e - 1 - (2 \times 3)$	λ_3^2	$(g - 1 - 3)(e - 1 - 3)$	$\sum_{k=4}^p \lambda_k^2$
...	...	...	...	...
$IPCA_p$	$g + e - 1 - (2 \times p)$	λ_p^2	-	-
Erro médio	$ge(m - 1)$	$SQ_{\text{erro médio}}$	-	-
Total	$gem - 1$	SQ_{TOTAL}	-	-

¹ GL: Graus de liberdade

² SQ: Soma de quadrados

³ $IPCA_k$: (interaction principal component analysis) modelo com k componentes, $k = 1, 2, \dots, p$.

2.2.4 Avaliação preditiva por validação cruzada

Em geral, é necessário o uso de procedimentos estatísticos computacionalmente intensivos para fazer previsões. Daí surge a importância que tem sido ultimamente dispensado aos métodos livres de distribuições teóricas como os baseados em reamostragem jackknife, bootstrap e validação cruzada. O critério preditivo de avaliação prioriza a capacidade de um modelo aproximar suas previsões a dados não incluídos na análise (simulando respostas futuras ainda não mensuradas).

Um modelo que seletivamente recupera o padrão e relega ruídos a um resíduo desconsiderado na predição de respostas, pode resultar em melhor precisão do que os próprios dados. Esse é o princípio subjacente à proposta de Gauch (1988) para seleção do modelo AMMI introduzida por ele como *avaliação preditiva*.

Dessa forma, através da validação cruzada, os dados de repetições, para cada combinação de genótipos e ambientes, são divididos, por um critério aleatório, em dois subconjuntos: (i) dados para o ajuste do modelo AMMI; e (ii) dados de validação. As respostas preditas do modelo AMMI, são comparadas com os dados de validação, calculando-se as diferenças entre esses valores. Logo, é obtida a soma de quadrados dessas diferenças e o resultado dividido pelo número de respostas preditas. À raiz quadrada desse resultado é chamado de diferença preditiva média (RMSPD), que é definida na expressão (3), Crosa et al. (1991) sugerem que o procedimento deve ser repetido 10 vezes, obtendo-se uma média dos resultados para cada membro da família de modelos.

Um pequeno valor de RMSPD indica sucesso preditivo do modelo, tal que o melhor modelo é aquele com o menor RMSPD. O modelo selecionado é então usado para analisar os dados de todas as m repetições, conjuntamente, em uma análise definitiva. (Dias, 2005).

Outros autores como Piepho (1994) sugere que o valor médio de RMSPD seja obtido a partir de 1000 randomizações diferentes e não 10 como propôs Crosa et al. (1991). O autor considera uma modificação da partição completamente aleatória dos dados (modelagem e validação) quando o ensaio é em blocos. Neste caso, ele recomenda sortear o bloco inteiro de um ensaio e não fazer componentes para cada combinação de genótipo e ambiente. Assim, a estrutura original de blocos é preservada. Contudo, apesar da coerência lógica desse tipo de proposta, estudos confirmando sua efetividade ainda não estão disponíveis. Gauch e Zobel (1996) sugerem que se faça o conjunto de dados de validação sempre com uma só observação para cada tratamento. Sendo assim, é mais provável para $m - 1$ dados, encontrar um modelo que mais se aproxime do ideal para analisar o conjunto completo dos m dados por tratamento.

Segundo Duarte e Vencovsky (1999), ao avaliar o modelo por validação cruzada, a análise AMMI deve partir das observações individuais propriamente ditas (dados de cada repetição dentro

de experimentos). Por outro lado, se o modelo for avaliado por um teste F a análise pode ser feita a partir das médias dos genótipos nos ambientes (experimentos), desde que se disponha dos quadrados médios residuais, obtidos nas análises de variâncias de cada experimento.

Dias e Krzanowski (2003) descrevem dois métodos que otimizam o processo de validação cruzada por validar o ajuste do modelo em cada um dos dados por vez e então combinar essa validação em uma medida simples e geral de ajuste. Esses métodos são a seguir descritos.

Método “leave-one-out”

Dias e Krzanowski (2003) propuseram dois métodos baseados em um procedimento “leave-one-out” completo, que otimiza o processo de validação cruzada. No que segue, assume-se que se deseja prever os elementos x_{ij} da matriz $\mathbf{X}$ por meio do modelo:

$$x_{ij} = \sum_{k=1}^n d_k u_{ik} v_{jk} + \varepsilon_{ij}$$

em que, d_i são a raiz quadrada dos autovalores da matriz $\mathbf{X}\mathbf{X}^T$, a i -ésima coluna $v_i = (v_{i1}, \dots, v_{ip})$ da matriz $\mathbf{V}_{p \times p}$ é o autovetor correspondente ao i -ésimo maior autovalor d_i^2 de $\mathbf{X}^T\mathbf{X}$ e a j -ésima coluna $u_j = (u_{1j}, \dots, u_{nj})^T$ da matriz $\mathbf{U}_{n \times p}$ é o autovetor correspondente ao j -ésimo maior autovalor d_j^2 de $\mathbf{X}\mathbf{X}^T$, ε_{ij} é o ruído.

Os métodos são aqueles apresentados em Krzanowski (1987) e Gabriel (2002), no qual prediz-se o valor $\hat{x}_{ij}^n$ de x_{ij} ($i = 1, \dots, g$; $j = 1, \dots, e$) para cada possível escolha de n (o número de componentes), e a medida de discrepância entre o valor atual e predito como

$$PRESS(n) = \sum_{i=1}^g \sum_{j=1}^e (x_{ij}^n - x_{ij})^2$$

Contudo, para evitar viés, os dados x_{ij} não devem ser usados nos cálculos de x_{ij}^n para cada i e j . Como conseqüência, apelo a alguma forma de validação cruzada é indicado, e os dois procedimentos diferem na forma com que eles lidam com isso (Dias, 2005). Ambos, entretanto, assumem que a DVS de $\mathbf{X}$ pode ser escrita como $\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^T$.

O procedimento de validação cruzada padrão subdivide $\mathbf{X}$ em um certo número de grupos, deleta-se cada grupo por vez a partir dos dados, avalia-se os parâmetros do modelo ajustados a partir dos dados remanescentes, e prediz-se o valor deletado (Wold, 1976, 1978). Krzanowski (1987) argumenta que a predição mais precisa resulta quando cada grupo deletado é tão pequeno

quanto possível, que no presente caso é um simples elemento de $\mathbf{X}$. Denota-se por $\mathbf{X}^{(-i)}$ o resultado de deletar a i -ésima linha de $\mathbf{X}$ e centralizar em torno das médias das colunas. Denota-se por $\mathbf{X}_{(-j)}$ o resultado de deletar a j -ésima coluna de $\mathbf{X}$ e centralizar em torno das médias das colunas, seguindo o esquema dado por Eastment and Krzanowski (1982). Então pode-se escrever

$$\begin{aligned}\mathbf{X}^{(-i)} &= \bar{\mathbf{U}} \bar{\mathbf{D}} \bar{\mathbf{V}}^T \text{ com } \bar{\mathbf{U}} = (\bar{u}_{pt}), \bar{\mathbf{V}} = (\bar{v}_{pt}) \text{ e } \bar{\mathbf{D}} = \text{diag}(\bar{d}_1, \dots, \bar{d}_l), \\ \mathbf{X}_{(-j)} &= \tilde{\mathbf{U}} \tilde{\mathbf{D}} \tilde{\mathbf{V}}^T \text{ com } \tilde{\mathbf{U}} = (\tilde{u}_{pt}), \tilde{\mathbf{V}} = (\tilde{v}_{pt}) \text{ e } \tilde{\mathbf{D}} = \text{diag}(\tilde{d}_1, \dots, \tilde{d}_{l-1}).\end{aligned}$$

Agora, considere-se o preditor

$$\hat{x}_{ij}^n = \sum_{t=1}^n \left(\tilde{u}_{it} \sqrt{\tilde{d}_t} \right) \left(\bar{v}_{tj} \sqrt{\bar{d}_t} \right) \quad (4)$$

Cada elemento no lado direito da expressão (4) é obtido da DVS de $\mathbf{X}$ centrada na média após omitir a i -ésima linha e a j -ésima coluna. Assim, o valor x_{ij} não é usado no cálculo da predição, e o máximo uso dos dados é feito com os outros elementos de $\mathbf{X}$. Os cálculos aqui são exatos, assim não há problema com a convergência como nos procedimentos de maximização que têm sido aplicados ao modelo AMMI, mas que não garantem a convergência (Dias e Krzanowski, 2003).

Gabriel (2002), tomou uma mistura de regressão e aproximação de uma matriz de posto inferior como a base para sua predição. O algoritmo para validação cruzada de aproximações de posto inferior proposto pelo autor é como segue: Para a matrix $\mathbf{X}$ (GE), se usa a partição

$$\mathbf{X} = \begin{bmatrix} x_{11} & \mathbf{X}_{1\cdot}^T \\ \mathbf{X}_{\cdot 1} & \mathbf{X}_{|11} \end{bmatrix}$$

e o ajuste aproximado da sub-matriz $\mathbf{X}_{|11}$ de posto n usando a DVS é:

$$\mathbf{X}_{|11} = \sum_{k=1}^n \mathbf{u}_{(k)} d_k \mathbf{v}_{(k)}^T = \mathbf{U} \mathbf{D} \mathbf{V}^T,$$

em que $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_n]$, $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_n]$ e $\mathbf{D} = \text{diag}(d_1, \dots, d_n)$.

Então, prediz-se x_{11} por $\hat{x}_{11} = \mathbf{X}_{1\cdot}^T \mathbf{V} \mathbf{D}^{-1} \mathbf{U}^T \mathbf{X}_{\cdot 1}$ e obtém-se o resíduo de validação cruzada $e_{11} = x_{11} - \hat{x}_{11}$.

Similarmente, obtém-se o valor ajustado da validação cruzada x_{ij} e os resíduos $e_{ij} = x_{ij} - \hat{x}_{ij}$ para todos os outros elementos x_{ij} , $i = 1, \dots, g$; $j = 1, \dots, e$; $(i, j) \neq (1, 1)$. Cada um irá requerer uma partição diferente de $\mathbf{X}$.

Esses resíduos e valores ajustados podem ser sumarizados por

$$PRESS(n) = \frac{1}{ge} \sum_{i=1}^g \sum_{j=1}^e e_{ij}^2 \quad \text{e} \quad PRECORR(n) = Corr(x_{ij}, \hat{x}_{ij} | \forall i, j),$$

respectivamente.

Com cada método, a escolha de n pode ser baseada em alguma função apropriada de

$$PRESS(n) = \frac{1}{ge} \sum_{i=1}^g \sum_{j=1}^e (x_{ij}^n - x_{ij})^2$$

Contudo, as características dessa estatística diferem para os dois métodos. O procedimento de Gabriel produz valores que primeiro decresce e então (usualmente) cresce com n . Por essa razão ele sugere que o valor ótimo de n seja aquele que produz o mínimo da função PRESS. O procedimento de Eastment-Krzanowski produz, geralmente, um conjunto de valores que é monotonicamente não-crescente com n (Dias, 2005). Por isso, sugerem o uso de

$$W_n = \frac{\frac{PRESS(n-1) - PRESS(n)}{D_n}}{\frac{PRESS(n)}{D_r}}$$

em que D_n é o número de graus de liberdade requeridos para ajustar o n -ésimo componente e D_r é o número de graus de liberdade remanescentes após ajustar o n -ésimo componente. Considerações sobre o número de parâmetros a serem estimados juntos com todas as restrições nos autovetores em cada estágio, mostra que $D_n = g + e - 2n$. D_r pode ser obtido por sucessivas subtrações, dando $(g - 1)e$ graus de liberdade na matriz centrada na média $\mathbf{X}$, isto é, $D_1 = (g - 1)e$ e $D_r = D_{r-1} - [g + e - (n - 1)2]$, $r = 2, 3, \dots, (g - 1)$, (Wold, 1978). W_n representa o aumento na informação preditiva suprida pelo n -ésimo componente, dividido pela informação preditiva média em cada um dos componentes remanescentes. Assim, importantes componentes devem produzir valores de W_n maiores que a unidade. Baseando-se a escolha de n em W_n pode ser vista como uma natural seleção de um melhor conjunto de variáveis regressoras ortogonais em análise de regressão múltipla (Dias e Krzanowski, 2003).

Ao nível computacional, a melhor precisão parece ser obtida quando as entradas (x_{ij}) em diferentes colunas de $\mathbf{X}$ são comparáveis em tamanho e existe relativamente pouca variação entre os d_i . O procedimento mais estável é, portanto, aquele no qual a média $\bar{x}_j$ e o desvio padrão s_j da coluna j ($j = 1, \dots, e$) são primeiro calculados dos valores presentes naquela coluna. As entradas existentes x_{ij} de $\mathbf{X}$ são então padronizadas para $x'_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}$, e estimativas são obtidas

pela aplicação de

$$\hat{x}_{ij} = \mathbf{X}_{i.}^T \mathbf{V} \mathbf{D}^{-1} \mathbf{U}^T \mathbf{X}_{.j}$$

aos dados padronizados, e então os valores finais são obtidos de

$$\hat{x}_{ij} = \bar{x}_j + s_j x'_{ij}$$

Voltando ao caso dos dados de genótipo-ambiente, fica claro aqui que $\mathbf{X}$ deve ser a matriz de interações previamente denotada por GE. Contudo, desde que se está simplesmente procurando pelo número apropriado de termos multiplicativos no modelo, e qualquer constante aditiva pode ser absorvida no componente ε_{ij} do modelo, pode-se aplicar o procedimento “leave-one-out” diretamente à matriz $\mathbf{Y}$ de dados. De fato, isso pode freqüentemente ser preferível dado os valores pequenos tomados por muitos elementos de GE.

Cornelius et al. (1993) compararam resultados de validação cruzada com aqueles obtidos após calcular a estatística PRESS nos modelos multiplicativos em dados MET (MultiEnvironment Trials) completos. A partição dos dados envolveu três repetições para modelagem e uma repetição para validação. Calcularam o RMSPD da estatística PRESS ajustando os valores de PRESS como $[PRESS/ge + 3s^2/4]^{1/2}$, em que g e e denotam o número de genótipos e ambientes no MET e s^2 é a variância residual conjunta dentro de ambientes.

O termo em s^2 é um ajuste para a diferença em variância da validação dos dados nas médias de caselas, para tornar os resultados comparáveis ao RMSPD da divisão 3 – 1 dos dados. Resultados em um MET com nove genótipos e vinte ambientes mostrou que PRESS é mais sensível a super ajuste do que os dados divididos (Dias, 2005). Neste trabalho será usado a metodologia de Eastment-Krzanowski, por apresentar melhores resultados, conforme (Dias, 2005).

2.2.5 Análise de Procrustes

Existem diferentes técnicas que recaem no âmbito da projeção ou da otimização, por exemplo, a análise de componentes principais que projeta o conjunto de dados em um espaço de alta dimensão em um subespaço do espaço original de baixa dimensão. Por outro lado, pode não existir um modelo geométrico de alta dimensão dos dados, então é construído diretamente um espaço de baixa dimensão no qual os pontos representam os indivíduos da amostra, otimizando algum critério da função. O escalonamento multidimensional (MS) é um bom exemplo desta técnica. As duas aproximações deixam como resultado espaços de baixa dimensão nos quais os indivíduos da amostra são representados por pontos, mas a projeção permite uma interpretação

do subespaço gerado feita em termos de dimensões do espaço original (Krzanowski, 2000).

É muito comum que sejam obtidas configurações no desenvolvimento de uma pesquisa e a comparação entre os resultados por meios gráficos é necessária. A possibilidade de comparação entre as configurações surge porque os pontos de todas as configurações referem-se às mesmas n entidades, mas é apenas a comparação entre os pontos correspondentes de duas configurações que é de interesse. A comparação de dois conjuntos de coeficientes de componentes principais poderia indicar se existem fontes comuns de variação ou não. No entanto, uma simples comparação dos coeficientes pode ser enganadora. Lembrar que as primeiras r componentes principais definem o subespaço r -dimensional do espaço original em que os indivíduos da amostra podem ser projetados com menos deslocamento.

A comparação de dois conjuntos de r componentes é equivalente à comparação de dois subespaços r -dimensionais correspondentes deste espaço comum. É possível que dois conjuntos de vetores sejam diferentes um do outro mas estejam definindo o mesmo subespaço. É necessária uma técnica analítica que proporcione uma medida numérica de quanto duas ou mais representações gráficas diferem. Tem-se duas opções: uma é assumir que os n pontos de cada um dos conjuntos de configurações referem-se às mesmas n entidades e tentar definir a quantidade numérica do desvio da coincidência dos conjuntos de pontos. A outra, é assumir que os p eixos do conjunto de configurações são os mesmos e tentar definir a quantidade numérica do desvio da coincidência dos conjuntos de subespaços definidos em relação a estes eixos (Krzanowski, 2000).

Comparação de duas configurações de n -pontos

Sejam $\mathbf{X}$ e $\mathbf{Y}$ duas matrizes com dimensão $(n \times p)$ e $(n \times q)$ que representam as coordenadas dos n pontos em cada uma das configurações. Em cada caso, as coordenadas do i -ésimo ponto associado aos eixos ortogonais no espaço Euclidiano estão dados pelos valores na i -ésima linha da matriz. A primeira configuração está em um espaço p -dimensional e o i -ésimo ponto tem coordenadas $(x_{i1}, x_{i2}, \dots, x_{ia})$, enquanto a segunda configuração está em um espaço q -dimensional e o i -ésimo ponto tem coordenadas $(y_{i1}, y_{i2}, \dots, y_{ib})$. Supondo primeiro que $p > q$, a segunda configuração está em um subespaço do espaço p -dimensional e pode ser tratado como se estivesse neste último espaço adicionando $(p - q)$ valores de coordenadas de zero a cada ponto, isto é, adicionando $(p - q)$ colunas de zeros ao lado direito de $\mathbf{Y}$, convertendo-a em uma matriz $(n \times p)$. Sem perda de generalidade, pode-se assumir que $p = q$, e que esta condição é atingida adicionando um número apropriado de colunas de zeros na menor matriz dentre as duas.

O objetivo é comparar as duas configurações, assumindo que as linhas das duas matrizes

referem-se à mesma entidade. Com a pressuposição de $p = q$, pode-se desenhar as duas configurações com os mesmos p eixos ortogonais. São calculados os centróides das duas configurações G e G' como a média das colunas de cada matriz. Segundo Krzanowski (2000) uma medida simples do grau de coincidência de duas configurações, é a soma dos quadrados das distâncias entre os pontos correspondentes,

$$M^2 = \sum_{i=1}^p \left\{ \sum_{j=1}^q (x_{ij} - y_{ij})^2 \right\} \quad (5)$$

O cálculo direto de M^2 entre duas configurações não é significativo, porque a posição, a orientação e a escala de uma configuração são arbitrárias em relação à outra. Quando se comparam duas configurações, significa uma comparação das relações internas entre os n pontos apresentados pelas duas representações. As relações internas entre os pontos de uma configuração não são modificadas por nenhuma das seguintes operações.

Operação 1 Um deslocamento de todos os pontos através de uma distância constante no mesmo sentido, significa, uma translação fixa de toda a configuração.

Operação 2 Um deslocamento fixo de todos os pontos através de um ângulo constante, mantendo a distância de cada ponto ao centróide, significa, uma rotação fixa de toda a configuração.

Operação 3 O estiramento ou encolhimento de todos os pontos através de uma constante em uma linha reta do ponto ao centróide da configuração, quer dizer, dilatação uniforme de toda a configuração.

Existe uma outra operação que mantém as relações internas entre os pontos de uma configuração, é chamada de reflexão dos pontos sobre um ponto, linha, plano etc. No entanto, a reflexão em q dimensões pode ser realizada pela rotação em $(q + 1)$ dimensões. Por exemplo, a rotação uni-dimensional dos pontos S , T sobre a origem O nos pontos S' , T' , pode ser obtida rotacionando a porção OST sobre O com um ângulo de 180° no plano completo. As reflexões estão consideradas na operação 2, após a adição de zero à coordenada de cada ponto na configuração. Antes de calcular M^2 , tem sentido fazer primeiro uma translação, rotação e dilatar as duas configurações a posições que sejam melhores para ajustar uma em relação à outra. Assim, M^2 é calculado como uma medida de “falta de ajuste” de uma configuração à outra, e proporciona uma comparação significativa das duas configurações. Agora, é evidente que as operações da 1 até a 3 realizadas nas duas configurações tem um rendimento insignificante “soma de quadrados

do resíduo” M^2 . O procedimento adequado é manter uma configuração fixa e igualar a outra a esta. Segundo Krzanowski (2000) deve-se fixar a configuração cujas coordenadas estão dadas por $\mathbf{X}$ e igualar a configuração com coordenadas $\mathbf{Y}$ a esta. A correspondência será obtida através da realização dos pontos 1 até o 3, na seqüência, de forma tal a tornar o valor final de M^2 para a correspondência das configurações tão pequena quanto possível. Em seguida, são apresentadas as três etapas da análise de procrustes segundo Krzanowski (2000).

Figura 1 – Reflexão em uma dimensão como rotação de duas dimensões

1. Correspondência por translação

Da definição (5), seja $\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$ e $\bar{y}_j = \frac{1}{n} \sum_{i=1}^n y_{ij}$ ($j = 1, 2, \dots, p$), então os centróides G_X e G_Y das duas configurações têm coordenadas $(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p)$ e $(\bar{y}_1, \bar{y}_2, \dots, \bar{y}_p)$ respectivamente.

Adicionando e subtraindo $\bar{x}_j, \bar{y}_j$ dentro da expressão (5) e agrupando os termos, obtém-se:

$$M^2 = \sum_{i=1}^n \left[\sum_{j=1}^p \{ (x_{ij} - \bar{x}_j) - (y_{ij} - \bar{y}_j) + (\bar{x}_j - \bar{y}_j) \}^2 \right]$$

Agora tomando os primeiros dois termos e expandindo o quadrado, e levando em conta que $\sum_{i=1}^n (x_{ij} - \bar{x}_j) = \sum_{i=1}^n (y_{ij} - \bar{y}_j) = 0$:

$$M^2 = \sum_{i=1}^n \sum_{j=1}^p \{ (x_{ij} - \bar{x}_j) - (y_{ij} - \bar{y}_j) \}^2 + n \sum_{j=1}^p (\bar{x}_j - \bar{y}_j)^2 \quad (6)$$

Mas $(x_{ij} - \bar{x}_j)$ e $(y_{ij} - \bar{y}_j)$ são elementos das matrizes $\mathbf{X}$ e $\mathbf{Y}$ depois de centrar com respeito à média e portanto, os centróides das coordenadas das configurações representadas por $\mathbf{X}$ e $\mathbf{Y}$

coincidem na origem dos eixos. Deste modo,

$$M^2 = M_O^2 + n \sum_{j=1}^p (\bar{x}_j - \bar{y}_j)^2 \quad (7)$$

em que M_O^2 é a soma de quadrados do resíduo entre as duas configurações depois da translação. Então, os centróides estão na origem. Não obstante, $\sum_{j=1}^p (\bar{x}_j - \bar{y}_j)^2$ é simplesmente a distância Euclidiana entre os centróides G_X e G_Y , de modo que:

$$M^2 = M_O^2 + n(G_X G_Y)^2$$

Por conseguinte, o melhor ajuste por translação é alcançado quando as duas configurações têm o mesmo centróide. Neste caso, $G_X G_Y = 0$ e $M^2 = M_O^2$. A forma mais fácil de garantir isso é centrar com respeito a média as duas matrizes $\mathbf{X}$ e $\mathbf{Y}$ no início.

2. Correspondência por rotação

É assumido que $\mathbf{X}$ e $\mathbf{Y}$ estão centradas com respeito a média. A rotação de $\mathbf{Y}$ relativa a $\mathbf{X}$ pode ser expressa como uma matriz ortogonal $\mathbf{Q}$, e depois desta rotação as coordenadas da configuração são dadas pelas linhas de $\mathbf{YQ}$.

Agora, pode ser escrito $M^2 = \sum_{i=1}^n \left\{ \sum_{j=1}^p (x_{ij} - y_{ij})^2 \right\}$ como $M^2 = \text{traço}(\mathbf{X} - \mathbf{Y})(\mathbf{X} - \mathbf{Y})^T$.

Expandindo o produto de matrizes obtém-se:

$$M^2 = \text{traço}(\mathbf{X}\mathbf{X}^T + \mathbf{Y}\mathbf{Y}^T - 2\mathbf{X}\mathbf{Y}^T) \quad (8)$$

Depois de rotacionar por $\mathbf{Q}$, $\mathbf{Y}$ se torna $\mathbf{YQ}$ e a soma de quadrados do resíduo se torna

$$M^2 = \text{traço}(\mathbf{X}\mathbf{X}^T + \mathbf{YQ}\mathbf{Q}^T\mathbf{Y}^T - 2\mathbf{XQ}^T\mathbf{Y}^T)$$

Como $\mathbf{Q}$ é uma matriz ortogonal, $\mathbf{Q}\mathbf{Q}^T = \mathbf{I}$ e então a expressão anterior se torna:

$$M^2 = \text{traço}(\mathbf{X}\mathbf{X}^T + \mathbf{Y}\mathbf{Y}^T - 2\mathbf{XQ}^T\mathbf{Y}^T) \quad (9)$$

Para fazer M^2 tão pequeno quanto possível, $\mathbf{Q}$ deve ser escolhida para que o traço $(2\mathbf{XQ}^T\mathbf{Y}^T)$ seja o maior possível. A rotação necessária $\mathbf{Q}$ está dada por $\mathbf{Q} = \mathbf{V}\mathbf{U}^T$, em que $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ é a

decomposição por valor singular da matriz $\mathbf{X}^T\mathbf{Y}$.

3. Correspondência por dilatação

Se as escalas em que são expressas as duas configurações diferem, uma etapa final do processo de correspondência é escalonar as coordenadas na matriz $\mathbf{Y}$ pelo fator c e estimar o valor de c que minimiza M^2 . Lembrando que $\mathbf{Y}$ tem sido rotacionado a $\mathbf{YQ}$, o efeito de uma mudança de escala pelo fator c é converter as coordenadas a $c\mathbf{YQ}$. Então, da expressão (9), têm-se que:

$$M^2 = \text{traço}(\mathbf{X}\mathbf{X}^T + c^2\mathbf{Y}\mathbf{Y}^T - 2c\mathbf{X}\mathbf{Q}^T\mathbf{Y}^T)$$

A mudança de escala não afeta a escolha da rotação. Assim, os resultados nessa etapa continuam sendo os mesmos estando incluído ou não c na expressão (9).

$$M^2 = c^2 \text{traço}(\mathbf{Y}\mathbf{Y}^T) - 2c \text{traço}(\mathbf{X}\mathbf{Q}^T\mathbf{Y}^T) + \text{traço}(\mathbf{X}\mathbf{X}^T) \quad (10)$$

Esta expressão é quadrática em c , e aplicando cálculo diferencial mostra-se que o valor mínimo de M^2 é obtido quando $c = \frac{\text{traço}(\mathbf{X}\mathbf{Q}^T\mathbf{Y}^T)}{\text{traço}(\mathbf{Y}\mathbf{Y}^T)}$. Note-se que,

$$\begin{aligned} \text{traço}(\mathbf{X}\mathbf{Q}^T\mathbf{Y}^T) &= \text{traço}(\mathbf{Q}^T\mathbf{Y}^T\mathbf{X}) \\ &= \text{traço}(\mathbf{U}\mathbf{V}^T\mathbf{V}\Sigma\mathbf{U}^T) \\ &= \text{traço}(\mathbf{V}^T\mathbf{V}\Sigma\mathbf{U}^T\mathbf{U}) \\ &= \text{traço}(\Sigma) \end{aligned}$$

Em que $\mathbf{U}$, $\mathbf{V}$ e Σ são as matrizes da decomposição singular de $\mathbf{X}^T\mathbf{Y}$ sendo $\mathbf{V}^T\mathbf{V} = \mathbf{I}$ e $\mathbf{U}^T\mathbf{U} = \mathbf{I}$. Então, pelas propriedades da decomposição por valores singulares, o valor ótimo do fator de dilatação c é,

$$c = \frac{\text{traço}(\Sigma)}{\text{traço}(\mathbf{Y}\mathbf{Y}^T)} \quad (11)$$

As três etapas do processo podem ser feitas independentemente uma das outras. Além disso, depois de ter feito as três etapas, o valor de M^2 resultante é o menor possível. Este valor é chamado de M_{min}^2 ; substituindo (11) em (10), e usando algumas propriedades da etapa 3, obtém-se

$$c^2 = \text{traço}(\mathbf{Y}\mathbf{Y}^T) + M_{min}^2 = \text{traço}(\mathbf{X}\mathbf{X}^T) \quad (12)$$

Esta identidade pode ser usada como base para a análise de variância, interpretada como a

partição da soma de quadrados total entre os vértices da configuração fixa na soma dos quadrados total entre os vértices da configuração ajustada mais a soma de quadrados do resíduo.

Infelizmente, o processo de ajustar uma configuração a outra não é simétrico, pois, o sistema de escala $\mathbf{Y}$ relativo a $\mathbf{X}$ não é o inverso do melhor sistema de escala $\mathbf{X}$ relativo a $\mathbf{Y}$. Para evitar problemas de interpretação, é feito comumente a padronização das duas configurações no início da análise para que o quadrado da distância total a seu centróide respectivo seja a unidade. Esta padronização garante que $\text{traço}(\mathbf{X}\mathbf{X}^T) = \text{traço}(\mathbf{Y}\mathbf{Y}^T) = 1$, e se isto é satisfeito, então a variância fica $c^2 + M_{min}^2 = 1$, em que $c = \text{traço}(\Sigma)$, um valor independente de rotacionar $\mathbf{Y}$ para ajustar $\mathbf{X}$ ou vice-versa.

2.2.6 Análise Multivariada da Variância - MANOVA

Geralmente quando se tem mais de uma variável medida por parcela nos delineamentos de experimentos é feita a Análise de Variância Multivariada - MANOVA. A Análise de Variância - ANOVA tem uma generalização quando as variáveis são vetores levando a análise de uma matriz de somas de quadrados e produtos cruzados. Considerando o caso de dupla entrada, é suposto que tem-se mge observações independentes geradas pelo modelo

$$\mathbf{Y}_{iju} = \boldsymbol{\mu} + \mathbf{g}_i + \mathbf{e}_j + (\mathbf{ge})_{ij} + \boldsymbol{\varepsilon}_{iju}, \text{ com } i = 1, 2, \dots, g, j = 1, 2, \dots, e, u = 1, 2, \dots, m$$

em que $\mathbf{g}_i$ é o efeito da i -ésima linha, $\mathbf{e}_j$ é o efeito da j -ésima coluna, $(\mathbf{ge})_{ij}$ o efeito da interação entre a i -ésima linha e a j -ésima coluna, $\boldsymbol{\varepsilon}_{iju}$ é o termo do erro que é assumido independente $N_p(\mathbf{0}, \Sigma)$ para todo i, j, u e m é o número de repetições. É necessário que o número de observações em cada casela (i, j) deva ser o mesmo, tal que a soma total de quadrados e matriz de produtos pode ser decomposta. O interesse está em testar a hipótese nula da igualdade de $\mathbf{g}_i$, igualdade de $\mathbf{e}_j$ e a igualdade de $(\mathbf{ge})_{ij}$ (Mardia et al., 2003).

Segundo Mardia et al. (2003), a matriz de somas de quadrados total e produtos cruzados pode ser particionada de maneira similar ao caso da análise univariada. Seja $\mathbf{T}$, $\mathbf{G}$, $\mathbf{E}$, e $\mathbf{ER}$ o total, linhas, colunas, e matrizes de erros das Somas de Quadrados e Produtos Cruzados - SSPC respectivamente, a seguinte identidade se mantém.

$$\mathbf{T} = \mathbf{G} + \mathbf{E} + \mathbf{ER} \quad (13)$$

em que

$$\begin{aligned}
 \mathbf{T} &= \sum_{i=1}^g \sum_{j=1}^e \sum_{u=1}^m (\mathbf{y}_{iju} - \bar{\mathbf{y}}_{...})(\mathbf{y}_{iju} - \bar{\mathbf{y}}_{...})^T \\
 \mathbf{G} &= em \sum_{i=1}^g (\bar{\mathbf{y}}_{i..} - \bar{\mathbf{y}}_{...})(\bar{\mathbf{y}}_{i..} - \bar{\mathbf{y}}_{...})^T \\
 \mathbf{E} &= gm \sum_{j=1}^e (\bar{\mathbf{y}}_{.j.} - \bar{\mathbf{y}}_{...})(\bar{\mathbf{y}}_{.j.} - \bar{\mathbf{y}}_{...})^T \\
 \mathbf{ER} &= \sum_{i=1}^g \sum_{j=1}^e \sum_{u=1}^m (\mathbf{y}_{iju} - \bar{\mathbf{y}}_{i..} - \bar{\mathbf{y}}_{.j.} + \bar{\mathbf{y}}_{...})(\mathbf{y}_{iju} - \bar{\mathbf{y}}_{i..} - \bar{\mathbf{y}}_{.j.} + \bar{\mathbf{y}}_{...})^T
 \end{aligned}$$

com

$$\bar{\mathbf{y}}_{i..} = \frac{1}{em} \sum_{j=1}^e \sum_{u=1}^m y_{iju}, \quad \bar{\mathbf{y}}_{.j.} = \frac{1}{gm} \sum_{i=1}^g \sum_{u=1}^m y_{iju} \quad \text{e} \quad \bar{\mathbf{y}}_{...} = \frac{1}{gem} \sum_{i=1}^g \sum_{j=1}^e \sum_{u=1}^m y_{iju}$$

Além disso, $\mathbf{ER}$ pode ser decomposto como:

$$\mathbf{ER} = \mathbf{I} + \mathbf{W} \tag{14}$$

em que

$$\mathbf{I} = m \sum_{i=1}^g \sum_{j=1}^e (\bar{\mathbf{y}}_{ij.} - \bar{\mathbf{y}}_{i..} - \bar{\mathbf{y}}_{.j.} + \bar{\mathbf{y}}_{...})(\bar{\mathbf{y}}_{ij.} - \bar{\mathbf{y}}_{i..} - \bar{\mathbf{y}}_{.j.} + \bar{\mathbf{y}}_{...})^T$$

e

$$\mathbf{W} = \sum_{i=1}^g \sum_{j=1}^e \sum_{u=1}^m (\mathbf{y}_{iju} - \bar{\mathbf{y}}_{ij.})(\mathbf{y}_{iju} - \bar{\mathbf{y}}_{ij.})^T \quad \text{com} \quad \bar{\mathbf{y}}_{ij.} = \frac{1}{m} \sum_{u=1}^m y_{iju}$$

Aqui $\mathbf{I}$ é a matriz SSPC devida a interação - $SSPC_{Int}$ e $\mathbf{W}$ é a matriz de resíduos SSPC - $SSPC_{res}$.

Testes para interações

De acordo com Mardia et al. (2003), das equações (13) e (14) pode-se particionar a matriz $\mathbf{T}$ em

$$\mathbf{T} = \mathbf{G} + \mathbf{E} + \mathbf{I} + \mathbf{W} \tag{15}$$

Sob a hipótese nula H_0 de que todos os g_i , e_j , e $(ge)_{ij}$ são zero, $\mathbf{T}$ deve ter a distribuição de Wishart $W_p(\Sigma, gem - 1)$.

Também pode-se escrever

$$\mathbf{W} = \sum_{i=1}^g \sum_{j=1}^e \mathbf{A}_{ij}$$

em que os $\mathbf{A}_{ij}$ são independentes e identicamente distribuídos i.i.d $W_p(\Sigma, m-1)$.

$$\mathbf{W} \sim W_p(\Sigma, ge(m-1)) \text{ e}$$

$$\mathbf{I} \sim W_p(\Sigma, (g-1)(e-1)) \text{ se os } (ge)_{ij} \text{ são iguais.}$$

Pode-se mostrar que as matrizes $\mathbf{G}$, $\mathbf{E}$, $\mathbf{I}$, e $\mathbf{W}$ são independentemente distribuídas uma da outra, e que a estatística LR para testar a igualdade dos termos da interação é:

$$\frac{|\mathbf{W}|}{|\mathbf{W} + \mathbf{I}|} = \frac{|\mathbf{W}|}{|\mathbf{E}|} \sim \Lambda(p, v_1, v_2)$$

em que p é o número de variáveis, $v_1 = ge(m-1)$, $v_2 = (g-1)(e-1)$ são os graus de liberdade, Λ refere-se à estatística lambda de Wilks e $|\cdot|$ corresponde ao determinante da matriz. A hipótese de não interação é rejeitada para valores baixos de Λ . Note-se que se $m = 1$ há uma observação por casela, então $\mathbf{W}$ tem zero graus de liberdade, e não pode ser feito nenhum teste para a interação.

Testes para efeitos principais

Segundo Mardia et al. (2003), se o efeito coluna não é importante, então pode-se mostrar que $\mathbf{E}$ tem distribuição $W_p(\Sigma, e-1)$. A estatística LR para testar a igualdade dos e_j independentemente de g_i e $(ge)_{ij}$ pode ser

$$\frac{|\mathbf{W}|}{|\mathbf{W} + \mathbf{E}|} \sim \Lambda(p, v_1, v_2)$$

em que $v_1 = ge(m-1)$ e $v_2 = e-1$. A igualdade é rejeitada para valores baixos de Λ .

Similarmente para testar a igualdade para os g efeitos linhas g_i independentemente dos efeitos colunas e_j , deve-se substituir $\mathbf{E}$ por $\mathbf{G}$ e intercambiar g e e no teste anterior. Note-se que se as interações são significativas então não faz sentido testar os efeitos para linhas e colunas; uma possibilidade neste caso é fazer os testes separadamente em cada uma das categorias ge linhas e colunas (Mardia et al. 2003).

De acordo com Mardia et al. (2003), pode-se decidir por ignorar o efeito da interação, seja

porque tem sido testadas e não são significativas, ou porque $m = 1$, ou por várias razões não é incluído no modelo o termo $(ge)_{ij}$. Neste caso é apropriado trabalhar com a matriz de erros $\mathbf{ER}$ do que com $\mathbf{W}$, o teste para os efeitos colunas fica

$$|\mathbf{ER}|/|\mathbf{ER} + \mathbf{G}| \sim \Lambda(p, v_1, v_2)$$

em que $v_1 = gem - g - e + 1$ e $v_2 = e - 1$. A igualdade dos e_j independentemente de g_i e $(ge)_{ij}$ é rejeitada para os valores baixos de Λ .

2.3 Materiais e métodos

2.3.1 Características dos dados

Os dados a serem utilizados no presente trabalho são do Instituto Agronômico de Campinas, e referem-se a dados de feijão comum (*Phaseolus vulgaris*), compostos por 19 genótipos avaliados em 18 ambientes e com 3 repetições, o delineamento experimental foi o aleatorizado em blocos. A parcela experimental foi constituída por quatro linhas de 4 metros de comprimento, espaçadas de 0,50 metros entre si, com 10 a 12 plantas viáveis por metro linear e a área útil da parcela correspondendo às duas linhas centrais. Os dados são de produtividade de grãos e tecnológicos, relativos à qualidade alimentícia do feijão; as variáveis respostas foram produtividade (kg/ha) e tempo de cozimento (min.).

2.3.2 Metodologia

Análise dos dados originais. O conjunto de dados é um experimento completo (sem dados perdidos), foi feita uma análise AMMI nos dados originais para cada uma das variáveis respostas, neste ponto foi encontrado o número de componentes multiplicativos necessários no modelo, os critérios para escolher o número de componentes foi o teste de Gollob, o teste de Cornelius F_R e a validação cruzada de Eastment e Krzanowski descritos anteriormente, logo, foram calculadas as respostas preditas para o modelo selecionado.

Representação gráfica - Biplot. Depois de selecionar o número de componentes no modelo levando em conta os três métodos mencionados e calculando as estimativas correspondentes, foram gerados os marcadores para genótipos e ambientes para cada variável resposta, assim como a representação gráfica deles, os eixos dos biplots apresentados representam a porcentagem de variabilidade retida por cada um dos eixos.

Análise de procrustes. Após escolher o modelo e encontrado os marcadores para genótipos e ambientes para cada variável resposta, decidiu-se fazer a análise de procrustes, isto para encontrar as similaridades entre as duas análises AMMI.

Análise dos dados originais com MANOVA. Como método de confirmação das hipóteses obtidas na análise AMMI e a análise de procrustes de uma maneira formal, foi feita uma análise de variância multivariada tradicional (MANOVA).

Os métodos descritos foram implementados no módulo IML (Interactive matrix programming) do pacote SAS (Statistical Analysis System).

2.4 Resultados

2.4.1 Análise dos dados originais

Para cada uma das variáveis respostas foi feita a análise de variância dando como resultado o ajuste dos efeitos principais por ANOVA (primeira etapa da análise AMMI). A correlação entre as variáveis respostas é -0,5052, foi obtida com os dados originais, o que indica uma correlação moderada inversa entre elas, enquanto a produtividade é alta o tempo de cozimento é baixo. Enquanto 0,0478 correlação entre os residuais, indica que os modelos selecionados estão deixando fora unicamente ruído.

Tabela 2 – Análise da variância conjunta calculada a partir das médias para a variável Produtividade

Fontes de variação	Graus Liberdade	Soma de Quadrados	Quadrado Médio	Teste F	Valor <i>p</i>
Genótipo (<i>G</i>)	18	6539004,2	363278	4,59	<0,0001
Ambiente (<i>E</i>)	17	233897361,6	13758668,3	173,67	<0,0001
Resíduo= $G \times E$	306	24242129,9	79222,6	-	-
Total	341	264678495,7	-	-	-

Nas Tabelas 2 e 3 observa-se que tanto os genótipos como os ambientes são estatisticamente significativos, também é de grande interesse a soma de quadrados $G \times E = 24242129,9$ e $G \times E = 7181,03395$ (objeto da decomposição DVS, na segunda etapa da análise), para a variável resposta produtividade, representou 9% da soma de quadrados total e para o tempo de cozimento 36%.

Tabela 3 – Análise da variância conjunta calculada a partir das médias para a variável Tempo de Cozimento

Fontes de variação	Graus Liberdade	Soma de Quadrados	Quadrado Médio	Teste F	Valor <i>p</i>
Genótipo (<i>G</i>)	18	1153,94242	64,10791	2,73	0,0002
Ambiente (<i>E</i>)	17	11513,60484	677,27087	28,86	<0,0001
Resíduo= $G \times E$	306	7181,03395	23,46743	-	-
Total	341	19848,58121	-	-	-

As estimativas das médias de genótipos e ambientes, assim como a média geral, ajustadas pelo modelo (sem interação), são dadas por o Tabela 4, nela pode-se observar que os genótipos com maior produtividade média são G11, G10, G15 e G13 e com menor os G8, G7, G1 e G9, enquanto ao tempo de cozimento os genótipos com maior tempo de cozimento em média são G15,

G1, G10 e G8 e com menor G7, G5, G19 e G13, já para os ambientes as maiores produtividades em média foram obtidas nos ambientes A5, A13, A12 e A1 e com menor os A17, A18, A16 e A6, para o tempo de cozimento em média os maiores tempos pertencem aos ambientes A17, A18, A9 e A7 e os menores a A12, A1, A11 e A8. É interessante notar que existem algumas diferenças entre as duas variáveis respostas, por exemplo, o genótipo G13 que para a produtividade é um dos que possui um valor alto, para o tempo de cozimento é um dos menores. Para os genótipos G7 e G1 a produtividade é pequena e o tempo de cozimento é dos mais altos, o mesmo acontece com os ambientes, com maior produtividade que apresentam menor tempo de cozimento, este é o caso de A1 e A12, enquanto para os ambientes A17 e A18 ocorre ao contrário.

Tabela 4 – Médias de produtividade e tempo de cozimento para Genótipos e Ambientes

Genótipo	Média ($\bar{Y}_{i.}$)		Ambientes	Média ($\bar{Y}_{.j}$)	
	Produtividade	Tempo C. ¹		Produtividade	Tempo C.
G1	2148,8802	32,4707	A1	3272,7193	21,4323
G2	2396,8422	31,0046	A2	2317,0614	30,1028
G3	2289,8026	28,1580	A3	1881,3596	31,4130
G4	2337,5195	28,3856	A4	2018,4211	31,3944
G5	2277,3450	27,1902	A5	4076,2676	25,1932
G6	2324,1702	29,5344	A6	1681,3597	29,6974
G7	2051,3359	26,2743	A7	1965,4386	35,2854
G8	2001,3272	31,2035	A8	2017,7874	23,0312
G9	2171,3037	29,1393	A9	2301,7983	35,5588
G10	2486,6924	32,2815	A10	2407,9437	28,4053
G11	2557,7411	29,7409	A11	2879,1491	22,2446
G12	2282,2393	29,3757	A12	3345,8333	20,1737
G13	2445,2254	27,6054	A13	3460,6065	29,3437
G14	2298,2628	28,8691	A14	2419,9561	23,2409
G15	2474,3609	33,1915	A15	2202,4035	33,3089
G16	2270,0485	28,1839	A16	1244,0877	31,0360
G17	2270,0319	29,4756	A17	828,2456	40,5781
G18	2371,4182	29,9026	A18	1165,2807	38,6161
G19	2335,9345	27,5165	Média Geral ($\bar{Y}_{..}$)	2304,7622	29,4475

¹ Tempo de Cozimento

O resíduo do ajuste dos efeitos principais à matriz de médias corresponde exatamente ao termo geral de interação $(\hat{ge})_{ij}$, isto é: $\hat{\varepsilon}_{ij} = Y_{ij} - \hat{Y}_{ij} \implies (\hat{ge})_{ij} = Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..}$. A aplicação dessa expressão à matriz de dados, uma vez disponíveis as estimativas de média, resulta na matriz de interações $(GE_{(g \times e)})$, para cada uma das variáveis respostas.

Observa-se na Tabela 5 que o primeiro eixo singular de interação para a variável produ-

Tabela 5 – Proporção da $SQ_{G \times E}$ explicada por cada eixo

Eixo Singular	Autovalor (λ_k^2)		Prop. $SQ_{G \times E}$ /Eixo		Prop. Acumulada	
	Prod. ¹	Tempo C. ²	Prod.	Tempo C.	Prod.	Tempo C.
1	5731965,0151	1578,3486	0,2364	0,2198	23,6446	21,9794
2	4012497,1756	1345,4351	0,1655	0,1874	40,1964	40,7154
3	3666128,8701	972,5948	0,1512	0,1354	55,3194	54,2593
4	2869403,5485	888,0084	0,1184	0,1237	67,1558	66,6253
5	2091924,6685	635,4937	0,0863	0,0885	75,7851	75,4749
6	1276137,7446	502,0010	0,0526	0,0699	81,0492	82,4656
7	1230525,9551	359,2838	0,0508	0,0500	86,1252	87,4688
8	982606,9046	246,4654	0,0405	0,0343	90,1785	90,9010
9	823909,1335	210,7389	0,0340	0,0293	93,5772	93,8356
10	646877,5389	137,2824	0,0267	0,0191	96,2456	95,7474
11	296955,2899	111,3167	0,0122	0,0155	97,4705	97,2975
12	218855,6842	70,3988	0,0090	0,0098	98,3733	98,2779
13	172497,2763	61,2275	0,0071	0,0085	99,0849	99,1305
14	140510,3132	38,5560	0,0058	0,0054	99,6645	99,6674
15	60292,1541	17,6614	0,0025	0,0025	99,9132	99,9134
16	18072,1035	6,1307	0,0007	0,0009	99,9877	99,9987
17	2970,5412	0,0910	0,0001	0,0000	100,0000	100,0000
Total ($\sum_k \lambda_k^2$)	24242129,9169	7181,0342	0,9998	1,0000	–	–

¹ Produtividade² Tempo de Cozimento

tividade captura 24% da $SQ_{G \times E}$, o segundo 17%, o terceiro 15%, o quarto 12%, e o quinto 9%, logo, um modelo *AMMI5*, por exemplo, explica 76% dessa soma de quadrados, dividindo-a em: $SQ_{G \times E[\text{padrão}]} = 18371919,2778$ (com os 76%) e $SQ_{G \times E[\text{ruídos}]} = 5870210,6391$ (com os 24% restantes).

Para a variável tempo de cozimento, tem-se que o primeiro eixo captura 22%, o segundo 19%, o terceiro 14%, o quarto 12% e o quinto 9% da $SQ_{G \times E}$, portanto, um modelo *AMMI5* explica 75% dessa soma de quadrados, dividindo-a em: $SQ_{G \times E[\text{padrão}]} = 5419,8806$ (com os 75%) e $SQ_{G \times E[\text{ruídos}]} = 1761,1536$ (com os 24% restantes).

Seleção do número de componentes no modelo

A Tabela 6 apresenta a análise conjunta de variância para a variável produtividade, incluindo o desdobramento da interação $G \times E$, atribuindo-se graus de liberdade aos componentes da interação ($IPCA_K$), conforme o sistema de Gollob (1968) e de Cornelius et al 1996. Também expressa os resultados do resíduo *AMMI* da interação, correspondente a cada membro da família

Tabela 6 – Análise da variância conjunta completa calculada a partir das médias usando os sistemas de Gollob e Cornelius para a variável produtividade

Fontes de variação	GL ¹	QM ² Gollob	F Gollob	Valor <i>p</i>	GL Cornelius	QM Cornelius	F _R Cornelius	Valor <i>p</i>
Genótipo (<i>G</i>)	18	363278,01	4,59	<0,0001	–	–	–	–
Ambiente (<i>E</i>)	17	13758668	173,67	<0,0001	–	–	–	–
Interação (<i>GE</i>)	306	79222,6468	2,1865	<0,0001	306	79222,6468	2,1865	<0,0001
<i>IPCA</i> ₁ ³	34	168587,2063	4,6528	<0,0001	272	68052,0768	1,8782	<0,0001
<i>IPCA</i> ₂	32	125390,5367	3,4606	<0,0001	240	60406,9489	1,6672	<0,0001
<i>IPCA</i> ₃	30	122204,2957	3,3727	<0,0001	210	51578,7565	1,4235	0,0006
<i>IPCA</i> ₄	28	102478,6982	2,8283	<0,0001	182	43747,9962	1,2074	0,0508
<i>IPCA</i> ₅	26	80458,6411	2,2206	0,0005	156	37629,5554	1,0385	0,3720
<i>IPCA</i> ₆	24	53172,4060	1,4675	0,0703	132	34803,5825	0,9605	0,6052
<i>IPCA</i> ₇	22	55932,9980	1,5437	0,0538	110	30577,6995	0,8439	0,8654
<i>IPCA</i> ₈	20	49130,3452	1,3559	0,1371	90	26454,8893	0,7301	0,9684
<i>IPCA</i> ₉	18	45772,7296	1,2633	0,2055	72	21625,4292	0,5968	0,9965
<i>IPCA</i> ₁₀	16	40429,8462	1,1158	0,3357	56	16252,7386	0,4486	0,9998
<i>IPCA</i> ₁₁	14	21211,0921	0,5854	0,8776	42	14599,9541	0,4029	0,9998
<i>IPCA</i> ₁₂	12	18237,9737	0,5033	0,9131	30	13144,7463	0,3628	0,9994
<i>IPCA</i> ₁₃	10	17249,7276	0,4761	0,9058	20	11092,2556	0,3061	0,9986
<i>IPCA</i> ₁₄	8	17563,7892	0,4847	0,8674	12	6777,8999	0,1871	0,9989
<i>IPCA</i> ₁₅	6	10048,6924	0,2773	0,9476	6	3507,1075	0,0968	0,9967
<i>IPCA</i> ₁₆	4	4518,0259	0,1247	0,9736	2	1485,2706	0,0410	0,9598
<i>IPCA</i> ₁₇	2	1485,2706	0,0410	0,9598	0	0	0	–
Erro médio _k	648	36233,24219	–	–	–	–	–	–

¹ GL: Graus de liberdade

² QM: Quadrado médio

³ *IPCA*_k: (interaction principal component analysis) modelo com *k* componentes, *k* = 1, 2, ..., 17.

de modelos. Assim, o resíduo de interação para o modelo *AMMI0* é toda a interação $G \times E$, com 306 gl; já para o modelo *AMMI1* é o restante da interação, após retirados os 34 gl.

Pelo teste *F*, aplicado segundo a regra de Gollob, 5 dos 17 eixos de interação são significativos, o que levaria à seleção do modelo *AMMI5*, enquanto pelo teste *F_R* de Cornelius, o modelo selecionado é o *AMMI4*, já que somente a partir de *IPCA*₄ o resíduo *AMMI* torna-se não significativo. Diante da maior simplicidade representativa do modelo e das propriedades do teste *F_R*, sugere-se, o modelo *AMMI4* como o melhor descritor do padrão de resposta diferencial dos genótipos aos ambientes.

A análise conjunta de variância para a variável tempo de cozimento é apresentada na Tabela 7, o desdobramento da interação $G \times E$ é incluso, segundo o sistema de Gollob (1968) e de Cornelius et al (1996) são atribuídos os graus de liberdade aos componentes da interação (*IPCA*_{*K*}). Para cada membro da família de modelos são expressos os resultados do resíduo *AMMI* da interação. Assim, o resíduo de interação para o modelo *AMMI0* é toda a interação $G \times E$, com 306 gl; já para o modelo *AMMI1* é o restante da interação, após retirados os 34 gl.

Tabela 7 – Análise da variância conjunta completa calculada a partir das médias usando os sistemas de Gollob e Cornelius para a variável tempo de cozimento

Fontes de variação	GL ¹	QM ² Gollob	F Gollob	Valor <i>p</i>	GL Cornelius	QM Cornelius	F _R Cornelius	Valor <i>p</i>
Genótipo (<i>G</i>)	18	64,1079	2,73	0,0002	–	–	–	–
Ambiente (<i>E</i>)	17	677,2708	28,86	<0,0001	–	–	–	–
Interação (<i>GE</i>)	306	23,4674	2,6612	<0,0001	306	23,4674	2,6612	<0,0001
<i>IPCA</i> ₁ ³	34	46,4220	5,2642	<0,0001	272	20,5981	2,3358	<0,0001
<i>IPCA</i> ₂	32	42,0448	4,7678	<0,0001	240	17,7385	2,0115	<0,0001
<i>IPCA</i> ₃	30	32,4198	3,6764	<0,0001	210	15,6412	1,7737	<0,0001
<i>IPCA</i> ₄	28	31,7146	3,5964	<0,0001	182	13,1684	1,4933	0,0002
<i>IPCA</i> ₅	26	24,4421	2,7717	<0,0001	156	11,2894	1,2802	0,0211
<i>IPCA</i> ₆	24	20,9167	2,3719	0,0003	132	9,5390	1,0817	0,2690
<i>IPCA</i> ₇	22	16,3311	1,8519	0,0105	110	8,1806	0,9277	0,6827
<i>IPCA</i> ₈	20	12,3233	1,3974	0,1157	90	7,2600	0,8233	0,8752
<i>IPCA</i> ₉	18	11,7077	1,3276	0,1632	72	6,1481	0,6972	0,9718
<i>IPCA</i> ₁₀	16	8,5801	0,9730	0,4849	56	5,4532	0,6184	0,9870
<i>IPCA</i> ₁₁	14	7,9512	0,9017	0,5569	42	4,6206	0,5240	0,9946
<i>IPCA</i> ₁₂	12	5,8666	0,6653	0,7855	30	4,1222	0,4675	0,9936
<i>IPCA</i> ₁₃	10	6,1227	0,6943	0,7303	20	3,1220	0,3540	0,9962
<i>IPCA</i> ₁₄	8	4,8195	0,5465	0,8215	12	1,9903	0,2257	0,9972
<i>IPCA</i> ₁₅	6	2,9436	0,3338	0,9192	6	1,0369	0,1176	0,9943
<i>IPCA</i> ₁₆	4	1,5327	0,1738	0,9518	2	0,0455	0,0052	0,9949
<i>IPCA</i> ₁₇	2	0,0455	0,0052	0,9949	0	0	0	–
Erro médio _k	648	8,8184	–	–	–	–	–	–

¹ GL: Graus de liberdade

² QM: Quadrado médio

³ *IPCA*_k: (interaction principal component analysis) modelo com *k* componentes, *k* = 1, 2, ..., 17.

Segundo a regra de Gollob e pelo teste *F*, 7 dos 17 eixos de interação são significativos, o que levaria à seleção do modelo *AMMI*7, enquanto pelo teste *F_R* de Cornelius, o modelo selecionado é o *AMMI*6, já que somente a partir de *IPCA*₆ o resíduo *AMMI* torna-se não significativo.

Diante da maior simplicidade representativa do modelo e das propriedades do teste *F_R*, sugere-se, o modelo *AMMI*4 como o melhor descritor do padrão de resposta diferencial dos genótipos aos ambientes.

A Tabela 8 apresenta os resultados para as variáveis respostas quando se aplica o procedimento de Eastment Krzanowski (1982). O método de Eastment e Krzanowski, através da estatística *W*, não selecionou componente para reter no modelo, isto significa, que o modelo seria aditivo para a produtividade e o tempo de cozimento, para a produtividade o maior valor da estatística *W* está associado ao componente 3 e para o tempo de cozimento ao componente 5.

Tabela 8 – Estatísticas PRESS, W e PRECORR para Validação Cruzada de Eastment e Krzanowski segundo as variáveis produtividade e tempo de cozimento

Variável Produtividade				Variável Tempo de Cozimento		
n	PRESS	W	PRECORR	PRESS	W	PRECORR
1	71439,3970	-0,0720	0,9527	23,4860	-0,9810	0,7766
2	78210,8210	-0,7582	0,9483	22,4904	0,3877	0,7892
3	70388,6180	0,9177	0,9538	21,8063	0,2591	0,7972
4	64448,4810	0,7151	0,9578	21,1870	0,2268	0,8081
5	59580,1840	0,5932	0,9611	20,0598	0,4079	0,8211
6	61116,5180	-0,1699	0,9603	19,8554	0,0696	0,8262
7	57665,7930	0,3747	0,9627	18,8255	0,3425	0,8360
8	55318,2520	0,2445	0,9643	18,5705	0,0791	0,8397
9	53425,6710	0,1864	0,9655	17,5492	0,3063	0,8499
10	48300,0340	0,5056	0,9688	17,0430	0,1415	0,8546
11	47463,9730	0,0752	0,9694	16,5437	0,1288	0,8595
12	46499,1660	0,0782	0,9701	16,2852	0,0598	0,8614
13	45519,5920	0,0704	0,9706	15,7814	0,1045	0,8659
14	44841,5550	0,0420	0,9711	15,3700	0,0743	0,8693
15	43626,8820	0,0636	0,9719	15,0424	0,0498	0,8721
16	43318,8320	0,0128	0,9721	14,8627	0,0218	0,8737
17	43249,6230	0,0021	0,9721	14,8596	0,0003	0,8737

Um passo seguinte na análise, após a escolha do modelo *AMMI*, é a predição de respostas para cada combinação de genótipos e ambientes. Considerando-se a seleção do modelo *AMMI4* para a variável produtividade e *AMMI6* para o tempo de cozimento, isso deverá ser feito a partir das expressões:

para a variável produtividade

$$\hat{Y}_{ij} = \bar{Y}_{i.} + \bar{Y}_{.j} - \bar{Y}_{..} + \sum_{k=1}^4 \lambda_k \gamma_{ik} \alpha_{jk}$$

$$\hat{Y}_{ij} = \bar{Y}_{i.} + \bar{Y}_{.j} - \bar{Y}_{..} + \lambda_1 \gamma_{i1} \alpha_{j1} + \lambda_2 \gamma_{i2} \alpha_{j2} + \lambda_3 \gamma_{i3} \alpha_{j3} + \lambda_4 \gamma_{i4} \alpha_{j4}$$

e para a variável tempo de cozimento

$$\hat{Y}_{ij} = \bar{Y}_{i.} + \bar{Y}_{.j} - \bar{Y}_{..} + \sum_{k=1}^6 \lambda_k \gamma_{ik} \alpha_{jk}$$

$$\hat{Y}_{ij} = \bar{Y}_{i.} + \bar{Y}_{.j} - \bar{Y}_{..} + \lambda_1 \gamma_{i1} \alpha_{j1} + \lambda_2 \gamma_{i2} \alpha_{j2} + \lambda_3 \gamma_{i3} \alpha_{j3} + \lambda_4 \gamma_{i4} \alpha_{j4} + \lambda_5 \gamma_{i5} \alpha_{j5} + \lambda_6 \gamma_{i6} \alpha_{j6}$$

Logo a estimativa $AMMI4$ da resposta de um genótipo i num ambiente j depende das estimativas de média do genótipo, do ambiente, da média geral, bem como de uma aproximação DVS de posto n para o termo específico de interação $\hat{g}e_{ij}$. Esse último termo reflete o *padrão* da interação $G \times E$ presente no conjunto de dados, sendo função dos n primeiros valores singulares, no caso da produtividade ($\lambda_1, \lambda_2, \lambda_3$ e λ_4), bem como de elementos dos 4 pares de vetores singulares ($\gamma_1, \gamma_2, \gamma_3, \gamma_4$ e $\alpha_1, \alpha_2, \alpha_3, \alpha_4$) e para o tempo de cozimento ($\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5$ e λ_6), e os 6 pares de vetores singulares ($\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6$ e $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6$). Assim, a partir das matrizes U, S e V é possível obter essa aproximação para qualquer combinação de genótipos e ambientes.

Apresenta-se a seguir, algumas das estimativas de resposta para cada uma das variáveis:

Variável produtividade, Genótipo 3 no ambiente 5

$$\begin{aligned}\hat{Y}_{35} &= \bar{Y}_{3.} + \bar{Y}_{.5} - \bar{Y}_{..} + \lambda_1 \gamma_{31} \alpha_{51} + \lambda_2 \gamma_{32} \alpha_{52} + \lambda_3 \gamma_{33} \alpha_{53} + \lambda_4 \gamma_{34} \alpha_{54} \\ &= 2289,8026 + 4076,2676 - 2304,7622 + [2394,1523(0,1606)(0,2595) \\ &\quad + 2003,1219(-0,0883)(0,0750) + 1914,7138(0,2073)(0,1548) \\ &\quad + 1693,9314(-0,11986)(0,01606)] = 4061,3080 + [144,7881] = 4206,0961 \quad kg/ha.\end{aligned}$$

Essa combinação de genótipo e ambiente tem sua resposta (4206,0961 kg/ha) determinada basicamente pelos efeitos principais (4061,3080 kg/ha), pois o termo de interação específica (144,7881 kg/ha) é pequeno.

Variável tempo de cozimento, Genótipo 3 no ambiente 5

$$\begin{aligned}\hat{Y}_{35} &= \bar{Y}_{3.} + \bar{Y}_{.5} - \bar{Y}_{..} + \lambda_1 \gamma_{31} \alpha_{51} + \lambda_2 \gamma_{32} \alpha_{52} + \lambda_3 \gamma_{33} \alpha_{53} + \lambda_4 \gamma_{34} \alpha_{54} + \lambda_5 \gamma_{35} \alpha_{55} + \lambda_6 \gamma_{36} \alpha_{56} \\ &= 28,1580 + 25,1932 - 29,4475 + [39,7284(-0,0982)(0,1926) \\ &\quad + 36,6801(-0,1980)(0,1421) + 31,1864(-0,1613)(-0,3888) \\ &\quad + 29,7994(0,1957)(0,0483) + 25,2090(-0,1810)(0,2129) \\ &\quad + 22,4053(0,1670)(-0,2413)] = 23,9036 + [-1,4212] = 22,4823 \quad min.\end{aligned}$$

Essa combinação de genótipo e ambiente mostra-se indesejável, pois apesar do aporte dos efeitos principais (23,9036 min), o componente negativo da interação específica (-1,4212 min), o que caracteriza uma falta de adaptação do genótipo $G3$ ao ambiente $A5$.

O cálculo dessas estimativas, matricialmente, torna-se conveniente e bastante simplificado. Isso porque com uma única operação é possível prever todas as respostas. O procedimento corresponde a obter uma estimativa da desconhecida matriz populacional de respostas verda-

deiras, com base no modelo escolhido. Interpretação nesse sentido pode ser realizada para todas as combinações de genótipos com ambientes.

2.4.2 Representação gráfica - Biplot

A última etapa da análise AMMI consiste na representação gráfica dos genótipos e ambientes no chamado *biplot*. Para isso, faz-se necessária a determinação de suas coordenadas para os eixos singulares de interação. A partir das mesmas matrizes U , S , V , resultantes da decomposição por valor singular - DVS da matriz GE , obtém-se novamente os resultados de interesse. De acordo com Gauch (1988), o primeiro eixo singular da análise AMMI captura a maior porcentagem de “padrão” e com acumulação subsequente das dimensões dos eixos, há uma diminuição na porcentagem de “padrão” e um incremento de “ruídos”. Com isso, apesar da porção pequena de $SQ_{G \times E}$ explicada pelos dois primeiros eixos (40,1%) para produtividade e (40,7%) para tempo de cozimento, espera-se capturar a maior parte do “padrão” devido a interação $G \times E$. Dessa forma, os escores de genótipos e de ambientes são plotados só até o segundo eixo, para cada uma das variáveis respostas.

Marcadores dos genótipos e dos ambientes para a variável produtividade, são obtidas as matriz de marcadores para genótipos **MGenProd**_{19×17} e a matriz de marcadores dos ambientes **MAmbProd**_{17×18}, mas dada a escolha do modelo AMMI4, são usadas apenas as quatro primeiras colunas da matriz **MGenProd**_{19×17} e as quatro primeiras linhas da matriz **MAmbProd**_{17×18}.

Tabela 9 – Coordenadas dos eixos para os gráficos biplot da variável produtividade

(continua)							
Pontos	Genótipos e Ambientes	Médias	lpca1	lpca2	lpca3	lpca4	Ordem p/ médias
1	G1	2148,8802	-13,7262	21,4873	-0,9652	6,7169	(17)
2	G2	2396,8422	15,5067	1,0958	-8,9703	16,2334	(5)
3	G3	2289,8026	7,8614	-3,9546	9,0745	-4,9331	(11)
4	G4	2337,5195	-0,8720	-12,7815	-7,3918	3,7561	(7)
5	G5	2277,3450	-0,6826	-9,2334	7,4929	11,9199	(13)
6	G6	2324,1702	-12,1503	14,5158	-5,3541	-21,9474	(9)
7	G7	2051,3359	-6,7877	-11,3469	7,9035	-4,6356	(18)
8	G8	2001,3272	-13,8244	0,2528	13,5228	-2,5459	(19)
9	G9	2171,3037	-7,7024	1,9688	3,8133	7,2844	(16)
10	G10	2486,6924	8,0595	-8,4371	-6,3007	-6,5078	(2)

Tabela 9 - Coordenadas dos eixos para os gráficos biplot da variável produtividade

Pontos	Genótipos e Ambientes	Médias	Ipca1	Ipca2	Ipca3	Ipca4	(conclusão)
							Ordem p/ médias
11	G11	2557,7411	25,7161	3,0357	-8,3895	-1,9692	(1)
12	G12	2282,2393	0,9777	-3,1722	8,5958	2,2312	(12)
13	G13	2445,2254	-0,2845	9,9941	-11,7048	8,6878	(4)
14	G14	2298,2628	-4,8543	0,9838	10,7326	9,8204	(10)
15	G15	2474,3609	19,8621	9,6044	19,6291	-4,8665	(3)
16	G16	2270,0485	-16,1049	-15,5235	-15,6623	4,9399	(14)
17	G17	2270,0319	-3,6950	-10,8945	5,5649	-10,1979	(15)
18	G18	2371,4182	-2,8962	16,5076	-7,5680	2,1996	(6)
19	G19	2335,9345	5,5973	-4,1025	-14,0226	-16,1863	(8)
20	A1	3272,7193	8,6849	-0,2000	-16,2342	8,4729	(4)
21	A2	2317,0614	-11,1793	24,0514	-12,8916	-9,0899	(8)
22	A3	1881,3596	-0,5982	-9,5742	-20,1551	-3,4012	(14)
23	A4	2018,4211	5,6392	0,5743	0,2431	0,7360	(11)
24	A5	4076,2675	12,6994	3,3584	6,7768	0,6613	(1)
25	A6	1681,3597	-24,8940	0,6767	0,1807	-18,3071	(15)
26	A7	1965,4386	3,9168	17,6745	5,2202	12,0701	(13)
27	A8	2017,7874	-16,4130	-8,4595	-6,9381	29,1167	(12)
28	A9	2301,7982	-1,7271	-5,3345	20,4714	-3,5822	(9)
29	A10	2407,9437	-5,3843	-8,6616	14,5130	1,1614	(7)
30	A11	2879,1491	18,2028	-12,8029	-5,2746	-7,5628	(5)
31	A12	3345,8333	8,9313	17,9901	-3,3621	0,1830	(3)
32	A13	3460,6065	20,6633	1,7810	6,6318	-2,3882	(2)
33	A14	2419,9561	1,0570	-16,1817	-11,1052	-10,0900	(6)
34	A15	2202,4035	4,7046	-1,2797	4,6868	0,0927	(10)
35	A16	1244,0877	-9,1440	4,5019	6,1053	-1,4413	(16)
36	A17	828,2456	-9,0040	-7,6659	5,5808	4,0996	(18)
37	A18	1165,2807	-6,1554	-0,4483	5,5510	-0,7309	(17)

As Figuras 2 e 3 ilustram as duas representações (os biplots AMMI1 e AMMI2, respectivamente) resultantes dos conjuntos de coordenadas, para o biplot AMMI1 (médias vs. $IPCA_1$) e para biplot AMMI2 ($IPCA_1$ vs $IPCA_2$). A partir delas são feitas, então, as devidas interpretações, procurando identificar genótipos e ambientes que menos contribuíram para a interação $G \times E$;

combinações de genótipos e ambientes desejáveis em termos de adaptabilidade; relações entre os eixos de interação e características genotípicas e ambientais conhecidas.

Figura 2 – Biplot AMMI1 para dados de produtividade de grãos (kg/ha), em feijoeiro, com dezanove genótipos (G) e dezessete ambientes (A). A figura captura 23,6% de variabilidade

Pela Figura 2, pode-se concluir que os genótipos mais estáveis e que menos contribuíram para a interação $G \times E$, captada pelo primeiro eixo ($IPCA_1$), foram G4, G7, G12, G13, G17 e G19; enquanto entre os ambientes nesse sentido são A3, A4, A7, A9, A12, A14, A15 e A18. Além de ser estáveis os genótipos indicam ser amplamente adaptados aos ambientes no teste. Os genótipos estáveis devem apresentar também um bom desempenho, neste caso uma alta produtividade, o que é avaliado através de suas médias. Assim entre os genótipos estáveis destaca-se o G13 com uma produtividade média elevada e entre os ambientes o A12.

Examinando a Figura 3, os genótipos e ambientes mais estáveis são G9, G12, G13 e G14; A4, A9, A15 e A18. Estes genótipos tiveram uma má classificação em produtividade média (G9 foi décimo nono, G12 foi décimo segundo e G14 foi décimo), apenas o G13 parece ter uma boa produtividade, ele foi o quarto. Para este conjunto, as produtividades média elevadas parecem estar associadas a adaptações específicas. Se observa que o primeiro eixo singular deve ser determinado por características contrastantes entre os ambientes A6 - A8 e o par A11 - A13, o segundo eixo parece resultante principalmente das diferenças entre A2, A8 e A11. Já para os genótipos, o primeiro eixo parece estar relacionado a aspectos determinantes da divergência entre os genótipos G11-G15 e G16, enquanto o segundo à divergência entre G1 e G16. Também

Figura 3 – Biplot AMMI2 para dados de produtividade de grãos (kg/ha), em feijoeiro, com dezoito genótipos (G) e dezessete ambientes (A). A figura captura 40,1% de variabilidade

observa-se a adaptabilidade dos genótipos nos ambientes, por exemplo, G16 com o A8, os G11 e G15 com os A13 e A11, os G1 e G6 com o A2. Corresponde ao melhorista identificar tais características para assim discernir melhor os mecanismos determinantes da interação.

Marcadores dos genótipos e dos ambientes para a variável tempo de cozimento, são obtidas as matriz de marcadores para genótipos **MGenTC**_{19×17} e a matriz de marcadores dos ambientes **MAmbTC**_{17×18}, mas dada a escolha do modelo AMMI6, são usadas apenas as seis primeiras colunas da matriz **MGenTC**_{19×17} e as quatro primeiras linhas da matriz **MAmbTC**_{17×18}.

Tabela 10 – Coordenadas dos eixos para os gráficos biplot da variável tempo de cozimento

(continua)									
Pontos	Genótipos e Ambientes	Médias	lpca1	lpca2	lpca3	lpca4	lpca5	lpca6	Ordem p/ médias
1	G1	32,4707	0,2727	-1,4580	1,3379	-1,1454	-1,3124	0,3517	(2)
2	G2	31,0046	0,1262	-0,5365	0,4291	1,1377	3,1036	0,1783	(5)
3	G3	28,1580	-2,1229	-1,1995	-0,9010	1,0686	-0,9089	0,7906	(15)
4	G4	28,3856	0,8496	0,6624	-0,5296	-1,2980	-0,3434	1,6383	(13)
5	G5	27,1902	0,4838	-1,0058	1,2259	-0,8930	-0,3608	0,2519	(18)
6	G6	29,5344	1,4440	1,7812	-1,7848	-1,1448	0,8588	-0,5959	(8)
7	G7	26,2743	1,0600	-0,5423	-0,2922	-1,2224	1,0656	-1,8462	(19)
8	G8	31,2035	2,3596	-0,9465	1,9664	-0,5586	-1,3938	-0,9789	(4)
9	G9	29,1393	0,8991	-0,2739	-1,1794	0,2840	1,1467	-0,0727	(11)

Tabela 10 - Coordenadas dos eixos para os gráficos biplot da variável tempo de cozimento (conclusão)

Pontos	Genótipos e Ambientes	Médias	Ipca1	Ipca2	Ipca3	Ipca4	Ipca5	Ipca6	Ordem p/ médias
10	G10	32,2815	0,8743	-0,3196	-1,7089	3,8969	-1,4880	-0,8764	(3)
11	G11	29,7409	-2,9560	-1,3804	-1,6213	-0,8550	0,3408	1,6764	(7)
12	G12	29,3757	0,3248	2,7015	0,9680	0,8369	-0,4750	1,3181	(10)
13	G13	27,6054	-0,6193	-0,5019	0,8359	0,1660	-0,9130	-0,0251	(16)
14	G14	28,8691	-0,2276	-0,6181	0,0894	-0,0722	0,0956	0,8650	(12)
15	G15	33,1915	1,5687	-1,7582	0,0986	0,2822	1,2276	0,4971	(1)
16	G16	28,1839	-1,9156	1,9176	-1,2406	-1,3725	-0,5774	-1,7422	(14)
17	G17	29,4756	1,1457	2,7871	0,0705	0,0105	-0,5575	1,1213	(9)
18	G18	29,9026	-2,3929	1,4001	2,9677	1,2677	1,2404	-0,8941	(6)
19	G19	27,5165	-1,1742	-0,7092	-0,7316	-0,3886	-0,7488	-1,6573	(17)
20	A1	21,4323	-0,5187	0,3105	-0,6493	-1,1808	-1,6291	-1,6978	(17)
21	A2	30,1028	0,0248	-0,0095	-0,9034	0,4147	0,2934	-0,4557	(9)
22	A3	31,4130	0,7249	-0,7438	-0,8897	0,1921	0,6296	-0,9187	(6)
23	A4	31,3944	1,0740	4,8121	-0,2684	-0,5133	-0,6279	1,4342	(7)
24	A5	25,1932	1,2145	0,8607	-2,1713	0,2639	1,0693	-1,1423	(13)
25	A6	29,6974	-2,0160	-1,8644	-1,5227	-0,4199	-3,0004	0,9570	(10)
26	A7	35,2854	0,8274	-0,4245	0,1096	4,6250	-0,9832	0,3279	(4)
27	A8	23,0312	0,0433	-0,2356	-0,2787	0,2085	0,9033	0,8286	(15)
28	A9	35,5588	-2,2921	0,6766	3,6872	0,5426	-0,4540	-0,9630	(3)
29	A10	28,4053	-2,5901	0,0933	0,8115	0,0145	2,2298	-0,2537	(12)
30	A11	22,2446	-0,4049	-0,5189	0,0876	-0,3644	1,0359	-0,1300	(16)
31	A12	20,1737	-0,6753	-0,5546	-0,5002	-0,6003	-0,1804	0,4322	(18)
32	A13	29,3437	-0,6528	0,8351	-0,1654	-0,8455	-0,2282	-0,9910	(11)
33	A14	23,2409	0,0461	-0,7682	-1,0859	0,5909	1,2977	-0,5178	(14)
34	A15	33,3089	0,3710	1,0486	0,8336	0,5224	-0,5140	0,6935	(5)
35	A16	31,0360	2,1699	-0,6815	1,0265	-1,4459	-0,8314	-1,3199	(8)
36	A17	40,5781	-0,9285	-0,9151	-0,2288	-1,1218	0,8721	2,8133	(1)
37	A18	38,6161	3,5826	-1,9208	2,1078	-0,8827	0,1174	0,9033	(2)

As Figuras 4 e 5 apresentam os biplots AMMI1 e AMMI2, respectivamente, para sua construção se usaram as seguintes coordenadas, para o biplot AMMI1 (médias vs. $IPCA_1$) e para biplot AMMI2 ($IPCA_1$ vs $IPCA_2$). São feitas as interpretações correspondentes, tentando identificar as contribuições dos genótipos e ambientes à interação $G \times E$; combinações de genótipos e ambientes desejáveis em termos de adaptabilidade; relações entre os eixos de interação e características genotípicas e ambientais conhecidas.

Da Figura 4, pode-se dizer que os genótipos mais estáveis e que menos contribuiriam para a interação $G \times E$, captada pelo primeiro eixo ($IPCA_1$), estão situados na faixa horizontal em

Figura 4 – Biplot AMMI1 para dados de tempo de cozimento (min.), em feijoeiro, com dezenove genótipos (G) e dezessete ambientes (A). A figura captura 21,9% de variabilidade

torno de zero em relação ao eixo $IPCA_1$, estes foram G2, G5, G9, G12, G13 e G14; enquanto entre os ambientes nesse sentido estão A1, A2, A3, A8, A9, A11, A12, A13 e A14. Além de ser estáveis os genótipos indicam ser amplamente adaptados aos ambientes no teste. É importante que os genótipos estáveis apresentem também um bom desempenho, neste caso um baixo tempo de cozimento, isto é avaliado através de suas médias. Assim, entre os genótipos estáveis destaca-se o G5 com um tempo de cozimento médio baixo e nos ambientes o A12, observa-se que o G13 ainda sendo um pouco maior em média que o G5 também proporciona um bom tempo de cozimento.

Observando a Figura 5, os genótipos e ambientes mais estáveis são G2, G5, G10 e G14; A1, A2, A7, A8, A11 e A14. Estes genótipos tiveram a seguinte classificação em tempo de cozimento médio, G2 foi décimo quinto, G5 foi décimo oitavo, G10 foi terceiro e G14 foi décimo segundo, G5 parece ter um bom tempo de cozimento. Se observa que o primeiro eixo singular pode estar determinado por características contrastantes entre os ambientes A6 - A9 - A10 e A18, o segundo eixo parece resultante principalmente das diferenças entre A4 e A6 - A8. Já para os genótipos, o primeiro eixo parece estar relacionado a aspectos determinantes da divergência entre os genótipos G11-G13-G18 e G8-G15, enquanto o segundo à divergência entre G3 - G15 e G12 - G17. Também observa-se a adaptabilidade dos genótipos nos ambientes, por exemplo, G8 e G15 com o A16, o G18 com A19, o G13 com o A10. Compete ao melhorista identificar tais características para

Figura 5 – Biplot AMMI1 para dados de tempo de cozimento (min.), em feijoeiro, com dezenove genótipos (G) e dezessete ambientes (A). A figura captura 40,7% de variabilidade

assim discernir melhor os mecanismos determinantes da interação.

Com os resultados encontrados nas duas análises, pode-se recomendar o genótipo G13 e o ambiente A12 como estáveis, com uma alta produtividade e um baixo tempo de cozimento.

2.4.3 Análise de procrustes

A estatística de procrustes M^2 definida na expressão 10 quantifica a diferença de duas configurações de pontos, neste caso os marcadores, quanto menor o valor da estatística, as configurações serão mais similares. Foi usada esta estatística para comparar os resultados após realizar as análises AMMI individuais para cada uma das variáveis respostas. A estatística M^2 foi calculada fazendo a rotação da matriz de marcadores da variável produtividade em relação à matriz de marcadores da variável tempo de cozimento, as matrizes inicialmente foram padronizadas, gerando assim o fator de dilatação $c = 1$. Neste trabalho procura-se avaliar a similaridade entre as análises individuais e fazer uma recomendação geral de genótipos e ambientes.

Depois de realizar as duas análises individuais, decidiu-se escolher um modelo que explique conjuntamente segundo as duas variáveis respostas. Para isto compararam-se as matrizes de marcadores levando em conta cada um dos modelos escolhidos, por exemplo, para a variável

produtividade o método de Cornelius selecionou um AMMI4 e para o tempo de cozimento um AMMI6, então foram aplicados os dois modelos nas duas variáveis e obtidas as respectivas matrizes de marcadores para genótipos e ambientes. Com estas matrizes de marcadores encontraram-se todas as combinações possíveis entre eles. Foram a seguir comparadas através da análise de procrustes, quanto menor valor da estatística de procrustes poderia indicar um melhor ajuste considerando duas variáveis respostas.

Tabela 11 – Análise de procrustes (M^2) para os marcadores de genótipos

Modelo para Produtividade	Modelo para Tempo de Cozimento	
	AMMI4	AMMI6
AMMI4	70,8369	92,0335
AMMI6	95,0152	96,3995

Na Tabela 11, observa-se os valores da estatística M^2 para os marcadores de genótipos segundo os modelos escolhidos pelo método de Cornelius, ou seja, AMMI4 para a variável produtividade e AMMI6 para o tempo de cozimento e as combinações entre eles, o menor valor da estatística foi encontrado quando as duas variáveis respostas foram modeladas com o modelo AMMI4, indicando que as matrizes de marcadores de genótipos são mais similares do que com a combinação dos outros modelos.

Tabela 12 – Análise de procrustes (M^2) para os marcadores de ambientes

Modelo para Produtividade	Modelo para Tempo de Cozimento	
	AMMI4	AMMI6
AMMI4	62,7695	75,7061
AMMI6	91,6764	90,3906

Na Tabela 12, encontram-se os valores da estatística M^2 para os marcadores de ambientes segundo os modelos escolhidos pelo método de Cornelius, do mesmo jeito que na tabela 11, neste caso o menor valor da estatística foi encontrado quando as duas variáveis respostas foram modeladas com o modelo AMMI4, indicando mais uma vez que as matrizes de marcadores de ambientes são mais similares do que com a combinação dos outros modelos.

Portanto, segundo a estatística de procrustes o modelo que gera maior similaridade entre as matrizes de marcadores tanto para genótipos quanto para ambientes é o modelo AMMI4, então poderia-se recomendar o genótipo G13 e o ambiente A12 para futuros estudos, pois apresentam um bom desempenho nas duas variáveis respostas.

2.4.4 Análise dos dados originais com MANOVA

Para o conjunto original de dados de 19 genótipos, 18 ambientes e duas variáveis respostas e 3 repetições, foi feita a análise de variância multivariada - MANOVA, gerando os seguintes resultados para os efeitos principais e a interação.

Tabela 13 – Análise de variância multivariada para comparar os efeitos e interação

Fontes de variação	Matriz de somas de quadrados e produtos cruzados - SSP	Graus de liberdade
Genótipos (G)	$\begin{bmatrix} 19616962,4360 & 54924,9130 \\ 54924,9130 & 3461,8280 \end{bmatrix}$	$g-1=18$
Ambientes (E)	$\begin{bmatrix} 701693571,1000 & -3459139,5140 \\ -3459139,5140 & 34540,8014 \end{bmatrix}$	$e-1=17$
Interação ($G \times E$)	$\begin{bmatrix} 72726478,9080 & -69763,2928 \\ -69763,2928 & 21543,0799 \end{bmatrix}$	$(g-1)(e-1)=306$
Resíduo	$\begin{bmatrix} 92711957,2130 & 25725,5827 \\ 25725,5827 & 4647,7757 \end{bmatrix}$	$ge(m-1)=684$
Total	$\begin{bmatrix} 886748969,6570 & -3448252,3112 \\ -3448252,3112 & 64193,4851 \end{bmatrix}$	$gem-1=1025$

Na Tabela 13 são apresentadas as matrizes de somas de quadrados e produtos cruzados - SSP, para cada efeito, SSP_{Gen} , SSP_{Amb} , SSP_{Int} e SSP_{res} , por exemplo, SSP_{Gen} contém na sua diagonal a soma de quadrados para o fator Gen para cada uma das p variáveis, neste caso $p = 2$ e os elementos fora da diagonal correspondem às somas de produtos para todos os pares de variáveis; estas matrizes são muito úteis na construção dos quatro testes estatísticos da MANOVA, para a construção destes testes são usadas diferentes combinações das matrizes e seus determinantes.

Tabela 14 – Testes de hipóteses multivariados para o efeito dos Genótipos

Estatística	Valor	Valor-F	GL Num	GL Den	Pr > F
Lambda de Wilks	0,4757	17,0700	36	1366	<0,0001
Traço de Pillai	0,5960	16,1300	36	1368	<0,0001
Traço de Hotelling-Lawley	0,9513	18,0300	36	1231,2	<0,0001
Maior raiz de Roy	0,7503	28,5100	18	684	<0,0001

Na Tabela 14, observa-se os resultados para os quatro testes multivariados, nos quais a hipótese nula é $H_{0Gen} = Gen_1 = Gen_2 = \dots = Gen_g$, em que $g = 19$, ou seja, que não existe diferença entre os níveis do fator genótipo, por meio dos testes pode-se rejeitar a hipótese nula, isto é, o genótipo tem um efeito significativo considerando as duas variáveis respostas produtividade e tempo de cozimento, simultaneamente.

Tabela 15 – Testes de hipóteses multivariados para o efeito dos Ambientes

Estatística	Valor	Valor-F	GL Num	GL Den	Pr > F
Lambda de Wilks	0,0222	229,2100	34	1366	<0,0001
Traço de Pillai	1,6122	167,2500	34	1368	<0,0001
Traço de Hotelling-Lawley	15,4370	309,7000	34	1224,8	<0,0001
Maior raiz de Roy	13,2910	534,7700	17	684	<0,0001

Na Tabela 15, encontram-se os valores para cada um dos quatro testes multivariados, aqui a hipótese nula é $H_{0Amb} = Amb_1 = Amb_2 = \dots = Amb_e$, em que $e = 18$, isto é, que não há diferença entre os níveis do fator ambiente, dados os valores dos testes pode-se rejeitar a hipótese nula, ou seja, o efeito que tem o ambiente sobre as variáveis respostas produtividade e tempo de cozimento é significativo.

Tabela 16 – Testes de hipóteses multivariados para o efeito da interação Genótipo $\times$ Ambiente

Estatística	Valor	Valor-F	GL Num	GL Den	Pr > F
Lambda de Wilks	0,0993	4,8500	612	1366	<0,0001
Traço de Pillai	1,2613	3,8200	612	1368	<0,0001
Traço de Hotelling-Lawley	5,4363	6,0600	612	1351,3	<0,0001
Maior raiz de Roy	4,6567	10,4100	306	684	<0,0001

Examinando a Tabela 16, pode-se concluir a partir dos resultados dos testes, que a hipótese nula $H_{0AmbGen} = GE_{11} = GE_{12} = \dots = GE_{ge}$ de não efeito de interação, é rejeitada, indicando que existe um efeito significativo da interação entre os genótipos e os ambientes sobre as duas variáveis respostas.

Dado que a análise MANOVA não proporciona uma boa interpretação sobre a interação, isto é, só permite saber se ela é ou não significativa, decidiu-se fazer uma análise de comparações post hoc para identificar quais das interações $(ge)_{ij}$ são significativamente diferentes das outras, nas variáveis respostas, produtividade e tempo de cozimento. Ao realizar as comparações encontrou-se que as interações para a variável produtividade não diferem significativamente e na variável tempo de cozimento somente diferem as interações G10A7 e a G12A1.

Também foi testado o efeito dos genótipos nos níveis dos ambientes, sendo significativos para os ambientes A1, A2, A3, A5, A6, A7, A8, A11, A12, A13 e A14, na variável resposta produtividade, indicando certa responsabilidade pela significância da interação e para a variável tempo de cozimento em todos os ambientes. Entretanto, o efeito dos ambientes nos níveis dos genótipos é significativo em todos os genótipos tanto para a variável produtividade como para o tempo de cozimento, coincidindo com o resultado das comparações post hoc realizadas.

Da análise AMMI seguida de procrustes, encontrou-se que o G13 e o A12, poderiam ser uma boa opção para futuros estudos, pois proporcionam estabilidade e um bom desempenho nas duas variáveis respostas, agora com os resultados obtidos da análise MANOVA, o A12 resultou ser um dos possíveis responsáveis pela interação, dado que o efeito dos ambientes nos níveis dos genótipos foi significativo para todos os níveis, poderia-se dizer que o G13 em combinação com o A12, proporcionam uma boa escolha, confirmando os resultados da análise AMMI.

3 CONCLUSÕES

Segundo os objetivos do trabalho, pode-se concluir:

- Segundo a análise AMMI aplicada aos dados originais, os modelos selecionados quando se aplicam os métodos de Gollob (1968), Cornelius (1996) e validação cruzada de Eastment Krzanowski (1982), permitem selecionar um modelo parcimonioso, para o caso da variável produtividade, o método de Gollob selecionou um modelo *AMMI5*, o de Cornelius um *AMMI4* e o de Eastment Krzanowski nenhum (*AMMI0*), o resultado do método de validação cruzada sugere um modelo aditivo mas levando na conta que a interação é significativa, opta-se por tanto, devido as propriedades do teste F_R , usar o modelo selecionado pelo método de Cornelius (*AMMI4*). Agora para a variável tempo de cozimento o método de Gollob selecionou um modelo *AMMI7*, o de Cornelius um *AMMI6* e o de Eastment Krzanowski nenhum (*AMMI0*), decidiu-se trabalhar com o modelo selecionado pelo método de Cornelius, isto é, um modelo *AMMI6*.
- Da representação gráfica através do biplot dos modelos selecionados para cada uma das variáveis respostas, conclui-se que uma boa recomendação poderia ser o genótipo G13 e o ambiente A12, pois proporcionam estabilidade e um bom desempenho nas variáveis respostas, ou seja, uma alta produtividade e um baixo tempo de cozimento.
- Pela análise de procrustes encontrou-se que o modelo que apresenta maior similaridade entre as matrizes de marcadores para genótipos e ambientes das duas variáveis respostas é o modelo *AMMI4*.
- A análise de MANOVA, confirmou os resultados encontrados na análise AMMI seguida de procrustes enquanto a genótipos e ambientes, o que indica que seria uma boa metodologia trabalhar desta maneira.

Apesar de que a análise MANOVA tenha confirmado os resultados da análise AMMI, recomenda-se realizar uma análise mais exaustiva da interação na MANOVA, tentando encontrar as interações responsáveis da significância da interação geral, por exemplo, usar a análise canônica de variáveis na MANOVA chamada CVA para o efeito da interação, como é apresentado em Lejeune e Calinski (2000).

Dado que as análises AMMI e procrustes apresentaram bom resultados conjuntamente, recomenda-se explorar a combinação das duas análises quando se tem mais de duas variáveis respostas, em que a análise procrustes já não compara matrizes se não subespaços.

REFERÊNCIAS

- BRO, R. **Multi-way analysis in the food industry. Models, algorithms, and applications**. 1998. 288p. Tese (Doutorado) - University of Amsterdam (NL) and Royal Veterinary and Agricultural University (DK), 1998.
- CROSSA, J. Statistical Analyses of multilocation trials. **Advances in Agronomy**, New York, v.44, p.55-85, 1990.
- CROSSA, J.; FOX, P. N.; PFEIFER, W. H.; RAJARAM, S.; GAUCH, H. G. AMMI adjustment for statistical analysis of an international wheat yield trial. **Theoretical and Applied Genetics**, New York, v.81, p.27-37, 1991.
- CORNELIUS, P. L.; CROSSA J.; SEYEDSADR M. S. Tests and estimators of multiplicative models for variety trials, 1993. In: DIAS, C. T. S.; KRZANOWSKI, W. J. **Model selection and cross-validation in additive main effect and multiplicative interaction (AMMI) models**. Crop Science, Madison, v.43, p.865-873, 2003.
- CORNELIUS, P. L.; CROSSA J.; SEYEDSADR M. S. Statistical tests and estimators of multiplicative models for genotype-by-environment interaction. In: KANG, M. S.; GAUCH, H. G. **Genotype-by-environment interaction**. Boca Raton: CRC Press, 1996. cap. 8. p.199-234.
- DIAS, C. T. S. **Métodos para a escolha de componentes em modelo de efeito principal aditivo e interação multiplicativa**. 2005. 73p. Tese (Livre-Docência no Departamento de Ciências Exatas) - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba. 2005.
- DIAS, C. T. S.; KRZANOWSKI, W. J. Model selection and cross-validation in additive main effect and multiplicative interaction (AMMI) models. **Crop Science**, Madison, v.43, p.865-873, 2003.
- DUARTE, J.B.; VENCOVSKY, R. **Interação genótipo × ambiente: uma introdução à análise “AMMI”**. Ribeirão Preto: Sociedade Brasileira de Genética, 1999. p.60 (Série Monografias).
- EASTMENT, H. T.; KRZANOWSKI, W. J. Cross-validatory choice of the number of components from a principal component analysis. **Technometrics**, Alexandria, v.24, p.73-77, 1982.
- FOX, P. N.; CROSSA, J.; ROMAGOSA, I. Multi-environment testing and genotype × environment interaction. In: KEMPTON, R.A.; FOX, P.N. (Ed.). **Statistical methods for plant variety evaluation**. New York: Chapman and Hall, 1997. cap. 8, p. 117-138.

GABRIEL, K. R. Le biplot-outil d'exploration de données multidimensionnelles. **Journal de la Societe Francaise de Statistique**, Paris, v.143, p.5-55, 2002.

GAUCH, H. Model selection and validation for yield trials with interaction. **Biometrics**, Washington, v.44, p.705-715, 1988.

GAUCH, H. **Statistical analysis of regional yield trials**: AMMI analysis of factorial designs. Amsterdam, Elsevier, 1992. 278p.

GAUCH, H. G.; ZOBEL, R. W. AMMI analysis of yield trials. In: KANG, M. S.; GAUCH, H. G. **Genotype-by-environment interaction**. Boca Raton: CRC Press, 1996. cap. 4. p.85-122.

GOLLOB, H. F. A statistical model which combines features of factor analytic and analysis of variance techniques. **Psychometrika**, Colorado Springs, v.33, p.73-115, 1968.

KRZANOWSKI, W. J. Cross-validation in principal component analysis. **Biometrics**, Washington, v.43, p.575-584, 1987.

KRZANOWSKI, W. J. **Principles of multivariate analysis**: A user's perspective. Oxford University Press, 2000. 586p.

LAVORANTI, O. J. **Estabilidade e adaptabilidade fenotípica através da reamostragem "bootstrap" no modelo AMMI**. 2003. 166p. Tese (Doutorado em Estatística e Experimentação Agrônômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba. 2003.

LEJEUNE, M; CALÍŃSKI, T. Canonical analysis applied to multivariate analysis of variance. **Journal of Multivariate Analysis**, Amsterdam, v.72, p.100-119, 2000.

MARDIA, K. V.; KENT, J.T.; BIBBY, J. M. **Multivariate analysis**. Amsterdam, Academic Press, 2003. 521p.

PIEPHO, H. P. Best linear unbiased prediction (BLUP) for regional yields trials: a comparison to additive main effects and multiplicative interaction (AMMI) analysis, 1994. In: DUARTE, J.B.; VENCOSKY, R. **Interação genótipo × ambiente**: uma introdução à análise "AMMI". Ribeirão Preto: Sociedade Brasileira de Genética, 1999. p.60 (Série Monografias).

PIEPHO, H. P. Robustness of statistical test for multiplicative terms in the additive main effects and multiplicative interaction model for cultivar trial. **Theoretical and Applied Genetics**, New York, v.90, p.438-443, 1995.

RAMALHO, M.A.P.; SANTOS, J.B.; ZIMMERMANN, M.J.O. **Genética quantitativa em plantas autógamas**: aplicações ao melhoramento do feijoeiro. Goiânia: UFG, 1993. 271p.

WEBER, W. E., WRICKE, G. e WESTERMANN, T. Selection of genotypes and prediction of performance by analysing genotype-by-environment interactions, 1996. In: DUARTE, J.B.; VENCOSKY, R. **Interação genótipo × ambiente**: uma introdução à análise “AMMI”. Ribeirão Preto: Sociedade Brasileira de Genética, 1999. p.60 (Série Monografias).

WOLD, S. Pattern recognition by means of disjoint principal component models, 1976. In: DIAS, C. T. S.; KRZANOWSKI, W. J. Model selection and cross-validation in additive main effect and multiplicative interaction (AMMI) models. **Crop Science**, Madison, v.43, p.865-873, 2003.

WOLD, S. Cross-validatory estimation of the number of components in factor and principal component models, 1978. In: DIAS, C. T. S.; KRZANOWSKI, W. J. Model selection and cross-validation in additive main effect and multiplicative interaction (AMMI) models. **Crop Science**, Madison, v.43, p.865-873, 2003.

ZOBEL, R. W.; WRIGHT, M. J. e GAUCH, H. G. Statistical analysis of yield trial, 1988. In: DUARTE, J.B.; VENCOSKY, R. **Interação genótipo × ambiente**: uma introdução à análise “AMMI”. Ribeirão Preto: Sociedade Brasileira de Genética, 1999. p.60 (Série Monografias).

APÊNDICES

APÊNDICE A - Dados originais para a variável produtividade

Tabela 17 – Dados originais para a variável produtividade

Genótipos	Ambientes								
	1	2	3	4	5	6	7	8	9
1	3143,3333	2630,8333	1466,6667	1900,0000	3937,9633	1975,0000	2025,0000	2111,1133	1787,5000
2	3550,0000	2129,1667	2220,8333	2108,3333	4268,5200	1041,6667	2438,3333	2388,8900	2047,5000
3	2738,3333	2304,1667	1733,3333	2150,0000	3793,5200	1237,5000	1873,3333	1759,2600	2573,3333
4	3845,0000	1991,6667	2029,1667	2033,3333	3999,0767	1929,1667	1685,8333	2291,6633	2311,6667
5	3396,6667	1925,0000	1695,8333	2041,6667	4350,9267	1583,3333	1895,0000	2393,5167	2494,1667
6	2925,0000	3175,0000	2091,6667	1833,3333	4095,3733	2450,0000	1931,6667	1638,8900	2350,8333
7	2795,0000	1770,8333	1837,5000	1679,1667	3544,4433	1570,8333	1572,5000	1643,5200	2283,3333
8	2590,0000	1650,0000	1562,5000	1800,0000	3550,0000	1695,8333	1905,0000	1486,1100	2261,6667
9	2835,0000	2362,5000	1758,3333	1579,1667	3680,5567	1600,0000	2296,6667	2282,4067	2291,6667
10	3435,0000	2408,3333	2333,3333	2258,3333	4319,4467	1608,3333	1759,1667	2055,5567	2542,5000
11	3930,0000	2333,3333	2508,3333	2479,1667	4825,9233	1416,6667	2365,0000	1689,8133	2268,3333
12	2788,3333	2150,0000	1558,3333	1950,0000	4128,7033	1566,6667	1856,6667	2217,5933	2254,1667
13	3645,0000	2691,6667	2075,0000	2208,3333	3820,3700	1666,6667	2438,3333	2375,0000	2145,0000
14	3376,6667	2129,1667	1425,0000	1941,6667	4138,8900	1820,8333	1847,5000	2453,7033	2585,8333
15	3418,3333	2387,5000	1525,0000	2295,8333	4820,3700	1437,5000	2433,3333	1481,4800	2866,6667
16	3240,0000	2454,1667	2416,6667	1950,0000	3907,4100	1700,0000	1556,6667	2625,0000	2064,1667
17	3270,0000	2008,3333	1662,5000	1950,0000	4246,2967	2079,1667	1570,0000	1777,7767	2353,3333
18	3650,0000	3155,8333	1808,3333	2150,0000	4182,4067	1454,1667	2202,5000	1995,3700	2286,6667
19	3610,0000	2366,6667	2037,5000	2041,6667	3838,8867	2112,5000	1690,8333	1671,2967	1965,8333

Genótipos	Ambientes								
	10	11	12	13	14	15	16	17	18
1	2279,6300	1833,0000	3675,0000	3147,6367	1866,6667	2195,8333	1046,6667	726,6667	931,3333
2	2442,5933	3079,1667	3820,8333	3764,6567	2645,8333	2329,1667	1160,0000	686,6667	1021,0000
3	2679,6300	3158,3333	3566,6667	3690,8700	2600,0000	2245,8333	1015,0000	1033,3333	1064,0000
4	2539,8167	3141,6667	3283,3333	3289,6267	2483,3333	2270,6667	1103,3333	1000,0000	847,0000
5	2498,1467	2766,6667	2904,1667	3449,1200	2241,6667	2183,3333	1073,3333	970,0000	1129,6667
6	2337,0400	2704,1667	3525,0000	3369,4267	2462,5000	1937,5000	1303,3333	606,6667	1097,6667
7	2608,3333	2829,1667	2858,3333	3016,0833	2058,3333	1750,0000	1171,6667	806,6667	1128,3333
8	2199,0767	2125,0000	3070,8333	2659,7033	2050,0000	1979,1667	1434,3333	913,3333	1091,3333
9	2376,8533	2662,5000	2900,0000	3532,9833	2200,0000	1820,8333	1151,6667	736,6667	1015,6667
10	2270,3700	3225,0000	3483,3333	3772,5900	3008,3333	2670,8333	1350,0000	886,6667	1373,3333
11	2307,4067	3420,8333	3820,8333	4217,5300	2633,3333	2454,1667	1218,3333	920,0000	1230,3333
12	2551,8533	3091,6667	3408,3333	3549,8233	2395,8333	2266,6667	1286,6667	830,0000	1229,0000
13	2120,3700	3166,6667	3795,8333	3474,3167	2341,6667	2195,8333	1461,6667	1033,3333	1359,0000
14	2288,8867	2533,3333	3204,1667	3722,5833	2091,6667	2279,1667	1380,0000	823,3333	1326,3333
15	2894,4433	3266,6667	3675,0000	4028,7033	2266,6667	2450,0000	1335,0000	616,6667	1339,3333
16	2634,2600	2725,0000	2862,5000	2785,0367	2866,6667	1908,3333	1073,3333	833,3333	1258,3333
17	2395,3700	2758,3333	2883,3333	3420,4633	2700,0000	2441,6667	1260,0000	900,0000	1184,0000
18	2182,4067	2887,5000	3412,5000	3081,8433	2241,6667	2241,6667	1631,6667	740,0000	1381,0000
19	2144,4433	3329,1667	3420,8333	3778,5267	2825,0000	2225,0000	1181,6667	673,3333	1133,6667

APÊNDICE B - Dados originais para a variável tempo de cozimento

Tabela 18 – Dados originais para a variável tempo de cozimento

Genótipos	Ambientes								
	1	2	3	4	5	6	7	8	9
1	26,9567	29,7367	32,6400	29,0600	21,3067	38,1433	36,2300	27,3133	39,0567
2	17,9900	32,4067	30,7533	29,1767	29,7867	22,7133	39,9867	29,7200	36,0400
3	20,6300	27,0867	31,9100	25,6033	20,6400	36,1900	40,7133	25,0233	30,8767
4	19,1733	25,7467	33,3900	40,1133	23,8833	31,9167	29,4733	18,8800	28,6800
5	16,4833	27,0800	26,5300	25,4033	21,7600	29,6200	29,4300	21,2033	36,4600
6	23,9867	31,5000	33,3900	40,7067	36,4633	24,7267	27,3133	22,9100	28,5867
7	15,9800	28,3467	30,3067	23,8733	23,8100	21,3867	26,3700	20,2000	27,3433
8	27,5100	30,2100	32,1500	30,4967	24,5833	27,0867	38,1167	21,2467	41,2067
9	21,2033	30,6200	37,9867	28,9967	28,2300	22,5367	37,2033	26,0667	29,1067
10	22,7867	36,6333	35,9567	31,8600	33,9600	36,7433	57,7867	23,4100	33,5367
11	20,5233	31,7467	27,3100	24,7167	23,0567	41,3000	29,3133	24,7767	34,1200
12	14,9800	29,1033	27,1600	46,5400	20,9533	27,1933	37,7367	25,8533	36,4267
13	19,5867	29,9000	25,6800	24,9300	20,6133	35,0767	32,9533	19,4200	44,5367
14	22,4867	29,4767	32,1333	25,9467	20,7233	26,9800	35,3967	22,2900	34,4733
15	18,2400	36,0300	40,1233	28,7533	30,6700	32,5533	40,4800	25,8500	34,2267
16	28,5567	31,1067	28,7033	36,6600	26,6200	31,6067	26,2400	18,6300	35,2467
17	23,2300	30,5567	25,6500	47,1467	28,6367	22,0000	37,7367	23,4133	36,0400
18	20,1433	27,3900	31,8600	32,4967	20,0067	22,1867	37,8467	20,6567	55,2367
19	26,7667	27,2767	33,2133	24,0133	22,9667	34,2900	30,0967	20,7300	34,4167

Genótipos	Ambientes								
	10	11	12	13	14	15	16	17	18
1	32,9200	21,6100	20,3700	30,2967	23,0767	37,8500	44,0100	46,3833	47,5133
2	40,4400	25,3800	21,7100	26,6467	29,6667	30,9433	28,2333	41,5667	44,9233
3	26,4133	19,3300	19,2967	28,4867	22,7133	29,6267	25,6367	40,8233	35,8433
4	26,6800	19,2600	18,4000	27,6633	20,9733	31,4167	29,0133	43,0100	43,2667
5	25,0900	20,1067	20,2133	25,8467	18,2700	30,7100	32,9867	37,9000	44,3300
6	23,9200	22,3533	17,5933	28,9033	26,6233	35,4767	30,7700	38,3233	38,0733
7	25,1500	20,5833	18,2933	31,1500	23,6867	28,4633	38,5733	33,2533	36,1667
8	19,9533	24,7700	24,4467	28,8200	23,2600	34,8200	40,3533	36,9800	55,6533
9	24,0033	23,4533	21,1067	27,1967	22,9267	31,3400	32,5767	44,6200	35,3333
10	25,3933	21,2033	21,2800	29,2433	29,8467	37,2700	31,8467	34,9567	37,3533
11	33,7533	26,1700	29,6100	27,5867	27,1233	31,1233	23,9433	50,2100	28,9533
12	30,2033	21,2133	19,0133	28,3867	19,5467	43,1600	31,5067	34,2900	35,4967
13	28,1567	21,7000	19,7033	28,3700	18,9100	29,3533	25,1900	40,5200	32,2967
14	23,3933	23,1567	18,1933	32,5567	22,7033	36,2200	27,0900	48,5567	37,8667
15	31,6967	28,1267	19,7633	31,0533	28,6300	35,7933	35,6500	46,0333	53,7733
16	32,8967	19,7167	19,1433	36,8467	19,2133	28,2133	26,6000	35,3600	25,9500
17	21,4000	19,1300	19,7267	30,6633	19,0300	33,9400	30,8733	43,5333	37,8533
18	39,3400	23,8433	17,2833	31,6167	22,8467	36,9467	26,2133	40,7367	31,5967
19	28,8967	21,5400	18,1533	26,1967	22,5300	30,2033	28,6167	33,9267	31,4633

APÊNDICE C - Estimativa ($G \times E$) modelo AMMI4 produtividade

Tabela 19 – Estimativa da interação $G \times E$ modelo AMMI4 para a variável produtividade

Genótipos	Ambientes								
	1	2	3	4	5	6	7	8	9
1	-50,9280	621,6355	-200,9057	-60,3553	-104,2491	233,0991	402,0494	245,7850	-134,7366
2	417,6250	-178,9168	105,8163	97,8424	150,5501	-684,0888	229,2166	271,1179	-274,4138
3	-120,0491	-255,1436	-132,9596	40,6365	144,7882	-106,4277	-51,2773	-302,1706	210,9583
4	146,8084	-236,5127	259,1021	-11,2912	-101,6097	-57,0396	-222,5714	283,0877	-95,0878
5	-24,7276	-419,3906	-102,7519	1,4419	18,9817	-206,1205	17,1197	384,3962	161,1257
6	-207,4650	753,4790	50,8519	-77,6362	-156,3479	713,1179	-83,8877	-525,2619	-87,4360
7	-224,2660	-256,7778	-30,8306	-46,2852	-73,8128	247,5879	-241,8320	17,5870	250,6550
8	-361,2173	9,4413	-258,0434	-76,4005	-84,7550	393,3681	-9,8178	56,8110	308,4776
9	-67,4742	18,0864	-115,8747	-36,0169	-60,5454	60,4090	112,4580	295,4071	54,7694
10	118,8310	-152,6408	225,0834	34,2820	27,0122	-88,3411	-228,9938	-206,6765	-74,5851
11	342,2464	-88,4213	131,3399	143,2740	278,6175	-603,5859	86,8165	-446,8892	-225,2992
12	-111,5156	-218,3189	-151,0515	7,4236	61,4906	-65,7792	19,5652	16,1133	183,2085
13	259,1596	315,4738	110,8466	7,6842	-43,6252	-147,3178	219,2880	254,2949	-323,5573
14	-133,3843	-149,6965	-256,2335	-16,9727	20,8842	-56,3355	172,9347	282,8265	187,6687
15	-189,3177	-199,8607	-482,9116	118,7136	414,2973	-395,3086	291,2767	-685,1281	333,7303
16	159,3554	-36,3123	457,1324	-99,9072	-359,5310	297,1439	-359,5838	648,1508	-227,6993
17	-206,6585	-199,7626	29,0409	-33,2467	-52,5436	272,3085	-301,0661	-182,7299	214,9500
18	113,0435	506,9779	-11,2613	-7,0725	-31,1736	41,6341	267,4621	24,4431	-245,8657
19	139,9341	166,6606	373,6104	13,8861	-48,4287	151,6762	-319,1568	-431,1643	-216,8628

Genótipos	Ambientes								
	10	11	12	13	14	15	16	17	18
1	-118,4154	-570,6616	268,4390	-267,8003	-419,2654	-95,9747	206,6726	-18,9791	64,5906
2	-204,3170	192,7810	191,3382	224,1132	-65,5186	31,0124	-215,0235	-131,5348	-157,5997
3	117,8950	183,1737	-32,3431	227,3605	21,3031	84,1185	-27,1749	-10,0489	7,3608
4	12,4879	158,3489	-212,1887	-98,7742	250,0931	-22,0416	-100,1109	79,9794	-32,6797
5	206,2412	-23,8816	-195,2170	-9,3259	-54,7922	44,8267	-6,7598	167,6118	41,2220
6	-163,5047	-212,7891	166,6071	-208,3046	33,1745	-102,8652	175,3953	-121,7303	54,6029
7	244,1500	15,0880	-292,1756	-96,9807	135,4410	19,1999	65,9185	173,2052	94,1281
8	265,5446	-306,9531	-164,8527	-189,4479	-143,1886	-2,2196	213,7787	187,5681	161,9067
9	88,2222	-240,6161	-44,8623	-147,7591	-155,8473	-20,2093	92,0764	105,4046	62,3723
10	-69,3172	337,1742	-59,8088	125,2650	280,6799	18,5803	-140,7669	-69,7320	-76,0457
11	-288,8018	488,3819	312,1376	485,8504	91,0961	77,5962	-269,8624	-309,7129	-204,7837
12	149,5540	-3,8034	-76,8261	66,2301	-65,6063	49,1524	26,0433	72,6321	41,4881
13	-244,8149	-137,0981	218,1948	-86,4513	-119,6981	-68,1808	-36,3891	-103,7576	-74,0517
14	184,7844	-231,8382	-59,9435	-50,8310	-239,3275	27,1146	100,1888	136,3227	81,8381
15	89,0935	171,8480	283,2935	569,3180	-303,3038	172,6977	-11,5228	-162,8691	-14,0463
16	-0,3971	-49,1543	-369,5459	-476,0925	358,2626	-128,8483	-25,3670	196,8544	15,5392
17	183,1780	119,9935	-249,5687	-34,4938	213,4826	21,6951	33,4140	106,0354	65,9718
18	-234,6692	-240,7811	296,9520	-85,8883	-208,3316	-70,0166	51,4236	-133,6862	-33,1904
19	-216,9135	350,7872	20,3702	54,0125	391,3464	-35,6377	-131,9339	-163,5628	-98,6232

APÊNDICE D - Estimativa ($G \times E$) AMMI6 tempo de cozimento

Tabela 20 – Estimativa da interação $G \times E$ modelo AMMI6 para a variável tempo de cozimento

Genótipos	Ambientes								
	1	2	3	4	5	6	7	8	9
1	1,4305	-2,2083	-1,2775	-5,1659	-5,9359	4,8866	-2,9006	-1,1506	2,9572
2	-7,2130	0,9218	2,1176	-4,8383	2,1752	-9,5268	2,6479	3,2008	-0,0335
3	-0,5896	0,6263	0,1516	-5,0395	-1,4213	7,8917	5,9930	0,5638	-2,4831
4	-0,5805	-0,8924	-1,3762	7,4737	0,1707	1,0019	-4,7648	0,8050	-5,5781
5	-0,1447	-1,6769	-0,6220	-3,6032	-3,8492	0,7320	-2,7312	-0,3872	2,1673
6	1,9273	1,6799	2,1780	9,7948	8,4588	-6,1803	-6,0910	0,1837	-9,1228
7	2,3133	0,9423	3,5637	-4,0822	4,3806	-5,1315	-6,2313	-0,5670	-3,2428
8	1,7977	-1,9034	0,5796	-2,7903	-2,7380	-2,5067	1,0355	-2,4098	2,4738
9	-1,8657	1,5777	2,7481	-1,0061	4,8011	-3,1354	0,8930	1,4670	-6,8914
10	-0,1324	3,1475	3,0088	-2,4629	4,9357	3,4248	19,8705	-0,6684	-4,8874
11	-0,2345	0,3860	-1,1635	-6,7530	-3,0340	11,9424	-5,7774	2,1678	-2,3695
12	-2,4104	-1,2850	-3,9843	14,8480	-1,1750	-4,8303	3,9980	-0,0545	4,0531
13	1,7366	-0,9906	-2,4292	-4,4672	-5,7289	6,5882	0,2055	-1,0176	8,1373
14	-1,6711	-0,4766	-0,5332	-2,0250	-1,9075	2,0463	-0,0604	0,8989	-0,4824
15	-4,6009	0,2173	2,7277	-7,0052	0,9973	-3,3607	2,3161	2,0343	-5,3047
16	7,9138	1,1101	-0,7379	6,0716	3,0281	2,8173	-8,8860	-2,4402	2,3089
17	-0,7824	-0,7319	-2,6843	16,5763	1,7630	-4,8718	0,7369	-0,1989	-1,3016
18	-2,2506	-1,4566	-3,5704	0,6590	-5,4625	-7,4149	2,1014	-0,6166	18,3605
19	5,3566	1,0128	1,3035	-6,1846	0,5419	5,6274	-2,3552	-1,8106	1,2392

Genótipos	Ambientes								
	10	11	12	13	14	15	16	17	18
1	-2,7888	-0,2245	1,0315	-0,6975	-2,8822	0,0075	5,2416	1,9048	7,7721
2	6,8629	3,0423	-1,1681	-2,4485	4,5596	-1,0352	-3,3806	2,2074	1,9085
3	-1,4507	-0,6394	1,3984	-1,9280	0,9139	-0,6650	-3,2842	2,1117	-2,1496
4	-3,7688	-0,8300	0,8730	-0,3615	-1,9556	1,2029	0,8484	4,4918	3,2405
5	-1,2335	0,3523	0,3279	-0,7709	-1,6626	0,0403	4,2523	1,5867	7,2227
6	-2,9728	-0,2811	-0,7954	2,2025	1,3829	-0,5369	1,8151	-2,2056	-1,4371
7	-0,2063	1,6160	-0,5252	1,5234	2,3992	-2,8856	5,6880	-3,3144	3,7597
8	-7,4720	-1,4050	-1,8886	-0,8956	-2,9313	1,2676	11,0424	-5,1176	13,8616
9	-0,7319	0,7687	-0,2740	-1,0505	3,2261	-1,4283	-0,3408	0,1626	1,0798
10	-6,7202	-3,1853	-2,0081	-2,6417	2,9670	0,7574	-2,8799	-8,2632	-4,2623
11	6,5341	2,2177	4,7492	0,0289	1,7537	-3,3548	-8,3976	10,3515	-9,0470
12	-1,1852	-2,4170	-2,0489	-0,0214	-3,9160	5,3558	-2,6977	-0,6401	-1,5891
13	4,1029	0,1901	1,3480	0,9214	-1,6938	-0,0784	-2,8542	1,1860	-5,1559
14	0,5970	0,4335	0,8517	-1,2005	0,0007	-0,1450	-1,0977	3,3544	1,4167
15	-1,5317	1,3902	-0,3095	-3,5202	2,8185	-1,3184	2,6186	2,2823	9,5491
16	3,2683	-0,1997	1,0258	6,0760	-0,8724	-1,3624	-1,9729	-3,5574	-13,5910
17	-4,1779	-2,6312	-1,7758	0,5752	-3,4628	4,4761	-0,3727	-0,9738	-0,1624
18	11,7479	1,4417	-2,0163	1,7716	-1,5868	2,4590	-4,7844	-2,5942	-6,7877
19	1,1267	0,3607	1,2043	2,4373	0,9419	-2,7566	0,5564	-2,9729	-5,6286

APÊNDICE E - Estimativa $\hat{Y}$ modelo AMMI4 produtividade

Tabela 21 – Estimativa $\hat{Y}$ modelo AMMI4 para a variável produtividade

Genótipos	Ambientes								
	1	2	3	4	5	6	7	8	9
1	3065,9093	2782,8149	1524,5720	1802,1838	3816,1365	1758,5768	2211,6060	2107,6904	2011,1797
2	3782,4244	2230,2247	2079,2559	2208,3435	4318,8977	1089,3509	2286,7352	2380,9853	2119,4645
3	3137,7106	2046,9582	1733,4405	2044,0980	4206,0962	1559,9724	1899,2017	1700,6572	2497,7970
4	3452,2850	2113,3060	2173,2190	2039,8871	4007,4151	1657,0774	1775,6245	2333,6323	2239,4678
5	3220,5745	1870,2537	1751,1906	1992,4458	4067,8321	1447,8220	1955,1411	2374,7664	2435,5068
6	3084,6623	3089,9484	1951,6196	1960,1929	3939,3277	2413,8856	1900,9590	1511,9335	2233,7702
7	2795,0270	1806,8574	1597,1028	1718,7096	3749,0285	1675,5213	1470,1803	1781,9482	2299,0270
8	2608,0670	2023,0677	1319,8813	1638,5855	3688,0776	1771,2928	1652,1858	1771,1634	2306,8409
9	3071,7866	2201,6893	1632,0265	1848,9457	3882,2637	1608,3101	1944,4381	2179,7360	2223,1092
10	3573,4805	2346,3508	2288,3732	2234,6333	4285,2100	1774,9487	1918,3750	1993,0411	2409,1434
11	3867,9446	2481,6190	2265,6784	2414,6740	4607,8639	1330,7527	2305,2340	1823,8771	2329,4780
12	3138,6807	2076,2196	1707,7852	2003,3217	4115,2353	1593,0575	1962,4809	2011,3778	2462,4839
13	3672,3421	2772,9984	2132,6694	2166,5684	4173,1056	1674,5051	2325,1897	2412,5455	2118,7042
14	3132,8356	2160,8655	1618,6267	1994,9489	4090,6524	1618,5247	2131,8739	2294,1145	2482,9676
15	3253,0003	2286,7995	1568,0468	2306,7335	4660,1635	1455,6498	2426,3140	1502,2580	2805,1273
16	3397,3611	2246,0355	2303,7784	1883,8002	3682,0229	1943,7899	1571,1412	2631,2246	2039,3853
17	3031,3305	2082,5685	1875,6702	1950,4440	3988,9937	1918,9378	1629,6422	1800,3271	2482,0179
18	3452,4188	2890,6953	1936,7543	2078,0045	4111,7500	1789,6498	2299,5567	2108,8865	2122,5885
19	3443,8256	2514,8943	2286,1423	2063,4794	4059,0111	1864,2082	1677,4541	1617,7954	2116,1077

Genótipos	Ambientes								
	10	11	12	13	14	15	16	17	18
1	2133,6463	2152,6055	3458,3904	3036,9242	1844,8087	1950,5469	1294,8783	653,3845	1073,9893
2	2295,7067	3164,0102	3629,2515	3776,7997	2446,5176	2325,4960	1121,1443	788,7909	1099,7610
3	2510,8791	3047,3633	3298,5306	3673,0074	2426,2997	2271,5625	1201,9532	803,2371	1157,6820
4	2453,1889	3070,2553	3166,4019	3394,5896	2702,8065	2213,1192	1176,7341	940,9823	1165,3582
5	2586,7677	2827,8504	3123,1992	3423,8635	2337,7468	2219,8131	1209,9108	968,4403	1179,0856
6	2263,8470	2685,7681	3531,8485	3271,7099	2472,5387	2118,9463	1438,8910	725,9233	1239,2916
7	2398,6675	2640,8109	2800,2315	3110,1995	2301,9709	1968,1771	1056,5800	748,0245	1005,9825
8	2370,0534	2268,7611	2877,5457	2967,7236	1973,3325	1896,7490	1154,4314	712,3787	1023,7524
9	2362,7074	2505,0745	3167,5126	3179,3889	2130,6504	2048,7357	1202,7057	800,1918	1094,1945
10	2520,5567	3398,2536	3467,9547	3767,8017	2882,5663	2402,9140	1285,2510	940,4438	1271,1653
11	2372,1208	3620,5099	3910,9499	4199,4358	2764,0311	2532,9787	1227,2043	771,5117	1213,4759
12	2534,9747	2852,8228	3246,4843	3504,3137	2331,8269	2229,0330	1247,6081	878,3548	1184,2458
13	2303,5919	2882,5142	3704,4914	3514,6184	2440,7212	2274,6859	1348,1619	864,9513	1231,6922
14	2586,2287	2640,8115	3279,3904	3403,2761	2174,1293	2223,0187	1337,7771	958,0689	1240,6194
15	2666,6359	3220,5959	3798,7256	4199,5232	2286,2511	2544,7000	1402,1637	834,9753	1320,8332
16	2372,8329	2795,2812	2941,5737	2949,8003	2743,5051	2038,8415	1184,0071	990,3863	1146,1063
17	2556,3914	2964,4123	3061,5343	3391,3824	2598,7084	2189,3682	1242,7714	899,5507	1196,5222
18	2239,9305	2705,0240	3709,4413	3441,3742	2278,2806	2199,0429	1362,1673	761,2154	1198,7463
19	2222,2024	3261,1087	3397,3758	3545,7913	2842,4748	2197,9381	1143,3261	695,8551	1097,8297

APÊNDICE F - Estimativa $\hat{Y}$ modelo AMMI6 tempo de cozimento

Tabela 22 – Estimativa $\hat{Y}$ modelo AMMI6 para a variável tempo de cozimento

Genótipos	Ambientes								
	1	2	3	4	5	6	7	8	9
1	25,8860	30,9178	33,1586	29,2517	22,2804	37,6071	35,4081	24,9039	41,5392
2	15,7764	32,5817	35,0877	28,1132	28,9255	21,7277	39,4905	27,7891	37,0824
3	19,5531	29,4395	30,2750	25,0653	22,4823	36,2995	39,9889	22,3054	31,7862
4	19,7898	28,1485	28,9748	37,8061	24,3019	29,6373	29,4587	22,7743	28,9187
5	19,0302	26,1686	28,5336	25,5338	19,0866	28,1720	30,2969	20,3867	35,4688
6	23,4465	31,8696	33,6779	41,2761	33,7389	23,6040	29,2813	23,3019	26,5229
7	20,5723	27,8718	31,8034	24,1389	26,4005	21,3926	25,8808	19,2909	29,1427
8	24,9860	29,9555	33,7485	30,3601	24,2111	28,9467	38,0769	22,3775	39,7886
9	19,2583	31,3722	33,8528	30,0800	29,6860	26,2537	35,8702	24,1900	28,3591
10	24,1338	36,0843	37,2557	31,7654	32,9628	35,9562	57,9899	25,1968	33,5054
11	21,4912	30,7822	30,5429	24,9348	22,4525	41,9331	29,8015	25,4925	33,4827
12	18,9501	28,7460	27,3569	46,1706	23,9464	24,7952	39,2116	22,9049	39,5401
13	21,3268	27,2700	27,1416	25,0850	17,6221	34,4434	33,6487	20,1714	41,8539
14	19,1827	29,0478	30,3013	28,7909	22,7072	31,1653	34,6466	23,3517	34,4979
15	20,5753	34,0641	37,8846	28,1331	29,9343	30,0806	41,3455	28,8095	33,9980
16	28,0825	29,9493	29,4115	36,2024	26,9576	31,2510	25,1359	19,3274	36,6040
17	20,6779	29,3989	28,7567	47,9987	26,9841	24,8536	36,0504	22,8604	34,2852
18	19,6368	29,1013	28,2976	32,5085	20,1857	22,7375	37,8419	22,8697	54,3743
19	24,8578	29,1846	30,7854	23,2788	23,8040	33,3937	30,9992	19,2896	34,8670

Genótipos	Ambientes								
	10	11	12	13	14	15	16	17	18
1	28,6397	25,0433	24,2284	31,6694	23,3819	36,3397	39,3008	45,5061	49,4114
2	36,8253	26,8440	20,5627	28,4523	29,3575	33,8308	29,2125	44,3426	42,0818
3	25,6650	20,3155	20,2825	26,1261	22,8652	31,3543	26,4622	41,4002	35,1770
4	23,5745	20,3526	19,9847	27,9202	20,2233	33,4499	30,8224	44,0079	40,7946
5	24,9144	20,3395	18,2443	26,3155	19,3209	31,0919	33,0309	39,9074	43,5815
6	25,5194	22,0504	19,4652	31,6331	24,7107	32,8589	32,9380	38,4593	37,2659
7	25,0257	20,6873	16,4752	27,6938	22,4668	27,2501	33,5507	34,0904	39,2026
8	22,6892	22,5956	20,0411	30,2041	22,0655	36,3325	43,8343	37,2165	54,2338
9	27,3651	22,7050	19,5914	27,9850	26,1588	31,5724	30,3869	40,4324	39,3877
10	24,5190	21,8932	20,9995	29,5359	29,0419	36,9003	30,9901	35,1488	37,1878
11	35,2328	24,7556	25,2163	29,6660	25,2879	30,2475	22,9318	51,2230	29,8625
12	27,1483	19,7558	18,0530	29,2504	19,2531	38,5930	28,2665	39,8661	36,9552
13	30,6660	20,5925	19,6795	28,4229	19,7049	31,3884	26,3396	39,9219	31,6181
14	28,4238	22,0996	20,4469	27,5648	22,6631	32,5855	29,3598	43,3540	39,4543
15	30,6175	27,3787	23,6081	29,5674	29,8033	35,7344	37,3985	46,6043	51,9092
16	30,4099	20,7812	19,9358	34,1560	21,1048	30,6829	27,7995	35,7570	23,7615
17	24,2554	19,6414	18,4259	29,9469	19,8061	37,8131	30,6913	39,6323	38,4818
18	40,6082	24,1413	18,6125	31,5704	22,1092	36,2230	26,7066	38,4390	32,2835
19	27,6009	20,6742	19,4469	29,8499	22,2517	28,6212	29,6613	35,6741	31,0565

APÊNDICE G - Programa para a análise AMMI para uma variável

```

dm 'log;clear;output;clear';
data dados;
input gen amb y1;
cards;
1 1 3143.3333
1 2 2630.8333
1 3 1466.6667
.
.
.
19 16 1181.6667
19 17 673.3333
19 18 1133.6667
;
proc print; run;

*****
*   Análise univariada para verificar normalidade variável resposta y1   *
*****;
proc univariate data=dados normal plot;
var y1;
run;

*****
*   Ajuste dos efeitos principais por ANOVA variável resposta y1   *
*****;
proc glm data=dados;
class gen amb;
model y1=gen amb;
lsmeans gen amb;
output out=geam1 R=GE1;
run;

```

```

*****
*   Análise univariada para verificar normalidade variável resposta y1   *
*****;

proc univariate data=geam1 normal plot;
var GE1;
run;

*****

*   Geração da matriz GXE variável resposta y1   *
*****;

proc sort data=dados;
by gen;
proc transpose data=dados out=medias1 (drop=_name_) prefix=A;
var y1;
by gen;
run;

proc print data=medias1 (drop=gen);
run;

proc sort data=dados;
by gen;
proc transpose data=geam1 out=GXE1 prefix=A;
var GE1;
by gen;
run;

proc print data=GXE1 (drop=gen);
run;

*****

*   Dados em forma matricial variável resposta y1   *
*****;

proc iml;
use work.medias1;
setin work.medias1;

```

```

read all var{A1 A2 A3 A4 A5 A6 A7 A8 A9 A10 A11 A12 A13 A14 A15
          A16 A17 A18} into X1;
print X1;

*****
*           Matriz GXE e somas de quadrados           *
*****;

ng=nrow(X1);
na=ncol(X1);
gm=X1[, :];
am=X1[ : ,];
gg=X1[ : ,];
XG=gm*J(ncol(gm),na,1);
XA=J(ng,nrow(am),1)*am;
MG=J(ng,na,1)*gg;
GXEY1=X1-XG-XA+MG;
corr=(ng*na)*(gg##2);
SQTRY1=SSQ(X1)-corr;
SQGENY1=na*(SSQ(gm-gg));
SQAMBY1=ng*(SSQ(am-gg));
SQGXEY1=SSQ(GXEY1);
GLGENY1=ng-1;
GLAMBY1=na-1;
GLGXEY1=GLGENY1*GLAMBY1;
QMGENY1=SQGENY1/GLGENY1;
QMAMBY1=SQAMBY1/GLAMBY1;
QMGXEY1=SQGXEY1/GLGXEY1;
FGENY1=QMGENY1/QMGXEY1;
VALORP_FGENY1=1-PROBF(FGENY1,GLGENY1,GLGXEY1);
FAMBY1=QMAMBY1/QMGXEY1;
VALORP_FAMBY1=1-PROBF(FAMBY1,GLAMBY1,GLGXEY1);
GLMODY1=(ng-1)+(na-1)+(ng-1)*(na-1);
print GXEY1, gm gg, am, QMGENY1 QMAMBY1 glgxeY1 SQGXEY1 QMGXEY1,
      FGENY1 VALORP_FGENY1, FAMBY1 VALORP_FAMBY1;

```

```

*****
*                               DVS da matriz GXE                               *
*****;
EIGV=EIGVAL(GXEY1*t(GXEY1));
NNUL1=EIGV<-1E-9;
NNUL2=EIGV>1E-9;
NNUL=NNUL1+NNUL2;
R=(NNUL^=0)[+,];
CALL SVD(EG,VS,EA,GXEY1);
S=DIAG(VS[1:R,]);
U=EG[,1:R];
VT=EA[,1:R];

*****
* Desdobramento da SQ(GXE) por AMMI através de DVS *
*****;
VS2=VS##2;
AUTV=VS2[1:R];
SAUTV=SUM(VS2);
PRP=INV(SAUTV)*VS2;
PROP=PRP[1:R];

RES1=SAUTV-VS2[1:1,];
PACUM=CUSUM(PROP)*100;
LINH=T(1:R);
RESUMY1=LINH|AUTV|PROP|PACUM;
COLU={'CP/EIXO' 'AUTOVALOR' 'PROP./CP' '%ACUMULADA'};
print RESUMY1[COLNAME=COLU format=12.4];

*****
*                               Teste de Gollob                               *
*****;
alpha=0.05;
glerr=648;
qmerr=36233.24219;
resulGolY1=J(R+1,5,0);

```

```

do i=0 to R by 1;
if i=0 then GLGolY1=GLGENY1*GLAMBY1;
else GLGolY1=(ng+na-1-(2*i));
if i=0 then Gol1=sum(VS2[i+1:R]);
else Gol1=(VS2[i]);
Gol2=glgoly1*qmerr;
QMGolY1=gol1/Glgoly1;
FGolY1=Gol1/Gol2;
valpGolY1=1-probf(FGolY1,GLGolY1,GLerr);
resulGolY1[i+1,1]=i;
resulGolY1[i+1,2]=GLGolY1;
resulGolY1[i+1,3]=QMGolY1;
resulGolY1[i+1,4]=FGolY1;
resulGolY1[i+1,5]=valpGolY1;

/*print i glgoly1 qmgoly1 fgoly1 valpgoly1[format=12.4];*/
end;

valorespGolY1=resulGolY1[,5]<alpha;
numcompGolY1=sum(valorespGolY1)-1;

print 'Resultados teste Gollob', resulGolY1[format=12.4];
print valorespGolY1 numcompGolY1;
/*print R, VS2;*/

*****
*           Teste de Cornelius           *
*****;

alpha=0.05;
glerr=648;
qmerr=36233.24219;
resulCorY1=J(R,5,0);

do i=0 to R-1 by 1;
GLCorY1=(ng-1-i)*(na-1-i);

```

```

Cor1=sum(VS2[i+1:R]);
Cor2=glcory1*qmerr;
QMCory1=cor1/GLcorY1;
FRCory1=Cor1/Cor2;
valpCorY1=1-probf(FRCory1,GLCorY1,GLerr);
resulCorY1[i+1,1]=i;
resulCorY1[i+1,2]=GLCorY1;
resulCorY1[i+1,3]=QMCory1;
resulCorY1[i+1,4]=FRCory1;
resulCorY1[i+1,5]=valpCorY1;
/*print i glcory1 qmcory1 frcory1 valpcory1[format=12.4];*/
end;

valorespCorY1=resulCorY1[,5]<alpha;
numcompCorY1=sum(valorespCorY1);

print 'Resultados teste Cornelius', resulCorY1[format=12.4];
print valorespCorY1 numcompCorY1;

*****
*Validação Cruzada - Eastment Krzanowski (1982)*
*****;
*****
*      Decomposição DVS da matriz GXE      *
*****;
CALL SVD(EG1,VS1,EA1,GXEY1);
SVC=DIAG(VS1[1:na,]);
UVC=EG1[,1:na];
VTVC=t(EA1[,1:na]);
AVC=UVC*SVC*VTVC;

*****
* Cria matrizes m, Precorr, Press_m, Dr_m, Dm_m e Wm de zeros *
*****;
z=min(na,ng)-1;
mY1=shape(0,z,1);

```

```

PRECORR_mY11=shape(0,z,1);
Press_mY1=shape(0,z,1);
Dr_mY1=shape(0,z,1);
Dm_mY1=shape(0,z,1);
WmY1=shape(0,z,1);

do m=1 to z;
XEY1=SHAPE(0,ng,na);
*****
* Submatriz GXE_i com eliminação da i-ésima linha de GXE *
*****;
do i=1 to ng;
  if i=1 then GXE_i=GXEY1[2:ng,];
  else if i=ng then GXE_i=GXEY1[1:ng-1,];
  else GXE_i=GXEY1[1:i-1,]/GXEY1[i+1:ng,];
  GXEce_i=GXE_i-(j(ng-1,1,1)*GXE_i[:,]);
*****
*Decomposição de valor singular com os valores de GXE_i centrados na média*
*****;
call SVD (EG2,VS2,EA2,GXEce_i);
Sbarra=DIAG(VS2);
Ubarra=EG2[,1:na];
VTbarra=t(EA2[,1:na]);
A_i=Ubarra*Sbarra*VTbarra;
*****
*Troca de sinais da matriz VTbarra para que fiquem iguais aos da matriz VTVC*
*****;
SUU=0; SUUT=0;
DO III=1 TO na;
  DO JJJ=1 TO na;
    IF VTVC[III,JJJ]<0 THEN SUU=-1;
    ELSE IF VTVC[III,JJJ]>0 THEN SUU=1;
    IF VTbarra[III,JJJ]<0 THEN SUUT=-1;
    ELSE IF VTbarra[III,JJJ]>0 THEN SUUT=1;
    IF SUU ^= SUUT THEN VTbarra[III,JJJ]=-VTbarra[III,JJJ];
  end;
end;

```



```

end;

*****
* Submatriz GXE_j com eliminação da j-ésima coluna de GXE *
*****;

do j=1 to na;
  if j=1 then GXE_j=GXEY1[,2:na];
  else if j=na then GXE_j=GXEY1[,1:na-1];
  else GXE_j=GXEY1[,1:j-1]||GXEY1[,j+1:na];
  GXEce_j=GXE_j-(j(ng,1,1)*GXE_j[:,]);
*****
*Decomposição de valor singular com os valores de GXE_j centrados na média*
*****;
call SVD (EG3,VS3,EA3,GXEce_j);
Stil=DIAG(VS3);
Util=EG3[,1:na-1];
VTtil=t(EA3[,1:na-1]);
A_j=Util*Stil*VTtil;
*****
* Troca de sinais da matriz Util para que fiquem iguais aos da matriz UVC *
*****;
SU=0; SUT=0;
DO II=1 TO ng;
DO JJ=1 TO (na-1);
IF UVC[II,JJ]<0 THEN SU=-1;
ELSE IF UVC[II,JJ]>0 THEN SU=1;
IF Util[II,JJ]<0 THEN SUT=-1;
ELSE IF Util[II,JJ]>0 THEN SUT=1;
IF SU ^= SUT THEN Util[II,JJ]=-Util[II,JJ];
END;
END;

*****
*Cálculo dos valores preditos XE*
*****;

do t=1 to m;
XEY1[i,j]=XEY1[i,j]+Util[i,t]*sqrt(Stil[t,t])*sqrt(SBarra[t,t])*VTBarra[t,j];
end; /* Fim do do para t */

```

```

    end; /* Fim do do para j */
end; /* Fim do do para i */

XXi=j(ng,na,0); do iiii=1 to ng; XXi[iiii,]=j(1,na,X1[iiii,:]); end;
XXj=j(ng,na,0); do jjjj=1 to na; XXj[,jjjj]=j(ng,1,X1[:,jjjj]); end;
XXbarra=j(ng,na,X1[:,:]);
Xchapeu=XEY1+XXi+XXj-XXbarra;

*****
*   Correlação   *
*****;
PRECORRY1=corr(colvec(GXEY1)||colvec(XEY1));
PRECORR_mY11[m]=PRECORR_mY11[m]+PRECORRY1[1,2];

*****
*   Função PRESS_m   *
*****;
diff=XEY1-GXEY1;
Press_mY1[m]=(1/(ng*na))*sum((XEY1-GXEY1)##2);

*****
*   Função W_m   *
*****;
Dm_mY1[m]=ng+na-(2*m);
if m=1 then
Dr_mY1[m]=(ng-1)*na;
else
Dr_mY1[m]=Dr_mY1[m-1]-(ng+na-((m-1)*2));

if m=1 then
    WmY1[m]=((((1/(ng*na))*sum((GXEY1)##2))-Press_mY1[m])/Dm_mY1[m])/
                (Press_mY1[m]/Dr_mY1[m]);
else WmY1[m]=((Press_mY1[m-1]-Press_mY1[m])/Dm_mY1[m])/(Press_mY1[m]/Dr_mY1[m]);

mY1[m]=mY1[m]+m;

```

```

end; /* Fim do do para m */

print 'Resultados de Validação Cruzada Eastment Krzanowski (1982)',
      mY1 Press_mY1 Precorr_mY11 WmY1 Dm_mY1 Dr_mY1;

*****
*   Respostas preditas para o modelo selecionado pelo teste de Cornelius   *
*****;
GEIY1=J(ng,na,0);

do j=1 to numcompCorY1 by 1;
  GEIY1j=S[j,j]*U[,j]*T(VT[,j]);
  GEIY1=GEIY1+GEIY1j;
end;

XPCorY1=XG+XA-MG+GEIY1;

print 'Matriz de aproximação DVS para a matriz GXE do modelo selecionado
      pelo teste de Cornelius', GEIY1[format=12.4];
print 'Matriz de respostas preditas segundo o modelo selecionado pelo teste
      de Cornelius', XPCorY1[format=12.4];

*****
*                               Marcadores para genótipos e ambientes          *
*****;
marcagenY1=U*SQRT(S);
marcaambY1=SQRT(S)*t(VT);
col1={'Ammi1' 'Ammi2' 'Ammi3' 'Ammi4' 'Ammi5' 'Ammi6' 'Ammi7' 'Ammi8' 'Ammi9'
      'Ammi10' 'Ammi11' 'Ammi12' 'Ammi13' 'Ammi14' 'Ammi15' 'Ammi16' 'Ammi17'};
lin1={'Ammi1' 'Ammi2' 'Ammi3' 'Ammi4' 'Ammi5' 'Ammi6' 'Ammi7' 'Ammi8' 'Ammi9'
      'Ammi10' 'Ammi11' 'Ammi12' 'Ammi13' 'Ammi14' 'Ammi15' 'Ammi16' 'Ammi17'};
print 'Marcadores para genótipos e ambientes',
      'Marcadores para genótipos' marcagenY1 [colname=col1 format=12.4],
      'Marcadores para ambientes' marcaambY1 [rowname=col1 format=12.4]/;

create marcagenY11 from marcagenY1[colname=col1];

```

```

append from marcagenY1;
close marcagenY1;

create marcaambY11 from marcaambY1[colname=col1];
append from marcaambY1;
close marcaambY11;

*****
*                               *
*                               *
*****;

GENY1=1:ng;
AMBY1=1:na;
OBJ=t(GENY1||AMBY1);
PROD=GM//t(AM);
IPCA1=marcagenY1[,1]//t(marcaambY1[1,]);
IPCA2=marcagenY1[,2]//t(marcaambY1[2,]);
IPCA3=marcagenY1[,3]//t(marcaambY1[3,]);
IPCA4=marcagenY1[,4]//t(marcaambY1[4,]);
ordgenY1=ng-rank(gm)+1;
ordambY1=na-t(rank(am))+1;
ordmedY1=ordgenY1//ordambY1;
biplotammi1Y1=obj||prod||ipca1||ordmedY1;
biplotammi2Y1=obj||prod||ipca1||ipca2||ordmedY1;
biplotammi3Y1=obj||prod||ipca1||ipca2||ipca3||ordmedY1;
biplotammi4Y1=obj||prod||ipca1||ipca2||ipca3||ipca4||ordmedY1;

var1={'Gen_Amb' 'Medias' 'Ipca1' 'Ord_Med'};
var2={'Gen_Amb' 'Medias' 'Ipca1' 'Ipca2' 'Ord_Med'};
var3={'Gen_Amb' 'Medias' 'Ipca1' 'Ipca2' 'Ipca3' 'Ord_Med'};
var4={'Gen_Amb' 'Medias' 'Ipca1' 'Ipca2' 'Ipca3' 'Ipca4' 'Ord_Med'};

print biplotammi1Y1[colname=var1 format=12.4],
      biplotammi2Y1[colname=var2 format=12.4],
      biplotammi3Y1[colname=var3 format=12.4],
      biplotammi4Y1[colname=var4 format=12.4];
quit;

```

APÊNDICE H - Programa para realizar a análise de procrustes

```

*****
*   Análise Procrustes para Genótipos   *
*****;

proc iml;
use work.marcagenY11;
read all into marcagenY1;
use work.marcagenY22;
read all into marcagenY2;

filmarY1=nrow(marcagenY1);
filmarY2=nrow(marcagenY2);

marcagenY1=marcagenY1[,1:4];
marcagenY2=marcagenY2[,1:6];

*****
*   Matrizes padronizadas   *
*****;

marcagenY1padr=standard(marcagenY1);
marcagenY1padr=marcagenY1padr||j(filmarY1,2,0);
marcagenY2padr=standard(marcagenY2);

*****
* Decomposição do valor singular *
*****;

Zpro=t(marcagenY1padr)*marcagenY2padr;
call svd(Vpro, Dpro, Upro,Zpro);
Dproc=diag(Dpro);
Q=Vpro*t(Upro);

R2proGen=fuzz(trace(marcagenY1padr*t(marcagenY1padr))+
              trace(marcagenY2padr*t(marcagenY2padr))-(2*trace(Dproc)));
print R2proGen;

```

```

*****
*   Análise Procrustes para Ambientes   *
*****;

proc iml;
  use work.marcaambY11;
  read all into marcaambY1;
  use work.marcaambY22;
  read all into marcaambY2;

  filmarY1=nrow(marcaambY1);
  filmarY2=nrow(marcaambY2);

  marcaambY1=marcaambY1[,1:4];
  marcaambY2=marcaambY2[,1:6];

  *****
  *   Matrizes padronizadas   *
  *****;
  marcaambY1padr=standard(marcaambY1);
  marcaambY1padr=marcaambY1padr||j(filmarY1,2,0);
  marcaambY2padr=standard(marcaambY2);

  *****
  * Decomposição do valor singular *
  *****;
  Zpro=t(marcaambY1padr)*marcaambY2padr;
  call svd(Vpro, Dpro, Upro,Zpro);
  Dproc=diag(Dpro);
  Q=Vpro*t(Upro);

  R2proAmb=fuzz(trace(marcaambY1padr*t(marcaambY1padr))+
                trace(marcaambY2padr*t(marcaambY2padr))-(2*trace(Dproc)));
  print R2proAmb;
  quit;

```