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

Os Métodos Biplot e Escalonamento Multidimensional nos
delineamentos experimentais

Édila Cristina de Souza

Tese apresentada para obtenção do título de Doutor
em Ciências. Área de concentração: Estatística e
Experimentação Agronômica

Piracicaba
2010

Édila Cristina de Souza
Licenciada em Matemática

**Os Métodos Biplot e Escalonamento Multidimensional nos
delineamentos experimentais**

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

Tese apresentada para obtenção do título de Doutor em
Ciências. Área de concentração: Estatística e Experi-
mentação Agronômica

**Piracicaba
2010**

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

Souza, Édila Cristina de
Os Métodos Biplot e Escalonamento Multidimensional nos delineamentos experimentais / Édila
Cristina de Souza. - - Piracicaba, 2010.
134 p. : il.

Tese (Doutorado) - - Escola Superior de Agricultura "Luiz de Queiroz", 2010.

1. Análise multivariada 2. Correlação genética e ambiental 3. Escalonamento multidimensional
4. Estabilidade 5. Fenótipos 6. Genética 7. Melão I. Título

CDD 635.61
S725m

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

Dedicatória

*Espera no senhor
 Mesmo quando a vida pedir de ti mais do que podes dar
 E o cansaço já fizer teu passo vacilar
 Mesmo se a solidão teu peito machucar
 E te der vontade de ir embora e tudo abandonar
 Espera no senhor.
 Há um Deus que te ama e ele tudo pode transformar
 Seu amor te sustentará, espera n'Ele
 E ele tudo fará, tudo fará. (Eliana Ribeiro)*

A
DEUS

força maior de todo ser humano.

Aos meus pais,
Adélia Catarina Souza e
Benedito Rondon Moraes Souza (in memoriam),

Aos meus irmãos Ronaldo e Nádia e aos meus sobrinhos Mônica, Pedror e Isabelly.

A minha amiga Greice Ximena, pelo carinho e incentivo.
 Agradeço sua energia que, positivamente, muitas batalhas me ajudou a ganhar.

AGRADECIMENTOS

Ao Professor Dr. Carlos Tadeu dos Santos pela orientação, a paciência, os ensinamentos, a amizade, confiança e o constante estímulo e apoio para a realização deste trabalho. Sua conduta sempre firme e sincera me fez enxergar os requisitos que a vida acadêmica exige de um profissional.

Aos professores do Departamento de Ciências Exatas da ESALQ/USP, Dr. César Gonçalves de Lima, Dra. Clarice Garcia Borges Demétrio, Dr. Décio Barbin, Dr. Edwin Ortega, Dr. Gabriel Adrián Sarriés, Dr. Gerson Barreto Mourão, Dra. Maria Cristina Stolf Nogueira, Dra. Roseli Aparecida Leandro, Dr. Silvio Sandoval Zocchi, Dra. Sônia Maria de Stefano Piedade e Dr. Vitor Ozaki.

Aos funcionários do Departamento de Ciências Exatas da ESALQ/USP, Jorge Alexandre Wiendl, Luciane Brajão, Solange de Assis Paes Sabadin, Eduardo Bonilha, Expedita Maria de Azevedo e Rosni Pinto, pela paciência, dedicação e disposição, propiciando condições para a realização deste trabalho.

Aos funcionários das bibliotecas Central, da Economia e da Genética, pela dedicação e auxílio na pesquisa bibliográfica deste trabalho.

Aos funcionários da segurança do campus e da limpeza do LCE, pelo incentivo incansável durante a minha permanência.

Aos amigos e colegas da ESALQ e da UFMT, agradeço pela ajuda, apoio, compreensão e amizade.

A todos que, direta ou indiretamente contribuíram para a realização deste trabalho, o meu MUITO OBRIGADA!!!

SUMÁRIO

RESUMO	9
ABSTRACT	11
LISTA DE FIGURAS	13
LISTA DE TABELAS	15
1 INTRODUÇÃO	17
2 REVISÃO BIBLIOGRÁFICA	19
2.1 Fatos históricos	19
2.1.1 Estrutura dos dados para o método Biplot	23
2.2 Definição do Biplot	23
2.2.1 Decomposição em Valores Singulares (DVS)	24
2.2.2 Enfoque Clássico	25
2.2.3 Construção do Biplot	29
2.2.4 Tipos de Biplots	35
2.2.5 Qualidade da aproximação do Biplot	43
2.3 Interação Genótipo-Ambiente ($G \times A$)	44
2.3.1 Regras de interpretação do Biplot para interação ($G \times A$)	48
2.4 MANOVA Biplot	50
2.4.1 MANOVA	51
2.4.2 MANOVA dupla entrada	54
2.4.3 MANOVA Biplot de dupla entrada	58
2.5 Análise de Distância Multifatorial	62
2.5.1 Escalonamento Multidimensional	62
3 MATERIAIS E MÉTODOS	65
3.1 Características dos dados utilizados	65
3.2 Metodologias de análises propostas	66
3.2.1 Modelo AMMI	66
3.2.1.1 Biplot AMMI	71
3.2.2 Modelo SREG	73
3.2.2.1 GGEbiplot	75

3.2.2.2 Transformação dos dados para GGEbiplot	76
3.3 Escalonamento Multidimensional	77
3.3.1 Algoritmos MDS	79
3.3.1.1 Escalonamento clássico	79
3.3.2 Funções e modelos MDS	81
3.3.3 Função <i>Stress</i>	82
3.3.4 Matrizes de Distância	82
3.3.5 Biplots	83
4 RESULTADOS	85
4.1 Análise da variável Produção utilizando os modelos AMMI e SREG	85
4.2 Análise da variável SST os modelos AMMI e SREG	91
4.3 Análise da variável Produção utilizando a técnica MDS	99
4.4 Análise da variável SST utilizando a técnica MDS	103
5 CONCLUSÕES	109
REFERÊNCIAS	111
ANEXOS	117

RESUMO

Os Métodos Biplot e Escalonamento Multidimensional nos delineamentos experimentais

O objetivo deste trabalho foi avaliar os métodos estatísticos de análise da interação de genótipos com ambientes ($G \times A$), enfatizando a adaptabilidade e a estabilidade fenotípica. As variáveis estudadas foram produção e teor de sólidos solúveis totais (SST) do melão do tipo Galia, testando 9 genótipos em 12 ambientes. O experimento foi conduzido no delineamento aleatorizado em blocos com 3 repetições, realizado no Pólo Agroindustrial Mossoró-Assu no Rio Grande do Norte. O desempenho dos cultivares foi analisado por meio da utilização de análises de variância, metodologias de adaptabilidade e estabilidade. Realizou-se as análises para a produção e o teor de sólidos solúveis, utilizando as metodologias AMMI (Additive Main Effects and Multiplicative Interaction) e SREG (Sites Regression), representando graficamente de forma simultânea os genótipos e ambientes através dos gráficos Biplot AMMI, GGEbiplot e Trilinear plot. A análise AMMI possui a vantagem de estudar detalhadamente a estrutura do efeito de interação, além de representar simultaneamente os escores dos efeitos da interação para cada fator. Na análise SREG, incorpora o efeito de genótipo e na maioria dos casos está altamente correlacionado com os escores do primeiro componente principal, possui a vantagem de permitir a avaliação gráfica direta do efeito de genótipo. Propõe-se, também a metodologia MDS (Multidimensional Scalling) para verificar as similaridades e dissimilaridades entre os ambientes, através de uma matriz de distâncias, representando geometricamente os dados no espaço bidimensional (Biplot) para cada variável estudada, em que pode-se observar as disparidades entre os ambientes, mostrando que esses apresentam características diferentes.

Palavras-chave: Biplot; Interação $G \times A$; Análise Multivariada; Estabilidade; Adaptabilidade

ABSTRACT

The Biplot Methods and Multidimensional Scaling in experimental designs

The objective of this study was to evaluate statistical methods of analysis of the interaction of genotypes with environments ($G \times A$), emphasizing the adaptability and stability phenotype. The variables studied were production and soluble solids contents (SST) Melon Galia type, testing 9 genotypes in 12 environments. The experiment was conducted in a randomized block with 3 replications, it was done at Pole Agroindustrial Mossoró-Assu in Rio Grande do Norte. The performance of cultivars was analyzed by using analysis of variance, methods of adaptability and stability. It carried out the analysis for the production and soluble solids, using the methodologies AMMI (Additive Main Effects and Multiplicative Interaction) and SREG (Sites Regression), graphing simultaneously the genotypes and environments through the AMMI Biplot graphs, GGE Biplot and trilinear plot. The AMMI analysis has the advantage of studying in detail the structure of the interaction effect, and represents both the scores of the interaction effects for each factor. The analysis SREG, incorporates the effect of genotype and in most cases is highly correlated with the scores of the first principal component, it has the advantage of allowing direct graphical assessment of the effect of genotype. It was also proposed the methodology MDS (Multidimensional Scalling) to check the similarities and dissimilarities between the environments, through a distance matrix, representing geometrically the data in two-dimensional space (Biplot) each variable studied, in wich one can be observed disparities environmental show different characteristics.

Keywords: Biplot; Interaction $G \times E$; Multivariate analysis; Stability; Adaptability

LISTA DE FIGURAS

Figura 1 - Projeção do vetor $\mathbf{a}_i$ sobre o vetor $\mathbf{b}_j$	28
Figura 2 - Produto interno para cada elemento da matriz $\mathbf{Y}$ de posto 2	30
Figura 3 - Produto interno para cada elemento de uma matriz de posto 2	31
Figura 4 - Comparação entre os elementos da primeira coluna da matriz produto, visualizando as projeções sobre o marcador B_1	33
Figura 5 - Comparação entre os elementos da primeira linha da matriz produto, visualizando as projeções sobre o marcador A_1	34
Figura 6 - Cosseno dos ângulos entre os vetores colunas representando a estimativa do coeficiente de correlação entre os vetores dos marcadores de coluna	35
Figura 7 - Cosseno dos ângulos entre os vetores linhas representando a estimativa do coeficiente de correlação entre os vetores dos marcadores de linha	36
Figura 8 - Foto do melão do tipo Galia	66
Figura 9 - Biplot AMMI com os dois primeiros componentes principais correspondente à representação dos nove genótipos (G) e doze ambientes (A) para a variável Produção	88
Figura 10 - Biplot AMMI utilizando as médias ajustadas correspondente à representação dos nove genótipos (G) e doze ambientes (A) da variável Produção	89
Figura 11 - Gráfico Trilinear construído pelos três escores do modelo AMMI para a variável Produção, representando os nove genótipos (G) e doze ambientes (A)	90
Figura 12 - GGEbiplot correspondente à representação dos nove genótipos (G) e doze ambientes (A) para a variável Produção	92
Figura 13 - Biplot AMMI com os dois primeiros componentes principais correspondente à representação dos nove genótipos (G) e doze ambientes (A) da variável SST	95
Figura 14 - Biplot AMMI utilizando as médias ajustadas correspondente à representação dos nove genótipos (G) e doze ambientes (A) para a variável SST	97
Figura 15 - Gráfico Trilinear construído pelos três escores do modelo AMMI para a variável SST, representando os nove genótipos (G) e doze ambientes (A)	98
Figura 16 - GGE Biplot correspondente à representação dos nove genótipos (G) e doze ambientes (A) para a variável SST	99

Figura 17 -Diagrama de Shepard para a variável produção	102
Figura 18 -Biplot com Escalonamento Multidimensional dos nove genótipos (G) e doze ambientes (A) para a variável produção	104
Figura 19 -Diagrama de Shepard para a variável SST	105
Figura 20 -Biplot com Escalonamento Multidimensional dos nove genótipos (G) e doze ambientes (A) para a variável SST	107

LISTA DE TABELAS

Tabela 1 - Resumo das características mais importantes dos Biplots considerados na representação da matriz de dados $\mathbf{Y}$ com posto (q)	41
Tabela 2 - Principais modelos utilizados na interação $G \times A$	47
Tabela 3 - Esquema das observações em tabelas de dupla entrada	54
Tabela 4 - Elementos para MANOVA de dupla entrada	58
Tabela 5 - Distribuição de parâmetros da maior raiz característica	59
Tabela 6 - Disposição dos dados em linhas (genótipos) e colunas (ambientes)	67
Tabela 7 - Quadro resumo da análise conjunta de variância obtida a partir das médias	70
Tabela 8 - Funções de estimação das observações conforme os marcadores do Biplot	72
Tabela 9 - Médias preditas da variável Produção	85
Tabela 10 -Análise de variância conjunta para a variável Produção	86
Tabela 11 -Resultados dos testes de Gollob da análise AMMI da variável Produção	86
Tabela 12 -Resultados da produção média e os escores para o Biplot da variável Produção	87
Tabela 13 -Resultados dos testes de Gollob da análise GGE para a variável Produção	91
Tabela 14 -Médias preditas da variável SST	93
Tabela 15 -Análise de variância conjunta para a variável SST	93
Tabela 16 -Resultados dos testes de Gollob da análise AMMI para a variável SST	94
Tabela 17 -Resultados da produção média e os escores para o Biplot da variável SST	96
Tabela 18 -Resultados dos testes de Gollob da análise GGE para a variável SST	98
Tabela 19 -Matriz de dissimilaridade entre os 12 ambientes para a variável Produção com os dados transformados	100
Tabela 20 -Valores das dimensões onde serão plotados os 9 genótipos e os 12 ambientes para a variável produção	101
Tabela 21 -Matriz de dissimilaridade entre os 12 ambientes para a variável SST	103
Tabela 22 -Valores das dimensões onde serão plotados os 9 genótipos e os 12 ambientes para a variável SST	106

1 INTRODUÇÃO

Os métodos gráficos são ferramentas capazes de mostrar de forma clara as tendências, as aglomerações, os pontos discrepantes, a hierarquização e a eliminação de efeitos pontuais ou marginais, que podem perturbar a percepção global. Com os recursos computacionais houve o desenvolvimento e interesse sobre os mesmos, e estes vêm sendo aplicados em várias áreas como a Ecologia, Lingüística, Economia, Marketing, Setor Financeiro e de Seguros, Sociologia, Medicina, Biologia etc (ESCOFIER; PAGÈS, 1990).

A análise estatística multivariada envolve um conjunto de métodos estatísticos e matemáticos, destinada a descrever e interpretar os dados que provém da observação de várias variáveis estudadas conjuntamente e algumas estruturas de correlação (CUADRAS, 2006).

Biplot é uma técnica multivariada proposta por Gabriel (1971), com o objetivo de representar graficamente uma matriz de dados, de tal forma que esta representação permita visualizar em um plano as relações e inter-relações entre as linhas e colunas desta matriz. Fatorando a matriz de dados original pela DVS (Decomposição em Valores Singulares) como a soma de produtos de matrizes que contém os marcadores de linhas e colunas que constituem os elementos para a representação gráfica, consegue-se uma avaliação visual da estrutura da matriz de dados.

Independente da forma de fatoração, as aplicações do método Biplot dividem-se em representação da matriz de dados com finalidade descritiva e diagnóstico de modelos. A técnica é bastante útil, pois o gráfico utilizado para representar simultaneamente as linhas e colunas de uma matriz de dados pode indicar a existência de agrupamentos entre as observações, assim como mostrar as variâncias e correlações entre as variáveis. Atualmente, todas as áreas do conhecimento utilizam essa metodologia. Esta técnica não se difundiu com a mesma velocidade de outras técnicas multivariadas, como Análise de Componentes Principais e Análise de Correspondências, devido a não incorporação nos softwares estatísticos (CÁRDENAS; GALINDO; VICENTE-VILLARDÓN, 2007).

A avaliação de cultivares e a identificação de mega-ambientes são os objetivos mais importantes em ensaios de avaliação de cultivares. As análises AMMI (Additive Main

Effects and Multiplicative Interaction) e SREG (Sites Regression), apresentam propriedades importantes e adequadas ao estudo da interação ($G \times A$), sendo recomendadas suas utilizações pelos melhoristas de plantas. Os gráficos Biplots captam a porção padrão da interação $G \times A$, eles mostram quais os genótipos e ambientes que menos contribuíram para a interação, os chamados genótipos estáveis e combinações de genótipos e ambientes desejáveis quanto a adaptabilidade.

O Escalonamento multidimensional ou MDS (Multidimensional Scaling) é um método multivariado, que verifica a similaridade e dissimilaridade de dados sobre um conjunto de objetos e podem estar inter-correlacionados. A MDS tenta modelar tais dados, como distâncias entre pontos em um espaço geométrico. A principal razão em utilizar a técnica MDS está na visualização gráfica da estrutura dos dados, que é mais fácil de entender do que uma matriz composta por números, além disso apresenta informações essenciais sobre os dados, por exemplo na biologia molecular referente à dados genômicos.

O presente trabalho tem o objetivo de propor a metodologia MDS (Multidimensional Scalling), determinando assim as associações entre os fatores observados nas análises em delineamentos experimentais e dessa forma comparar com os métodos AMMI e SREG, por meio do gráfico Biplot, estabelecendo as condições de verificar quais os métodos apresentam desempenho satisfatório e suas especificidades.

2 REVISÃO BIBLIOGRÁFICA

2.1 Fatos históricos

O método Biplot, desenvolvido por Gabriel (1971) é uma representação de baixa dimensão de uma matriz de dados $\mathbf{Y}_{n \times p}$, em que as linhas e colunas da matriz são representadas conjuntamente por pontos no espaço bi ou tri-dimensional. As coordenadas dos pontos são determinadas por marcadores de linha e coluna. Os valores singulares da DVS são distribuídos para os marcadores de linhas e colunas de diferentes maneiras, dependendo se o interesse está na relação entre linhas (JK-Biplot), ou entre colunas (GH-Biplot) ou entre linhas e colunas (SQRT-Biplot).

Bradu e Gabriel (1978) demonstraram que a colinearidade entre os vetores linhas e colunas em tabela de dupla entrada, aproximada sobre o gráfico Biplot, pode dar indícios para o ajuste de alguns modelos bilineares do tipo multiplicativo. Gabriel e Zamir (1979) estudaram a interação dos modelos bilineares através de uma aproximação de posto inferior. Gabriel (1981) comentou as vantagens do método Biplot em relação à seleção e diagnóstico de modelos que se ajustam a matriz de dados multivariados. Cox e Gabriel (1982) fizeram uma comparação com a análise exploratória de dados, demonstrando que os Biplots são melhores para descrever um conjunto de dados.

Galindo (1986) propõe uma alternativa de representação simultânea para as linhas e colunas (HJ-Biplot), que possui a desvantagem de não reproduzir os dados originais. Alguns investigadores o utiliza devido à alta qualidade de representação para as linhas e colunas e as relações linha-coluna, como na Análise de Correspondência.

Vicente-Villardón (1992) desenvolveu o Biplot Generalizado como uma alternativa, considerando a importância dos diferentes indivíduos (linhas) e variáveis (colunas) da matriz de dados. Todos os resultados são generalizados através de ponderações feitas por matrizes simétricas definidas positivas para as linhas e/ou colunas. Greenacre (1993) estendeu esta generalização utilizando o método de Análise de Correspondências e Gabriel (1995) também a utilizou na Análise de Correspondências Múltiplas. Vasquez (1995) realizou a formulação algébrica dos Biplots Generalizados deduzindo suas propriedades, baseando-se em estimação por mínimos quadrados generalizados.

Com enfoque diferente, Gower e Harding (1988) e Gower (1992) desenvolveram os Biplots não lineares. Utilizados para obter o ajuste de trajetórias não lineares e representar as variáveis que se projetam sobre as representações obtidas através de coordenadas principais. Gower e Hand (1996) definiram os Biplots de interpolação e predição. Com a interpolação é possível sobrepor novos indivíduos projetando-os sobre o subespaço da representação, enquanto que com a predição é possível inferir valores das variáveis originais. Gower (2003) unificou a geometria dos seus Biplots.

Yan (2001) utilizou o Biplot genótipo $\times$ ambiente, enfocando-o como modelos bilineares para a análise de interação genótipo $\times$ ambiente (linhas-colunas) na área da agronomia, possibilitando o estudo da estabilidade dos genótipos a serem testados em vários ambientes. Martín-Rodríguez *et al.* (2002) desenvolveram o Meta Biplot, que permite a comparação de vários Biplot, sendo aplicado ao estudo com muitos indivíduos e variáveis ao longo do tempo em diversas situações experimentais.

Varela (2002) utilizou o Biplot como ferramenta para explicar a interação de ordem superior associada a um modelo linear, dando ênfase às interações de segunda e terceira ordem, para as tabelas de dupla e tripla entrada, respectivamente. Utilizou o Biplot para identificar as linhas e as colunas responsáveis pela ausência/presença de interação e no diagnóstico de modelos, identificando os diferentes tipos de interações presentes em uma tabela de múltiplas entradas.

No mundo empresarial os investidores necessitam saber sobre o desempenho financeiro dos seus empreendimentos. Le Roux, Gardner e Olivier (2003) aplicaram a metodologia CVA (Canonical variate analysis) Biplot, que otimiza a separação das variáveis preditoras em classes, através da distância euclidiana, em indústrias sul-africanas, discriminando-as em quatro classes. O uso do Biplot revelou problemas financeiros que permaneciam escondidos, mostrando a insolvência financeira das empresas.

Vairinhos e Galindo (2004) desenvolveram Biplots para *data mining* com padrões de classificação que permitem aproximar conjuntos de indivíduos e observações através de gráficos. Amaro *et al.* (2004) definiram a generalização do Manova-Biplot de dupla entrada, no contexto do modelo linear geral multivariado. Realizaram uma representação ponderada de posto inferior da matriz de médias, permitindo estudar e comparar as

separações entre grupos, a influência das variáveis na separação, o efeito da interação, testando hipóteses e construindo regiões de confiança para as médias ou combinações de médias.

Ramírez *et al.* (2005) propuseram o Biplot para detectar multicolinearidade entre as variáveis baseado na decomposição espectral da inversa da matriz de correlações. Quando as variáveis são binárias (presença/ausência), a representação clássica do Biplot não é adequada. Vicente-Villardón *et al.* (2006) sugeriram o Biplot Logístico, o qual é uma representação baseada em modelos de regressão logística.

Vários *softwares* estatísticos foram desenvolvidos para a construção do Biplot e os mais conhecidos são:

- 1) **Canoco** (2002) desenvolvido por Cajo J. F. ter Braak, é o software mais popular que permite uma boa modelagem estatística de dados ecológicos e biológicos. Por exemplo as relações entre as comunidades vegetais e animais e seu ambiente, os efeitos do impacto sobre o meio ambiente e os efeitos de tratamentos complexos ecotoxicológicos;
- 2) **BiPlot** (2002) criado por Lipkovich e Smith, desenvolveram uma macro para o Excel, implementado na linguagem Visual Basic. O programa calcula a Decomposição em Valores Singular (DVS) da matriz de dados originais ou da matriz de dados transformados. A construção do Biplot é baseada em técnicas multivariadas como: Análise de Componentes Principais, Análise de Correspondência, Escala Métrica Multidimensional, Análise Discriminante Canônica, Análise de Correlação Canônica e Análise de Redundância;
- 3) **XLS-Biplot** (2005) desenvolvido por Frederic Udina. O software oferece ao usuário dois tipos de ambientes. Um ambiente altamente interativo, no qual o usuário é capaz de definir e alterar as opções gráficas. O gráfico resultante pode ser disponibilizado em vários formatos, incluindo o formato para impressão. Pode ser executado no Windows ou Linux;
- 4) **GGEbiplot** (2006) desenvolvido por Yan e Kang, utiliza a metodologia **GGE** (Genótipo mais interação Genótipo-Ambiente), é usado por agrônomos e geneticistas nos estudos de cultivares. O software possui algumas características especiais como análise de dados

multivariados, imputação, transformação de dados, mapeamento de QTL, análise de tabela de tripla-entrada etc;

- 5) **SAS** (2007), utiliza o conhecimento da álgebra matricial, pode-se construir programas com o procedimento PROC IML. Há também macros desenvolvidas, que utilizam a primeira e segunda componente singular, simplificando a construção do Biplot. O módulo SAS/INSIGHT permite a construção do Biplot bi e tridimensional de uma forma amigável, para atender a leigos no assunto;
- 6) **SAS/JMP** (2007), desenvolvido pelo **SAS Institute**, trabalha com entradas de dados via planilha e possui um menu de seleção com opções para análise. Faz análises estatísticas e de forma interativa pode-se visualizar gráficos multidimensionais (2D e 3D), é possível também identificar as diferenças de comportamento entre os segmentos analisados, permitindo tomar decisões ao nível de ciclo, visualizando, conhecendo os detalhes e otimizando os produtos e serviços.
- 7) **R** é um *software* livre desenvolvido por vários contribuidores e possui alguns pacotes como:
 - a) **Agricolae** (2007) desenvolvido por Felipe de Mendiburu e Reinhard Simon, utilizado em pesquisas agrícolas e em vários tipos de análises de delineamentos tais como, análise do modelo AMMI com gráfico Biplot;
 - b) **BiplotGUI** (2008) foi desenvolvido por Anthony la Grange para ajudar os usuários leigos em programação a construir o Biplot de maneira interativa e
 - c) **Bpca** (2008) desenvolvido por José Cláudio Faria com implementação dos métodos clássicos do Biplot: SQRT, JK, GH e HJ. A apresentação gráfica pode ser escolhida em duas formas: 2D (estática) e 3D (rotação);
- 8) **Classical Biplot**(2009) é um pacote desenvolvido por José Luis Vicente Villardón. Para compilá-lo necessita-se do MATLAB. Existe uma versão completa do programa denominada de MULTBILOT (MULTivariate Analysis using BILOT).

2.1.1 Estrutura dos dados para o método Biplot

Na aplicação dos métodos Biplot, deve-se considerar a finalidade da análise e a matriz de dados utilizada (CÁRDENAS; GALINDO e VICENTE-VILLARDÓN, 2007). Quando a finalidade for descritiva, os dados são organizados em matrizes retangulares, por exemplo $\mathbf{Y}_{n \times p}$, com n indivíduos e p variáveis (quantitativas ou categóricas), formada por i linhas ($i = 1, \dots, n$) e as j colunas ($j = 1, \dots, p$), respectivamente. O elemento genérico y_{ij} , corresponde ao valor do i -ésimo indivíduo da j -ésima variável. Em alguns casos, estas matrizes podem constar de dados longitudinais, ou seja, possuir t tempos distintos nas linhas ou colunas.

Se a finalidade for analisar a interação ou independência de duas variáveis categóricas, os dados são apresentados em tabelas de contingência de dupla entrada ou uma matriz $\mathbf{X}_{I \times J}$, constituída por frequências absolutas dos n indivíduos sobre as duas variáveis medidas e cada uma delas com categorias diferentes ($i = 1, \dots, I; j = 1, \dots, J$), pois cada célula da tabela contém a frequência cruzada. O elemento x_{ij} corresponde ao número de indivíduos que possuem simultaneamente as i -ésimas e j -ésimas categorias de ambas variáveis, sendo o total de indivíduos $N = \sum_{i=1}^I \sum_{j=1}^J x_{ij}$.

Quando a finalidade for analisar a evolução dos indivíduos ou a estrutura de grupos de variáveis, os dados se encontram em tabelas de múltiplas entradas, formada por arranjo de ordem três ou matrizes de três modos (modo linha, modo coluna e modo tubo), que não são exatamente tabela de tripla entrada, já que os modos não estão totalmente cruzados.

2.2 Definição do Biplot

Seja $\mathbf{Y}_{n \times p}$ uma matriz de dados, em que as n linhas correspondem aos indivíduos e as p colunas as variáveis medidas sobre os indivíduos. O Biplot da matriz $\mathbf{Y}$ é uma representação gráfica feita por meio de vetores chamados de marcadores $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ para as linhas de $\mathbf{Y}$ e os marcadores $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_p$ para as colunas de $\mathbf{Y}$, de tal forma que o produto interno de $\mathbf{a}_i^T \mathbf{b}_j$, para $i = 1, \dots, n$ e $j = 1, \dots, p$, seja igual ou próximo aos elementos de y_{ij} da matriz original $\mathbf{Y}$.

Se considerarmos os marcadores $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ como linhas da matriz $\mathbf{A}$ e os

marcadores $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_p$ como linhas da matriz $\mathbf{B}$, a decomposição da matriz $\mathbf{Y}$ é dada por:

$$\mathbf{Y} \approx \mathbf{A}\mathbf{B}^T. \quad (1)$$

A estrutura da matriz $\mathbf{Y}$ é visualizada, representando os marcadores no espaço euclidiano, em duas ou três dimensões.

A decomposição em geral, não é única. Existem várias formas de decomposição de uma matriz. O método mais conhecido de aproximação para uma matriz de posto inferior é a Decomposição em Valores Singulares, segundo Gabriel (1971) e Greenacre (1984).

2.2.1 Decomposição em Valores Singulares (DVS)

A Decomposição em Valores Singulares (DVS) é um conceito algébrico sobre fatoração de matrizes, proposto por Good (1969). Seja $\mathbf{Y} \in \mathbb{R}^{n \times p}$ uma matriz de posto r . A DVS de $\mathbf{Y}$ consiste em escrevê-la como o produto de três matrizes (JOHNSON; WICHERN, 2002),

$$\mathbf{Y}_{(r)n \times p} = \mathbf{U}_{(r)n \times r} \mathbf{\Lambda}_{(r)r \times r} \mathbf{V}_{(r)r \times p}^T \quad (2)$$

em que:

- i) As colunas das matrizes $\mathbf{U}$ e $\mathbf{V}$ são ortonormais, isto é $\mathbf{U}_{(r)r \times n}^T \mathbf{U}_{(r)n \times r} = \mathbf{V}_{(r)r \times p}^T \mathbf{V}_{(r)p \times r} = \mathbf{I}_{(r)}$,
- ii) $\mathbf{U}$ é uma matriz de autovetores da matriz $\mathbf{Y}\mathbf{Y}^T$,
- iii) $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_r)$ é uma matriz diagonal formada pelos r maiores valores singulares de $\mathbf{Y}$, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > 0$, os quais correspondem às raízes quadradas dos autovalores não nulos da matriz $\mathbf{Y}\mathbf{Y}^T$ ou $\mathbf{Y}^T\mathbf{Y}$,
- iv) $\mathbf{V}$ é uma matriz de autovetores da matriz $\mathbf{Y}^T\mathbf{Y}$.

A DVS também pode ser expressa como uma soma de r matrizes de posto 1,

$$\begin{aligned} \mathbf{Y}_{(r)} &= \lambda_1 \mathbf{u}_1 \mathbf{v}_1^T + \lambda_2 \mathbf{u}_2 \mathbf{v}_2^T + \dots + \lambda_r \mathbf{u}_r \mathbf{v}_r^T \\ &= \sum_{i=1}^r \lambda_i \mathbf{u}_i \mathbf{v}_i^T \end{aligned} \quad (3)$$

em que os vetores colunas $[\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r]$ constituem uma base ortonormal do espaço vetorial gerado pelas colunas da matriz $\mathbf{Y}$, com n elementos cada um e $[\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r]$ constituem uma base ortonormal do espaço vetorial gerado pelas linhas da matriz $\mathbf{Y}$ com p elementos cada um (KRZANOWSKI, 2000).

2.2.2 Enfoque Clássico

A análise Biplot é utilizada em qualquer tipo de variáveis (contínuas ou discretas), quando a finalidade é aproximar os dados originais e realizar uma análise simultânea das relações entre indivíduos e/ou variáveis. A fundamentação teórica se baseia na aproximação da matriz $\mathbf{Y}$ de ordem $n \times p$ de posto (r) por outra matriz de posto (q), em que $q < r$, através de sua DVS. Seja

$$\mathbf{Y}_{(r)n \times p} \approx \mathbf{Y}_{(q)n \times p} = \mathbf{U}_{(q)n \times q} \mathbf{\Lambda}_{(q)q \times q} \mathbf{V}_{(q)q \times p}^T = \mathbf{A}_{(q)n \times q} \mathbf{B}_{(q)q \times p}^T \quad (4)$$

em que:

$\mathbf{Y}_{(r)}$ é a matriz de dados de ordem $(n \times p)$ e posto r ,

$\mathbf{Y}_{(q)}$ é a matriz aproximada da matriz de dados $\mathbf{Y}_{(r)}$ de posto q , sendo $(q < r)$,

$$\mathbf{A}_{(q)n \times q} = \mathbf{U}_{(q)n \times q} \mathbf{\Lambda}_{(q)q \times q}^s$$

$$\mathbf{B}_{(q)q \times p}^T = \mathbf{\Lambda}_{(q)q \times q}^{1-s} \mathbf{V}_{(q)q \times p}^T$$

s é uma constante que pertence ao intervalo $[0, 1]$.

Gabriel (1971) desenvolveu a técnica do Biplot Clássico, também chamado de Biplot Componente Principal, que propõe representar os dois conjuntos de vetores, ou seja, as linhas e colunas de qualquer matriz de dados, por exemplo $\mathbf{Y}_{(r)}$ conforme a expressão (3) em uma matriz $\mathbf{Y}_{(2)}$ que tenha posto 2 (KRZANOWSKI, 2000), então

$$\begin{aligned} \mathbf{Y}_{(2)} &= \lambda_1 \mathbf{u}_1 \mathbf{v}_1^T + \lambda_2 \mathbf{u}_2 \mathbf{v}_2^T \\ &= \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \end{pmatrix}. \end{aligned}$$

Os vetores $\mathbf{u}_i$ são dados por $\mathbf{u}_i^T = (u_{i1}, u_{i2}, \dots, u_{in})$ e os vetores $\mathbf{v}_i$ são dados por $\mathbf{v}_i^T = (v_{i1}, v_{i2}, \dots, v_{ip})$ para $i = 1, 2$. Assim $\mathbf{Y}_{(r)}$ é aproximada por:

$$\mathbf{Y}_{(2)} = \begin{pmatrix} u_{11} & u_{21} \\ u_{12} & u_{22} \\ \vdots & \vdots \\ u_{1n} & u_{2n} \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1p} \\ v_{21} & v_{22} & \cdots & v_{2p} \end{pmatrix}.$$

Para obter o Biplot, é necessário escrever $\mathbf{Y}_{(2)}$ como o produto de $\mathbf{A}_{n \times 2} \mathbf{B}_{2 \times p}^T$. Existem várias maneiras de fatoração, Gabriel (1971) sugeriu:

$$\mathbf{Y} = \begin{pmatrix} u_{11}\sqrt{\lambda_1} & u_{21}\sqrt{\lambda_2} \\ u_{12}\sqrt{\lambda_1} & u_{22}\sqrt{\lambda_2} \\ \vdots & \vdots \\ u_{1n}\sqrt{\lambda_1} & u_{2n}\sqrt{\lambda_2} \end{pmatrix} \begin{pmatrix} v_{11}\sqrt{\lambda_1} & v_{12}\sqrt{\lambda_1} & \cdots & v_{1p}\sqrt{\lambda_1} \\ v_{21}\sqrt{\lambda_2} & v_{22}\sqrt{\lambda_2} & \cdots & v_{2p}\sqrt{\lambda_2} \end{pmatrix}, \text{ para } s = 1/2, \quad (5)$$

$$\mathbf{Y} = \begin{pmatrix} u_{11}\lambda_1 & u_{21}\lambda_2 \\ u_{12}\lambda_1 & u_{22}\lambda_2 \\ \vdots & \vdots \\ u_{1n}\lambda_1 & u_{2n}\lambda_2 \end{pmatrix} \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1p} \\ v_{21} & v_{22} & \cdots & v_{2p} \end{pmatrix}, \text{ para } s = 1 \text{ ou} \quad (6)$$

$$\mathbf{Y} = \begin{pmatrix} u_{11} & u_{21} \\ u_{12} & u_{22} \\ \vdots & \vdots \\ u_{1n} & u_{2n} \end{pmatrix} \begin{pmatrix} v_{11}\lambda_1 & v_{12}\lambda_1 & \cdots & v_{1p}\lambda_1 \\ v_{21}\lambda_2 & v_{22}\lambda_2 & \cdots & v_{2p}\lambda_2 \end{pmatrix}, \text{ para } s = 0. \quad (7)$$

Denota-se a i -ésima linha de $\mathbf{A}$ por $\mathbf{a}_i^T$ e a j -ésima coluna de $\mathbf{B}^T$ por $\mathbf{b}_j^T$; na fatoração (5) tem-se $\mathbf{a}_i^T = (u_{1i}\sqrt{\lambda_1} \quad u_{2i}\sqrt{\lambda_2})$ e $\mathbf{b}_j^T = (v_{1j}\sqrt{\lambda_1} \quad v_{2j}\sqrt{\lambda_2})$. O Biplot consiste em plotar $n + p$ vetores $\mathbf{a}_i^T (i = 1, \dots, n)$ e $\mathbf{b}_j^T (j = 1, \dots, p)$ em um plano, dando uma representação de np elementos de $\mathbf{Y}$ através de produtos correspondentes aos efeitos de linha e coluna. Então, para a fatoração (5), temos que $\mathbf{a}_i$ e $\mathbf{b}_j$ são representados por vetores que partem da origem até o ponto das coordenadas dadas por $(u_{1i}\sqrt{\lambda_1} \quad u_{2i}\sqrt{\lambda_2})$ e $(v_{1j}\sqrt{\lambda_1} \quad v_{2j}\sqrt{\lambda_2})$, respectivamente.

Os vetores $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ são chamados de efeitos ou marcadores de linhas de $\mathbf{Y}_{(q)}$ e os vetores $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_p$ são chamados de efeitos ou marcadores de colunas de $\mathbf{Y}_{(q)}$. Cada elemento $y_{ij(q)}$ de $\mathbf{Y}_{(q)}$ é representado pelo produto interno de

$$y_{ij(q)} = \mathbf{a}_{i(q)}^T \mathbf{b}_{j(q)}, \text{ em que } i = 1, \dots, n; j = 1, \dots, p, \quad (8)$$

que corresponde aos marcadores de linha e marcadores de coluna.

No Biplot, o produto interno entre dois vetores $\mathbf{a}_i^T \mathbf{b}_j$ pode ser entendido como o produto do comprimento de um dos vetores vezes o comprimento da projeção do outro, conforme a Figura 1. Pode-se verificar também, quais vetores linhas ou colunas estão relacionados a outros vetores linhas ou colunas (mesma direção ou direção inversa dos vetores correspondentes) e quais vetores linhas ou colunas são perpendiculares (quando $\cos(\theta) = 0$, então $\mathbf{a}_i^T \mathbf{b}_j = 0$), dando idéia no geral de correlação ou associação entre estas entidades. A projeção do vetor $\mathbf{a}_i$ sobre $\mathbf{b}_j$ utilizando o conceito do processo de ortogonalização de Gram-Schmidt é dada por

$$Proj(\mathbf{a}_i/\mathbf{b}_j) = \frac{(\mathbf{a}_i^T \mathbf{b}_j)}{\mathbf{b}_j^T \mathbf{b}_j} \mathbf{b}_j = \frac{(\mathbf{a}_i^T \mathbf{b}_j)}{\|\mathbf{b}_j\|} \frac{1}{\|\mathbf{b}_j\|} \mathbf{b}_j = \frac{(\mathbf{a}_i^T \mathbf{b}_j)}{\|\mathbf{b}_j\|} \mathbf{u} = \frac{(\mathbf{a}_i^T \mathbf{b}_j)}{\|\mathbf{b}_j\|}$$

em que $\|\mathbf{b}_j\| = \sqrt{\mathbf{b}_j^T \mathbf{b}_j}$ denomina-se de norma euclidiana ou comprimento do vetor $\mathbf{b}_j$, a distância entre os vetores $\mathbf{a}_i$ e $\mathbf{b}_j$ é definida por $d(\mathbf{a}_i, \mathbf{b}_j) = \|\mathbf{a}_i - \mathbf{b}_j\|$. Seja $\mathbf{b}_j \neq 0$, o vetor $\mathbf{u} = \frac{\mathbf{b}_j}{\|\mathbf{b}_j\|}$, é um vetor de base ortonormal de comprimento unitário $\|\mathbf{u}\| = \left\| \frac{\mathbf{b}_j}{\|\mathbf{b}_j\|} \right\| = \frac{1}{\|\mathbf{b}_j\|} \|\mathbf{b}_j\| = 1$. Então, o comprimento da projeção será definido por

$$L_{Proj(\mathbf{a}_i/\mathbf{b}_j)} = \frac{|\mathbf{a}_i^T \mathbf{b}_j|}{\|\mathbf{b}_j\|} = \|\mathbf{a}_i\| \left| \frac{\mathbf{a}_i^T \mathbf{b}_j}{\|\mathbf{a}_i\| \|\mathbf{b}_j\|} \right| = \|\mathbf{a}_i\| |\cos(\theta)|$$

em que θ é o ângulo entre $\mathbf{a}_i$ e $\mathbf{b}_j$ podendo ser expresso em função do produto interno e do comprimento dos vetores, obtido através da lei dos cossenos.

A matriz $\mathbf{Y}_{(q)}$ pode ser fatorada de diversas formas, como foi mostrado nas expressões (5), (6) e (7). A cada fatoração corresponde um Biplot diferente. A escolha da fatoração é importante. Na prática a expressão (5) corresponde a uma fatoração geral sem ênfase nas linhas ou colunas, a expressão (6) é uma fatoração com ênfase nas linhas e a expressão (7) é uma fatoração com ênfase nas colunas.

Figura 1 - Projeção do vetor $\mathbf{a}_i$ sobre o vetor $\mathbf{b}_j$

Ao empregar o Biplot para inspeção de relações entre linhas e colunas da matriz $\mathbf{Y}$ e/ou entre colunas, deve-se impor uma métrica particular e fazer a fatoração resultante como um único Biplot. Então, se desejarmos as relações entre linhas de $\mathbf{Y}$ para representar as correspondentes relações dos vetores $\mathbf{a}_i$, deve-se impor os seguintes requisitos:

$$\mathbf{Y}\mathbf{Y}^T = \mathbf{A}\mathbf{A}^T \text{ e } \mathbf{B}^T\mathbf{B} = \mathbf{I}_{(2)},$$

que são satisfeitos pela equação (6), ou seja

$$\mathbf{Y}\mathbf{Y}^T = \mathbf{A}\mathbf{B}^T(\mathbf{A}\mathbf{B}^T)^T = \mathbf{A}\mathbf{B}^T\mathbf{B}\mathbf{A}^T = \mathbf{A}\mathbf{A}^T,$$

para quaisquer duas linhas de $\mathbf{y}_i$ e $\mathbf{y}_j$ de $\mathbf{Y}$,

$$\mathbf{y}_i^T \mathbf{y}_j = \mathbf{a}_i^T \mathbf{a}_j,$$

$$\|\mathbf{y}_i\| = \|\mathbf{a}_i\|,$$

$$\|\mathbf{y}_i - \mathbf{y}_j\| = \|\mathbf{a}_i - \mathbf{a}_j\| \text{ e}$$

$$\cos(\mathbf{y}_i, \mathbf{y}_j) = \cos(\mathbf{a}_i, \mathbf{a}_j).$$

De forma similar, as relações entre colunas de $\mathbf{Y}$ são representadas pela correspondente relação dos vetores $\mathbf{b}$ se $\mathbf{Y}^T\mathbf{Y} = \mathbf{B}\mathbf{B}^T$, a qual requer $\mathbf{A}^T\mathbf{A} = \mathbf{I}_{(2)}$, que é satisfeita pela equação (7), ou seja $\mathbf{Y}^T\mathbf{Y} = (\mathbf{A}\mathbf{B}^T)^T\mathbf{A}\mathbf{B}^T = \mathbf{B}\mathbf{A}^T\mathbf{A}\mathbf{B}^T = \mathbf{B}\mathbf{B}^T$.

A fatoração (1) para matrizes de posto 3 pode ser representada por bi-modelo constituído dos vetores $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ e $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_p$ de origem comum. A interpretação para modelos tridimensionais é análoga a do Biplot bi-dimensional. Para matrizes de posto 3 ou maior, necessita-se de uma melhor aproximação para construção do Biplot (GABRIEL, 1971).

2.2.3 Construção do Biplot

Se o posto da matriz $\mathbf{Y}_{(q)}$ for igual a 2, então o Biplot representa de forma exata, a matriz sobre o plano. Quando o posto de $\mathbf{Y}_{(q)}$ for maior que 2, então a fatoração do Biplot terá uma representação gráfica aproximada da matriz original $\mathbf{Y}_{(r)}$, pois será escolhida as duas primeiras colunas de $\mathbf{U}_{(q)}$, as duas primeiras colunas de $\mathbf{V}_{(q)}$ e os dois primeiros valores singulares da matriz diagonal $\mathbf{\Lambda}_{(q)}$.

Os elementos da matriz $\mathbf{Y}_{(q)}$ podem ser obtidos conforme a equação $Y_{ij(q)} = \mathbf{a}_{i(q)}^T \mathbf{b}_{j(q)}$. Por exemplo, seja $\mathbf{Y}_{(2)_{4 \times 3}} = \mathbf{A}_{(2)_{4 \times 2}} \mathbf{B}_{(2)_{2 \times 3}}^T$

$$\begin{pmatrix} 20 & -9 & 6 \\ 6 & 12 & -15 \\ -10 & -6 & 9 \\ 8 & -12 & 12 \end{pmatrix} = \begin{pmatrix} 4 & 3 \\ -3 & 3 \\ 1 & -3 \\ 4 & 0 \end{pmatrix} \begin{pmatrix} 2 & -3 & 3 \\ 4 & 1 & -2 \end{pmatrix}.$$

O elemento Y_{11} da matriz $\mathbf{Y}_{(2)}$, é calculado pela fatoração, isto é, pelo produto entre a primeira linha da matriz $\mathbf{A}$ definida por $\mathbf{a}_1$ e a primeira coluna da matriz $\mathbf{B}$ definida por $\mathbf{b}_1$, tem-se

$$Y_{11} = a_{11}b_{11} + a_{12}b_{21} = (4)(2) + (3)(4) = 20. \quad (9)$$

Um ponto do plano pode ser identificado com o par de coordenadas (x, y) .

	x	y			x	y	
A_1	4	3	e		B_1	2	4
A_2	-3	3			B_2	-3	1
A_3	1	-3			B_3	3	-2
A_4	4	0					

Dessa maneira, pode-se representar em um plano cartesiano cada ponto das matrizes $\mathbf{A}$ e $\mathbf{B}$, conforme a Figura 2.

A distância entre a origem O do Biplot e o marcador de linha A_1 é $\overrightarrow{OA_1}$ e a distância da origem O do Biplot e o marcador de coluna B_1 é $\overrightarrow{OB_1}$. Assim, para cada ponto do plano, está associado um único vetor.

Figura 2 - Produto interno para cada elemento da matriz $\mathbf{Y}$ de posto 2

Fazendo as seguintes relações, tem-se:

$$\begin{aligned} x_1 &= a_{11} = \overrightarrow{OA_1} \cos(\beta) \\ y_1 &= a_{12} = \overrightarrow{OA_1} \sin(\beta) \end{aligned}$$

e

$$\begin{aligned} x_2 &= b_{11} = \overrightarrow{OB_1} \cos(\alpha_{11} + \beta) \\ y_2 &= b_{21} = \overrightarrow{OB_1} \sin(\alpha_{11} + \beta). \end{aligned}$$

em que: (β) é o ângulo formado entre os vetores $\overrightarrow{OA_1}$ e $\overrightarrow{OA_4}$; (α_{11}) é o ângulo formado entre os vetores $\overrightarrow{OA_1}$ e $\overrightarrow{OB_1}$ e $(\alpha_{11} + \beta)$ é o ângulo formado entre os vetores $\overrightarrow{OB_1}$ e $\overrightarrow{OA_4}$. Representando graficamente essas relações na Figura 3, pode-se observar o produto interno dos marcadores das linhas e das colunas.

Substituindo na equação (9), temos:

$$Y_{11} = x_1 x_2 + y_1 y_2 = \overrightarrow{OA_1} \overrightarrow{OB_1} [\cos \beta \cos(\alpha_{11} + \beta) + \sin \beta \sin(\alpha_{11} + \beta)]$$

e utilizando a propriedade da diferença de cossenos das funções trigonométricas, pode-se

Figura 3 - Produto interno para cada elemento de uma matriz de posto 2

simplificar para

$$Y_{11} = \overrightarrow{OA_1} \cdot \overrightarrow{OB_1} \cos[\beta - (\alpha_{11} + \beta)] = \overrightarrow{OA_1} \cdot \overrightarrow{OB_1} \cos(-\alpha_{11}),$$

como $\cos(\alpha_{11}) = \cos(-\alpha_{11})$, então

$$Y_{11} = \overrightarrow{OA_1} \cos(\alpha_{11}) \overrightarrow{OB_1}. \quad (10)$$

Portanto, o elemento da primeira linha e da primeira coluna da matriz $\mathbf{Y}$ é o produto de vetores da primeira linha $\overrightarrow{OA_1}$ com o vetor da primeira coluna $\overrightarrow{OB_1}$ e o cosseno do ângulo entre eles (α_{11}). Os vetores podem ser calculados por:

$$\begin{aligned} \overrightarrow{OA_1} &= \sqrt{x_1^2 + y_1^2} = \|\mathbf{a}_1\| = \sqrt{4^2 + 3^2} = 5 \\ \overrightarrow{OB_1} &= \sqrt{x_2^2 + y_2^2} = \|\mathbf{b}_1\| = \sqrt{2^2 + 4^2} = 4,471 \end{aligned}$$

e os ângulos α_{11} e β podem ser escritos como:

$$\begin{aligned} \alpha_{11} + \beta &= \arctan\left(\frac{y_2}{x_2}\right) = \arctan\left(\frac{4}{2}\right) = 63,4^\circ \text{ e} \\ \beta &= \arctan\left(\frac{y_1}{x_1}\right) = \arctan\left(\frac{3}{4}\right) = 36,9^\circ. \end{aligned}$$

Generalizando a equação (10) para qualquer elemento da linha i e coluna j , temos:

$$Y_{ij} = \overrightarrow{OA_i} \cos(\alpha_{ij}) \overrightarrow{OB_j}. \quad (11)$$

Essa é a propriedade do produto interno, em que Y_{ij} é o elemento da matriz $\mathbf{Y}$ na linha i e coluna j , $\overrightarrow{OA_i}$ é a distância entre a origem O do Biplot e o marcador de linha A_i , $\overrightarrow{OB_j}$ é a distância da origem O do Biplot e o marcador coluna B_j e α_{ij} é o ângulo entre os vetores $\overrightarrow{OA_i}$ e $\overrightarrow{OB_j}$.

As distâncias entre os vetores $\overrightarrow{OA_i}$ e $\overrightarrow{OB_j}$ são valores positivos, mas o $\cos(\alpha_{ij})$ pode ser positivo ou negativo dependendo do ângulo. O ângulo α_{ij} é determinado pelo sinal de Y_{ij} , se

i) $Y_{ij} = 0 \Rightarrow \alpha_{ij} = 90^\circ$,

ii) $Y_{ij} > 0 \Rightarrow \alpha_{ij} < 90^\circ$,

iii) $Y_{ij} < 0 \Rightarrow \alpha_{ij} > 90^\circ$.

O gráfico Biplot não mostra somente as linhas da matriz $\mathbf{A}$ e as colunas da matriz $\mathbf{B}$, mas também a matriz $\mathbf{Y}$ resultante do produto interno entre $\mathbf{A}$ e $\mathbf{B}$. Além disso, a equação (11) permite uma visualização da relação entre as linhas e colunas em diferentes perspectivas. Uma das aplicações da propriedade do produto interno do Biplot é a comparação visual dos elementos linha com uma única coluna (YAN; KANG, 2002). Reescrevendo a equação (11), temos:

$$Y_{ij} = [\overrightarrow{OA_i} \cos(\alpha_{ij})] \overrightarrow{OB_j} = \overrightarrow{OP_{ij}} \overrightarrow{OB_j} \quad (12)$$

em que $\overrightarrow{OP_{ij}}$ é a projeção do vetor da linha i de $\overrightarrow{OA_i}$ sobre o vetor da coluna j de $\overrightarrow{OB_j}$. Então, $\overrightarrow{OB_j}$ é comum para todos os elementos na coluna B_j e a equação (12) pode ser reescrita como,

$$\frac{Y_{ij}}{\overrightarrow{OB_j}} = \overrightarrow{OP_{ij}}.$$

A magnitude de $\frac{Y_{ij}}{\overrightarrow{OB_j}}$ dos elementos linha do marcador coluna B_j pode ser visualizada pela comparação de suas projeções sobre o vetor coluna $\overrightarrow{OP_{ij}}$. Por exemplo, na Figura 4, compara-se os quatro marcadores de linha A_1, A_2, A_3 e A_4 , projetando-os sobre o vetor $\overrightarrow{OB_1}$. As

projeções são denotadas por P_{11}, P_{21}, P_{31} e P_{41} , respectivamente. Elas são ordenadas conforme os valores dos elementos Y_{11}, Y_{21}, Y_{31} e Y_{41} . Portanto, a ordem das projeções é $P_{11} > P_{41} > P_{21} > 0 > P_{31}$, cujo P_{31} é negativo, uma vez que está do lado oposto do marcador B_1 , conforme a ordem dos valores de $Y_{11} > Y_{41} > Y_{21} > 0 > Y_{31}$.

Figura 4 - Comparação entre os elementos da primeira coluna da matriz produto, visualizando as projeções sobre o marcador B_1

Outra aplicação da propriedade é a comparação visual dos elementos coluna com uma única linha, isto é, $\overrightarrow{OP_{ij}}$ é a projeção do vetor da coluna j de $\overrightarrow{OB_j}$ sobre o vetor da linha i de $\overrightarrow{OA_i}$. $\overrightarrow{OA_i}$ é comum para todos os elementos na linha A_i , então a equação (11) pode ser escrita como,

$$Y_{ij} = \overrightarrow{OA_i} [\cos(\alpha_{ij}) \overrightarrow{OB_j}] = \overrightarrow{OA_i} \overrightarrow{OP_{ij}} \Rightarrow \frac{Y_{ij}}{\overrightarrow{OA_i}} = \overrightarrow{OP_{ij}}.$$

Na Figura 5, comparam-se os três marcadores de coluna B_1, B_2 e B_3 , projetando-os sobre o vetor $\overrightarrow{OA_1}$. As projeções são denotadas por P_{11}, P_{12} e P_{13} , respectivamente. São ordenadas conforme os valores dos elementos Y_{11}, Y_{12} e Y_{13} . Portanto, a ordem das projeções é $P_{11} > P_{13} > 0 > P_{12}$, cujo P_{12} é negativo, pois está do lado oposto do marcador A_1 , conforme a ordem dos valores $Y_{11} > Y_{13} > 0 > Y_{12}$.

Figura 5 - Comparação entre os elementos da primeira linha da matriz produto, visualizando as projeções sobre o marcador A_1

Todas as interpretações anteriormente discutidas podem ser feitas de forma similar, tomando agora as projeções de grupos dos vetores colunas sobre os vetores linhas da matriz $\mathbf{Y}$ e vice-versa.

Na Figura 6, verifica-se a relação entre os três marcadores de coluna através dos ângulos entre os seus vetores. O cosseno do ângulo formado entre dois vetores é igual ao coeficiente de correlação entre as variáveis correspondentes (é também o valor do produto interno dos vetores que unem dois pontos da nuvem em $\mathbb{R}$). Portanto, se os dois vetores possuem a mesma orientação, o coeficiente de correlação será próximo de 1. Se dois vetores são quase perpendiculares, a correlação será próxima de zero. Mas, se os dois vetores forem orientados em direções opostas, o coeficiente de correlação será próximo de -1. Os conceitos de comprimento e ângulos permitem que se façam interpretações das estatísticas de maneira geométrica, auxiliando na compreensão dos seus significados (JOHNSON; WICHERN, 2002).

Na Figura 7, pode-se visualizar as relações entre os vetores dos marcadores de linha através dos ângulos entre eles. Dos ângulos, podemos predizer que A_1 e A_4 na matriz $\mathbf{Y}$ possui uma correlação altamente positiva (37° , ângulo agudo), A_2 e A_3 possui uma correlação

Figura 6 - Cosseno dos ângulos entre os vetores colunas representando a estimativa do coeficiente de correlação entre os vetores dos marcadores de coluna

altamente negativa (153° , ângulo obtuso). Estas predições são confirmadas pelos coeficientes de correlações entre os marcadores de linha e foram calculadas baseadas nos dados da matriz $\mathbf{Y}$ (YAN; KANG, 2002).

Esta propriedade angular do Biplot é muito útil, pois ajuda a visualizar as inter-relações: entre os níveis do fator linha; entre os níveis do fator coluna e entre os níveis dos fatores linha e coluna de um conjunto de dados de dupla entrada. A relação entre cosseno de um ângulo entre dois vetores e a correlação é demonstrado na próxima seção.

2.2.4 Tipos de Biplots

A escolha dos marcadores de linhas e colunas a partir da aproximação de posto inferior obtido, pode ser feita de diversas maneiras, por essa razão vários autores propõem distintas escolhas e estudam suas propriedades conforme a fatoração escolhida.

A representação é realizada pela fatoração da matriz $\mathbf{Y}$ em um produto de matrizes, tal que:

$$\mathbf{Y} \approx (\mathbf{U}\mathbf{\Lambda}^s)(\mathbf{\Lambda}^{1-s}\mathbf{V}^T) = \mathbf{AB}^T, \quad (13)$$

Figura 7 - Cosseno dos ângulos entre os vetores linhas representando a estimativa do coeficiente de correlação entre os vetores dos marcadores de linha

sendo que as matrizes $\mathbf{A}$ e $\mathbf{B}$ possuem posto q com as respectivas coordenadas dos n vetores marcadores de linhas $\mathbf{a}_i$ e dos p vetores marcadores de colunas $\mathbf{b}_j$, em que $i = 1, \dots, n$ e $j = 1, \dots, p$.

Na expressão (13), a constante $s \in [0, 1]$. Assim, Gabriel (1971) propõe três representações de Biplots: GH-Biplot, JK-Biplot e SQRT-Biplot e Galindo (1986) propõe uma modificação com o HJ-Biplot.

- 1) Se $s = 0$, temos o **GH-Biplot** ou CMP-Biplot (Column Metric Preserving). Este Biplot é obtido pela escolha dos marcadores de colunas, conforme as expressões $\mathbf{A} = \mathbf{U}$ e $\mathbf{B} = \mathbf{V}\mathbf{\Lambda}$. Pode-se fazer um ajuste nos marcadores associados (multiplicando e dividindo por um fator de escala) conforme abaixo:

$$\mathbf{A}_{(q)} = \sqrt{(n-1)}\mathbf{U}_{(q)} \quad \text{e} \quad \mathbf{B}_{(q)} = \frac{1}{\sqrt{(n-1)}}\mathbf{V}_{(q)}\mathbf{\Lambda}_{(q)} \quad (14)$$

Impondo como restrição métrica $\mathbf{A}^T\mathbf{A} = \mathbf{I}$ e $\mathbf{Y}(\mathbf{Y}^T\mathbf{Y})^{-1}\mathbf{Y}^T = \mathbf{A}\mathbf{A}^T$. Os produtos escalares das colunas da matriz $\mathbf{Y}^T\mathbf{Y}$ coincidem com os produtos escalares dos marcadores $\mathbf{B}\mathbf{B}^T$.

A matriz de dados originais $\mathbf{Y}$ será representada pela matriz centrada por coluna $\mathbf{Y}_c = \mathbf{Y} - \frac{1}{n}\mathbf{1}\mathbf{1}^T\mathbf{Y}$, para que a matriz de variâncias e covariâncias $\mathbf{S}$ seja proporcional a matriz $(\mathbf{Y}^T\mathbf{Y})_{p \times p}$,

$$\begin{aligned}\mathbf{S} &= \frac{1}{(n-1)}\mathbf{Y}^T\mathbf{Y} = \frac{1}{(n-1)}(\mathbf{A}\mathbf{B}^T)^T(\mathbf{A}\mathbf{B}^T) = \frac{1}{(n-1)}\mathbf{B}\mathbf{A}^T\mathbf{A}\mathbf{B}^T \\ &= \mathbf{B}\mathbf{U}^T\mathbf{U}\mathbf{B}^T = \mathbf{B}\mathbf{B}^T\end{aligned}$$

assim, a decomposição espectral da matriz de covariâncias é também uma decomposição em valores singulares:

$$\mathbf{S} = \frac{1}{(n-1)}\mathbf{Y}^T\mathbf{Y} = \frac{1}{(n-1)}\mathbf{V}\mathbf{\Lambda}^2\mathbf{V}^T, \quad (15)$$

então a melhor aproximação da matriz de covariâncias de posto (q) é:

$$\mathbf{S} \approx \mathbf{S}_{(q)} = \frac{1}{(n-1)}\mathbf{V}_{(q)}\mathbf{\Lambda}_{(q)}\mathbf{\Lambda}_{(q)}\mathbf{V}_{(q)}^T = \mathbf{B}_{(q)}\mathbf{B}_{(q)}^T.$$

Como se está aproximando uma matriz $\mathbf{Y}_{(n \times p)}$ de posto (r) para uma matriz de menor posto (q) , perde-se informação, pois a representação do Biplot é aproximada. A maneira de medir essas perdas consiste em calcular a qualidade de representação do ajuste para a aproximação da matriz de variâncias e covariâncias que é definida por:

$$\left(\sum_{k=1}^q \lambda_k^4 / \sum_{k=1}^r \lambda_k^4 \right) \cdot 100,$$

em que λ_k são os autovalores da matriz $\mathbf{Y}$. Este valor é maior que o obtido para os elementos da matriz original. No GH-Biplot obtém-se uma máxima qualidade de representação para as colunas da matriz $\mathbf{Y}$ e uma mínima qualidade de representação para as linhas.

Existem algumas propriedades importantes no GH-Biplot:

- 1.1) o produto escalar de um vetor por ele mesmo é soma de quadrados de seus elementos e os elementos da diagonal da matriz $\mathbf{S}$ contém as variâncias $\mathbf{s}_j$ de cada uma das variáveis

$$\mathbf{S}_{jj} = \mathbf{b}_j^T \mathbf{b}_j;$$

- 1.2) a correlação entre duas variáveis é aproximada pelo cosseno do ângulo formado pelos correspondentes marcadores coluna:

$$\mathbf{b}_i^T \mathbf{b}_j = \|\mathbf{b}_i\| \|\mathbf{b}_j\| \cos(\mathbf{b}_i, \mathbf{b}_j)$$

portanto

$$\cos(\mathbf{b}_i, \mathbf{b}_j) = \frac{\mathbf{b}_i^T \mathbf{b}_j}{\|\mathbf{b}_i\| \|\mathbf{b}_j\|} \approx \frac{s_{ij}}{s_{ii} s_{jj}} = r_{ij};$$

- 1.3) a distância de Mahalanobis entre duas linhas da matriz $\mathbf{Y}$ coincide com a distância euclidiana entre os marcadores de linha, proporcionando uma melhor aproximação para a matriz de covariâncias do que a distância de Mahalanobis entre pontos do marcador linha.

Cada elemento da matriz $\mathbf{Y}$ pode ser escrito como $Y_{ij} = \mathbf{a}_i^T \mathbf{b}_j$, de forma que a linha i de $\mathbf{Y}$ pode ser escrita como $\mathbf{B}\mathbf{a}_i$ e a linha j pode ser escrita como $\mathbf{B}\mathbf{a}_j$. Então a distância de Mahalanobis entre duas linhas i e j da matriz $\mathbf{Y}$ pode aproximar-se como:

$$\begin{aligned} d^2(\mathbf{y}_i, \mathbf{y}_j) &= \sum_{k=1}^p (\mathbf{y}_{ik} - \mathbf{y}_{jk})^T \mathbf{S}^{-1} (\mathbf{y}_{ik} - \mathbf{y}_{jk}) \\ &= (\mathbf{B}\mathbf{a}_i - \mathbf{B}\mathbf{a}_j)^T \mathbf{S}^{-1} (\mathbf{B}\mathbf{a}_i - \mathbf{B}\mathbf{a}_j) \\ &= (\mathbf{a}_i - \mathbf{a}_j)^T \mathbf{B}^T \mathbf{S}^{-1} \mathbf{B} (\mathbf{a}_i - \mathbf{a}_j) \end{aligned}$$

substituindo pelas equações (14) e (15), temos:

$$\begin{aligned} d^2(\mathbf{y}_i, \mathbf{y}_j) &= (\mathbf{a}_i - \mathbf{a}_j)^T \frac{1}{\sqrt{n-1}} (\mathbf{V}\mathbf{\Lambda})^T ((n-1)\mathbf{V}\mathbf{\Lambda}^{-2}\mathbf{V}^T) \frac{1}{\sqrt{n-1}} (\mathbf{V}\mathbf{\Lambda}) (\mathbf{a}_i - \mathbf{a}_j) \\ &= (\mathbf{a}_i - \mathbf{a}_j)^T \mathbf{\Lambda} \mathbf{V}^T \mathbf{V} \mathbf{\Lambda}^{-2} \mathbf{V}^T \mathbf{V} \mathbf{\Lambda} (\mathbf{a}_i - \mathbf{a}_j) \\ &= (\mathbf{a}_i - \mathbf{a}_j)^T (\mathbf{a}_i - \mathbf{a}_j) \end{aligned}$$

Em dimensão reduzida tem-se que:

$$d^2(\mathbf{y}_i, \mathbf{y}_j) = \sum_{k=1}^p (\mathbf{y}_{ik} - \mathbf{y}_{jk})^T \mathbf{S}^{-1} (\mathbf{y}_{ik} - \mathbf{y}_{jk}) = (\mathbf{a}_i - \mathbf{a}_j)^T (\mathbf{a}_i - \mathbf{a}_j).$$

Através das propriedades dos produtos escalares calculados com a métrica associada à inversa da matriz de covariâncias

$$\mathbf{Y} \mathbf{S}^{-1} \mathbf{Y} = \mathbf{A}_{(q)} \mathbf{A}_{(q)}^T,$$

pode-se conseguir em dimensão reduzida, uma aproximação do produto escalar com a métrica de Mahalanobis.

- 2) Se $s = 1$, temos o **JK-Biplot** ou RMP-Biplot (Row Metric Preserving). Considera-se a matriz $\mathbf{Y}$ centrada por colunas. Este Biplot é obtido pela escolha dos marcadores de linhas, conforme as expressões abaixo:

$$\mathbf{A}_{(q)} = \mathbf{U}_{(q)}\mathbf{\Lambda}_{(q)} \quad \text{e} \quad \mathbf{B}_{(q)} = \mathbf{V}_{(q)}$$

Tendo a restrição métrica $\mathbf{B}^T\mathbf{B} = \mathbf{I}$ e $\mathbf{Y}(\mathbf{Y}^T\mathbf{Y})^{-1}\mathbf{Y}^T = \mathbf{B}\mathbf{B}^T$. A aproximação dos produtos escalares em dimensão reduzida é considerada ótima no sentido dos mínimos quadrados de:

$$\mathbf{Y}\mathbf{Y}^T = \mathbf{A}\mathbf{B}^T\mathbf{B}\mathbf{A}^T = \mathbf{A}\mathbf{V}^T\mathbf{V}\mathbf{A}^T = \mathbf{A}\mathbf{A}^T.$$

A decomposição espectral da matriz de produtos escalares entre as linhas é também sua decomposição em valores singulares:

$$\mathbf{Y}\mathbf{Y}^T = \mathbf{U}\mathbf{\Lambda}^2\mathbf{U}^T$$

então a melhor aproximação é a de posto q :

$$\mathbf{Y}\mathbf{Y}^T \approx \mathbf{U}_{(q)}\mathbf{\Lambda}_{(q)}^2\mathbf{U}_{(q)}^T = \mathbf{A}_{(q)}\mathbf{A}_{(q)}^T$$

que coincide com a aproximação que se obtém da matriz $\mathbf{Y}$. Então, a distância euclidiana entre duas linhas de $\mathbf{Y}$ coincide com a distância euclidiana entre os marcadores de $\mathbf{A}$.

A qualidade da representação é melhor para as linhas do que para as colunas. Os marcadores para as colunas são as projeções dos eixos originais (base canônica em espaço p -dimensional) no espaço das componentes principais. As propriedades e demonstrações são semelhantes ao GH-Biplot.

- 3) Se $s = 1/2$, temos o **SQRT-Biplot** (Square Root Biplot) ou Biplot Simétrico: Este Biplot é obtido pela escolha dos marcadores de linhas e de colunas, conforme as expressões abaixo:

$$\mathbf{A}_{(q)} = \mathbf{U}_{(q)}\mathbf{\Lambda}_{(q)}^{1/2} \quad \text{e} \quad \mathbf{B}_{(q)} = \mathbf{V}_{(q)}\mathbf{\Lambda}_{(q)}^{1/2}$$

Impondo como restrição métrica $\mathbf{A}^T \mathbf{A} \neq \mathbf{I}$ e $\mathbf{B}^T \mathbf{B} \neq \mathbf{I}$. Este tipo de Biplot é utilizado quando o objetivo da análise consiste na aproximação dos elementos Y_{ij} da matriz $\mathbf{Y}$, por exemplo, no caso de modelos de diagnóstico em tabelas de contingência, em que tanto as linhas como as colunas são igualmente importantes.

- 4) O **HJ-Biplot** (Row-Column Metric Preserving): Quando as linhas e colunas têm a mesma importância e querem interpretar as características de ambas mantendo a relação entre elas, são utilizadas interpretações simétricas como Análise Fatorial de Correspondência, em que interpretam as posições das linhas, as posições das colunas e as relações linha-coluna através dos fatores.

Galindo (1986) propõe através da DVS a escolha dos marcadores:

$$\mathbf{A}_{(q)} = \mathbf{U}_{(q)} \mathbf{\Lambda}_{(q)} \quad \text{e} \quad \mathbf{B}_{(q)} = \mathbf{V}_{(q)} \mathbf{\Lambda}_{(q)},$$

Verifica-se que com essa fatoração, os dados originais não se reproduzem, ou seja $\mathbf{Y} \neq \mathbf{A}\mathbf{B}^T$. Neste caso, o objetivo é ter uma máxima qualidade de representação do ajuste para as linhas e colunas de $\mathbf{Y}$, em que é incorporado a cada matriz de marcadores, a matriz diagonal $\mathbf{\Lambda}$, que possibilita que a decomposição espectral tanto para a matriz de variâncias e covariâncias entre colunas como para a matriz de distâncias euclidianas entre linhas, coincida com a decomposição em valores singulares de $\mathbf{Y}$. A qualidade do ajuste tanto para as linhas como para as colunas é a mesma, sendo melhor que os Biplots desenvolvidos por Gabriel(1971).

Como as coordenadas das linhas do HJ-Biplot coincidem com as do JK-Biplot e as coordenadas das colunas são as mesmas do GH-Biplot, o HJ-Biplot possui as mesmas propriedades dos respectivos Biplot desenvolvidos por Gabriel (1971). A proximidade entre indivíduos é interpretada como similaridade, os ângulos entre dois vetores que unem duas variáveis com a origem, como correlação. Projetando os indivíduos sobre os vetores que representam as variáveis é possível determinar a distribuição aproximada que os indivíduos apresentam sobre estas variáveis (CÁRDENAS; GALINDO; VICENTE-VILLARDÓN, 2007).

A vantagem de se utilizar HJ-Biplot, é que ele proporciona uma melhor qualidade

na representação simultânea tanto para as linhas (indivíduos) como para as colunas (variáveis), cujo objetivo não é a aproximação dos elementos da matriz de dados.

A Tabela 1 apresenta uma comparação das representações das coordenadas dos Biplots desenvolvidos por Gabriel (GH-Biplot, JK-Biplot e SQRT-Biplot) e por Galindo (HJ-Biplot) com as respectivas qualidades de representação dos ajustes para os marcadores das linhas e colunas.

Tabela 1 - Resumo das características mais importantes dos Biplots considerados na representação da matriz de dados $\mathbf{Y}$ com posto (q)

	Representação Simultânea	Coordenadas		Qualidade de representação	
		linhas	colunas	linhas	colunas
$s = 0$	GH-Biplot	$\mathbf{U}$	$\mathbf{\Lambda V}$	$\frac{q}{r}$	$\frac{\sum_{k=1}^q \lambda_k^4}{\sum_{k=1}^r \lambda_k^4}$
$s = 1$	JK-Biplot	$\mathbf{U\Lambda}$	$\mathbf{V}$	$\frac{\sum_{k=1}^q \lambda_k^4}{\sum_{k=1}^r \lambda_k^4}$	$\frac{q}{r}$
$s = 1/2$	SQRT-Biplot	$\mathbf{U\Lambda^{1/2}}$	$\mathbf{\Lambda^{1/2}V}$	$\frac{\sum_{k=1}^q \lambda_k^4}{\sum_{k=1}^r \lambda_k^4}$	$\frac{\sum_{k=1}^q \lambda_k^4}{\sum_{k=1}^r \lambda_k^4}$
	HJ-Biplot	$\mathbf{U\Lambda}$	$\mathbf{\Lambda V}$	$\frac{\sum_{k=1}^q \lambda_k^4}{\sum_{k=1}^r \lambda_k^4}$	$\frac{\sum_{k=1}^q \lambda_k^4}{\sum_{k=1}^r \lambda_k^4}$

Através dos Biplots clássicos surgiram novas alternativas e contribuições dos métodos Biplot com várias aplicações práticas, seguindo as últimas linhas de pesquisa em análise multivariada, dos quais pode-se citar:

- a) Biplot MLG** (Biplot em Modelo Linear Generalizado): Para a aproximação dos Biplots, vários autores se fundamentam em generalizações heurísticas dos métodos de estimação utilizados nos modelos bilineares, considerados como extensões dos modelos lineares

generalizados, segundo NELDER e WEDDERBUM (1972):

$$\eta = g[E(\mathbf{y})] = g(\mu) = \mathbf{AB}^T \Rightarrow g(\mu_{ij}) = \mathbf{a}_i^T \mathbf{b}_j$$

assim, as matrizes dos marcadores $\mathbf{A}$ e $\mathbf{B}$ na fatorização clássica, conforme equação (4), são em modelos lineares generalizados as matrizes de parâmetros desconhecidos, em que a distribuição das p variáveis da matriz $\mathbf{Y}$ é uma das distribuições pertencentes à família exponencial (normal, binomial, poisson etc), e os valores esperados denotados por $\mu = E(\mathbf{y})$ das variáveis observadas, se relacionam com os preditores lineares η através da função de ligação $g(\cdot)$, como por exemplo a identidade, a logit, a probit, etc, em que $g(\mu_{ij})$ resulta em uma forma linearizada da função de ligação (CÁRDENAS; GALINDO; VICENTE-VILLARDÓN, 2001).

Existe dois tipos de Biplots ajustados por MLG:

- 1) quando a função de ligação utilizada é diferente da identidade ($g(\cdot) \neq I$) é denominado de “Biplot de Regressão não linear”;
- 2) quando a função de ligação for igual a identidade ($g(\cdot) = I$) é denominado de “Biplot de Regressão linear”, segundo a terminologia utilizada por Gower e Hand (1996).

Como alternativa, Vicente-Villardón *et al.* (2006) utilizaram Análise de Correspondências Múltiplas para variáveis respostas binárias (ausência/presença) com distribuição binomial, consideraram o enfoque dos Biplots de Predição e aplicaram sobre os Biplots ajustados através da Regressão Bilinear Generalizada. Na forma matricial o modelo linear Biplot Logístico tem a seguinte função de ligação:

$$\text{logit}(\boldsymbol{\pi}) = \log\left(\frac{\boldsymbol{\pi}}{1 - \boldsymbol{\pi}}\right) = \mathbf{1b}_0^T + \mathbf{AB}^T,$$

em que $\boldsymbol{\pi}$ é a matriz das probabilidades esperadas, $\mathbf{1}$ é o vetor de uns, $\mathbf{b}_0^T$ é o vetor de constantes, $\mathbf{A}$ e $\mathbf{B}^T$ são as matrizes formada pelos marcadores de linhas e colunas da matriz $\mathbf{Y}$. A inclusão do vetor de constantes é pelo fato de não ser possível centralizar a matriz de dados constituída por valores binários.

- b) **Biplot PMD** (Biplots para mineração de dados): O processo de procurar valiosas informações em grandes massas de dados é conhecido por mineração de dados (*data*

mining), cujo exemplo é o projeto genoma humano, que já armazenou centenas de gigabytes de dados. A mineração de dados é considerada um assunto interdisciplinar, que representa a confluência de várias idéias, inclusive da análise exploratória de dados dentre outras. Essa técnica deve-se satisfazer os três princípios básicos: incrementabilidade, escalabilidade e interpretabilidade.

Vairinhos e Galindo (2004) desenvolveram a interface gráfica para sistemas de *data mining* usando o conceito de Biplot. O surgimento de bases de dados gigantes, com milhares de entradas, torna-se inviável a realização de possíveis interpretações. Surgem problemas, tais como, limitações de memória dos computadores e a dificuldade de realizar sem erros lógicos ou omissões, as operações referentes à interpretação. Do reconhecimento destes fatos resultou a necessidade de desenvolver sistemas capazes de sugerir possíveis interpretações ao analista, que o auxiliem nas operações de interpretação desses resultados.

A concepção de algoritmos com essa capacidade exige uma formulação matemática do problema, que permite aproximar conjuntos de indivíduos ou observações através de construção de sistemas visuais adaptados do conceito Biplot.

2.2.5 Qualidade da aproximação do Biplot

Na interpretação de um Biplot é necessário considerar a qualidade da aproximação denominada de “qualidade global de representação”, pois o posto da matriz de dados original $\mathbf{Y}$ é geralmente superior a dois, portanto não é possível fazer uma representação exata sobre o plano. A qualidade global de representação é definida de forma similar ao coeficiente de determinação calculado na análise de regressão, ou seja, o quociente entre a soma de quadrados dos elementos de $\mathbf{Y}_{(q)}$ e a soma de quadrados dos elementos de $\mathbf{Y}$,

$$CR_{Global} = \frac{(\lambda_i)^2}{\sum_{k=1}^q (\lambda_k)^2} = \frac{\text{variabilidade explicada pelo eixo } i}{\text{variabilidade total}}.$$

Se esse indicador não for suficiente para avaliar os ajustes individuais de cada linha e de cada coluna, deve-se considerar as contribuições do fator ao elemento linha e ao

elemento coluna (qualidades de representação) que determinam a parte da variabilidade de cada elemento explicado por esse fator. A interpretação desses efeitos é similar ao da análise de correspondência através da inércia explicada ou variância absorvida (CÁRDENAS *et al*, 2003).

Considerando-se que todos os elementos possuam o mesmo peso, então a inércia de uma variável (elemento coluna) com o centro da origem é igual ao quadrado da distância deste ponto ao centro da origem. A contribuição do fator ao elemento coluna é definido por:

$$CR_{coluna\ j} = \frac{b_{jk}^2}{\sum_{k=1}^q b_{jk}^2},$$

para a j -ésima variável, k -ésimo eixo com q dimensões. Essas contribuições permitem saber quais são as variáveis responsáveis pela ordenação dos indivíduos sobre as projeções de cada um dos eixos do Biplot.

A contribuição do fator ao elemento linha é definido como o quociente entre o quadrado das coordenadas da i -ésima linha e a soma de quadrados das coordenadas dos n pontos da linha sobre os q eixos ($q < r$),

$$CR_{linha\ i} = \frac{a_{ik}^2}{\sum_{k=1}^q a_{ik}^2},$$

calcula a variabilidade da i -ésima linha ao k -ésimo eixo.

Na interpretação de um gráfico Biplot pode-se utilizar outros indicadores que permitam ressaltar a importância das diferentes formas de transformações dos dados (centrados, padronizados etc) que influenciam sobre as distâncias entre indivíduos e entre variáveis.

2.3 Interação Genótipo-Ambiente ($G \times A$)

O cultivo extensivo em ambientes diversificados provoca uma resposta diferencial dos genótipos. A seleção de genótipos com alta produtividade é o objetivo dos programas de melhoramento genético. Neste sentido, a interação genótipos e ambientes ($G \times A$),

representa o componente da variação fenotípica resultante do comportamento diferencial apresentado pelos genótipos quando submetidos a mais de um tipo de ambiente.

Para avaliar um grupo de genótipos testados em vários ambientes realizam-se análises da adaptabilidade e estabilidade, tornando possível a identificação de cultivares com comportamento previsível em função do estímulo ambiental, considerando os diferentes locais, épocas ou anos de plantio, diversos níveis tecnológicos etc.

A adaptabilidade é a capacidade do genótipo aproveitar os efeitos ambientais, garantindo o alto nível de produtividade. A estabilidade está relacionada com a manutenção da produtividade ou da previsibilidade nos diversos ambientes. Existem diversas metodologias propostas para investigar a adaptabilidade e a estabilidade fenotípica. A diferença entre elas está nos conceitos e procedimentos biométricos utilizados para medir a interação $G \times A$. A escolha de um método de análise depende dos dados experimentais, do número de ambientes, da precisão requerida e do tipo de informação necessária (LAVORANTI, 2003).

Embora existam várias metodologias, para a avaliação de cultivares e identificação de mega-ambientes, que os pesquisadores atuais continuam buscando instrumentos estatísticos mais eficientes. Entre as metodologias tradicionais, pode-se destacar os métodos baseados em:

1. Análise de variância: PLAISTED e PETERSON (1959); WRICKE (1965); SHUKLA (1972); ANNICCHIARICO (1992) e MAGARI & KANG (1997);
2. Análise de regressão linear simples: FINLAY e WILKINSON (1963); EBERHART e RUSSELL (1966); PERKINS e JINKS (1968) e TAI (1971);
3. Análise de regressão linear bi-segmentada: VERMA et al. (1978); SILVA e BARRETO (1986); CRUZ et al. (1989) e STORCK & VENCOVSKY (1994);
4. Análise não lineares: CHAVES et al. (1989); TOLER e BURROWS (1998); SILVA (1998) e ROSSE e VENCOVSKY (2000);
5. Análises não paramétricas, como a ordem de classificação genotípica: LIN e BINNS (1988); HUEHN (1990) e FOX et al. (1990);

6. Análises multivariadas que integram a análise comum de variância (método univariado) com a análise de componentes principais (método multivariado): AMMI (Additive Main Effects and Multiplicative Interaction), SREG (Sites Regression), GREG (Genotypes Regression model), SHMM (Shifted Multiplicative Model) e COMM (Complete Multiplicative Model) (CORNELIUS et al., 1996).

Segundo DIAS (2005), a maioria dos dados coletados em experimentos na agricultura são de natureza multivariada, pois vários atributos são medidos em cada um dos indivíduos incluídos nos experimentos. As técnicas multivariadas mais utilizadas usadas para analisar tais dados incluem análise de componentes principais (ACP); variáveis canônicas ou análise discriminante; análise de correlação canônica e análise de agrupamento.

Dentre as diversas metodologias já propostas para o estudo da interação $G \times A$, os modelos estatísticos multiplicativos são muito úteis para estudar padrões de performance de genótipos por ambientes e fazer previsões a respeito da performance média de genótipos à ambientes específicos. A Tabela 2 mostra os principais modelos utilizados nas análise estatística de interação $G \times A$ que obtêm o gráfico biplot por meio da técnica DVS (Decomposição em valores singulares) aplicada à matriz de interação $G \times A$ (KROONENBERG, 2007).

Em relação as novas análises estatísticas propostas para a interação $G \times A$, a análise AMMI se destaca, pois interpreta a interação $G \times A$ como componente multiplicativo e a submete à análise por componentes principais. Proposto por Mandel (1971), as vantagens deste método são de:

- a) permitir uma análise mais detalhada da interação genótipo-ambiente;
- b) garantir a seleção dos genótipos, capitalizando suas interações positivas com os ambientes;
- c) auxiliar na realização do zoneamento agrônômico com a finalidade de selecionar os locais de teste;
- d) auxiliar na identificação de genótipos adaptados ao ambiente com alta produtividade;
- e) possibilitar fácil interpretação gráfica dos resultados, nos chamados Biplots (representação gráfica simultânea dos genótipos e ambientes);
- f) propiciar estimativas mais precisas das respostas genotípicas.

Tabela 2 - Principais modelos utilizados na interação $G \times A$

Modelo	Modelo ajustado
AMMI (Additive Main Effects and Multiplicative Interaction)	$Y_{ij} = \mu + g_i + a_j + \bar{\varepsilon}_{ij}$
$Y_{ij} = \mu + g_i + a_j + (ga)_{ij} + \bar{\varepsilon}_{ij}$	
$(ga)_{ij} = \sum_{k=1}^p \lambda_k \gamma_{ik} \alpha_{jk}$ pela DVS	
$Y_{ij} = \mu + g_i + a_j + \sum_{k=1}^p \lambda_k \gamma_{ik} \alpha_{jk} + \bar{\varepsilon}_{ij}$	
GREG (Genotype Regression)	$\bar{Y}_{ij} = \mu + g_i + \bar{\varepsilon}_{ij}$
$\bar{Y}_{ij} = \mu + g_i + \sum_{k=1}^p \lambda_k \gamma_{ik} \alpha_{jk} + \bar{\varepsilon}_{ij}$	
SREG (Sites Regression) ou GGE	$\bar{Y}_{ij} = \mu + a_i + \bar{\varepsilon}_{ij}$
$\bar{Y}_{ij} = \mu + a_i + \sum_{k=1}^p \lambda_k \gamma_{ik} \alpha_{jk} + \bar{\varepsilon}_{ij}$	
SHMM (Shifted Multiplicative Model)	$\bar{Y}_{ij} = \mu + \bar{\varepsilon}_{ij}$
$\bar{Y}_{ij} = \mu + \sum_{k=1}^p \lambda_k \gamma_{ik} \alpha_{jk} + \bar{\varepsilon}_{ij}$	
COMM (Completely Multiplicative)	$\bar{Y}_{ij} = \bar{\varepsilon}_{ij}$
$\bar{Y}_{ij} = \sum_{k=1}^p \lambda_k \gamma_{ik} \alpha_{jk} + \bar{\varepsilon}_{ij}$	

Outra abordagem para estudos da interação $G \times A$, proposta por Yan e Kang (2002), é o método utilizado para avaliar a adaptabilidade e a estabilidade fenotípica baseado em análise gráfica denominada de GGEbiplot (efeitos principais dos genótipos e interação $G \times A$). Este Biplot é construído plotando-se os dois componentes principais de uma ACP (Análise de Componentes Principais) utilizando Modelos de Regressões Locais (SREG). A primeira componente, quando se encontra altamente correlacionada com o efeito principal do genótipo, representa a proporção do rendimento que se deve somente às características do genótipo. A segunda componente representa a parte do rendimento devida à interação genótipo-ambiente.

Neste gráfico pode-se visualizar três aspectos importantes:

- a) a relação genótipo $\times$ ambiente, permitindo o agrupamento de genótipos e ambientes

com comportamentos semelhantes, mostrando o genótipo com maior potencial e sua identificação em cada subgrupo de ambientes (mega-ambientes);

b) a inter-relação entre ambientes, facilitando a identificação do melhor ambiente na avaliação das cultivares e indicando qual ambiente pode ser menos favorável;

c) a inter-relação entre genótipos, facilitando a comparação dos genótipos e do ordenamento para os parâmetros de rendimento e estabilidade (YAN, 2002).

2.3.1 Regras de interpretação do Biplot para interação ($G \times A$)

Um ponto importante dos atuais gráficos para o Biplot é que os eixos das coordenadas vertical e horizontal possuam a mesma escala. Uma das propriedades do Biplot é que os produtos internos de um genótipo (marcador de linha) e um ambiente (marcador de coluna) são aproximadamente idênticos aos valores da tabela de dupla entrada. A interpretação do Biplot quanto a interação $G \times A$ é feita observando-se a magnitude e os sinais dos escores dos genótipos e dos ambientes para os eixos que representam a interação. Kroonenberg, 2007 apresenta algumas interpretações para tabela $G \times A$:

a) Geral (sem considerar as escalas dos eixos das coordenadas):

a.1) os genótipos são perfeitamente mostrados como pontos e os ambientes como vetores;

a.2) se dois vetores de genótipos formarem um ângulo pequeno, significa que eles possuem o mesmo padrão de resposta sob os ambientes;

a.3) se dois vetores de ambientes formarem um ângulo pequeno significa que eles estão fortemente associados.

b) Ambientes centrados:

b.1) o Biplot mostra a tabela do efeito principal do genótipo mais o efeito de interação;

b.2) os genótipos divergem da média para cada ambiente;

- b.3)** a origem representa os valores médios de cada ambiente, isto é, representa o genótipo o qual possui um valor médio em cada ambiente. Este genótipo médio possui valor zero na matriz de dados centrados;
- b.4)** um genótipo com uma grande distância em relação a origem, indica que o genótipo possui alto efeito de interação;
- b.5)** extensa projeção de um genótipo em um vetor ambiente, significa que este genótipo diverge para a média dos ambientes;

c) Genótipos e ambientes centrados:

- c.1)** o Biplot mostra a interação de uma tabela de dupla entrada, em que existe no mínimo $\min(g, a)$ dimensões ou eixos de coordenadas;
- c.2)** os genótipos e os ambientes divergem para suas médias;
- c.3)** a origem representa os valores médios para cada ambiente e cada genótipo por todos os ambientes;
- c.4)** um genótipo (ambiente) com uma grande distância da origem têm alto efeito de interação com pelo menos um ambiente (genótipo);
- c.5)** extensa projeção de um genótipo em um vetor ambiente, significa que este genótipo diverge para a média do ambiente, e vice-versa.

d) Escala do Componente Principal: $\mathbf{U}$, $\mathbf{V}$ e Λ (Biplot Componente Principal):

d.1) Ambiente centrado:

- i)** o cosseno do ângulo entre dois quaisquer vetores ambientes aproxima de sua correlação, ocorrendo a igualdade significa que o ajuste esta perfeito;
- ii)** os comprimentos dos vetores ambientes são aproximadamente proporcionais ao padrão dos ambientes divergentes com proporcionalidade exata se o ajuste for perfeito;
- iii)** o produto interno entre dois vetores ambientes aproxima sua covariância, ocorrendo a igualdade significa que o ajuste esta perfeito;

- iv) a distância euclidiana entre dois genótipos não aproxima as distâncias entre suas linhas da matriz original, mas de sua distância padronizada, que é a raiz quadrada da distância de Mahalanobis.
- v) os vetores ambientes podem ser mais longos que os genótipos, fazendo com que o gráfico tenha uma aparência deselegante; então multiplica-se todas as coordenadas do ambiente por uma constante arbitrária, que fará com que os comprimentos dos vetores dos genótipos e ambientes sejam comparados. Porém, não é obrigado usar a constante, uma vez que ela é uma medida *ad-hoc*.

d.2) Ambiente padronizado:

- i) o comprimento dos vetores ambientes indica quão bem os ambientes estão representados pelo gráfico com um perfeito ajuste para todos os vetores de comprimentos iguais.
- ii) o produto interno entre dois ambientes (cosseno do ângulo entre eles) aproxima de suas correlações com igualdade se o ajuste estiver perfeito.

e) Escala simétrica $U\Lambda^{1/2}$ e $V\Lambda^{1/2}$:

- e.1) se o ângulo entre dois vetores ambientes for pequeno, significa que eles são altamente correlacionados, mas a correlação deles não pode ser deduzida pelo gráfico; a associação é semelhante para os genótipos;
- e.2) devido a escala simétrica dos ambientes e genótipos, ambos são locados no mesmo espaço e o produto interno é facilmente avaliado.

2.4 MANOVA Biplot

A idéia da MANOVA Biplot surgiu com o trabalho de Gabriel (1972), na análise de matrizes de dados meteorológicos, demonstrando detalhadamente a precipitação em Israel. A MANOVA Biplot proporciona uma visão clara da estrutura multivariada dos dados, mostrando as inter-relações complexas entre os grupos e entre as variáveis. O Biplot construído representa aproximadamente os dados de forma simultânea, separando os grupos por círculos (regiões de confiança).

2.4.1 MANOVA

Nos delineamentos experimentais, geralmente se dispõem de várias variáveis respondidas, buscando as diferenças entre os vários níveis de fatores (grupos). A análise adequada para esta situação é a análise de variância multivariada (MANOVA), que é uma extensão da análise de variância simples (ANOVA). A diferença entre as duas está no fato de que a ANOVA avalia os efeitos de médias dos fatores para uma variável, enquanto a MANOVA é usada para investigar se os vetores de médias dos fatores de estudo são os mesmos, e se não, qual componente de média difere significativamente.

Algumas pressuposições da estrutura dos dados devem ser obedecidas para validade da inferência estatística, tais como: as amostras dos tratamentos devem ser independentes; todos os tratamentos possuem covariância comum e cada tratamento tem distribuição normal multivariada.

As limitações de se utilizar as técnicas multivariadas está na complexidade da apresentação dos resultados e na interpretação dos mesmos, principalmente quando se tem várias variáveis, devido às inter-relações entre elas e à variabilidade de características associadas aos modelos. Assim, muitos pesquisadores preferem utilizar análises univariadas para cada uma das variáveis separadamente, com consequentes erros e omissões na interpretação de seus resultados (AMARO; VICENTE-VILLARDÓN; GALINDO-VILLARDÓN, 2004).

Para superar esses problemas têm-se utilizado a metodologia Biplot, com o intuito de representar os resultados de uma MANOVA. O método MANOVA-Biplot para tabela de uma entrada foi desenvolvido por Gabriel (1972, 1995), Biplot Canônico (VICENTE-VILLARDÓN, 1992; GOWER e HAND, 1996) e MANOVA Biplot (AMARO; VICENTE-VILLARDÓN; GALINDO-VILLARDÓN, 2004), permitindo uma representação ponderada da matriz de médias e as direções com o máximo poder de discriminação entre os grupos. Esta técnica serve para ilustrar as principais diferenças entre os grupos de indivíduos, as variáveis responsáveis pelas diferenças e fazer inferências sobre as variáveis originais através de círculos de confiança situados sobre os pontos que representam os grupos.

Para entender a estrutura de dados e o modelo, suponha que temos n observações independentes com p variáveis observadas $Y_1, Y_2, \dots, Y_p$, sob diversas condições experimentais (níveis de um fator). O modelo MANOVA possui a seguinte forma (JOHNSON;

WICHERN, 2002):

$$\mathbf{Y} = \mathbf{Z} \boldsymbol{\beta} + \boldsymbol{\varepsilon} = [\mathbf{Z}\tilde{\boldsymbol{\beta}}_1, \mathbf{Z}\tilde{\boldsymbol{\beta}}_2, \dots, \mathbf{Z}\tilde{\boldsymbol{\beta}}_p] + [\tilde{\boldsymbol{\varepsilon}}_1, \tilde{\boldsymbol{\varepsilon}}_2, \dots, \tilde{\boldsymbol{\varepsilon}}_p] \quad (16)$$

então, a matriz das observações é dada por

$$\mathbf{Y}_{(n \times p)} = [\tilde{\mathbf{y}}_1, \tilde{\mathbf{y}}_2, \dots, \tilde{\mathbf{y}}_p],$$

de dimensão $(n \times p)$ e posto $r < p$, em que $\tilde{\mathbf{y}}_j = (y_{1j}, y_{2j}, \dots, y_{nj})^T$ é um vetor com as n observações da variável Y_j , $j = 1, 2, \dots, p$. A matriz de delineamento é

$$\mathbf{Z}_{(n \times q)} = [\tilde{\mathbf{z}}_1, \tilde{\mathbf{z}}_2, \dots, \tilde{\mathbf{z}}_q].$$

Quando $\mathbf{Z}$ representa uma matriz de q variáveis independentes observadas em cada n indivíduos, a expressão (16) é chamada de Modelo de Regressão Multivariada. Quando $\mathbf{Z}$ é uma matriz de delineamento (composta por 0's e 1's) é chamada de Modelo Linear Geral. O cálculo é o mesmo para ambos os casos, supondo que a primeira coluna de $\mathbf{Z}$ é um vetor de uns $\mathbf{1}$, permitindo o efeito da média global (MARDIA, 1979).

A matriz de parâmetros desconhecidos é

$$\boldsymbol{\beta}_{(q \times p)} = [\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_2, \dots, \tilde{\boldsymbol{\beta}}_p]$$

e a matriz dos erros aleatórios

$$\boldsymbol{\varepsilon}_{(n \times p)} = [\tilde{\boldsymbol{\varepsilon}}_1, \tilde{\boldsymbol{\varepsilon}}_2, \dots, \tilde{\boldsymbol{\varepsilon}}_p],$$

em que $\boldsymbol{\varepsilon} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$.

A estimação dos parâmetros é análoga ao caso univariado, $\hat{\boldsymbol{\beta}}$ é o estimador de mínimos quadrados de $\boldsymbol{\beta}$ que minimiza $\|\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}\| = \text{tr}[(\mathbf{Y} - \mathbf{Z}\boldsymbol{\beta})^T (\mathbf{Y} - \mathbf{Z}\boldsymbol{\beta})]$,

$$\hat{\boldsymbol{\beta}} = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{Y} = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T (\tilde{\mathbf{y}}_1, \tilde{\mathbf{y}}_2, \dots, \tilde{\mathbf{y}}_p).$$

O teste de hipótese multivariado é definido por $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$, em que $\mathbf{C}$ tem dimensões $(g \times q)$ e posto $g \leq r$, em que r é o posto da matriz $\mathbf{Z}$. A matriz $\mathbf{C}$ possui algumas formas distintas, inclusive a de matriz identidade, pois é formada por coeficientes de contrastes para as médias. Assim, é possível a construção de Biplots para as distintas

hipóteses. Mardia (1979) define que as linhas da matriz $\mathbf{C}$ efetuam afirmações sobre o efeito na regressão das combinações lineares das variáveis independentes.

Seja $\mathbf{D} = \mathbf{C}\boldsymbol{\beta}$, o estimador de $\mathbf{D}$ é definido por $\hat{\mathbf{D}} = \mathbf{C}\hat{\boldsymbol{\beta}} = \mathbf{C}(\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{Y}$, então a matriz de soma de quadrados e produtos “entre” níveis do fator é dada por

$$\mathbf{H} = \underbrace{\mathbf{Y}^T \mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{C}^T}_{\hat{\mathbf{D}}^T} \underbrace{[\mathbf{C} (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{C}^T]}_{\mathbf{R}^{-1}} \underbrace{\mathbf{C} (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{Y}}_{\hat{\mathbf{D}}}$$

e a matriz de soma de quadrados e produtos “dentro” dos níveis do fator é definida por

$$\mathbf{E} = \mathbf{Y}^T [\mathbf{I} - \mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T] \mathbf{Y},$$

sendo que $\mathbf{R} = \mathbf{C} (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{C}^T$.

A representação do Biplot da matriz $\mathbf{Y}$ pode ser obtida a partir da Decomposição em Valores Singulares (DVS), conforme a equação (2) em que $\mathbf{Y} = \mathbf{U} \boldsymbol{\Lambda} \mathbf{V}^T$ e pela Decomposição em Valores Singulares Generalizados (DVSG), tem-se $\mathbf{Y}^* = \mathbf{R}^{-1/2} \hat{\mathbf{D}} \mathbf{E}^{-1/2}$, então

$$\mathbf{R}^{-1/2} \hat{\mathbf{D}} \mathbf{E}^{-1/2} = \mathbf{U} \boldsymbol{\Lambda} \mathbf{V}^T$$

portanto,

$$\hat{\mathbf{D}} = \underbrace{\mathbf{R}^{1/2} \mathbf{U} \boldsymbol{\Lambda}}_{\mathbf{A}} \underbrace{\mathbf{V}^T \mathbf{E}^{1/2}}_{\mathbf{B}}$$

Tomando como marcadores de linhas (grupos) a matriz

$$\mathbf{A} = \mathbf{R}^{1/2} \mathbf{U} \boldsymbol{\Lambda}$$

e os marcadores para as colunas (variáveis)

$$\mathbf{B} = \mathbf{E}^{1/2} \mathbf{V}^T.$$

A aproximação da matriz de parâmetros é dada por

$$\mathbf{A} \mathbf{B}^T = \mathbf{R}^{1/2} \mathbf{U} \boldsymbol{\Lambda} \mathbf{V}^T \mathbf{E}^{1/2} = \mathbf{R}^{1/2} \mathbf{R}^{-1/2} \hat{\mathbf{D}} \mathbf{E}^{-1/2} \mathbf{E}^{1/2} = \hat{\mathbf{D}}.$$

Essa decomposição é denominada de Decomposição em Valores Singulares Generalizada da matriz de estimadores dos parâmetros $\hat{\mathbf{D}}$ utilizando $\mathbf{R}^{-1/2}$ e $\mathbf{E}^{-1/2}$ como

métricas, dessa forma o primeiro autovalor da decomposição do Biplot coincide com o da MANOVA e consegue-se obter as direções de separações entre os níveis do fator. Em geral, o Biplot obtido é de difícil interpretação, pois precisa de uma representação alternativa. Nas representações das médias conserva-se a propriedade de que o eixo maximiza o teste F para as hipóteses estudadas (AMARO; VICENTE-VILLARDÓN; GALINDO-VILLARDÓN, 2004).

2.4.2 MANOVA dupla entrada

Na MANOVA de dupla entrada/bidimensional pode-se testar várias hipóteses para os efeitos principais e para as interações; associado a cada uma das hipóteses pode-se construir distintas representações canônicas que permitem o estudo separado de cada uma delas.

Seja y_{ijkh} representando a k -ésima observação sobre a h -ésimo grupo obtido sob i -ésimo nível do fator da primeira entrada de classificação (linha) e o j -ésimo nível do fator da segunda entrada (coluna). Considere $\mathbf{y}_{ijk}^T = [y_{ijk1}, \dots, y_{ijkp}]$, os dados obtidos de um delineamento de dupla entrada com r linhas e c colunas organizados conforme a Tabela 3.

Tabela 3 - Esquema das observações em tabelas de dupla entrada

Tratamentos	Tratamentos Colunas		
Linhas	1	$\dots$	c
1	$\mathbf{y}_{111}^T$	$\dots$	$\mathbf{y}_{1c1}^T$
	$\vdots$	$\dots$	$\vdots$
	$\mathbf{y}_{11n}^T$	$\dots$	$\mathbf{y}_{1cn}^T$
$\vdots$	$\dots$	$\dots$	$\dots$
r	$\mathbf{y}_{r11}^T$	$\dots$	$\mathbf{y}_{rc1}^T$
	$\vdots$	$\dots$	$\vdots$
	$\mathbf{y}_{r1n}^T$	$\dots$	$\mathbf{y}_{rcn}^T$

Morrison (1978) define o modelo linear para cada observação por:

$$y_{ijkh} = \mu_h + \underbrace{\alpha_{ih}}_{\text{fator 1}} + \underbrace{\tau_{jh}}_{\text{fator 2}} + \eta_{ijh} + \varepsilon_{ijkh}$$

em que:

μ_h : constante (média geral) da h -ésimo grupo , $h = 1, \dots, p$;

α_{ih} : efeito do i -ésimo nível do fator 1 ou linha da h -ésimo grupo, $i = 1, \dots, r$;

τ_{jh} : efeito do j -ésimo nível do fator 2 ou coluna da h -ésimo grupo, $j = 1, \dots, c$;

η_{ijh} : efeito da interação do i -ésimo nível do fator 1 e j -ésimo nível do fator 2 da h -ésimo grupo;

ε_{ijkh} : erro experimental associado a cada observação, pressuposto a $N_p(\mathbf{0}, \mathbf{\Sigma})$ e $k = 1, \dots, n$.

A matriz de parâmetros pode ser escrita como:

$$\boldsymbol{\beta} = \begin{pmatrix} \alpha_{11} \cdots \alpha_{r1} & \tau_{11} \cdots \tau_{c1} & \eta_{111} \cdots \eta_{1c1} \cdots \eta_{r11} \cdots \eta_{rc1} & \mu_1 \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_{1p} \cdots \alpha_{rp} & \tau_{1p} \cdots \tau_{cp} & \eta_{11p} \cdots \eta_{1cp} \cdots \eta_{r1p} \cdots \eta_{rcp} & \mu_p \end{pmatrix}$$

A matriz de delineamento pode ser expressa na seguinte forma:

$$\mathbf{Z} = \begin{pmatrix} \mathbf{j} \cdots \mathbf{0} & \mathbf{j} \cdots \mathbf{0} & \mathbf{j} \cdots \mathbf{0} & \cdots & \mathbf{0} \cdots \mathbf{0} & \mathbf{j} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ \mathbf{j} \cdots \mathbf{0} & \mathbf{0} \cdots \mathbf{j} & \mathbf{0} \cdots \mathbf{j} & \cdots & \mathbf{0} \cdots \mathbf{0} & \mathbf{j} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ \mathbf{0} \cdots \mathbf{j} & \mathbf{j} \cdots \mathbf{0} & \mathbf{0} \cdots \mathbf{0} & \cdots & \mathbf{j} \cdots \mathbf{0} & \mathbf{j} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ \mathbf{0} \cdots \mathbf{j} & \mathbf{0} \cdots \mathbf{j} & \mathbf{0} \cdots \mathbf{0} & \cdots & \mathbf{0} \cdots \mathbf{j} & \mathbf{j} \end{pmatrix}$$

em que $\mathbf{j}$ é o vetor coluna de n uns, enquanto $\mathbf{0}$ é o vetor coluna contendo n zeros. O primeiro bloco horizontal de tais vetores em $\mathbf{Z}$ corresponde aos n vetores de observações em cada c níveis do fator coluna aparecendo com o primeiro fator linha. O modelo é repetido para r blocos dos vetores colunas. As colunas $r+c+1, \dots, r+c+rc$ da matriz apresenta os parâmetros

de efeito de interação η_{ijh} , enquanto a última coluna de uns considera os parâmetros de nível geral μ_h .

Na forma matricial, tem-se:

$$\mathbf{Y}_{(rcn) \times p} = \mathbf{Z}_{(rcn) \times (r+1)(c+1)} \boldsymbol{\beta}_{(r+c)(c+1) \times (p)} + \boldsymbol{\varepsilon}_{(rcn) \times (p)}.$$

As matrizes para os testes de hipóteses dos efeitos das linhas é dada por

$$\mathbf{C}_{(1)} = \begin{pmatrix} 1 & 0 & \cdots & 0 & -1 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 & -1 & \cdots & 0 \end{pmatrix}$$

e dos efeitos das colunas é definida por

$$\mathbf{C}_{(2)} = \begin{pmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & -1 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 1 & \cdots & 0 & -1 & \cdots & 0 \\ 0 & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & -1 & \cdots & 0 \end{pmatrix}$$

A matriz $\mathbf{C}_{(1)}$ possui dimensões $(r-1) \times [(r+1)(c+1)]$ formada pela matriz de contraste mais zeros nos elementos restantes. A matriz $\mathbf{C}_{(2)}$ possui dimensões $(c-1) \times [(r+1)(c+1)]$ formada por zeros nas primeiras r colunas seguida da matriz de contraste com c colunas e zeros nas últimas $rc+1$ colunas. O teste de hipótese para a matriz de interação, Morrison (1978) denota o efeito η_{ijh} , agrupando primeiro por linhas e depois por colunas dentro de cada linha nas colunas da matriz de parâmetros: $\eta_{11h}, \dots, \eta_{1ch}, \dots, \eta_{r1h}, \dots, \eta_{rch}$. Se $\mathbf{C}$ possui dimensões $(c-1) \times c$, então a matriz de interação pode ser escrita como

$$\mathbf{C}_{(3)} = \begin{pmatrix} \mathbf{0} & \mathbf{C} & \mathbf{0} & \cdots & \mathbf{0} & -\mathbf{C} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{C} & \cdots & \mathbf{0} & -\mathbf{C} & \mathbf{0} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{C} & -\mathbf{C} & \mathbf{0} \end{pmatrix}$$

com as dimensões $(r-1)(c-1) \times (r+1)(c+1)$. A primeira coluna é formada por sub-matrizes nulas com dimensão $(c-1) \times (r+c)$, enquanto as próximas r colunas deve ser $(c-1) \times c$

conforme a matriz $\mathbf{C}$. A última coluna é formada por sub-matrizes nulas e o posto da matriz $\mathbf{C}_{(3)}$ é de $(r-1)(c-1)$.

Denota-se $\boldsymbol{\alpha}_i = (\alpha_{i1}, \dots, \alpha_{ip})^T$, $\boldsymbol{\tau}_j = (\tau_{j1}, \dots, \tau_{jp})^T$ e $\boldsymbol{\eta}_{ij} = (\eta_{ij1}, \dots, \eta_{ijp})^T$ e as hipóteses de interesse são:

$$H_o : \alpha_1 = \dots = \alpha_r, \text{ não há efeito dos marcadores de linha;}$$

$$H_o : \tau_1 = \dots = \tau_c, \text{ não há efeito dos marcadores de coluna;}$$

$$H_o : \eta_{11} = \dots = \eta_{rc}, \text{ não há efeito de interação.}$$

Para escrever essas hipóteses na forma matricial compacta, $\mathbf{CB} = \mathbf{0}$, é necessário escrever uma matriz $\mathbf{C}$ distinta segundo a hipótese que se quer provar (MORRISON, 1978). Então, para testar as hipóteses dos efeitos das linhas e das colunas é necessário testar $\mathbf{C}_{(3)}\mathbf{B} = \mathbf{0}$, ou seja, verificar se a hipótese do efeito de não interação é verdadeira. Por essa razão, realiza-se na prática primeiramente o teste de hipótese da interação, desde que estas rejeições impliquem, que os testes dos fatores das linhas e das colunas devem ser feitos separadamente para cada categoria das c colunas e r linhas.

O procedimento para obter as generalizações das matrizes das somas de quadrados para análise de variância multivariada de duas entradas é o mesmo procedimento para os casos univariados. Morrison (1978) define os totais das observações como:

- a) Células: $C_{ijh} = \sum_{k=1}^n y_{ijkh};$
- b) Fator Coluna: $T_{jh} = \sum_{i=1}^r \sum_{k=1}^n y_{ijkh};$
- c) Fator Linha: $R_{ih} = \sum_{j=1}^c \sum_{k=1}^n y_{ijkh};$
- d) Total: $G_h = \sum_{i=1}^r \sum_{j=1}^c \sum_{k=1}^n y_{ijkh}.$

Estes totais são calculados para todas combinações dos tratamentos. Então, as matrizes de Resíduos e das somas de quadrados e produtos cruzados são calculadas conforme a Tabela 4, em que $u, v = 1, \dots, p$, para as cinco matrizes.

Tabela 4 - Elementos para MANOVA de dupla entrada

Fontes de Variação	Soma de Quadrados (SQ)	
	Matriz	Elemento geral
Linhas	$\mathbf{H}_1$	$h_{1uv} = \frac{1}{cn} \sum_{i=1}^r R_{iu} R_{iv} - \frac{G_u G_v}{rcn}$
Colunas	$\mathbf{H}_2$	$h_{2uv} = \frac{1}{rn} \sum_{j=1}^c T_{ju} T_{jv} - \frac{G_u G_v}{rcn}$
Interação	$\mathbf{H}_3$	$h_{3uv} = t_{uv} - h_{1uv} - h_{2uv} - e_{uv}$
Resíduo	$\mathbf{E}$	$e_{uv} = \sum_{i=1}^r \sum_{j=1}^c \sum_{k=1}^n y_{ijk} y_{ijkv} - \frac{1}{n} \sum_{i=1}^r \sum_{j=1}^c C_{iju} C_{ijv}$
Total	$\mathbf{T}$	$t_{uv} = \sum_{i=1}^r \sum_{j=1}^c \sum_{k=1}^n y_{ijk} y_{ijkv} - \frac{G_u G_v}{rcn}$

Para provar estas hipóteses, usa-se o Princípio de União-Intersecção de Roy, baseado nos autovalores não nulos da matriz $\mathbf{H}\mathbf{E}^{-1}$ sendo que:

$$\mathbf{H} = \hat{\mathbf{D}}^T \mathbf{R}^{-1} \hat{\mathbf{D}} \quad \text{e} \quad \mathbf{E} = \mathbf{Y}^T [\mathbf{I} - \mathbf{Z}(\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T] \mathbf{Y}.$$

Assim, os produtos $\mathbf{H}_1 \mathbf{E}^{-1}$ para efeito linha, $\mathbf{H}_2 \mathbf{E}^{-1}$ para efeito coluna e $\mathbf{H}_3 \mathbf{E}^{-1}$ para efeito interação, temos:

$$\mathbf{H}_i = \mathbf{Y}^T \mathbf{Z}_i (\mathbf{Z}_i^T \mathbf{Z}_i)^{-1} \mathbf{C}_i^T [\mathbf{C}_i (\mathbf{Z}_i^T \mathbf{Z}_i) \mathbf{C}_i^T]^{-1} \mathbf{C}_i (\mathbf{Z}_i^T \mathbf{Z}_i)^{-1} \mathbf{Z}_i^T \mathbf{Y}, \quad i = 1, 2, 3$$

em que $\mathbf{Z}_1$ possui dimensão $(rcn) \times r$, $\mathbf{Z}_2$ possui dimensão $(rcn) \times c$ e $\mathbf{Z}_3$ possui dimensão $(rcn) \times (rc)$.

A partir destes produtos, pode-se calcular as maiores raízes características c_{1s}, c_{2s}, c_{3s} e as estatísticas $\frac{c_{1s}}{1 + c_{1s}}$ conforme a Tabela 5 (MORRISON, 1978), aplicando o critério da maior raiz característica.

2.4.3 MANOVA Biplot de dupla entrada

Na MANOVA Biplot de dupla entrada temos dois fatores de variação, com i e j níveis respectivamente. AMARO, VICENTE-VILLARDÓN e GALINDO-VILLARDÓN

Tabela 5 - Distribuição de parâmetros da maior raiz característica

Fontes de Variação	Estatística	Parâmetros		
		s	m	n
Linhas	$\frac{c_{1s}}{1 + c_{1s}}$	$\min(r - 1, p)$	$\frac{ r - 1 - p - 1}{2}$	$\frac{rc(n - 1) - p - 1}{2}$
Colunas	$\frac{c_{2s}}{1 + c_{2s}}$	$\min(c - 1, p)$	$\frac{ c - 1 - p - 1}{2}$	$\frac{rc(n - 1) - p - 1}{2}$
Interação	$\frac{c_{3s}}{1 + c_{3s}}$	$\min[(r - 1)(c - 1), p]$	$\frac{ (r - 1)(c - 1) - p - 1}{2}$	$\frac{rc(n - 1) - p - 1}{2}$

(2004) propõem quatro tipos de Biplot para representar os resultados de uma MANOVA para um delineamento de dupla entrada. Todos foram construídos a partir do Biplot da matriz de parâmetros do modelo MANOVA. Cada um dos Biplot tem um particular objetivo e o que os diferencia basicamente é a forma de escolher as matrizes $\mathbf{R}$ e $\hat{\mathbf{D}}$ na decomposição inicial. Lembrando que $\mathbf{C}_1$, $\mathbf{C}_2$ e $\mathbf{C}_3$ são as matrizes de contrastes apropriadas.

i) Biplot Total:

Equivale a tomar todos os tratamentos (combinações de níveis dos fatores) como MANOVA de entrada unidimensional (GABRIEL, 1995).

$$\mathbf{R} = (\mathbf{Z}^T \mathbf{Z})^{-1}; \quad \hat{\mathbf{D}} = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Y}; \quad \mathbf{C} = \mathbf{I},$$

A MANOVA Biplot de dupla entrada obtido desta forma possui as mesmas propriedades da MANOVA Biplot unidimensional. Permite estudar as semelhanças e as diferenças entre níveis de tratamentos, que neste caso são as combinações de níveis dos fatores; verifica as variáveis mais importantes nas comparações entre os grupos, a direção da separação entre grupos e as qualidades de representação. Em geral, não é a forma mais adequada para estudar a interação, os efeitos das linhas ou colunas, devido o gráfico mostrar os efeitos combinados. Neste tipo de Biplot, o eixo é a combinação de variáveis que maximiza a parte controlada da MANOVA.

ii) Biplot das linhas:

Neste caso temos as seguintes matrizes

$$\mathbf{R} = \mathbf{C}_1(\mathbf{Z}_1^T \mathbf{Z}_1)^{-1} \mathbf{C}_1^T; \quad \hat{\mathbf{D}} = \mathbf{C}_1(\mathbf{Z}_1^T \mathbf{Z}_1)^{-1} \mathbf{Z}_1^T \mathbf{Y};$$

Denota-se a matriz das médias das linhas por $\bar{\mathbf{Y}}_{li} = (\mathbf{Z}_1^T \mathbf{Z}_1)^{-1} \mathbf{Z}_1^T \mathbf{Y}$. Os marcadores para os níveis do fator são $\bar{\mathbf{A}}_{li} = \bar{\mathbf{Y}}_{li} \mathbf{E}^{-1/2} \mathbf{V}$ e os marcadores para as variáveis (colunas) $\mathbf{B} = \mathbf{E}^{-1/2} \mathbf{V}$. As propriedades são:

- a) O Biplot de linhas é equivalente a obtenção das variáveis canônicas para a hipótese de efeito de linhas; os eixos são combinações lineares das variáveis que maximizam o teste F para o efeito do fator linha.
- b) O Biplot permite estudar semelhanças e diferenças entre os níveis do fator construídos segundo o fator de classificação por linhas, quando o efeito for significativo.
- c) O produto escalar entre os marcadores de linha para $\hat{\mathbf{D}}$ aproxima dos produtos escalares para as linhas na métrica $\mathbf{E}^{-1}$, coincide com a distância de Mahalanobis.
- d) Pode-se construir círculos de confiança ao redor dos centróides dos níveis do fator, de tal forma que se os círculos não se interceptam então o efeito do fator linha é significativo.
- e) Pode-se interpretar a importância das variáveis na separação dos níveis do fator, de forma similar ao Biplot unidimensional.

iii) Biplot das colunas:

Neste caso substitui as matrizes por

$$\mathbf{R} = \mathbf{C}_2(\mathbf{Z}_2^T \mathbf{Z}_2)^{-1} \mathbf{C}_2^T; \quad \hat{\mathbf{D}} = \mathbf{C}_2(\mathbf{Z}_2^T \mathbf{Z}_2)^{-1} \mathbf{Z}_2^T \mathbf{Y};$$

Denota-se a matriz das médias das colunas por $\bar{\mathbf{Y}}_{co} = (\mathbf{Z}_2^T \mathbf{Z}_2)^{-1} \mathbf{Z}_2^T \mathbf{Y}$. Os marcadores para os níveis do fator são $\bar{\mathbf{A}}_{co} = \bar{\mathbf{Y}}_{co} \mathbf{E}^{-1/2} \mathbf{V}$ e os marcadores para as variáveis (colunas) $\mathbf{B} = \mathbf{E}^{1/2} \mathbf{V}$. O Biplot de colunas possui propriedades semelhantes ao Biplot de linhas, neste caso faz comparações entre os grupos do fator coluna.

O produto escalar dos marcadores para as colunas de $\hat{\mathbf{D}}$ aproxima da matriz de soma de quadrados e produtos dentro dos níveis do fator $\mathbf{E}$

$$\mathbf{B}\mathbf{B}^T = \mathbf{E}^{1/2} \mathbf{V}\mathbf{V}^T \mathbf{E}^{1/2} = \mathbf{E}.$$

iv) Biplot de interação:

Neste caso substitui as matrizes $\mathbf{R}$ e $\hat{\mathbf{D}}$ por

$$\mathbf{R} = \mathbf{C}_3(\mathbf{Z}_3^T \mathbf{Z}_3)^{-1} \mathbf{C}_3^T; \quad \hat{\mathbf{D}} = \mathbf{C}_3(\mathbf{Z}_3^T \mathbf{Z}_3)^{-1} \mathbf{Z}_3^T \mathbf{Y}.$$

Os marcadores para os centroídes dos grupos são $\bar{\mathbf{A}} = \bar{\mathbf{Y}}_{in} \mathbf{E}^{-1/2} \mathbf{V}$, em que $\bar{\mathbf{Y}}_{in}$ é a matriz das médias das Células e pode ser expressa como $\bar{\mathbf{Y}}_{in} = (\mathbf{Z}_3^T \mathbf{Z}_3)^{-1} \mathbf{Z}_3^T \mathbf{Y}$. Os marcadores para as variáveis são $\mathbf{B} = \mathbf{E}^{1/2} \mathbf{V}$ e os marcadores para o conjunto de contrastes são $\mathbf{A} = \mathbf{R}^{1/2} \mathbf{U} \mathbf{\Lambda}$.

As propriedades são:

- a) O Biplot obtido desta forma é semelhante a obtenção das variáveis canônicas para as hipóteses de interação, os eixos são combinações lineares de variáveis que maximizam o teste F de Snedecor para a interação. Isso nos permite a interpretação da interação com máxima resolução, a diferença para o Biplot Total é que ele combina com os efeitos principais;
- b) A qualidade do ajuste para a representação dos contrastes é $\sum_{j=1}^d \lambda_j^2 / \sum_{j=1}^r \lambda_j^2$ e a qualidade do ajuste para a representação das médias não se calcula, pois o propósito da representação é a interpretação da interação;
- c) Para cada um dos contrastes, a distância ao quadrado do marcador é aproximadamente a distância de Mahalanobis ao vetor nulo;
- d) Na projeção das médias, a distância ao quadrado do marcador que representa uma das médias é aproximadamente a distância de Mahalanobis ao vetor nulo, sendo que esta aproximação não é a melhor em relação as médias. Pode ser interpretada como a parte da distância explicada pela interação,

$$\bar{\mathbf{A}} \bar{\mathbf{A}}^T = \bar{\mathbf{Y}}_{in} \mathbf{E}^{-1/2} \mathbf{V} \mathbf{V}^T \mathbf{E}^{-1/2} \bar{\mathbf{Y}}_{in} = \bar{\mathbf{Y}}_{in} \mathbf{E}^{-1} \bar{\mathbf{Y}}_{in};$$

- e) A interação será máxima para aqueles vetores paralelos ao vetor de interação resultante e a interação será mínima para os vetores perpendiculares.

2.5 Análise de Distância Multifatorial

2.5.1 Escalonamento Multidimensional

Escalonamento Multidimensional ou MDS (Multidimensional Scaling) é o método que representa as medidas de proximidade (similaridades ou dissimilaridades) entre pares de objetos como distâncias euclidianas em um espaço multidimensional de baixa dimensão, permitindo a visualização da estrutura dos dados originais.

O MDS têm suas origens na psicometria (mensuração de fenômenos psicológicos), onde foi proposto para ajudar a compreender as decisões das pessoas sobre a semelhança entre os membros de um conjunto de objetos. A história do MDS pode ser dividida em quatro fases, cada uma corresponde aproximadamente uma década. Os primeiros estudos que envolvem a técnica de MDS, foram desenvolvidos por Torgenson (1952), iniciou os estudos para variáveis quantitativas, definindo o problema de escalonamento dimensional e utilizando-o como solução. Posteriormente Shepard (1962), ampliou os estudos para as variáveis qualitativas, fazendo uma formulação bastante precisa do escalonamento multidimensional quando demonstrou, empiricamente, que conhecendo uma ordenação das distâncias entre pontos, poderia encontrar uma configuração de pontos em um espaço euclidiano de baixa dimensão, cujas interdistâncias euclidianas reproduziam praticamente a ordenação original.

Kruskal (1964), foi o responsável pelo aprimoramento da técnica, criou os algoritmos, iniciando pela transformação da matriz de proximidades e a matriz de distâncias, obtendo coordenadas aleatórias, que proporcionam os “estímulos”, passando pela comparação das proximidades com as distâncias, obtendo as disparidades.

Guttman (1968) fez uma abordagem diferente de Kruskal, na criação do MDS não-métrico, definindo uma função de perda chamada de “coeficiente de alienação, basicamente equivale a função *stress* definida por Kruskal, conduzindo a um algoritmo diferente para a minimização. Guttman e Lingoes desenvolveram uma série de programas computacionais para MDS não-métrica.

Em 1966, Gower propôs o método de Análise de Coordenadas Principais, pode-se considerar como um método métrico de escalonamento multidimensional, resolvendo os problemas dos processos iterativos das técnicas não-métricas. Durante a década de 80 à

90, numerosos pesquisadores continuaram a busca por algoritmos que conseguisse alcançar uma configuração final de pontos cujas distâncias fossem o mais próximo possível das dissemelhanças.

Os algoritmos existentes utilizados para solucionar o MDS, não garantem a obtenção de mínimos globais, por isso é necessário aplicar outros métodos. Por exemplo, nos últimos anos, tem-se aplicado métodos de Optimização Estocástica, já que na maioria destes métodos demonstraram a convergência assintótica para ótimo global. Os mais conhecidos são *Simulated Annealing* e o Algoritmo Genético. O método de *Simulated Annealing*, baseia-se no algoritmo de Metropolis, que utiliza as técnicas de Monte Carlo para gerar as cadeias de Markov. Os Algoritmos Genéticos são utilizados para obter o máximo de uma função positiva (BORG e GROENEN, 2005).

A MDS possui quatro efeitos, na sua utilização como:

- a) um método que representa as similaridades/dissimilaridades dos dados, como as distâncias em um espaço de baixa dimensão, a fim de tornar esses dados acessíveis à inspeção visual e à exploração;
- b) uma técnica que permite testar com certo critério as diferenças entre os objetos de interesse, são espelhados nas correspondentes diferenças empíricas desses objetos.
- c) uma abordagem analítica que permite descobrir as dimensões que fundamentam as decisões de similaridade/dissimilaridade.
- d) um modelo psicológico que explique a capacidade da dissimilaridade em termos de uma norma que imite um tipo particular de função de distância.

Os propósitos diferentes do MDS e a existência de uma enorme variedade de modelos geométricos, conduziu a uma confusão desnecessária sobre como o MDS deve ser utilizado. Atualmente, a utilização de técnicas de redução dimensional vêm mostrando um grande avanço devido aos Bancos de Dados Muito Grande (BDMG) em diversas áreas como: engenharia (identifica as similaridades entre várias cidades), ciências sociais (estuda o comportamento dos candidatos políticos), ciências da computação (mineração e visualização dos dados) e ciências biológicas. Na biologia molecular, utilizam-se as técnicas de redução dimensional para analisar sequências de genoma, pois são dados altamente complexos e extensos.

Assim, a MDS é uma ferramenta de auxílio, tanto no processo de transformação dos dados, como no processo de visualização dos dados.

Existem dois tipos de modelos para esta técnica:

- a) Métrico: é usado quando a transformação aplicada às similaridades (ou dissimilaridades) preserva as propriedades métricas das distâncias porque pertence a escalas de razão ou intervalo;
- b) Não-métrico: é usado quando se têm informação de dados avaliados segundo uma escala ordinal. Não presume uma relação linear entre as proximidades e as distâncias, estabelecendo uma relação crescente entre ambas (BORG e GROENEN, 2005).

Diversos programas computacionais foram desenvolvidos para análise de dados por MDS, por exemplo:

- a) ALSCAL é o mais flexível, compatível à vários modelos, convergente (o que é desejável) e mais rápido. É o programa que está disponível nos principais softwares estatísticos, tais como SAS e SPSS. No entanto, o algoritmo otimiza o ajuste das distâncias ao quadrado para as dissimilaridades, que não é o critério de otimização mais desejável. ALSCAL possui aspectos mais descritivos do que inferenciais;
- b) MULTISCALE é baseado no princípio da máxima verossimilhança, é o único que permite testes de significância estatística utilizados para fins inferenciais. Fornece ao usuário uma seleção de modelos e é menos flexível quanto aos tipos de dados;
- c) SMACOF é o mais rápido, otimiza o ajuste das distâncias de dissimilaridades por um algoritmo convergente, é mais flexível nos tipos de dados analisados. Está atualmente em desenvolvimento, por isso não está disponível em qualquer software (YOUNG e HAMMER, 1987).

3 MATERIAIS E MÉTODOS

3.1 Características dos dados utilizados

Os dados a serem utilizados são relativos a experimentos realizado no Pólo Agroindustrial Mossoró-Assu, situado no Estado do Rio Grande do Norte que é o principal produtor e exportador de melão (*Cucumis melo L.*) do Brasil. O melão do tipo Galia é caracterizado pela forma arredondada do fruto, coloração amarela, polpa esverdeada com rendilhamento e aroma suave, conforme Figura 8. Foram coletados no período de agosto a outubro, nos anos de 2000, 2001 e 2002, totalizando doze ambientes:

- a) Agropólo Mossoró-Assu: MO-00 Mossoró, 2000 (A1); MO-01 Mossoró, 2001 (A2); MO-02 Mossoró, 2002 (A3);
- b) Baraúna: BA-00, 2000 (A4); BA-01, 2001 (A5); BA-02, 2002 (A6);
- c) Assu: AS-00, 2000 (A7); AS-01, 2001 (A8); AS-02, 2002 (A9);
- d) Alto do Rodrigues: AR-00, 2000 (A10); AR-01, 2001 (A11); AR-02, 2002 (A12).

Os híbridos simples avaliados do melão do tipo Galia foram: GPS 400 (G1), DRG 1531 (G2), DRG 1537 (G3), DRG 1531 (G4), Solar King (G5), Galileu (G6), Supra (G7), Num 1502 (G8) e Arava (G9). As variáveis medidas foram PRODUÇÃO ($t.ha^{-1}$) e SST (teor de sólidos solúveis, %). Adotou-se o delineamento experimental aleatorizado em blocos com três repetições.

No melhoramento genético do meloeiro, é relevante considerar os aspectos relacionados ao ciclo, à resistência a doenças e à tolerância a pragas, à capacidade produtiva e à qualidade de fruto. Genótipos precoces são desejáveis porque agregam valor econômico e agrônômico, ofertando o produto em menor prazo e com economia de defensivos no controle de pragas ou infecções de microorganismos patogênicos. Além disso, as plantas precisam ser altamente produtivas e prolíficas para permitirem maiores ganhos ao produtor. Essas características devem estar associadas àquelas que conferem qualidade ao fruto (VIANA, 2006).

O rendilhado do melão está relacionado ao teor de sólidos solúveis, responsável pelo sabor e o seu aspecto visual, que o diferencia dos outros tipos de melões. Sua qualidade

nutricional também tem contribuído favoravelmente para o seu consumo, pois é considerado pouco calórico e boa fonte de sódio, potássio, vitaminas A e C e beta-caroteno. A qualidade em frutos de melão envolve atributos relacionados às características da polpa, como sólidos solúveis, aparência interna e externa da fruta, espessura de polpa (SILVA, et al, 2005).

Figura 8 - Foto do melão do tipo Galia

Devido às diferentes condições de ambientes em que os híbridos de melão são avaliados no Rio Grande do Norte, espera-se que ocorra uma acentuada interação genótipos $\times$ ambientes e que a mesma tenha papel importante na manifestação fenotípica, essa interação é decorrente do comportamento diferencial dos genótipos nos diferentes ambientes. Para identificar os híbridos estáveis, diversos métodos têm sido propostos, os quais diferem entre si em função dos conceitos de estabilidade, restrições impostas e procedimentos biométricos adotados para estimação.

3.2 Metodologias de análises propostas

3.2.1 Modelo AMMI

Suponha que um conjunto de $(g)_i$ genótipos ou marcadores de linhas ($i = 1, \dots, g$) tenha sido testado experimentalmente em $(a)_j$ ambientes ou marcadores de colunas ($j = 1, \dots, a$) e as observações para uma variável são representadas por Y_{ij} . A média

de cada combinação de genótipo e ambiente, obtida de r repetições num experimento (um conjunto balanceado de dados) ou blocos pode ser representada conforme a Tabela 6 que possui dimensão $n \times p$. Os vetores marginais são os vetores de médias das linhas e vetores de médias das colunas dos elementos da matriz de dados $\mathbf{Y}$ e $\bar{Y}_{..}$ é a média geral (DIAS e KRZANOWSKI, 2003).

Tabela 6 - Disposição dos dados em linhas (genótipos) e colunas (ambientes)

Genótipos	Ambientes					médias
	1	2	3	$\cdots$	p	
1	$\bar{Y}_{11}$	$\bar{Y}_{12}$	$\bar{Y}_{13}$	$\cdots$	$\bar{Y}_{1p}$	$\bar{Y}_{1\bullet}$
2	$\bar{Y}_{21}$	$\bar{Y}_{22}$	$\bar{Y}_{23}$	$\cdots$	$\bar{Y}_{2p}$	$\bar{Y}_{2\bullet}$
3	$\bar{Y}_{31}$	$\bar{Y}_{32}$	$\bar{Y}_{33}$	$\cdots$	$\bar{Y}_{3p}$	$\bar{Y}_{3\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$
n	$\bar{Y}_{n1}$	$\bar{Y}_{n2}$	$\bar{Y}_{n3}$	$\cdots$	$\bar{Y}_{np}$	$\bar{Y}_{n\bullet}$
médias	$\bar{Y}_{\bullet 1}$	$\bar{Y}_{\bullet 2}$	$\bar{Y}_{\bullet 3}$	$\cdots$	$\bar{Y}_{\bullet p}$	$\bar{Y}_{\bullet\bullet}$

O modelo linear para este experimento, segundo Shafii e Price (1998) é dado por:

$$\bar{Y}_{ij} = \mu + g_i + a_j + (ga)_{ij} + \bar{\varepsilon}_{ij}$$

em que:

$\bar{Y}_{ij}$: resposta média observada do i -ésimo genótipo no j -ésimo ambiente;

μ : média geral;

g_i : efeito do i -ésimo genótipo estimado por $\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet}$, sendo que $i = 1, \dots, g$, $\bar{Y}_{i\bullet}$ é a média da i -ésima linha e $\bar{Y}_{\bullet\bullet}$ é a média geral;

a_j : efeito do j -ésimo ambiente estimado por $\bar{Y}_{\bullet j} - \bar{Y}_{\bullet\bullet}$, sendo que $j = 1, \dots, a$, $\bar{Y}_{\bullet j}$ é a média da j -ésima coluna e $\bar{Y}_{\bullet\bullet}$ é a média geral;

$(ga)_{ij}$: efeito da interação do i -ésimo genótipo no j -ésimo ambiente estimado por $\bar{Y}_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..}$;

$\bar{\varepsilon}_{ij}$: erro médio assumindo independente e distribuído como uma $N(0, \frac{\sigma^2}{r})$, sendo r o número de repetições.

O modelo AMMI (Additive Main effects and Multiplicative Interaction model) possui a finalidade de modelar a estrutura de dados de efeitos principais aditivos de genótipos $(g)_i$ e de ambientes $(a)_j$ de forma tradicional e os efeitos multiplicativos da interação genótipo-ambiente $(ga)_{ij}$ pela análise de componentes principais. O objetivo principal é de selecionar modelos que expliquem o padrão selecionado para a interação ($SQ_{G \times A}$ Padrão) descartando os ruídos presentes nos dados ($SQ_{G \times A}$ Ruídos) e proporcionando uma melhoria na capacidade preditiva das respostas genotípicas (FREITAS, 2007).

O modelo AMMI ajuda a identificar os genótipos superiores que foram adaptados no chamado zoneamento agrônômico, com fins de recomendação regionalizada e de seleção de locais de teste. É um modelo mais parcimonioso do que o modelo de análise de variância convencional, para descrever a interação $G \times A$, é também considerado como um dos melhores modelos em termos de interpretação da interação $G \times A$, do que em relação a regressão linear simples com base nas médias ambientais, pois a interação $G \times A$ pode ser modelada em mais de uma dimensão (DUARTE; VENCOVSKY, 1999).

Para verificar a variabilidade genética entre os níveis do fator linha (genótipos) e a precisão experimental, faz-se a Análise de Variância de cada ambiente (análises individuais). Caso sejam detectados diferenças entre os níveis do fator linha, realiza-se uma análise de variância conjunta. Constatando a presença de interação genótipo-ambiente (teste F significativo), analisa-se a estabilidade produtiva utilizando-se o modelo AMMI:

$$\begin{aligned} \bar{Y}_{ij} &= \underbrace{\mu + g_i + a_j}_{\text{parte aditiva}} + \underbrace{\sum_{k=1}^p \lambda_k \gamma_{ik} \alpha_{jk}}_{(ga)_{ij}: \text{parte multiplicativa}} + \bar{\varepsilon}_{ij} \\ \bar{Y}_{ij} &= \mu + g_i + a_j + \overbrace{\sum_{k=1}^l \lambda_k \gamma_{ik} \alpha_{jk} + \sum_{k=l+1}^p \lambda_k \gamma_{ik} \alpha_{jk}} + \bar{\varepsilon}_{ij} \end{aligned}$$

com $l < p$ e $\sum_{k=l+1}^p \lambda_k \alpha_{ik} \gamma_{jk} = \rho_{ij}$ é o ruído de interação, em que:

$\bar{Y}_{ij}$: resposta média do i -ésimo genótipo no j -ésimo ambiente;

μ : média geral das observações;

g_i : efeito principal do i -ésimo genótipo (linha) $i = (1, \dots, g)$;

a_j : efeito principal do j -ésimo ambiente (coluna), $j = (1, \dots, a)$;

λ_k : é a raiz quadrada do k -ésimo autovalor da matriz $(\mathbf{GA})(\mathbf{GA})^T$ e $(\mathbf{GA})^T(\mathbf{GA})$, de iguais autovalores não nulos, cujo λ_k^2 é o k -ésimo autovalor da matriz resultante desse produto;

α_{jk} : elemento correspondente ao j -ésimo ambiente no k -ésimo autovetor de $(\mathbf{GA})^T(\mathbf{GA})$ associado a λ_k^2 ;

γ_{ik} : elemento correspondente ao i -ésimo genótipo do k -ésimo autovetor de $(\mathbf{GA})(\mathbf{GA})^T$ associado a λ_k^2 ; $\mathbf{GA} = (\widehat{ga})_{ij}$ é a matriz de interações obtidas como resíduo do ajuste aos efeitos principais, aplicadas à matriz de médias, $k = (1, \dots, p)$, em que p é o número de raízes características não nulas, $p = 1, 2, \dots, \min(g-1, a-1)$;

ρ_{ij} : ruído presente nos dados associado ao termo $(ga)_{ij}$ da interação do i -ésimo genótipo no j -ésimo ambiente ou resíduo AMMI;

$\bar{\varepsilon}_{ij}$: erro experimental médio, sendo $\bar{\varepsilon}_{ij} \sim N(0, \frac{\sigma^2}{r})$, em que r o número de repetições.

O valor $(\widehat{ga})_{ij}$ estimado é definido por $\mathbf{Z} = (z_{ij}) = (\widehat{ga})_{ij} = \sum_{k=1}^l \lambda_k \gamma_{ik} \alpha_{jk} = \bar{Y}_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..}$, sendo que a matriz $\mathbf{Z}$ é de posto l . A matriz de interação $\mathbf{GA}$ é padronizada pelo termo $(ga)_{ij}$ representado por $(ga)_{ij} = \sum_{k=1}^l \lambda_k \gamma_{ik} \alpha_{jk} + \rho_{ij}$, sob as restrições $\sum_{i=1}^g g_i = \sum_{j=1}^a a_j = \sum_{i=1}^g (ga)_{ij} = \sum_{j=1}^a (ga)_{ij} = 0$. Na metodologia AMMI a matriz de interação $\mathbf{GA}$ é representada pela soma das l parcelas, cada uma é resultante da multiplicação de λ_k , com a mesma unidade de Y_{ij} , por um efeito genotípico γ_{ik} e um efeito ambiental α_{jk} , ambos adimensionais, isto é, $\sum_{k=1}^l \lambda_k \gamma_{ik} \alpha_{jk}$, sendo que l equivale ao número de termos da interação. O termo λ_k possui a informação referente a k -ésima parcela da interação genótipo-ambiente

e os efeitos γ_{ik} e α_{jk} representam os pesos do i -ésimo genótipo e do j -ésimo ambiente, na parcela da interação.

Assim, essa soma de quadrados é desdobrada em componentes relativos a cada termo da DVS, ou cada eixo da interação da ACP (Análise de Componentes Principais). A análise de variância é esquematizada pelo quadro completo da análise conjunta obtida a partir das médias, conforme Tabela 7 (DUARTE e VENCOSKY, 1999).

Tabela 7 - Quadro resumo da análise conjunta de variância obtida a partir das médias

Fonte de variação	Graus de liberdade	Soma de Quadrados
Genótipos ($\mathbf{G}$)	$g - 1$	SQ_G
Ambientes($\mathbf{A}$)	$a - 1$	SQ_A
Interação($\mathbf{G} \times \mathbf{A}$)	$(g - 1)(a - 1)$	$SQ_{G \times A}$
ACP 1	$g + a - 1 - (2 \times 1)$	λ_1^2
ACP 2	$g + a - 1 - (2 \times 2)$	λ_2^2
ACP 3	$g + a - 1 - (2 \times 3)$	λ_3^2
$\vdots$	$\vdots$	$\vdots$
ACP p	$g + a - 1 - (2 \times p)$	λ_p^2
Erro Médio/r	$ga(r - 1)$	$SQ_{\text{Erro Médio}}$
Total	$gar - 1$	SQ_{Total}

A estatística do teste F é definida por:

$$F_{R,l} = \frac{SQ_{(G \times A)} - \sum_{k=1}^n \lambda_k^2}{f_2 \, QM_{\text{Erro Médio}}}$$

com $f_2 = (g - 1 - l)(a - 1 - l)$ com n termos multiplicativos incluídos no modelo. Piepho (1995) define o teste F utilizado, como um teste robusto para a escolha do modelo AMMI que foi o proposto por Gollob. A estatística $F_{R,l}$ sob a hipótese nula de que não haja mais do que n termos determinando a interação, tem uma distribuição F aproximada com f_2 graus de liberdade e $GL_{\text{Erro Médio}}$. Sob essa hipótese, a $SQ_{G \times A}$ para o resíduo é aproximadamente uma variável qui-quadrado (DUARTE E VENCOSKY, 1999).

O objetivo da análise é de resumir grande parte da interação genótipo-ambiente no menor número possível de eixos (dois ou três no máximo), levando em consideração a soma de quadrados da interação padrão ($SQ_{G \times A \text{ Padrão}}$) com poucos graus de liberdade, resultando em um modelo reduzido, descartando o resíduo adicional ($SQ_{G \times A \text{ Ruídos}}$). Um dos procedimentos consiste em determinar os graus de liberdade associados à parcela da ($SQ_{G \times A}$) relacionada a cada membro da família de modelos AMMI ($\text{AMMI}_{(0)}$, $\text{AMMI}_{(1)}$, $\dots$, $\text{AMMI}_{(l)}$). No modelo $\text{AMMI}_{(0)}$ nenhum eixo é ajustado (modelo aditivo sem interação); $\text{AMMI}_{(1)}$ ajusta-se apenas o primeiro eixo, e assim sucessivamente, até $\text{AMMI}_{(l)}$ (modelo completo).

Obtém-se, o quadrado médio ($Q.M.$) correspondente a cada parcela ou modelo, um teste F para avaliar a significância de cada componente em relação ao $QM_{\text{ErroMédio}}$. Assim, o ponto de parada que determina a seleção do modelo depende da significância do teste $F_{R,l}$ para os sucessivos eixos de interação.

A adaptabilidade é predita em função das respostas para cada combinação de genótipos e ambientes, sobre o modelo AMMI selecionado pelo ACP_k (eixo da interação da ACP - Análise de Componentes Principais) conforme o nível de significância de $F_{R,l}$. A expressão da resposta esperada do i -ésimo genótipo no j -ésimo ambiente ($\bar{Y}_{ij}$)

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

em que l é o número de eixos significativos para a decomposição da interação genótipo-ambiente.

Duarte e Vencovsky (1999) avaliam a estabilidade dos genótipos e ambientes através da análise da magnitude e do sinal dos escores do Biplot, correspondentes ao modelo AMMI selecionado, sendo considerados estáveis os genótipos e/ou ambientes cujos escores são baixos (próximos de zero), que são próprios dos genótipos e ambientes que pouco ou quase nada contribuem para a interação.

3.2.1.1 Biplot AMMI

O Biplot é uma ferramenta que permite diagnosticar de forma mais detalhada situações em que a interação significativa é acompanhada pela não significância dos efeitos aditivos dos genótipos e ambientes ou quando a interação é influenciada por pontos discrepantes

(SHAFII e PRICE, 1998).

O Biplot permite estimar conforme as projeções dos vetores, os valores médios dos genótipos (linhas) e ambientes (colunas), assim como os efeitos principais e interações de uma tabela de dupla entrada conforme a Tabela 8, em que $\mathbf{g}_\bullet = \frac{\sum_i g_i}{n}$ representa a média dos genótipos (linhas), $\mathbf{a}_\bullet = \frac{\sum_j a_j}{n}$ representa a médias dos ambientes (colunas). Esses resultados são utilizados quando se está interessado em diagnosticar modelos a partir de um Biplot e na interpretação de modelos associados a um delineamento experimental. O gráfico além de representar simultaneamente as linhas e as colunas referentes da matriz de dados, pode indicar também a existência de agrupamentos entre as observações e mostrar as variâncias e as correlações entre as variáveis.

Tabela 8 - Funções de estimação das observações conforme os marcadores do Biplot

Medidas	Função	Marcadores
Média de uma linha	$\bar{Y}_{i\bullet}$	$\mathbf{g}_i^T \mathbf{a}_\bullet$
Média de uma coluna	$\bar{Y}_{\bullet j}$	$\mathbf{g}_\bullet^T \mathbf{a}_j$
Média geral	$\bar{Y}_{\bullet\bullet}$	$\mathbf{g}_\bullet^T \mathbf{a}_\bullet$
Efeito linha	$\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet}$	$(\mathbf{g}_i - \mathbf{g}_\bullet)^T \mathbf{a}_\bullet$
Efeito coluna	$\bar{Y}_{\bullet j} - \bar{Y}_{\bullet\bullet}$	$\mathbf{g}_\bullet^T (\mathbf{a}_j^T - \mathbf{a}_\bullet)$
Efeito resíduo	$\bar{Y}_{ij} - \bar{Y}_{i\bullet} - \bar{Y}_{\bullet j} + \bar{Y}_{\bullet\bullet}$	$(\mathbf{g}_i - \mathbf{g}_\bullet)^T (\mathbf{a}_j - \mathbf{a}_\bullet)$

O Biplot baseia-se na aproximação DVS (Decomposição em Valores Singulares) de uma matriz, por outra de posto inferior, conforme a definição de fatoração dada por Gabriel (1971). Na análise AMMI, a DVS é aplicada à matriz de interações genótipo-ambiente, que será representada por meio de sua aproximação de posto q definida por:

$$\widehat{\mathbf{GA}}_{(g \times a)} = \sum_{k=1}^q \lambda_k \gamma_k \boldsymbol{\alpha}_k^T = \tilde{\mathbf{U}} \tilde{\boldsymbol{\Lambda}} \tilde{\mathbf{V}}^T$$

em que:

$\tilde{\mathbf{U}}_{(g \times q)}$ possui em suas colunas apenas os q primeiros autovetores $\gamma_{(g \times 1)}$ da matriz resultante do produto $(\mathbf{GA})(\mathbf{GA})^T$;

$\tilde{\Lambda}_{(q)}$ é a matriz diagonal com os primeiros valores singulares não nulos $\lambda_1, \lambda_2, \dots, \lambda_q$ da matriz resultante do produto de $(\mathbf{GA})(\mathbf{GA})^T$ ou $(\mathbf{GA})^T(\mathbf{GA})$;

$\tilde{\mathbf{V}}_{(q \times g)}$ possui em suas linhas os q primeiros autovetores de $\alpha_{(1 \times a)}^T$ da matriz resultante do produto de $(\mathbf{GA})^T(\mathbf{GA})$.

A expressão pode ser escrita como:

$$\widehat{\mathbf{GA}}_{(g \times a)} = (\tilde{\mathbf{U}}\tilde{\Lambda}^{1/2})(\tilde{\Lambda}^{1/2}\tilde{\mathbf{V}}^T) = \tilde{\mathbf{G}}_{(g \times q)}\tilde{\mathbf{A}}_{(q \times a)}^T = \sum_{k=1}^q (\lambda_k^{1/2}\gamma_k^{1/2})(\lambda_k^{1/2}\alpha_k^T)$$

em que $\tilde{\mathbf{G}}_{(g \times q)}$ tem g vetores denominados marcadores de linhas, com q componentes cada um (coordenadas dos genótipos para cada eixo selecionado) e $\tilde{\mathbf{A}}_{(q \times a)}^T$ terá a marcadores de colunas, com q componentes cada um (coordenadas dos ambientes para os q eixos).

Na análise AMMI, os gráficos Biplot utilizados para o estudo da interação $G \times A$, construídos em duas dimensões, geralmente apresentam no eixo das abscissas os efeitos principais (médias dos genótipos e dos ambientes) e no eixo das ordenadas os escores do primeiro eixo de interação (ACP_1). Para a representação em três dimensões utiliza-se (ACP_1) vs. (ACP_2) vs. Médias.

3.2.2 Modelo SREG

O conceito SREG (Sites Regression), origina-se da análise de ensaios multi-ambientes (MET-MultiEnvironment Trials) no melhoramento de plantas para testar adaptação geral e específica de cultivares. A variação fenotípica observada ou desempenho de um cultivar (P) consiste do efeito principal do genótipo (G), do efeito principal do ambiente (A) e efeitos ambientais nos genótipos ou interação genótipo-ambiente (GA),

$$P = G + A + GA \Rightarrow P - A = G + GA. \quad (17)$$

A análise de dados, segundo o modelo SREG (Sites Regression), utiliza o gráfico conhecido em inglês como GGE (Genotype-Genotype Environment), o qual foi proposto para explorar as respostas dos genótipos a ambientes específicos. O modelo multiplicativo SREG é semelhante ao modelo multiplicativo AMMI, a diferença está nos efeitos principais dos genótipos que são considerados junto com o efeito da interação $G \times A$ e no AMMI são estimados como efeitos aditivos.

A equação (17) está escrita em termos de componentes de variância. Quando os efeitos são medidos, tornam-se:

$$\bar{Y}_{ij} = \mu + G_i + A_j + (GA)_{ij} \quad \text{ou} \quad \bar{Y}_{ij} - \mu - A_j = G_i + (GA)_{ij}$$

em que:

$\bar{Y}_{ij}$: é a resposta média do i -ésimo genótipo e j -ésimo ambiente;

μ : é a média geral;

G_i : é o efeito principal do i -ésimo genótipo, $i = (1, \dots, g)$;

A_j : é o efeito principal do j -ésimo ambiente, $j = (1, \dots, a)$;

$(GA)_{ij}$: é o efeito da interação do i -ésimo genótipo no j -ésimo ambiente.

Separando a mistura GGA em dois termos multiplicativos, temos:

$$\bar{Y}_{ij} - \mu - A_j = g_{i1}a_{1j} + g_{i2}a_{2j} + \varepsilon_{ij} \quad (18)$$

em que:

g_{i1} e a_{1j} : são os escores principais para o i -ésimo genótipo no j -ésimo ambiente, respectivamente;

g_{i2} e a_{2j} : são os escores secundários para i -ésimo genótipo no j -ésimo ambiente, respectivamente;

ε_{ij} : é o resíduo não explicado pelos escores primários e secundários.

Aplicando a DVS na equação (18), conforme a seção 2.2.1, tem-se:

$$\bar{Y}_{ij} - \mu - A_j = \lambda_1 u_{i1} v_{j1} + \lambda_2 u_{i2} v_{j2} + \varepsilon_{ij}$$

em que:

λ_1 e λ_2 : são os maiores autovalores do primeiro e segundo componentes principais ACP_1 e ACP_2 , respectivamente;

u_{i1} e u_{i2} : são os autovetores do i -ésimo genótipo para ACP_1 e ACP_2 , respectivamente;

v_{j1} e v_{j2} : são os autovetores do j -ésimo ambiente para ACP_1 e ACP_2 , respectivamente.

3.2.2.1 GGEbiplot

No GGEbiplot, quando um valor da componente principal ACP_1 está perto de 1, significa que os genótipos apresentam médias altas de produtividade. Se o segundo componente principal ACP_2 está próximo de zero, os genótipos são considerados como os mais estáveis e ideais. O gráfico também pode indicar a existência de um ambiente ideal, quando o componente principal ACP_1 é um valor próximo de 1, significa alta capacidade de discriminação dos genótipos e se a segunda componente principal ACP_2 está próxima de zero, indica os genótipos mais representativos dos ambientes. Assim, pode-se afirmar que o primeiro componente principal representa a produtividade e a segunda componente principal representa a estabilidade. Essas propriedades ocorrem quando o ACP_1 dos cultivares são altamente correlacionados com os efeitos dos cultivares.

O gráfico consiste em fazer um polígono unindo os pontos dos genótipos mais afastados da origem em cada quadrante, sendo que todos os outros genótipos estarão incluídos dentro do polígono formado. Posteriormente, é traçada uma perpendicular para cada lado do polígono com relação à origem e, assim, os ambientes e genótipos são separados em setores. Os genótipos em cada setor são os que apresentam o melhor desenvolvimento nos ambientes incluídos naquele setor.

Ambientes/épocas localizados mais distantes da origem, discriminam melhor os cultivares do que aqueles localizados próximos à origem. Os polígonos formados indicam a presença ou ausência de interação do genótipo e ambiente e sugere também a ausência ou existência de multi-ambientes entre os ambientes/épocas analisados.

O multi-ambiente é definido como um grupo de ambientes/épocas que possui o melhor cultivar em comum e que possua uma certa repetibilidade durante os anos. Pode ser dividido em simples ou complexo. Simples, quando a ausência de interação genótipo e ambiente cruzado para todos os ambientes/épocas e complexa quando envolve interação genótipo e ambiente cruzado não se repetem durante os anos. Para o simples, um ou poucos ambientes/épocas seriam suficientes para uma efetiva avaliação dos cultivares e para o complexo, distintos locais/épocas de testes são requeridos para selecionar cultivares superiores dentro de toda a região durante os anos.

3.2.2.2 Transformação dos dados para GGEbiplot

A equação (18), refere-se ao modelo para ambiente centrado, pois $(\mu + A_j)$ são subtraídos de cada um dos valores observados Y_{ij} . Existe dois problemas em utilizar este modelo:

- i) ele só pode ser utilizado para analisar dados MET (multi-ambientes), que tenha uma única característica, por exemplo, produção é uma característica que está em todos os ambientes;
- ii) em conjunto de dados, cujos ambientes são heterogêneos.

Para solucionar estes problemas, Yan e Kang (2002), propuseram duas transformações. A primeira é a transformação de dados para dimensionar o ambiente centralizado, tem-se:

$$\frac{\bar{Y}_{ij} - \mu - A_j}{d_j} = g_{i1}a_{1j} + g_{i2}a_{2j} + \varepsilon_{ij}$$

em que d_j é o desvio fenotípico padronizado do j -ésimo ambiente:

$$d_j = \sqrt{\frac{1}{m-1} \sum_{i=1}^m (Y_{ij} - \bar{Y}_{.j})^2}$$

em que m é o número de genótipos testados, Y_{ij} é a resposta do i -ésimo genótipo e j -ésimo ambiente e $\bar{Y}_{.j}$ é a média do j -ésimo ambiente. A única característica para analisar dados MET é que este modelo pressupõe que todos os ambientes são igualmente importantes. Opõe-se, portanto na possibilidade de detectar eventuais diferenças entre os ambientes.

A segunda transformação de dados para dimensionar o ambiente centralizado, tem-se:

$$\frac{\bar{Y}_{ij} - \mu - A_j}{s_j} = g_{i1}a_{1j} + g_{i2}a_{2j} + \varepsilon_{ij}$$

em que s_j é o erro padrão do j -ésimo ambiente:

$$s_j = \sqrt{\frac{1}{m(r-1)} \sum_{i=1}^m \sum_{k=1}^r (Y_{ijk} - \bar{Y}_{ij.})^2}$$

em que r é o número de repetições ($k = 1, \dots, r$) no j -ésimo ambiente e Y_{ijk} é a resposta do i -ésimo genótipo e j -ésimo ambiente na k -ésima repetição. Este modelo, somente pode ser utilizado para dados que possuem réplicas.

A metodologia SREG utilizada na construção do gráfico GGEbiplot tem sido reconhecida como uma metodologia revolucionária na abordagem com análise gráfica bidimensional para melhoramento e pesquisa genética, incluindo, também, análise de estabilidade e adaptabilidade na recomendação de cultivares, destacando-se das demais técnicas pela sua alta eficiência estatística na explicação da interação $G \times A$.

As análises para verificar a estabilidade produtiva e teor de sólidos solúveis dos genótipos e ambientes para o modelo AMMI e SREG, foram realizadas por meio do pacote computacional SAS/STAT v9.1 utilizando o procedimento PROC GLM, conforme Anexo A.

3.3 Escalonamento Multidimensional

Multidimensional Scaling (MDS), Escalonamento Multidimensional ou Análises de Proximidades, é um método que trata de representar as medidas de proximidade (similaridades ou dissimilaridades) entre pares de objetos como distâncias em um espaço multidimensional em dimensão reduzida, permitindo assim a inspeção visual da estrutura dos dados.

Os algoritmos MDS tornaram-se populares com o surgimento dos computadores, o qual apresenta uma solução analítica sem interações. O principal ganho da utilização da técnica MDS, é que ela pode ser utilizada tendo variáveis de qualquer escala.

A abordagem em MDS, é definida como, considere a matriz de dissimilaridade $\Delta = [\delta_{ij}] \in \mathbb{R}$ de ordem $(n \times n)$, em que δ_{ij} representa uma medida de proximidade entre o i -ésimo e o j -ésimo objeto. Um algoritmo de redução dimensional obtém a configuração dos pontos (vetores de coordenadas) denominado por $\mathbf{x}_i = (x_{i1}, \dots, x_{iq})$ de ordem $(n \times q)$ em uma dimensão menor, ou seja $(n > q)$, também deve-se verificar se a matriz de distância euclidiana $\mathbf{D} = [d_{ij}]$ de ordem $(n \times n)$, sendo que $d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|$, em que $i = (1, \dots, n)$ e $j = (1, \dots, q)$, obtida a partir desse conjunto de pontos, se aproxime ao máximo da matriz de dissimilaridade original, ou seja $\mathbf{D} \approx \Delta$ (TIMM, 2002).

Representando matricialmente o problema em MDS, em que a partir da matriz de dissimilaridade Δ , através de um algoritmo MDS, obtém-se uma matriz de coordenadas $\mathbf{X}$. A i -ésima linha da matriz $\mathbf{X}$ representa o vetor de coordenadas do i -ésimo objeto. A distância euclidiana d_{ij} calculada entre as i -ésimas e j -ésimas linhas aproxima-se da dissimilaridade

entre os i -ésimos e j -ésimos objetos, ou seja $d_{ij} \approx \delta_{ij}$.

$$\underbrace{\Delta_{(n \times n)} = \begin{pmatrix} 0 & \delta_{12} & \cdots & \delta_{1n} \\ \delta_{21} & 0 & \cdots & \delta_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ \delta_{n1} & \delta_{n2} & \cdots & 0 \end{pmatrix}}_{\text{Matriz de dissimilaridade}} \Rightarrow \text{MDS} \Rightarrow \underbrace{\mathbf{X}_{(n \times q)} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1q} \\ x_{21} & x_{22} & \cdots & x_{2q} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nq} \end{pmatrix}}_{\text{Vetores de coordenadas de linhas}}$$

A validação feita em MDS, é a partir da matriz de coordenadas $\mathbf{X}$, em que pode-se obter a matriz de distâncias $\mathbf{D}$. Usa-se então uma função de *Stress* para medir a diferença entre a matriz de dissimilaridade original e a matriz de distância calculada. A escolha do algoritmo utilizado na MDS não depende da determinação do número de dimensões, tarefa importante para um bom desempenho da análise dos dados e interpretação dos resultados.

$$\mathbf{X}_{(n \times q)} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1q} \\ x_{21} & x_{22} & \cdots & x_{2q} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nq} \end{pmatrix} \Rightarrow$$

$$\underbrace{\mathbf{D}_{(n \times n)} = \begin{pmatrix} 0 & d_{12} & \cdots & d_{1n} \\ d_{21} & 0 & \cdots & d_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ d_{n1} & d_{n2} & \cdots & 0 \end{pmatrix}}_{\text{Matriz de distâncias}} \quad \underbrace{\Delta_{(n \times n)} = \begin{pmatrix} 0 & \delta_{12} & \cdots & \delta_{1n} \\ \delta_{21} & 0 & \cdots & \delta_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ \delta_{n1} & \delta_{n2} & \cdots & 0 \end{pmatrix}}_{\text{Matriz de dissimilaridade}}$$

Processo de validação $d_{ij} \approx \delta_{ij}$

Existem outras distâncias além da distância Euclidiana, por exemplo, a distância de Mahalanobis, a distância Euclidiana ponderada, medida de City block, medida Minkowski etc, podem ser mais apropriadas em alguns casos, onde dados originais podem ser ordinais ou nominais. Em algumas situações os dados são descritos por funções (contínuas ou discretas), neste caso pode ser utilizada uma medida de correlação (COX e COX, 2001).

3.3.1 Algoritmos MDS

Existem diversos algoritmos para a resolução da metodologia MDS, tais como: CMDS (Classical Multidimensional Scaling) como algoritmo clássico; FASTMAP, LANDMARK, como algoritmos rápidos; Nyström e SMACOF (Scaling by MAjorizing a COnvex Function) são algoritmo de otimização.

3.3.1.1 Escalonamento clássico

Considere a matriz de dados $\mathbf{Y}_{(n \times p)}$, em que as n linhas correspondem aos indivíduos da amostra e as p colunas são as variáveis. Gower (1971) define que a distância entre todos os pares de indivíduos ou dissimilaridades pode ser calculada através do conjunto de distâncias positivas d_{ij} . Estas distâncias devem satisfazer as seguintes condições, $d_{ij} = d_{ji}$ e $d_{ii} = d_{jj} = 0$. Então a matriz $\mathbf{D}_{(n \times n)}$ de distâncias ou dissimilaridades terá o (i, j) -ésimo elemento igual a $\frac{1}{2}d_{ij}^2$ e $\mathbf{X}_{(n \times q)}$ é a matriz que contém as coordenadas dos n pontos em q dimensões, tal que a origem das coordenadas é a média dos pontos (COX e COX, 2001).

A distância euclidiana entre os i -ésimos e j -ésimos objetos é dada por:

$$d_{ij}^2 = (\mathbf{x}_i - \mathbf{x}_j)^T (\mathbf{x}_i - \mathbf{x}_j). \quad (19)$$

Seja $\mathbf{E}$, a matriz de produto interno das coordenadas, em que $\mathbf{E} = (e_{ij}) = \mathbf{x}_i^T \mathbf{x}_j$, então as distâncias euclidianas ao quadrado podem ser escritas em termos dos elementos da matriz $\mathbf{E}$ e conforme a equação (19), tem-se

$$\begin{aligned} d_{ij}^2 &= \mathbf{x}_i^T \mathbf{x}_i + \mathbf{x}_j^T \mathbf{x}_j - 2\mathbf{x}_i^T \mathbf{x}_j \\ &= e_{ii} + e_{jj} - 2e_{ij}. \end{aligned} \quad (20)$$

O posto da matriz $\mathbf{E}$ é igual a q , para obter uma solução única é imposta uma restrição localizada, a mais conveniente é considerar que o centróide dos pontos reside na origem, isto

é, $\sum_{i=1}^n x_{is} = 0$, em que $s = 1, \dots, q$. Para calcular $\mathbf{E}$, considera-se que

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n d_{ij}^2 &= \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i^T \mathbf{x}_i + \mathbf{x}_j^T \mathbf{x}_j \\ \frac{1}{n} \sum_{j=1}^n d_{ij}^2 &= \mathbf{x}_i^T \mathbf{x}_i + \frac{1}{n} \sum_{j=1}^n \mathbf{x}_j^T \mathbf{x}_j \\ \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n d_{ij}^2 &= \frac{2}{n} \sum_{i=1}^n \mathbf{x}_i^T \mathbf{x}_i. \end{aligned}$$

Substituindo o último resultado na expressão (20), tem-se:

$$\begin{aligned} e_{ij} &= \mathbf{x}_i^T \mathbf{x}_j \\ &= -\frac{1}{2} \left(d_{ij}^2 - \frac{1}{n} \sum_{i=1}^n d_{ij}^2 - \frac{1}{n} \sum_{j=1}^n d_{ij}^2 + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n d_{ij}^2 \right) \\ &= a_{ij} - a_{i.} - a_{.j} + a_{..}, \end{aligned}$$

em que $a_{ij} = -\frac{1}{2}d_{ij}^2$ e

$$a_{i.} = \frac{1}{n} \sum_{j=1}^n a_{ij}, \quad a_{.j} = \frac{1}{n} \sum_{i=1}^n a_{ij}, \quad a_{..} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n a_{ij}.$$

Define-se a matriz $\mathbf{A} = [a_{ij}]$, então o produto interno da matriz $\mathbf{E}$ é dado por

$$\mathbf{E} = \mathbf{H} \mathbf{A} \mathbf{H}$$

em que $\mathbf{H}$ é a matriz centrada

$$\mathbf{H} = \left(\mathbf{I} - \frac{\mathbf{1}\mathbf{1}^T}{n} \right)$$

sendo que $\mathbf{1} = (1, 1, \dots, 1)^T$ com n coordenadas. Considerando que $\mathbf{E}$ pode ser expressa como $\mathbf{E} = \mathbf{X} \mathbf{X}^T$, em que $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_q]$, é a matriz $n \times q$ das coordenadas. O posto de $\mathbf{E}$ será:

$$\text{posto}(\mathbf{E}) = \text{posto}(\mathbf{X} \mathbf{X}^T) = \text{posto}(\mathbf{X}) = q.$$

Dessa forma, a matriz $\mathbf{E}$ é simétrica, positiva semi-definida com posto q , apresenta q autovalores não negativos.

Portanto, a matriz de distâncias $\mathbf{D}$, pode ser escrita como,

$$\mathbf{D} = \mathbf{E}\mathbf{1}\mathbf{1}^T + \mathbf{1}\mathbf{1}^T\mathbf{E} - 2\mathbf{X}\mathbf{X}^T$$

e a matriz de produto interno $\mathbf{E}$, pode ser expressa como

$$\mathbf{E} = \mathbf{X}\mathbf{X}^T = \left(\mathbf{I} - \frac{\mathbf{1}\mathbf{1}^T}{n}\right) \mathbf{D} \left(\mathbf{I} - \frac{\mathbf{1}\mathbf{1}^T}{n}\right). \quad (21)$$

Qualquer solução da equação (21) irá produzir coordenadas das observações. Existem várias maneiras de se fazer isso, no entanto a mais utilizada é a Análise de Coordenadas Principais (KRZANOWSKI e MARRIOTT, 1994). Esta técnica baseia-se em encontrar autovalores e autovetores da matriz do lado esquerdo da equação (21), cuja as colunas de $\mathbf{X}$ contém os autovetores correspondentes aos autovalores em ordem decrescente. Caso algum desses vetores seja negativo, então faz-se a transformação das distâncias d_{ij} para

$$d_{ij}^* = \sqrt{d_{ij}^2 + h}, \quad \forall i, j$$

em que h é o valor absoluto do menor autovalor. Dessa forma a matriz de inter-distâncias é Euclidiana.

3.3.2 Funções e modelos MDS

MDS é definido por uma função $f : p_{ij} \rightarrow d_{ij}(\mathbf{X})$ denota-se como sendo as proximidades p_{ij} aproximam-se das distâncias $d_{ij}(\mathbf{X})$ de um espaço $\mathbf{X}$, $f : p_{ij}$ especifica um mapeamento ou uma transformação de p_{ij} . Assim, um modelo MDS é uma proposição na qual a partir de medidas de proximidades e depois de alguma transformação de f obtêm-se as distâncias entre pontos na matriz $\mathbf{X}$:

$$f(p_{ij}) = d_{ij}(\mathbf{X}).$$

Os modelos MDS necessitam que cada valor de proximidade seja mapeado exatamente em suas correspondentes distâncias. Em geral, não se trata exatamente de satisfazer f , mas procura-se satisfazer uma configuração (em uma dimensão de baixa ordem), onde as distâncias obtidas aproximam-se de f , o tanto quanto seja possível. A condição referente a “tão próximo quanto possível” é quantificada por uma medida de ajuste entre as distâncias d_{ij} e as de dissimilaridades, denominada de *stress* (BORG e GROENEN, 2005).

3.3.3 Função *Stress*

Os modelos de MDS exigem que cada valor de proximidade seja mapeado exatamente em suas correspondentes distâncias. As proximidades sempre contêm ruído devido à imprecisão na medição, nos erros de amostragem. Não se consegue satisfazer $f(p_{ij}) = d_{ij}(\mathbf{X})$, mas a melhor aproximação “tão igual quanto” possível $f(p_{ij}) \approx d_{ij}(\mathbf{X})$ pode ser quantificada por uma medida de ajuste entre as distâncias d_{ij} e $f(p_{ij})$ definida como função *stress*. Quanto menor for o valor do *stress* melhor é o ajuste da matriz de distância reproduzida à matriz de distância observada.

A função *stress* representa a soma de todos os erros sobre os pares (i, j) . Essa informação pode perder a exatidão de uma representação MDS, devido a dependência da normalização de suas dissimilaridades. O erro ao quadrado é definido por

$$e_{ij}^2 = [f(p_{ij}) - d_{ij}(\mathbf{X})]^2.$$

Somando os e_{ij}^2 sobre todos os pares (i, j) gera uma medida de qualidade do ajuste, *badness-of-fit* para a representação completa do MDS, denominado de *raw stress*

$$\sigma(\mathbf{X}) = \sum_{(i,j)} [f(p_{ij}) - d_{ij}(\mathbf{X})]^2.$$

A função *stress* são índices que avaliam a desigualdade (admissivelmente transformada) das proximidades e distâncias correspondentes. É semelhante ao coeficiente de correlação, exceto pelo fato dele medir a qualidade do ajuste.

A melhor maneira de verificar a correspondência entre d_{ij} e δ_{ij} é o diagrama de Shepard, cada ponto têm as seguintes coordenadas (δ_{ij}, d_{ij}) . É um diagrama de dispersão que representa as proximidades (eixo horizontal) em relação as disparidades ou as distâncias (eixo vertical), que detecta os “outliers”. O valor do *Stress* pode ser interpretado como um coeficiente de correlação que mede a falta de ajuste.

3.3.4 Matrizes de Distância

Existem três tipos de matrizes utilizadas nos modelos MDS: Matriz de dissimilaridade, Matriz de similaridade e Matriz de distância Euclidiana.

Considera-se a matriz de dissimilaridade $\Delta = \delta_{ij}$ de ordem $n \times n$, simétrica e quadrada, em que seus elementos representam as medidas de proximidade. Os valores de cada entrada δ_{ij} dessa matriz, são positivos e indicam a dissimilaridade entre os i -ésimos e o j -ésimos objetos, sendo que os valores da diagonal principal são iguais a zero.

Se os dados são apresentados em forma de vetores, é necessário definir uma função de dissimilaridade entre os pares de objetos. Borg e Groenen, 2005, definem a matriz de dissimilaridade entre dois vetores como a soma das dissimilaridades dos seus atributos: $\Delta(x_i, x_j) = \sum_p \Delta(x_i, x_j)$. Utilizam-se pesos w , quando os atributos possuem diferenças, assim $\Delta(x_i, x_j) = \sum_p w \Delta(x_i, x_j)$

A matriz de similaridade $\mathbf{S} = s_{ij}$ de ordem $n \times n$, cujos elementos representam medidas de proximidade que indicam o quanto dois objetos são similares. Esta matriz apresenta as mesmas propriedades da matriz de dissimilaridade. As medidas de correlação, por exemplo, podem ser interpretadas como similaridade entre objetos. O MDS aceita medidas de similaridade mediante uma transformação para medidas de dissimilaridade (BORG e GROENEN, 2005).

A matriz de distância euclidiana, é uma matriz de dissimilaridade, cujos elementos representam a norma euclidiana. Esta matriz é formada a partir de um conjunto de dados. Para calcular as distâncias, utiliza-se a equação generalizada (21) (BORG e GROENEN, 2005).

3.3.5 Biplots

Gower e Harding (1988) aplicaram a definição do Biplot para o escalonamento multidimensional e Gower e Hand (1996) descreveram-na com mais detalhes. Esta técnica pode ser usada para melhorar a análise de distância e utilizaram os resultados algébricos da ACP (Análise de componentes principais) chamando de Análise de Coordenadas Principais.

A análise de Coordenadas Principais pode ser entendida como um caso mais geral da Análise de Componentes Principais que usa uma variedade de diferentes medidas de distância ou de similaridade. Os dados a serem analisados podem ser uma mistura de valores contínuos ou medições binárias (COX e COX, 2001).

Suponha que a matriz $\mathbf{E}$ da equação (21) foi obtida pelo escalonamento métrico

clássico. A matriz $\mathbf{E}$ pode ser escrita em termos de decomposição espectral:

$$\mathbf{E} = \mathbf{X}\mathbf{X}^T = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T,$$

em que:

$\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ é uma matriz diagonal de autovalores de $\mathbf{E}$, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$;

$\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$, é a matriz de autovetores normalizados e $\mathbf{v}_i^T \mathbf{v}_i = 1$

Como existem $n - q$ vetores nulos, $\mathbf{E}$ pode-se ser reescrita como,

$$\mathbf{E} = \mathbf{V}_1 \mathbf{\Lambda}_1 \mathbf{V}_1^T,$$

em que $\mathbf{\Lambda}_1 = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_q)$ e $\mathbf{V}_1 = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_q)$.

Assim, como $\mathbf{E} = \mathbf{X}\mathbf{X}^T$, a matriz de coordenadas $\mathbf{X}$ é dada por

$$\mathbf{X} = \mathbf{V}_1 \mathbf{\Lambda}_1^{1/2},$$

em que $\mathbf{\Lambda}_1^{1/2} = \text{diag}(\lambda_1^{1/2}, \lambda_2^{1/2}, \dots, \lambda_q^{1/2})$. Caso a matriz de distâncias entre os pontos não seja euclidiana, devido a matriz $\mathbf{E}$ não ser uma matriz positiva semi-definida, com autovalores negativos. Neste caso, existem duas possibilidades: descartar esses autovalores e seus respectivos autovetores na formação das matrizes $\mathbf{V}_1$ e $\mathbf{\Lambda}_1$; ou adicionar uma constante apropriada c às dissimilaridades e repetir o processo novamente (COX e COX, 2001).

Assim, a Análise de Componentes Principais (ACP) equivale a Análise de Coordenadas Principais, quando a distância euclidiana é utilizada para medir a dissimilaridade, as distâncias entre os pontos representam as observações aproximadas no biplot equivalente as distâncias euclidianas entre as observações originais (COX e COX, 2001).

Os vetores (eixos) indicam apenas a direção, e não dando importância o seu comprimento. Gower e Hand (1996) incentivaram a análise destas variáveis como eixos verdadeiros e desenhá-los no biplot inteiro, juntamente com as marcas de escala adequada.

As análises e a rotina computacional para o MDS, conforme o Anexo B, foram implementadas através do sistema de análise estatístico R versão 2.10.1 (The R Foundation for Statistical Computing, 2009).

4 RESULTADOS

4.1 Análise da variável Produção utilizando os modelos AMMI e SREG

Inicialmente foi analisado a variável resposta Produção, utilizando as metodologias AMMI e GGE. Na Tabela 9 apresenta-se as Médias preditas da variável Produção. A maior média da variável Produção está no G9 no ambiente A12 e a menor média está no G3 no ambiente A5.

Tabela 9 - Médias preditas da variável Produção

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
G1	30,65	21,47	18,87	24,02	18,68	21,16	28,96	28,85	21,44	21,45	18,55	30,59
G2	29,34	23,19	17,86	22,96	20,45	16,63	25,65	30,45	27,77	23,04	23,06	25,28
G3	30,99	26,69	20,31	24,01	14,52	20,52	26,41	35,64	31,11	23,11	26,80	34,18
G4	32,26	26,12	16,03	28,31	17,91	22,71	32,59	31,33	33,35	26,94	16,10	32,38
G5	28,99	30,55	22,34	32,27	18,51	22,73	31,32	37,75	26,92	24,30	23,31	31,68
G6	26,75	26,17	20,14	28,24	23,67	18,83	33,42	35,96	26,96	21,83	24,56	40,14
G7	33,74	28,87	18,30	27,49	24,56	22,11	34,32	37,79	32,49	31,07	19,73	47,38
G8	31,97	27,29	20,83	29,26	23,33	24,02	34,56	39,71	30,43	28,69	24,12	47,28
G9	31,56	28,88	21,71	25,37	20,97	22,76	30,66	39,03	40,33	28,74	24,15	49,26

A Tabela 10 corresponde ao esquema de análise de variância para o modelo AMMI da variável Produção, com os graus de liberdade dos quadrados médios. Verifica-se ao nível de 1% de significância que o efeito de Ambiente, o efeito de Genótipo e o efeito interação Genótipo $\times$ Ambiente são significativos. As somas de quadrados (SQ) dos respectivos efeitos correspondem a 4,78%, 9,26% e 14,43% da soma de quadrados total. O coeficiente de variação é de 16,63% e a média da variável Produção é de 27,22 $t.ha^{-1}$.

Na Tabela 11 apresenta os resultados do teste de Gollob, que determina o significado de cada um dos termos do modelo AMMI para a variável Produção (Gollob, 1968). Na primeira coluna é a soma de quadrados corrigida para o número de repetições ($S.Q.$), na segunda coluna mostra o percentual total (%) da soma dos quadrados da interação $G \times A$ explicada por cada um dos termos do modelo AMMI e na terceira coluna é apresentado o percentual acumulado. As demais colunas fornecem os graus de liberdade ($g.l.$), o quadrado médio ($Q.M.$), o valor do teste F e o valor da probabilidade associada ao teste F para cada

Tabela 10 - Análise de variância conjunta para a variável Produção

Fonte de variação	g.l.	S.Q.	Q.M.	F	Pr>F
Blocos d. Ambiente (B d. A)	24	628,45	26,19	1,28	0,1833
Ambiente (A)	11	10114,62	919,51	44,87	<,0001
Genótipo (G)	8	1780,53	222,66	10,86	<,0001
Interação ($G \times A$)	8	2774,80	31,53	1,54	0,0073
Resíduo	192	3934,41	20,50		
Total	323	19232,81			
Média ($t.ha^{-1}$)	27,22		CV (%)	16,63	

um dos termos da análise AMMI, respectivamente. Neste caso, os dois primeiros termos são significativos ao nível de 1%, sendo que o primeiro contribuiu com 49,03% e o segundo com 20,17% perfazendo 69,20% da variabilidade total.

Tabela 11 - Resultados dos testes de Gollob da análise AMMI da variável Produção

S.Q.	%	%Acumulada	g.l.	Q.M	F	Pr>F
1360,49	49,0304	49,030	18	75,5829	3,92152	0,00000
559,71	20,1711	69,202	16	34,9816	1,81497	0,03071
407,08	14,6707	83,872	14	29,0771	1,50863	0,10959
240,00	8,6494	92,522	12	20,0003	1,03769	0,41540
123,34	4,4450	96,967	10	12,3341	0,63994	0,77860
54,36	1,9589	98,926	8	6,7945	0,35252	0,94395
19,88	0,7165	99,642	6	3,3136	0,17192	0,98410
9,93	0,3579	100,000	4	2,4828	0,12882	0,97182
0,00	0,0000	100,000	2	0,0000	0,00000	1,00000

Com respeito ao nível de significância, vale salientar que, com a utilização de 1% em vez de 5%, reduz-se a probabilidade de ocorrer o erro tipo I, isto é, de aceitar o modelo AMMI com maior número de eixos. Uma das premissas de resposta da análise AMMI é a de que nos primeiros eixos concentra-se a maior porcentagem “padrão” (LAVORANTI, 2003).

A Tabela 12, mostra os resultados da produção média, a ordem decrescente para a produção média e os escores dos genótipos e ambientes para os três primeiros termos

AMMI ($Dim1$, $Dim2$, $Dim3$) e os códigos utilizados nos gráficos Biplot.

Tabela 12 - Resultados da produção média e os escores para o Biplot da variável Produção

Código	Produção Média	Ordem p/médias	Dim 1	Dim 2	Dim 3
G1	23,7261	(9)	-1,36343	-0,71567	-1,09622
G2	24,1003	(8)	-2,01629	0,67030	0,48838
G3	25,7731	(7)	-0,34815	1,12496	1,51640
G4	26,3622	(6)	-0,60000	1,84150	-1,77579
G5	27,5564	(4)	-1,74616	-0,72336	0,76326
G6	27,2231	(5)	0,27793	-2,17959	0,49545
G7	29,8197	(3)	1,91289	-0,09187	-1,43354
G8	30,1239	(2)	1,41020	-1,13698	-0,29433
G9	30,2836	(1)	2,47301	1,21071	1,33639
A1	30,6933	(4)	-0,69292	1,05217	-0,96067
A2	26,5807	(7)	-0,47387	0,24459	0,50997
A3	19,5993	(12)	-1,02959	-0,39658	1,08338
A4	26,8815	(6)	-1,14702	-0,71552	-0,50516
A5	20,2911	(11)	-0,21758	-1,16003	-0,84710
A6	21,6985	(10)	-0,96401	0,54327	-0,53102
A7	30,8767	(3)	-0,15503	-0,86008	-1,40732
A8	35,1689	(2)	0,41321	-0,48484	1,13091
A9	30,0885	(5)	0,99001	2,60378	0,76004
A10	25,4637	(8)	0,30677	0,93331	-1,00216
A11	21,7081	(9)	-0,91902	-0,81846	1,78760
A12	37,5741	(1)	3,88906	-0,94162	-0,01848

Para a construção do gráfico da Figura 9, utilizou-se os valores das colunas $Dim1$ e $Dim2$ da Tabela 12. Neste Biplot, genótipos e ambientes estáveis são aqueles cujos pontos situam-se próximos à origem, ou seja, os valores dos escores são quase nulos nos dois eixos de interação ($Dim 1$ e $Dim 2$), segundo Duarte e Vencovsky, 1999. O genótipo que menos contribuiu para a interação $G \times A$ foi G3 e o ambiente A2 foi o mais estável, cujo o ponto situa-se próximo à origem, ou seja os escores são os de menor valor para os dois eixos

de interação ($Dim1$ e $Dim2$). Os genótipos G7 e G9 estão adaptados aos ambientes A9 e A12, pois os pontos relativos a esses genótipos apontam para uma direção comum e os vetores partindo da origem com ângulo pequeno entre eles, que pode ser orientada pelos marcadores dos ambientes A9 e A12. Os genótipos G3 e G4 também estão adaptados aos ambientes A1 e A10 por interpretação semelhante. A falta de adaptação do genótipo G2 ao ambiente A12 é nítida nos dois gráficos (marcadores apontando em direções opostas).

Os aspectos característicos dos genótipos e ambientes podem ser determinados pelos eixos de interação, observa-se que o primeiro eixo deve ser determinado por características contrastantes entre os ambientes A1 e A6 com o ambiente A12. Na Figura 9, pode-se observar que os genótipos e os ambientes mais estáveis (com pequena contribuição na $SQ_{(G \times A)}$ são aqueles cujos pontos estão próximos da origem, ou seja, com os escores praticamente nulos para os dois eixos de interação $Dim1$ e $Dim2$, é o caso do genótipo G3 e o ambiente A2.

Figura 9 - Biplot AMMI com os dois primeiros componentes principais correspondente à representação dos nove genótipos (G) e doze ambientes (A) para a variável Produção

Na Figura 10, utilizou-se o eixo das abscissas para representar os efeitos principais (médias dos genótipos e dos ambientes) e o eixo das ordenadas para representar os escores dos genótipos e ambientes para o referido eixo de interação. Genótipos e ambientes estáveis são aqueles cujos pontos situam-se próximos de zero (DUARTE e VENCovsky, 1999). Pode-se verificar quais os genótipos que possuem produção inferior ou superior à produção média global através da linha vertical de referência. Os genótipos G1, G2, G3 e G4 possuem produção inferior a produção média global que é de $27,22 \text{ t.ha}^{-1}$, enquanto que os genótipos G7, G8 e G9 possuem produção superior e os ambientes com pouco rendimento são A2, A3, A4, A5, A6, A10 e A11 e os ambientes que apresentaram o melhor rendimento são A1, A7, A8, A9 e A12. Os pontos situados em torno de zero em relação ao eixo horizontal *Dim1*, correspondem aos genótipos G3 e G6 e ambientes A2, A5, A7, A8 e A10 mais estáveis. Os genótipos G2 e G9 são os que possuem as produções médias mais divergentes, pelo fato de estar na posição oposta.

Figura 10 - Biplot AMMI utilizando as médias ajustadas correspondente à representação dos nove genótipos (G) e doze ambientes (A) da variável Produção

Nas Figuras 9 e 10, pode-se verificar que o ambiente A2 foi o que menos contribuiu para a interação $G \times A$ e o genótipo que também menos contribuiu foi o genótipo G3.

O gráfico Trilinear plot (Trinomial plot ou Diagrama Ternary) é um gráfico exclusivo para uma tabela de contigência com três colunas ou três variáveis. Ele mostra as frequências e as proporções, cada observação é um ponto dentro do triângulo equilátero cuja coordenada corresponde à proporção relativa em cada uma das colunas. Os três vértices representam os extremos, quando 100% ocorre em uma das três colunas; um ponto está exatamente no centro, quando a proporção for igual a 1/3 em todas as três colunas (FRIENDLY, 2001). Nota-se que na Figura 11 os genótipos G2, G4, G6, G7 e G8 e os ambientes A10 e A12 não são proporcionais, pois eles estão nos extremos do gráfico para a variável Produção.

Figura 11 - Gráfico Trilinear construído pelos três escores do modelo AMMI para a variável Produção, representando os nove genótipos (G) e doze ambientes (A)

Na Tabela 13 apresenta os resultados do teste de Gollob, que determina o significado de cada um dos termos do modelo GGE para a variável Produção (Gollob, 1968). Neste caso, os três primeiros termos são significativos ao nível de 1%, sendo que o primeiro

contribuiu com 64,80%, o segundo com 12,58% e o terceiro 9,10% perfazendo 86,50% da variabilidade total. Pode-se verificar que o modelo GGE tem uma maior contribuição na variabilidade total que o modelo AMMI.

Tabela 13 - Resultados dos testes de Gollob da análise GGE para a variável Produção

S.Q.	%	%Acumulada	g.l.	Q.M	F	Pr >F
2952,04	64,8040	64,804	18	164,002	8,50902	0,00000
573,04	12,5796	77,384	16	35,815	1,85821	0,02579
414,78	9,1053	86,489	14	29,627	1,53715	0,09979
353,31	7,7561	94,245	12	29,443	1,52760	0,11587
132,33	2,9049	97,150	10	13,233	0,68658	0,73641
83,83	1,8402	98,990	8	10,478	0,54365	0,82272
34,09	0,7483	99,738	6	5,681	0,29475	0,93897
11,92	0,2617	100,000	4	2,980	0,15461	0,96077
0,00	0,0000	100,000	2	0,000	0,00000	1,00000

O gráfico da Figura 12 refere-se ao GGEbiplot, proposto por Yan (2001), foi construído um polígono com base nos genótipos que estão localizados nos pontos mais afastados. Os genótipos G6, G7, G8 e G9 tiveram um bom desempenho em todos os ambiente A1. Os ambientes A9 e A12 por estarem mais distantes da origem, são os que apresentam maior contribuição para a interação.

4.2 Análise da variável SST os modelos AMMI e SREG

Na Tabela 14 apresenta-se as Médias preditas da variável SST, cuja maior média predita está no G9 no ambiente A6 e a menor média está no G8 no ambiente A5.

A Tabela 15 corresponde a análise de variância conjunta dos dados observados para a variável SST, verifica-se ao nível de 1% de significância que o efeito de Ambiente, o efeito de Genótipo e o efeito interação Genótipo $\times$ Ambiente são significativos. As somas de quadrados (SQ) dos respectivos efeitos correspondem a 38,57%, 4,34% e 31,68% da soma de quadrados total. O coeficiente de variação é de 10,71% e a média da variável SST é de 11,34.

Na Tabela 16 apresenta-se os resultados do teste de Gollob, que determina o significado de cada um dos termos do modelo AMMI para a variável SST (Gollob, 1968).

Figura 12 - GGEbiplot correspondente à representação dos nove genótipos (G) e doze ambientes (A) para a variável Produção

Tabela 14 - Médias preditas da variável SST

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
G1	8,69	11,31	9,66	10,89	12,73	12,47	10,23	14,35	9,43	9,21	11,42	8,57
G2	8,83	11,19	10,32	11,20	13,71	13,13	11,27	11,66	8,52	9,59	11,41	9,14
G3	9,39	11,76	10,20	12,27	11,65	11,50	12,08	12,52	9,16	10,56	11,33	10,14
G4	10,42	10,32	11,57	11,87	12,68	12,85	11,04	11,11	9,13	9,26	12,69	10,31
G5	11,00	13,34	12,06	12,90	8,73	14,18	12,46	13,32	9,42	10,35	12,58	8,91
G6	9,48	12,32	11,09	12,47	9,71	13,34	11,45	11,63	9,82	9,34	12,25	11,28
G7	9,89	12,98	11,41	12,70	7,65	14,59	13,22	14,74	9,49	11,49	12,13	11,02
G8	9,17	12,95	12,51	12,17	6,51	14,66	14,51	12,81	11,44	13,16	12,10	10,45
G9	10,08	13,05	10,88	13,38	7,04	15,37	13,69	12,02	10,45	12,68	12,66	11,09

Tabela 15 - Análise de variância conjunta para a variável SST

Fonte de variação	g.l.	S.Q.	Q.M.	F	Pr(>F)
Blocos d. Ambiente (B d. A)	24	52,73	2,20	1,49	0,0740
Ambiente (A)	11	509,95	46,36	31,44	<,0001
Genótipo (G)	8	57,42	7,18	4,87	<,0001
Interação ($G \times A$)	88	418,81	4,76	3,23	<,0001
Resíduo	192	283,07	1,47		
Total	323	1321,98			
Média (%)	11,34		CV (%)	10,71	

Na primeira coluna é a soma de quadrados corrigida para o número de repetições (*S.Q.*), na segunda coluna mostra o percentual do total (%) da soma dos quadrados da interação $G \times A$ explicada por cada um dos termos do modelo AMMI e na terceira coluna é apresentado o percentual acumulado. As demais colunas fornecem os graus de liberdade (*g.l.*), o quadrado médio (*Q.M.*), o valor do teste F e o valor da probabilidade associada ao teste F para cada um dos termos da análise AMMI, respectivamente. Neste caso, os dois primeiros termos são significativos ao nível de 1%, sendo que o primeiro contribuiu com 70,29% e o segundo com 11,35% perfazendo 81,65% da variabilidade total.

A Tabela 17, mostra os resultados da produção média, a ordem decrescente da média e os escores dos genótipos e ambientes para os três primeiros termos AMMI

Tabela 16 - Resultados dos testes de Gollob da análise AMMI para a variável SST

S.Q.	%	%Acumulada	g.l.	Q.M	F	Pr >F
294,417	70,2983	70,298	18	16,3565	10,8697	0,00000
47,547	11,3528	81,651	16	2,9717	1,9748	0,01595
28,686	6,8493	88,500	14	2,0490	1,3616	0,17407
16,797	4,0106	92,511	12	1,3997	0,9302	0,51757
10,708	2,5568	95,068	10	1,0708	0,7116	0,71309
10,435	2,4916	97,559	8	1,3044	0,8668	0,54533
5,991	1,4305	98,990	6	0,9985	0,6636	0,67917
4,230	1,0100	100,000	4	1,0575	0,7028	0,59084
0,000	0,0000	100,000	2	0,0000	0,0000	1,00000

(*Dim1*, *Dim2*, *Dim3*) e os códigos utilizados para a geração do gráfico Biplot para a variável SST.

Na Figura 13 os ambientes mais próximos da origem, portanto com escores próximos de zero, são os ambientes A1, A3, A4, A9 e A11 e os genótipos são G3 e G5. Estes ambientes e genótipos com maior proximidade da origem, contribuíram muito pouco com a interação $G \times A$. O comprimento do vetor de ambiente ou genótipo relaciona-se à variância associada ao ambiente e ao genótipo. Assim, os ambientes A5 e A8 e os genótipos G1, G2, G4, G7 e G9 apresentaram as maiores variâncias. Outro ponto de interesse aplicado é a tentativa de identificar aspectos característicos dos genótipos e dos ambientes, determinando os eixos da interação *Dim1* e *Dim2*. Os ambientes A1, A8 e A11, de forma não tão elevada e o genótipo G6 são os que menos contribuíram para a interação $G \times A$ captada pelo eixo *Dim1*. Os ambientes A5, A6, A7, A9 e A10 e os genótipos G2 e G3, são os que menos contribuíram para a interação captada pelo eixo *Dim2*.

Por terem os vetores quase colineares, os ambientes A6 e A7 demonstraram correlação positiva entre eles, porém demonstraram correlação negativa com o genótipo G3, devido a distância extrema entre eles. Os vetores que se apresentaram perpendiculares demonstraram falta de correlação entre eles, ocorrendo esta configuração com os ambiente A5 e A8, os ambientes A7 e A8 e os ambientes A10 e A12, assim como os genótipos G1 e G7.

Figura 13 - Biplot AMMI com os dois primeiros componentes principais correspondente à representação dos nove genótipos (G) e doze ambientes (A) da variável SST

Tabela 17 - Resultados da produção média e os escores para o Biplot da variável SST

Código	Produção Média	Ordem p/médias	Dim 1	Dim 2	Dim 3
<i>G1</i>	10.7469	(9)	1.19753	1.23842	-0.09071
<i>G2</i>	10.8317	(8)	1.29987	-0.13395	-0.49124
<i>G3</i>	11.0469	(7)	0.60904	0.07413	-0.58657
<i>G4</i>	11.1047	(6)	1.08647	-1.05509	0.28398
<i>G5</i>	11.6044	(4)	-0.42642	0.29552	1.18581
<i>G6</i>	11.1808	(5)	0.04071	-0.67686	0.41619
<i>G7</i>	11.7761	(3)	-0.91686	0.72325	0.32656
<i>G8</i>	11.8697	(1)	-1.56822	0.02601	-0.82777
<i>G9</i>	11.8669	(2)	-1.32212	-0.49143	-0.21624
<i>A1</i>	9.6615	(11)	0.16996	-0.31478	0.81888
<i>A2</i>	12.1370	(5)	-0.43673	0.39886	0.29277
<i>A3</i>	11.0781	(7)	-0.15624	-0.35397	0.33681
<i>A4</i>	12.2052	(4)	-0.16114	-0.28924	0.35828
<i>A5</i>	10.0456	(10)	2.82062	-0.03815	-0.44686
<i>A6</i>	13.5652	(1)	-0.54323	-0.07434	0.24354
<i>A7</i>	12.2163	(3)	-0.78055	-0.06665	-0.57346
<i>A8</i>	12.6856	(2)	0.12189	1.69494	0.25181
<i>A9</i>	9.6507	(12)	-0.23713	-0.01798	-0.46490
<i>A10</i>	10.6289	(8)	-0.82070	0.12033	-0.99816
<i>A11</i>	12.0615	(6)	0.13683	-0.36500	0.41775
<i>A12</i>	10.1022	(9)	-0.11358	-0.69402	-0.23646

Na Figura 14, pode-se verificar quais os genótipos que possuem produção inferior ou superior à produção média global através da linha vertical de referência. Os genótipos *G1*, *G2*, *G3*, *G4* e *G6* possuem produção inferior a produção média global que é de 11,34, enquanto que os genótipos *G5*, *G7*, *G8* e *G9* possuem produção superior e os ambientes com pouco rendimento são *A1*, *A3*, *A5*, *A9*, *A10* e *A12* e os ambientes que apresentaram o melhor rendimento são *A2*, *A4*, *A6*, *A7*, *A8*, e *A11*. Os pontos situados em torno de zero em relação ao eixo horizontal *Dim1*, correspondem aos genótipos *G5* e *G6* e ambientes *A1*, *A2*, *A3*, *A4*,

A8, A9, A11 e A12 são os mais estáveis. Os ambientes A5 e A6 são os que possuem SST mais divergentes.

Figura 14 - Biplot AMMI utilizando as médias ajustadas correspondente à representação dos nove genótipos (G) e doze ambientes (A) para a variável SST

O gráfico Trilinear plot desenvolvido por Friendly (2001) representa os três escores do modelo AMMI, conforme os valores da Tabela 17. Nota-se que na Figura 15 os genótipos G2, G4, G8 e G9 e o ambiente A5 não são proporcionais, pois eles estão nos extremos do gráfico para a variável SST.

Na Tabela 18 apresenta os resultados do teste de Gollob, que determina o significado de cada um dos termos do modelo GGE para a variável SST (Gollob, 1968). Neste caso, os dois primeiros termos são significativos ao nível de 1%, sendo que o primeiro contribuiu com 73,26% e o segundo com 10,04% perfazendo 83,30% da variabilidade total. Pode-se verificar que o modelo GGE tem uma maior contribuição na variabilidade total que o modelo AMMI.

A Figura 16, refere-se ao gráfico GGE Biplot correspondente a variável SST,

Figura 15 - Gráfico Trilinear construído pelos três escores do modelo AMMI para a variável SST, representando os nove genótipos (G) e doze ambientes (A)

Tabela 18 - Resultados dos testes de Gollob da análise GGE para a variável SST

S.Q.	%	%Acumulada	g.l.	Q.M	F	Pr>F
348,901	73,2630	73,263	18	19,3834	12,8812	0,00000
47,820	10,0413	83,304	16	2,9887	1,9862	0,01521
29,328	6,1583	89,463	14	2,0948	1,3921	0,15860
17,160	3,6034	93,066	12	1,4300	0,9503	0,49771
10,969	2,3033	95,369	10	1,0969	0,7289	0,69676
10,442	2,1925	97,562	8	1,3052	0,8674	0,54487
7,353	1,5440	99,106	6	1,2255	0,8144	0,55976
4,258	0,8941	100,000	4	1,0645	0,7074	0,58765
0,000	0,0000	100,000	2	0,0000	0,0000	1,00000

pode-se verificar que não há correlação entre os ambientes A5 e A8 por estarem perpendiculares. Os ambientes mais próximos da origem, portanto com escores próximos de zero, são os ambientes A1, A3, A4, A9 e A11 e os genótipos são G3, G5 e G6. Estes ambientes e genótipos com maior proximidade da origem, contribuíram muito pouco com a interação

$G \times A$. O comprimento do vetor de ambiente ou genótipo relaciona-se à variância associada ao ambiente e ao genótipo.

A análise SREG (Sites Regression) por incorporar o efeito de genótipo e por estar na maioria dos casos altamente correlacionado com os escores do primeiro componente principal, possui a vantagem de permitir a avaliação gráfica direta do efeito de genótipo (YAN, 2001).

Figura 16 - GGE Biplot correspondente à representação dos nove genótipos (G) e doze ambientes (A) para a variável SST

4.3 Análise da variável Produção utilizando a técnica MDS

A técnica MDS (Multidimensional Scaling) é uma forma de “rearranjar” os objetos (ou variáveis) de uma maneira eficiente, a detectar as menores dimensões significativas para explicar as similaridades ou dissimilaridades (distâncias) entre eles. Ele realmente se move em torno de objetos no espaço definido pelo número requerido de dimensões, e verifica o quão bem as distâncias entre os objetos podem ser reproduzidos pela nova configuração.

Em termos mais técnicos, ele usa um algoritmo de minimização de função que avalia diferentes formações, com o objetivo de maximizar a bondade de ajuste (ou minimizar “a falta de ajuste”).

A técnica estatística multivariada Escalonamento Multidimensional pode representar as percepções e as semelhanças dos génotipos ou ambientes, essas relações são apresentadas como relações geométricas entre os pontos em um espaço multidimensional de baixa dimensão. Para a análise da variável produção foi utilizada a matriz de médias preditivas. Inicialmente, os dados foram transformados, calculando o $(\ln[\mathbf{Y}])$ e depois fez-se a centralização por colunas, conforme o Anexo B. Esta transformação foi utilizada para diminuir o valor do *Stress*, que ficou em 0,3728, indicando um bom ajuste.

Na Tabela 19 encontra-se a matriz de medidas de dissimilaridade entre os 12 genótipos da variável produção da matriz de dados transformados conforme o Anexo B. Para variáveis contínuas, a medida de dissimilaridade mais utilizada é a distância Euclidiana entre dois objetos, formando-se então uma matriz de distâncias. Pode-se verificar que os menores valores indica os ambientes mais parecidos, enquanto os valores maiores indica diferença entre os ambientes. Por exemplo, os ambientes A2 e A4; A1 e A7 possuem os menores valores entre eles, indica que são quase semelhantes em relação a produção.

Tabela 19 - Matriz de dissimilaridade entre os 12 ambientes para a variável Produção com os dados transformados

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
A1	0,0000	0,5562	1,4126	0,5600	1,3711	1,1412	0,3311	0,5363	0,4487	0,6356	1,1436	0,8191
A2	0,5562	0,0000	0,9626	0,2514	0,9810	0,7344	0,5463	0,8531	0,5195	0,3390	0,7244	1,1211
A3	1,4126	0,9626	0,0000	1,0137	0,5526	0,4503	1,4242	1,7685	1,3806	0,9178	0,4713	1,9912
A4	0,5600	0,2514	1,0137	0,0000	0,9897	0,7606	0,4671	0,8617	0,6499	0,4409	0,8200	1,1407
A5	1,3711	0,9810	0,5526	0,9897	0,0000	0,5894	1,3289	1,7331	1,3467	0,8330	0,7245	1,8972
A6	1,1412	0,7344	0,4503	0,7606	0,5894	0,0000	1,1518	1,5417	1,1281	0,6146	0,6336	1,7367
A7	0,3311	0,5463	1,4242	0,4671	1,3289	1,1518	0,0000	0,4978	0,5361	0,6709	1,1756	0,7283
A8	0,5363	0,8531	1,7685	0,8617	1,7331	1,5417	0,4978	0,0000	0,6411	1,0280	1,4408	0,4684
A9	0,4487	0,5195	1,3806	0,6499	1,3467	1,1281	0,5361	0,6411	0,0000	0,5865	1,0969	0,8404
A10	0,6356	0,3390	0,9178	0,4409	0,8330	0,6146	0,6709	1,0280	0,5865	0,0000	0,7682	1,2168
A11	1,1436	0,7244	0,4713	0,8200	0,7245	0,6336	1,1756	1,4408	1,0969	0,7682	0,0000	1,6941
A12	0,8191	1,1211	1,9912	1,1407	1,8972	1,7367	0,7283	0,4684	0,8404	1,2168	1,6941	0,0000

Na Tabela 20 encontra-se os valores para as posições das dimensões (Dim 1 e

Dim 2) onde serão plotados os 9 genótipos e os 12 ambientes da variável produção.

Tabela 20 - Valores das dimensões onde serão plotados os 9 genótipos e os 12 ambientes para a variável produção

	Dim 1	Dim 2
G1	0.2576	-0.1360
G2	0.2285	-0.1223
G3	0.3243	-0.3496
G4	0.3762	0.1434
G5	0.2594	-0.2962
G6	0.3013	-0.0567
G7	0.3932	0.2830
G8	0.3413	0.0564
G9	0.3736	0.1784
A1	0.4310	-0.2582
A2	-0.0072	-0.0256
A3	-0.9618	-0.1283
A4	0.0121	-0.2047
A5	-0.8477	0.4150
A6	-0.6690	0.0908
A7	0.4626	-0.0686
A8	0.8536	-0.1531
A9	0.3801	0.3049
A10	-0.1130	0.2098
A11	-0.6207	-0.3937
A12	1.0800	0.2117

No Anexo B, encontra-se os valores da medida de *Stress*. É uma medida de eficiência do ajuste da curva (goodness-of-fit) mais usada para avaliar se a proposta de arranjo dos dados reproduz eficientemente a matriz de distância observada denominada de *Raw Stress*. Quanto menor o valor, melhor é o ajuste da matriz de dados reproduzidas a matriz de dados observadas.

O Diagrama de Shepard, é uma forma de analisar a precisão do ajuste e detectar possíveis anomalias na representação multidimensional. Neste diagrama, os coeficientes e as distâncias são reproduzidas em pares. Um ajuste perfeito é indicado através de uma

diagonal no diagrama, que reflete uma função crescente ou decrescentes sem interrupções ou descontinuidades. O Diagrama de Shepard é obtido plotando-se as distâncias reproduzidas contra as distâncias observadas.

Na Figura 17 cada círculo do Diagrama de Shepard representa um par de amostras. O eixo horizontal indica a dissimilaridade inter-amostra e o eixo vertical indica a distância correspondente inter-ponto. Os pontos azuis na linha amarela, indica disparidades. Quanto mais pontos na linha, melhor o ajuste global. Os cinco piores pares de ponto de ajuste ou ambientes são identificados no diagrama, neste caso para a variável produção são: A5 e A6; A4 e A11; A7 e A9; A6 e A11; A1 e A7.

Figura 17 - Diagrama de Shepard para a variável produção

Na Figura 18, refere-se ao Biplot com Escalonamento Multidimensional dos nove genótipos e dos doze ambientes. Na representação, as distâncias entre os pontos estão diretamente relacionadas com dissimilaridades entre os ambientes. Os ambientes A2 e A4; A1 e A7; A2 e A10 possuem os menores valores, portanto são mais semelhantes entre eles, do que

em relação aos outros ambientes, conforme a Tabela 19 de medidas de dissimilaridades. Os ambientes A3 e A5 em relação ao ambiente A12 são os mais distantes, portanto apresentam maior diferença entre eles em relação a produção.

4.4 Análise da variável SST utilizando a técnica MDS

Inicialmente, os dados foram transformados, calculando o $(\ln[\mathbf{Y}])$ e depois fez-se a centralização por colunas, conforme o Anexo B. Esta transformação foi utilizada para diminuir o valor do *Stress*, que ficou em 0.1722. Na análise da variável SST, utilizando a técnica MDS, a Tabela 21 encontra-se a matriz de medidas de dissimilaridade entre os 12 ambientes para a variável SST da matriz de dados transformados conforme Anexo B. Pode-se verificar que os menores valores indica os ambientes mais parecidos, enquanto os valores maiores indica diferença entre os ambientes. Por exemplo, os ambientes A2, A4, A7 e A11 possuem os menores valores entre eles, indica que são quase semelhantes em relação a produção de teor de sólidos solúveis (SST).

Tabela 21 - Matriz de dissimilaridade entre os 12 ambientes para a variável SST

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
A1	0,0000	0,7361	0,4573	0,7206	0,8891	1,0503	0,7698	0,8922	0,3213	0,4974	0,6845	0,3463
A2	0,7361	0,0000	0,3699	0,1704	1,2114	0,3867	0,2024	0,3067	0,7211	0,4963	0,2412	0,6401
A3	0,4573	0,3699	0,0000	0,3593	1,0396	0,6433	0,3746	0,5507	0,4736	0,3706	0,3117	0,4102
A4	0,7206	0,1704	0,3593	0,0000	1,1610	0,3751	0,2258	0,3531	0,7453	0,5361	0,1398	0,6168
A5	0,8891	1,2114	1,0396	1,1610	0,0000	1,4227	1,2823	1,2036	1,0041	1,1741	1,0839	0,9847
A6	1,0503	0,3867	0,6433	0,3751	1,4227	0,0000	0,3889	0,4045	1,0441	0,7981	0,3974	0,9381
A7	0,7698	0,2024	0,3746	0,2258	1,2823	0,3889	0,0000	0,4074	0,7339	0,4439	0,2948	0,6391
A8	0,8922	0,3067	0,5507	0,3531	1,2036	0,4045	0,4074	0,0000	0,8864	0,6887	0,3607	0,8186
A9	0,3213	0,7211	0,4736	0,7453	1,0041	1,0441	0,7339	0,8864	0,0000	0,3741	0,7129	0,3091
A10	0,4974	0,4963	0,3706	0,5361	1,1741	0,7981	0,4439	0,6887	0,3741	0,0000	0,5498	0,3893
A11	0,6845	0,2412	0,3117	0,1398	1,0839	0,3974	0,2948	0,3607	0,7129	0,5498	0,0000	0,6009
A12	0,3463	0,6401	0,4102	0,6168	0,9847	0,9381	0,6391	0,8186	0,3091	0,3893	0,6009	0,0000

Na Figura 19, o eixo horizontal indica a dissimilaridade inter-amostra e o eixo vertical indica a distância correspondente inter-ponto. Os cinco piores pares do ponto de ajuste ou ambientes que são identificados no diagrama de Shepard, neste caso para a variável SST tem-se: A9 e A12; A1 e A12; A8 e A11; A1 e A5; A3 e A12.

Na Tabela 22 encontra-se os valores das coordenadas (Dim 1 e Dim 2) onde

Figura 18 - Biplot com Escalonamento Multidimensional dos nove genótipos (G) e doze ambientes (A) para a variável produção

Figura 19 - Diagrama de Shepard para a variável SST

serão plotados os 9 genótipos e os 12 ambientes para a variável SST conforme a Tabela 9.

Tabela 22 - Valores das dimensões onde serão plotados os 9 genótipos e os 12 ambientes para a variável SST

	Dim 1	Dim 2
G1	-0,1450	0,4913
G2	-0,1149	0,4636
G3	-0,1454	0,2517
G4	-0,0578	0,3351
G5	-0,3705	0,1420
G6	-0,2368	0,1372
G7	-0,4535	-0,0023
G8	-0,5068	-0,2999
G9	-0,4681	-0,2072
A1	0,4584	-0,1315
A2	-0,2908	0,0105
A3	0,0341	-0,0457
A4	-0,2367	0,0881
A5	0,8143	0,5656
A6	-0,5897	0,1458
A7	-0,3285	-0,1083
A8	-0,3145	0,3381
A9	0,3321	-0,3979
A10	-0,0084	-0,3726
A11	-0,1510	0,1458
A12	0,2809	-0,2379

Na Figura 20, refere-se ao Biplot com Escalonamento Multidimensional dos nove genótipos e dos doze ambientes para a variável SST. Na representação, as distâncias entre os pontos estão diretamente relacionadas com dissimilaridades entre os ambientes. Os ambientes A4 e A11; A2 e A4, portanto são mais semelhantes entre eles, do que em relação aos outros ambientes, conforme a Tabela 22 de medidas de dissimilaridades. Os ambientes A5 e A6 são os mais distantes, portanto apresentam maior diferença entre eles.

Figura 20 - Biplot com Escalonamento Multidimensional dos nove genótipos (G) e doze ambientes (A) para a variável SST

5 CONCLUSÕES

Os resultados obtidos nessa pesquisa permitem extrair as seguintes conclusões:

- a) A metodologia de MDS para os ambientes, mostrou-se eficiente em verificar os locais pela similaridade do desempenho dos cultivares, como técnica de redução dimensional se tentou preservar ao máximo as distâncias originais. O Biplot construído representou cada objeto por um ponto (ambiente) e cada variável um vetor (genótipo), permitindo dessa maneira avaliar: as semelhanças dentro do conjunto de ambientes; as semelhanças dentro do conjunto de genótipos e as relações entre os ambientes e genótipos;
- b) A análise AMMI e SREG explicam bem a interação $G \times A$, possibilitando uma fácil interpretação gráfica dos resultados das análises estatísticas e identificam de forma segura os melhores genótipos a serem cultivados nos ambientes avaliados;
- c) A utilização de novos métodos multivariados em experimentos agrônômicos, como a MDS, mostrou-se adequada, fornecendo resultados coerentes.

PROPOSTAS FUTURAS DE PESQUISAS

Como continuidade deste estudo, seria interessante a realização de pesquisas que:

- a) implemente as estruturas *three-way* utilizando o modelo PARAFAC (Parallel Factor Analysis);
- b) aplique a metodologia multivariada do Biplot nas áreas de Controle Estatístico de Qualidade e Análise Bayesiana;
- c) utilize o Biplot em novas técnicas como por exemplo, os métodos de inteligência artificial (Redes Neurais Artificiais, Algoritmos Genéticos e Lógica Difusa) que são baseados no desenvolvimento de sistemas computacionais inteligentes e poderão ampliar ainda mais este campo de pesquisa.

REFERÊNCIAS

- AMARO, R. I.; VICENTE-VILLARDÓN, J. L.; GALINDO-VILLARDÓN, M. P.; Manova Biplot para arreglos de tratamiento con dos factores basado en modelos lineales generales multivariantes. **Interciencia**, Caracas, n. 001, v. 29, p. 26–32, 2004.
- ANNICCHIARICO, P.; Cultivar adaptation and recommendation from alfalfa trials in Northern Italy. **Journal of Genetics and Breeding**, New Jersey, v. 46, p. 269–278, 1992.
- BORG, I.; GROENEN, P. J. F.; **Modern Multidimensional Scaling Theory and Applications**. New York: Springer Series in Statistics, 2005. 614 p.
- BRADU, D.; GABRIEL K. R.; The biplot as a diagnostic tool for model of two-way tables. **Technometrics**, Alexandria, v. 20, p. 47–68, 1978.
- CANOCO. **Plant Research International**. Version 4.5. Disponível em: <<http://www.pri.wur.nl/uk/products/canoco>> (2002). Acesso em: 13 maio 2009.
- CÁRDENAS, O.; GALINDO, P.; VICENTE-VILLARDÓN, J. L.; Estimación Máximo Verosímil en aproximaciones Biplots. **Revista Venezolana de Análisis de Coyuntura**, Caracas, n. 002, v. 07, p. 333–347, 2001.
- CÁRDENAS, O.; NOGUERA, C.; GALINDO, P.; VICENTE-VILLARDÓN, J. L.; El uso de información externa en aproximaciones Biplots. **Revista Venezolana de Análisis de Coyuntura**, Caracas, n. 002, v. 09, p. 257–276, 2003.
- CÁRDENAS, O.; NOGUERA, C.; GALINDO, P.; VICENTE-VILLARDÓN, J. L.; Alternativa a la regresión con componentes principales basado en Biplot de regresión. **Interciencia**, Caracas, n. 003, v. 31, p. 160–167, 2006.
- CÁRDENAS, O.; GALINDO, P.; VICENTE-VILLARDÓN, J. L.; Los Métodos Biplot: evolución y aplicaciones. **Revista Venezolana de Análisis de Coyuntura**, Caracas, n. 001, v. 13, p. 279–303, 2007.
- CHAVES, L. J.; VENCOSKY, R.; GERALDI, I. O.; Modelo não-linear aplicado ao estudo da interação genótipo x ambiente em milho. **Pesquisa Agropecuária Brasileira**, Brasília v. 24, n. 2, p. 259–268, 1989.
- CORNELIUS, P. L.; CROSSA, J.; SEYEDSADR, M. S.; Statistical test and estimator of multiplicative model for genotype-by-environment interaction. In: KANG, M.S.; GAUCH Jr, H. G. (Ed.). **Genotype-by-Environment Interaction**. New York: Boca raton, 1996. p. 199–234.
- COX, T. F.; COX, M. A. A.; **Multidimensional Scaling**. London: Chapman & Hall, 2001. 332 p.
- COX, C.; GABRIEL, K. R.; Some comparisons of biplot and pencil-and-paper E.D.A. methods. In: LAUNER, R. L.; SIEGEL A. F. (Ed.). **Modern data analysis**. London: Academic Press, 1982. p. 45–82.
- CUADRAS, C.M.; **Models Estadísticos Multivariantes**. Disponível em: <<http://www.ub.edu/stat/personal/cuadras/cuad2.html>> Acesso em: 8 out. 2009.

CRUZ, C. D.; TORRES, R. A. de.; VENCovsky, R.; Na alternative approach to the stability analysis proposed by Silva e Barreto. **Revista Brasileira de Genética**, Ribeirão Preto, v. 12, p. 567-580, 1989.

DIAS, C. T. S.; KRZANOWSKI, W. J.; Model Selection and Cross Validation in Additive Main Effect and Multilicative Interaction Models. **Crop Science**, Madison, v.43, p.865-873, 2003.

DIAS, C.T. S.; **Métodos para escolha de componentes em modelo de efeito principal aditivo e interação multiplicativa (AMMI)**. 2005. 80 p. Tese (Livre-docência) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba. 2005.

DUARTE, J.B.; VENCovsky, R.; **Interação genótipos $\times$ ambientes - uma introdução à análise "AMMI"**. Ribeirão Preto: Sociedade Brasileira de Genética, 1999, 60p.

EBERHART, S. A.; RUSSELL, W. A.; Stability parameters for comparing varieties. **Crop Science**, Madison, v. 6, p. 36-40, 1966.

ESCOFIER, B.; PAGÈS, J.; **Anályses factorielles simples et multiples: objectifs, méthodes et interprétati**. Paris: Dunod, 1990, 284 p.

FINLAY, K. W.; WILKINSON, G. M.; The analisys of adaptation in a Plant-Breeding Programme. **Australian Journal of Agricultural Research**, Victoria, v. 14, n. 5, p. 742-754, 1963.

FREITAS, J. G.; CRISÓSTOMO, J. R.; SILVA, F. P.; PITOMBEIRA, J. B.; TÁVORA, J. A. F.; Interação entre genótipo e ambiente em híbridos de melão Amarelo no Nordeste do Brasil. **Revista Ciência Agrônômica**, Fortaleza, v. 38, n. 2, p. 176-181, 2007

FRIENDLY, M.; **Visualizing categorical data**. Cary, SAS Publishing, 2001. 436 p.

FOX, P. N.; SKOVMAND, B. K.; THOMPSON, H. J.; BRAUN, H. J.; CORMIER, R.; Yield and adaptation of hexaploid spring triticales. **Euphytica**, New York, v. 47, n. 1, p. 57-64, 1990.

GABRIEL, K. R.; The biplot graphic display of matrices with application to principal component analysis. **Biometrika**, Great Britain, v. 58, n. 03, p. 453-467, 1971.

GABRIEL, K.R.; Analysis of meteorological data by means of canonical decomposition and Biplots. **Journal of Applied Meteorology**, Washington, v. 11, n. 7, p. 1071-1077, 1972.

GABRIEL, K. R.; ZAMIR, S.; Lower rank approximation of matrices by least squares with any choice of weights. **Technometrics**, Alexandria, v. 21, n. 04, p. 489-498, 1979.

GABRIEL, K. R.; Biplot display of multivariate matrices for inspection of data and diagnosis. In: BARNETT, V. (Ed.). **Interpreting Multivariate Data**, London: Wiley, 1981. p. 147-173.

GABRIEL, K. R.; Biplot display of multivariate categorical data, with Comments on Multiple Correspondence Analysis. In: KRZANOWSKI, W. J. (Ed.). **Recent Advances in Descriptive Multivariate Analysis**, Oxford: Clarendon Press, 1995. p. 190-226.

GABRIEL, K. R.; MANOVA Biplots for two-way contingency tables. In: KRZANOWSKI, W. J. (Ed.). **Recent Advances in Descriptive Multivariate Analysis**, Oxford: Clarendon Press, 1995. p. 227-268.

GALINDO, M. P.; Una alternativa de representación simultánea: HJ-Biplot. **Questão**, Catalunya, n. 01, p. 13–23, 1986.

GGE BILOT. **Beta Tester**. Version 5.2 (2006). Disponível em: <<http://www.ggebiplot.com>>. Acesso em: 19 jun. 2009.

GOLLOB, H. F.; A statistical model which combines features of factor analytic and analysis of variance. **Psychometrika**, Greensboro, v. 33, p. 73–115, 1968.

GOOD, I.J.; Some applications of the singular value decomposition of a matrix. **Technometrics**, Alexandria, v. 11, p. 823–831, 1969.

GOWER, J. C.; A General Coefficient of Similarity and some of its Properties **Biometrics**, Great Britain, n. 27, p. 857–871, 1971.

GOWER, J. C.; HARDING S.; Nonlinear biplots **Biometrika**, Great Britain, n. 75, p. 445–455, 1988.

GOWER, J. C.; Generalized Biplots. **Biometrika**, Great Britain, n. 79, p. 475–493, 1992.

GOWER, J. C.; HAND, D. J.; **Biplots**. London: Chapman & Hall, 1996. 277 p.

GOWER, J. C.; KRZANOWSKI, W. J.; Analysis of distance for structured multivariate data and extensions to multivariate analysis of variance **Applied Statistics**, London, n. 4, p. 505–519, 1999.

GOWER, J. C.; Unified Biplot Geometry. In: FERLIGOJ, A.; MRVAR, A. (Ed.). **Developments in Applied Statistics**. Ljubljana: FDV, 2003. p. 04–22.

GREENACRE, M. J.; Biplots in correspondence analysis. **Journal Applied Statistics**, London, v. 20, n. 02, p. 251–269, 1993.

GREENACRE, M. J.; **Theory and Applications of Correspondence Analysis**. London: Academic Press, 1984. 364 p.

GUTTMAN, L.; A general nonmetric technique for finding the smallest coordinate space for a configuration of points. **Psychometrika**, Greensboro, v. 33, p. 469–506, 1968.

HUEHN, M.; Nonparametric measures of phenotypic stability. Part 1 and 2: Theory and Applications. **Euphytica**, Wageningen, v. 47, n. 3, p. 189–194; 195–201, 1990.

JMP. **SAS Institute, Inc.** Version 7.0 (2007). Disponível em: <<http://www.jmp.com>>. Acesso em: 21 out. 2009.

JOHNSON, R. A.; WICHERN, D. W.; **Applied Multivariate Statistical Analysis**. New Jersey: Prentice-Hall, Fifth Edition, 2002. 767 p.

KROONENBERG, P. M.; **Introduction to Biplots for GxE Tables**., Disponível em: <<http://www.fsw.leidenuniv.nl/kroonenb/document/biplot.zip>>. Acesso em: 17 maio 2007.

KRZANOWSKI, W. J.; MARRIOTT, F. H. C.; **Multivariate Analysis**. Part 1: Distributions, Ordination and Inference. London: Kendall's Library of Statistics, 1994. 300p.

KRZANOWSKI, W. J.; **Principles of Multivariate Analysis: A User's Perspective**., Oxford: Oxford University Press, 2000. 563 p.

KRZANOWSKI, W. J.; Multifactorial analisys of distance in studies of ecological community structure. **Journal of Agricultural, Biological and Environmental Statistics**, Alexandria, v. 7, p. 222-232, 2002.

KRZANOWSKI, W. J.; Biplots for Multifactorial Analisys of Distance. **Biometrics**, Washington, v. 60, p. 517-524, 2004.

KRUSKAL J. B.; Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. **Psychometrika**, Greensboro, v. 29, p. 1-27, 1964.

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

Le ROUX, N. J.; GARDNER S.; OLIVIER, P.; Biplots for displaying multidimensional financial performance data graphically. **SA Journal Accounting Resseach**, Kengray, v. 17, n. 1, p. 41-64, 2003.

Le ROUX, N. J.; GARDNER S.; Analysing Your Multivariate Data as a Pictorial: A Case for Applying Biplot Methodology? **International Statistical Review**, Leidschenveen, v. 73, n. 3, p. 365-387, 2005.

LIN, C. S.; BINNS, M. R.; A superiority measure of cultivar performance for cultivar x location data. **Canadian Journal of Plant Science**, Ottawa, v. 68, p. 193-198, 1988.

LIPKOVICH, I.; SMITH E.; Biplot and Singular Value Decomposition Macros for Excel. **Journal of Statistical Software**, Los Angeles, v. 7, n. 5, p. 1-15, 2002.

MAGARI, R.; KANG, M. S.; SAS-STABLE: Stability analysis of balanced and unbalanced data. **Agronomy Journal**, Stanford, v. 89, p. 929-932, 1997.

MANDEL, J.; A new analysis of variance model for non-additive data. **Technometrics**, Alexandria, v. 13, n. 1, p. 1-18, 1971.

MANLY, B. J.; **Métodos Estatísticos Multivariados**. Porto Alegre: Bookman, 2008. 229 p.

MARDIA, K. V.; BIBBY, J. M.; **Multivariate Analysys**, Londres: Academic Press, 1979. 521 p.

MARTÍN-RODRÍGUEZ, J.; GALINDO, P.; VICENTE-VILLARDÓN, J. L.; Comparison and integration of subspaces from a biplot perspective. **Journal of Statistical Planning and Inference**, Maryland Heights, v. 2, n. 102, p. 411-423, 2002.

MINITAB. **Minitab Inc.** Version 15 (2007). Disponível em: <<http://www.minitab.com>>. Acesso em: 21 out. 2009.

MORRISON D. F.; **Multivariate Statistical Methods**, Londres: McGraw-Hill. Second Edition, 1978. 415 p.

NELDER, J. A.; WEDDERBUM R. W.; Generalized Linear Models. **Journal of the Royal Statistical Society**, London, v. 135, p. 370-384, 1972.

PERKINS, J. M.; JINKS, J. L.; Environmental and genotype-environmental components of variability. IV. Non-linear interactions for multiple inbred lines. **Heredity**, Edinburgh, v. 23, p. 525-35, 1968.

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

PLAISTED, R. L.; PETERSON, L. C.; A technique for evaluating the ability of selections to yield consistently in different locations or seasons. **American Potato Journal**, New York, v. 36, n. 11, p. 381-385, 1959.

R: Language and Environment for Statistical Computing. **The R Foundation for Statistical Computing**. R Development Core Team. Vienna, Austria.(2008) ISBN 3-900051-07-0. Disponível em: <<http://www.R-project.org>>. Acesso em: 19 dez. 2009.

RAMÍREZ, G.; VASQUEZ, M.; CAMARDIEL, A.; PÉREZ, B.; GALINDO, P.; Detección grafica de la multicoliearidad mediante el h-plot de la inversa de la matriz de correlaciones. **Revista Colombiana de Estadística**, Bogotá, v. 28, n. 2, p. 207-219, 2005.

RENCHE, A. C.; **Methods of Multivariate Analysis**. 2nd ed. New York: John Wiley & Sons, 2002. 738 p.

ROSSE, L. N.; VENCOSKY, R.; Modelo de regressão não-linear aplicado ao estudo da estabilidade fenotípica de genótipos de feijão no Estado do Paraná. **Bragantia**, Campinas, v. 59, n. 1, p. 99-107, 2000.

SAS. **SAS Institute Inc.** (2008) Disponível em: <<http://www.sas.com>>. Acesso em: 14 out. 2009.

SHAFII, B; PRICE W. J.; Analysis of Genotype-by-Environment Interaction Using the Additive Main Effects and Multiplicative Interaction Model and Stability Estimates. **Journal of Agricultural, Biological, and Environmental Statistics**, Alexandria, v. 3, n. 3, p. 335-345, 1998.

SILVA, J. G. C.; BARRETO, J. N.; An application of segmented linear regression to the study of genotypes environment interaction. **Biometrics**, Texas, v. 41, n. 4, p. 1093, 1986.

SHEPARD, R. N.; The analysis of proximities multidimensional scaling with an unknown distance function. **Psychometrika**, Greensboro, v. 27, p. 125-149, 1962.

SILVA, J. G. C.; Análise de adaptabilidade por regressão segmentada com estimação da junção dos segmentos. **Pesquisa Agropecuária Brasileira**, Brasília, v. 33, n. 7, p. 1013-1029, 1998.

SILVA, L. A.; INNECCO, R.; COSTA, J. T. A.; MELO, F. I. O.; MALUF, W. R.; PEDROSA, J. F.; Estudo de aspectos quantitativos e qualitativos de frutos de genótipos de melão. **Revista Ciência Agronômica**, Fortaleza, v.36, n.3, p. 310-315, 2005.

STATISTICA. **StatSoft Inc.** Version 8 (2007). Disponível em: <<http://www.statsoft.com>>. Acesso em: 14 out. 2009.

SHUKLA, G. K.; Some statistical aspects of partitioning genotype-environment components of variability. **Heredity**, Edinburgh, v. 29, n. 2, p. 237-45, 1972.

STORCK, L.; VENCovsky, R.; Stability analysis based on a bi-segmented discontinuous model with measurement errors in the variables. **Revista Brasileira de Genética**, Ribeirão Preto, v. 17, n. 1, p. 75-81, 1994.

TAI, G. C. C.; Genotypic stability analysis and its application to potato trials. **Crop Science**, Madison, v. 11, n. 2, p. 184-90, 1971.

TIMM, N. H.; **Applied multivariate analysis**. New York, Springer-Verlag, 2002. 718 p.

TOLER, J. E.; BURROWS, P. M.; Genotypic performance over environmental arrays: a non-linear grouping protocol. **Journal of Applied Statistics**, Abingdon, v. 25, n. 1, p. 131-143, 1998.

TORGERSON, W. S.; Multidimensional scaling: I. Theory and method. **Psychometrika**, Greensboro, v.17, p.401-419, 1952.

VARAS, M. J.; VICENTE-TAVERA S.; MOLINA E.; VICENTE-VILLARDÓN J. L.; Role of canonical biplot method in the study of building stones: an example from Spanish monumental heritage. **Environmetrics**, Malden, v. 16, p. 405-419, 2005.

VAIRINHOS, V. M.; GALINDO M. P.; Biplots PMD - Data Mining Centrada em Biplots, apresentação de um protótipo. **Associação Portuguesa de Classificação e Análise de Dados**, Lisboa, p.01-18, 2004.

VARELA, M.; **Los métodos Biplot como herramieta de análisis de interacción de orden superior en un modelo lineal/bilineal**. Tesis Doctoral, USAL, Salamanca, España, 2002.

VERMA, M. M.; CHAHAL, G. S.; MURTY, B. R.; Limitations of conventional regression analysis, a proposed modification. **Theoretical and Applied Genetics**, Berlin, v. 53, n. 1, p. 89-91, 1978.

VIANA, S. S.; SILVA, L. M. R.; LIMA, P. V. P. S.; LEITE, L. A. S.; Competitividade do Ceará no mercado internacional de frutas: o caso do melão. **Revista Ciência Agronômica**, Fortaleza, v.37, n.1, p.25-31, 2006.

VICENTE-VILLARDÓN, J. L.; **Una alternativa a las técnicas factoriales basada en una generalización de los métodos Biplot**. Tesis Doctoral, USAL, Salamanca, España, 1992.

VICENTE-VILLARDÓN, J. L.; GALINDO-VILLARDÓN, M. P.; BLÁSQUEZ-ZABALLOS, A.; Logistic Biplots. In: GREENACRE, M.; BLASIUS, J. (Ed.). **Multiple Correspondence Analysis and related methods**, London: Chapman & Hall, 2006. p. 491-509.

WRICKE, G.; Zur berechnung der okovalenz bei sommerweizen und hafer. **Pflanzenzucht**, Berlin, v. 52, p. 127-138, 1965.

YAN, W.; GGE Biplot. A windows application for graphical analysis of multienvironment trial data and other types two-way data. **Agronomy Journal**, Stanford, v. 93, p. 1111-1118, 2001.

YAN, W.; KANG, M. S.; **GGE Biplot Analysis. A Graphical Tool for Breeders, Geneticists and Agronomists**. Florida: CRC Press, 2002. 271 p.

YOUNG, F. W.; HAMMER R. M.; **Multidimensional Scaling: History, theory and applications**. New Jersey: Lawrence Erlbaum Associates, 1987. 318 p.

ANEXOS

ANEXO A - Programa em SAS para o modelo AMMI da variável Produção

```

*****;
* PROGRAMA UTILIZADO PARA O MODELO AMMI;
*****;
options ps = 5000 ls=78 nodate;
data MELAO;
input AMB $ GEN $ BLOC PROD SST;
cards;
A1 G1 1 33.91 7.85
A1 G1 2 26.31 10.14
A1 G1 3 31.73 8.09
A1 G2 1 29.42 10.61
A1 G2 2 29.87 6.68
A1 G2 3 28.74 9.19
A1 G3 1 23.02 9.13
      :
      :
A12 G6 1 44.07 9.45
A12 G6 2 29.31 12.26
A12 G6 3 47.04 12.12
A12 G7 1 31.84 11.24
A12 G7 2 57.95 10.78
A12 G7 3 52.36 11.04
A12 G8 1 52.61 11.10
A12 G8 2 50.51 10.59
A12 G8 3 38.71 9.67
A12 G9 1 50.28 10.40
A12 G9 2 52.77 11.38
A12 G9 3 44.72 11.50
;
* análise da variável resposta PROD;
proc glm data=MELAO outstat=SAIDA1;
class BLOC AMB GEN;
model PROD=BLOC(AMB) AMB GEN AMB*GEN/solution ss3;
run;
* Análisis de Varianza utilizando todos os datos';
proc glm data=MELAO outstat=SAIDA2;
class BLOC AMB GEN;
model PROD = BLOC AMB GEN AMB*GEN / ss4;
run;
* Calcula o MSE e DFE;
DATA SAIDA3;
SET SAIDA2;
IF _SOURCE_ = 'ERROR';
SSE=SS;
DFE=DF;
MSE=SSE/DFE;
KEEP DFE MSE;
DATA SAIDA4;
SET SAIDA2;
IF _SOURCE_ = 'BLOC';
NREP=DF+1;

```



```

* Calcula os resíduos através das médias dos GxA;
PROC SORT DATA=MELAO;
  BY AMB GEN;
PROC MEANS DATA=MELAO ;
  BY AMB GEN;
  VAR PRODM;
  OUTPUT OUT = MEDIAS MEAN=PRODM;
PROC PRINT DATA=MEDIAS;
RUN;
* Para calcular os resíduos da interação (GA), modifica-se o modelo para calcular os;
* resíduos incluindo o efeito genótipos (G+GA);
PROC GLM DATA=MEDIAS;
  CLASS AMB GEN;
  MODEL PRODM = AMB GEN / SS4 ;
  OUTPUT OUT=SAIRES R=RESID;
DATA SAIRES;
  SET SAIRES;
PROC PRINT DATA=SAIRES;
RUN;
PROC SORT DATA=SAIRES;
  BY GEN AMB;
PROC TRANSPOSE DATA=SAIRES OUT=SAIRES2;
  BY GEN;
  ID AMB;
  VAR RESID;
PROC PRINT DATA=SAIRES2;
RUN;
* Calcula a DVS usando o proc IML para calcular o AMMI;
PROC IML;
USE SAIRES2;
READ ALL INTO RESID;
NGEN=NROW(RESID);
NAMB=NCOL(RESID);
USE SAIDA3;
READ VAR {MSE} INTO MSEM;
READ VAR {DFE} INTO DFEM;
USE SAIDA4;
READ VAR {NREP} INTO NREP;
CALL SVD (U,L,V,RESID);      * Calcula DVS para o G e A;
MINIMO=MIN(NGEN,NAMB);
L=L[1:MINIMO,];
SS=(L##2)*NREP;              * Soma de quadrados para os termos AMMI;
SUMA=SUM(SS);
PORCENT=((1/SUMA)#SS)*100;    * Calcula a proporção da variabilidade e os g.l. para o modelo AMMI;
MINIMO=MIN(NGEN,NAMB);
PORCENTA=0;
  DO I = 1 TO MINIMO;
    DF=(NGEN-1)+(NAMB-1)-(2*I-1);
    DFA=DFA/DF;
    PORCEACU=PORCENT[I,];
    PORCENTA=PORCENTA+PORCEACU;
    PORCENAC=PORCENAC//PORCENTA;
  END;
DFE=J(MINIMO,1,DFEM);

```

```

MSE=J(MINIMO,1,MSEM);
SSDF=SS||PORCENT||PORCENAC||DFA||DFE||MSE;
L12=L##0.5;                                * Calcula os escores para os genótipos e ;
                                           * ambientes usando raiz quadrada dos autovalores;

SCOREG1=U[,1]#L12[1,];
SCOREG2=U[,2]#L12[2,];
SCOREG3=U[,3]#L12[3,];
SCOREE1=V[,1]#L12[1,];
SCOREE2=V[,2]#L12[2,];
SCOREE3=V[,3]#L12[3,];
SCOREG=SCOREG1||SCOREG2||SCOREG3;
SCOREE=SCOREE1||SCOREE2||SCOREE3;
SCORES=SCOREG//SCOREE;
CREATE SUMAS FROM SSDF;                    * cria arquivos que serão utilizados no Biplot;
APPEND FROM SSDF;
CLOSE SUMAS;
CREATE SCORES FROM SCORES;
APPEND FROM SCORES ;
CLOSE SCORES;
* cria o polígono;
d1=scoreg[,1:2][cvexhull(scoreg[,1:2])[loc(cvexhull(scoreg[,1:2])>0),],] ;
d=d1//d1[1,];
xxx=J(nrow(d)-1,1,0);
yyy=J(nrow(d)-1,1,0);
ppp={0 1,1 0};
do i=1 to nrow(d)-1 ;
    dd=d[i:i+1,];
    if dd[1,1]>dd[2,1] then ddd=ppp*dd;
    else ddd=dd;
    p=(ddd[2,2]-ddd[1,2])/(ddd[2,1]-ddd[1,1]) ;
    if p<0 then ss=1 ;
    else ss=-1 ;
    r=tan((180-90-abs(atan(p))*180/3.14156))*3.14156/180)*ss ;
    aa=(ddd[1,2]+ddd[2,2])/2-p*(ddd[1,1]+ddd[2,1])/2;
    xx=aa/(r-p) ;
if abs(r)<1 then xxx[i,]=40;
else xxx[i,]=40/abs(r);
if xx<0 then xxx[i,]=-xxx[i,] ;
else xxx[i,]=xxx[i,];
yyy[i,]=xxx[i,]*r;
end;
kk=xxx||yyy;
xx1={V1 V2};
create pol from d[colNAME=xx1];
append from d ;
close pol;
xx2={V3 V4};
create perp from kk[colNAME=xx2];
append from kk ;
close perp;
data pol; set pol; TYPE="pol";
data perp; set perp; TYPE="per";
* Verifica a significancia do modelo AMMI;
DATA SSAMMI;

```

```

SET SUMAS;
SSAMMI =COL1;
PORCENT =COL2;
PORCENAC=COL3;
DFAMMI =COL4;
DFE =COL5;
MSE =COL6;
DROP COL1 - COL6;
MSAMMI=SSAMMI/DFAMMI;
F_AMMI=MSAMMI/MSE;
PROBF=1-PROBF(F_AMMI,DFAMMI,DFE);
TITLE 'Resultados da soma de quadrados do modelo AMMI';
PROC PRINT DATA=SSAMMI NOOBS;
VAR SSAMMI PORCENT PORCENAC DFAMMI MSAMMI F_AMMI PROBF;
run;
*** Gera informação para a construção do Biplot ***;
* Para variável Genótipo;
PROC SORT DATA=MELAO;
BY GEN;
PROC MEANS DATA = MELAO NOPRINT;
BY GEN ;
VAR PROD;
OUTPUT OUT = MEDIAG MEAN=PROD;
DATA NAMEG;
SET MEDIAG;
TYPE = 'GEN';
NAME = GEN;
KEEP TYPE NAME PROD;
* Para variável ambiente;
PROC SORT DATA=MELAO;
BY AMB;
PROC MEANS DATA = MELAO NOPRINT;
BY AMB;
VAR PROD;
OUTPUT OUT = MEDIAE MEAN=PROD;
DATA NAMEE;
SET MEDIAE;
TYPE = 'AMB';
NAME1 = AMB;
NAME = COMPRESS(NAME1);
KEEP TYPE NAME PROD;
DATA NAMETYPE; * Contém o tipo e nome para genótipos e ambientes;
SET NAMEG NAMEE;
PROC PRINT DATA=NAMETYPE;
RUN;
* Gera o arquivo com Tipo, Nome e Escores para o Biplot;
DATA BILOT ;
MERGE NAMETYPE SCORES;
DIM1=COL1;
DIM2=COL2;
DIM3=COL3;
DROP COL1-COL3;
data biplot ;
set biplot pol perp;

```

```

proc print data=biplot;
run;
TITLE 'Resultados dos escores para o Biplot';
PROC PRINT DATA=BIPLOT NOOBS;
    VAR TYPE NAME PROD DIM1 DIM2 DIM3;
RUN;
* Gera arquivo tipo Annotate para saída SAS/GRAPH para o Biplot;
Data labels;
    set biplot;
    retain xsys '2' ysys '2' ;
    length function text $8 ;
    text = name ;
    if type = 'GEN' then do;
        color='red';
        size = 1.4;
        style = 'hwcgm001';
        x = dim1;
        y = dim2;
        if dim1 >=0
            then position='5';
            else position='5';
        function = 'LABEL';
        output;
    end;
    if type = 'AMB' then DO;
        color='black';
        size = 1.4;
        style = 'hwcgm001';
        x = 0.0;
        y = 0.0;
        function='MOVE';
        output;
        x = dim1;
        y = dim2;
        function='DRAW';
        output;
        if dim1 >=0
            then position='5';
            else position='5';
        function='LABEL';
        output;
    end;
Proc gplot data=biplot;
Plot dim2*dim1 v2*v1/ overlay Annotate=labels frame
    Vref=0.0 Href = 0.0
    cvref=black chref=black
    lvref=3 lhref=3
    vaxis=axis2 haxis=axis1
    vminor=1 hminor=1 nolegend;
symbol1 v=none c=black h=0.7 ;
symbol2 v=none c=blue i=j line=3 ;
axis2
    length = 4.9 in
    order = (-3.0 to 4.0 by 1.0)

```

```

label=(f=hwcgm001 h=1.0 a=90 r=0 'Fator 2')
value=(h=0.8)
minor=none;
axis1
length = 4.9 in
order = (-3.0 to 4.0 by 1.0)
label=(f=hwcgm001 h=1.0 'Fator 1')
value=(h=0.8)
minor=none;
Title f=hwcgm001 h=1.0 'AMMI Biplot';
run;

```

OBS: Neste programa foi utilizado a variável resposta PROD. Para calcular a variável SST, deve-se mudar PROD para SST. Assim, também como o modelo definido deve-se mudar no PROC GLM. Para a construção do gráfico Biplot com as médias ajustadas, muda-se no procedimento PROC GPLOT a variável dim2 por PROD, também deve-se observar que o tamanho do comprimento dos eixos (length), assim como a ordem (order) deverá ser alterada.

Programa em SAS para a construção do gráfico Trilinear Plot da variável Produção

```

* antes rode a macro triplot que está no site http://www.math.yorku.ca/SCS/vcd/triplot.html;
* e a macro gensym que está no site http://www.math.yorku.ca/SCS/vcd/gensym.html.
* triplot prod;
options ps = 5000 ls=78 nodate;
data produc;
input TYPE $ NAME $ PRODT DIM1T DIM2T DIM3T;
label TYPE='type' NAME='name' PRODT='prod' DIM1T='dim1' DIM2T='dim2';
Prod=int(PRODT+3); Dim1=int(DIM1T+3);
Dim2=int(DIM2T+3); Dim3=int(DIM3T+3);
datalines;
GEN G1 23.7261 -1.36343 -0.71567 -1.09622
GEN G2 24.1003 -2.01629 0.67030 0.48838
GEN G3 25.7731 -0.34815 1.12496 1.51640
GEN G4 26.3622 -0.60000 1.84150 -1.77579
GEN G5 27.5564 -1.74616 -0.72336 0.76326
GEN G6 27.2231 0.27793 -2.17959 0.49545
GEN G7 29.8197 1.91289 -0.09187 -1.43354
GEN G8 30.1239 1.41020 -1.13698 -0.29433
GEN G9 30.2836 2.47301 1.21071 1.33639
AMB A1 30.6933 -0.69292 1.05217 -0.96067
AMB A2 26.5807 -0.47387 0.24459 0.50997
AMB A3 19.5993 -1.02959 -0.39658 1.08338
AMB A4 26.8815 -1.14702 -0.71552 -0.50516
AMB A5 20.2911 -0.21758 -1.16003 -0.84710
AMB A6 21.6985 -0.96401 0.54327 -0.53102
AMB A7 30.8767 -0.15503 -0.86008 -1.40732
AMB A8 35.1689 0.41321 -0.48484 1.13091
AMB A9 30.0885 0.99001 2.60378 0.76004
AMB A10 25.4637 0.30677 0.93331 -1.00216
AMB A11 21.7081 -0.91902 -0.81846 1.78760
AMB A12 37.5741 3.88906 -0.94162 -0.01848
;

```

```
run;  
proc print data=produc;  
var name prod dim1 dim2 dim3;  
run;  
%triplot (data=produc,  
    var = Dim1 Dim2 Dim3, class=name, id=name, gridby=20, LABHT=1.2);
```

OBS: Os valores são referentes aos resultados do programa anterior.

ANEXO B - Rotina para utilização do pacote BiplotGUI no software R

O pacote BiploGUI utiliza da interface gráfica para a construção, manipulação e interpolação dos Biplots no software R na representação de uma matriz de dados originais em duas ou três dimensões. As amostras são representadas por pontos e determinadas pela escolha das técnicas: Análise de Componentes Principais e Escalonamento Multidimensional, utiliza-se também de várias transformações e medidas de dissimilaridade como avaliação.

```
#####
# Programa Melao
#
#####
#
# 1º) Instale o software R e depois o pacote BiplotGUI;
# 2º) Leia o pacote BiplotGUI

library(BiplotGUI)

# 3º) Leia o arquivo de dados, indicando o diretorio onde ele está salvo. Os dados
# podem ser inseridos no teclado ou ser importados para R e guardados como uma matriz
# ou um frame de dados. Neste caso, os dados "melao" podem ser visto no R digitando os
# comandos:
#
dadmelaio<-read.table("D:/Edila/melao/melao.txt", header=T)
attach(dadmelaio)
dadmelaio

# 4º) Para inicializar a interface gráfica com os dados, digite o comando
Biplots(dadmelaio)

# Inicialmente aparecerá uma janela, com o gráfico Biplot construído na metodologia
ACP (Análise de componentes Principais) à esquerda. No topo do lado direito, aparecerá
as variáveis preditivas.

# 5º) Na barra de ferramentas, a opção "Format", pode-se alterar cor, tamanho, grupos etc.

# 6º) Escolha a opção "Points" -> "Dissimilarity Metric" -> "Pythagoras" para calcular
as dissimilaridades utilizando a distância euclidiana.

# 7º) Escolha a opção MDS
```

Resultados referente a variável produção

```
* Matriz de dados originais da variável produção
#####
      G1    G2    G3    G4    G5    G6    G7    G8    G9
A1  30.65 29.34 30.99 32.26 28.99 26.75 33.74 31.97 31.56
A2  21.47 23.19 26.69 26.12 30.55 26.17 28.87 27.29 28.88
A3  18.87 17.86 20.31 16.03 22.34 20.14 18.30 20.83 21.71
A4  24.02 22.96 24.01 28.31 32.27 28.24 27.49 29.26 25.37
A5  18.68 20.45 14.52 17.91 18.51 23.67 24.56 23.33 20.97
A6  21.16 16.63 20.52 22.71 22.73 18.83 22.11 24.02 22.76
```

```

A7  28.96 25.65 26.41 32.59 31.32 33.42 34.32 34.56 30.66
A8  28.85 30.45 35.64 31.33 37.75 35.96 37.79 39.71 39.03
A9  21.44 27.77 31.11 33.35 26.92 26.96 32.49 30.43 40.33
A10 21.45 23.04 23.11 26.94 24.30 21.83 31.07 28.69 28.74
A11 18.55 23.06 26.80 16.10 23.31 24.56 19.73 24.12 24.15
A12 30.59 25.28 34.18 32.38 31.68 40.14 47.38 47.28 49.26

```

* Matriz de dados transformados da variável produção calculados apartir dos dados originais

	G1	G2	G3	G4	G5	G6
A1	0.27384375	0.22410910	0.19620516	0.23649617	0.069285492	0.006788732
A2	-0.08213230	-0.01112152	0.04682955	0.02536941	0.121699234	-0.015132024
A3	-0.21121556	-0.27227909	-0.22634604	-0.46286993	-0.191286658	-0.277037854
A4	0.03009790	-0.02108910	-0.05898900	0.10588320	0.176472555	0.060993672
A5	-0.22133548	-0.13685979	-0.56193241	-0.35197268	-0.379354325	-0.115537317
A6	-0.09667631	-0.34363438	-0.21605939	-0.11452654	-0.173979813	-0.344294398
A7	0.21712665	0.08970090	0.03628331	0.24667359	0.146591420	0.229408782
A8	0.21332108	0.26124332	0.33600919	0.20724421	0.333320024	0.302661469
A9	-0.08353058	0.16911363	0.20006990	0.26972587	-0.004795946	0.014608545
A10	-0.08306427	-0.01761083	-0.09719399	0.05628027	-0.107189101	-0.196460570
A11	-0.22831913	-0.01674315	0.05094247	-0.45851263	-0.148782998	-0.078626638
A12	0.27188424	0.07517090	0.29418126	0.24020905	0.158020116	0.412627602

	G7	G8	G9
A1	0.158296751	0.085800698	0.07534472
A2	0.002415667	-0.072476930	-0.01339646
A3	-0.453486266	-0.342603010	-0.29877279
A4	-0.046565024	-0.002775869	-0.14297840
A5	-0.159268223	-0.229257179	-0.33345293
A6	-0.264357331	-0.200110448	-0.25154117
A7	0.175340948	0.163699679	0.04641310
A8	0.271657191	0.302605780	0.28778480
A9	0.120545023	0.036431699	0.32054983
A10	0.075855397	-0.022448635	-0.01825590
A11	-0.378247006	-0.195955893	-0.19226140
A12	0.497812872	0.477090108	0.52056661

*

* A matriz Delta de distâncias inter-pontos

	A1	A2	A3	A4	A5	A6	A7
A1	0.0000000	0.76269204	0.4329508	0.72910841	0.7827661	1.0843463	0.7873621
A2	0.7626920	0.0000000	0.3298143	0.09463573	1.2367109	0.3281062	0.1247184
A3	0.4329508	0.32981428	0.0000000	0.30215115	0.9912012	0.6526538	0.3680167
A4	0.7291084	0.09463573	0.3021511	0.0000000	1.1543777	0.3577398	0.2169063
A5	0.7827661	1.23671093	0.9912012	1.15437772	0.0000000	1.4655087	1.3267616
A6	1.0843463	0.32810616	0.6526538	0.35773981	1.4655087	0.0000000	0.3644592
A7	0.7873621	0.12471845	0.3680167	0.21690635	1.3267616	0.3644592	0.0000000
A8	0.9046036	0.32847554	0.5186570	0.26183789	1.1515692	0.3357768	0.4467220
A9	0.2948124	0.74505581	0.4614297	0.74835673	1.0774870	1.0704085	0.7214474
A10	0.5254919	0.47603013	0.3296578	0.51427885	1.2479226	0.7789404	0.4151223
A11	0.6697576	0.19451903	0.2665173	0.10323404	1.0527106	0.4386837	0.3099996
A12	0.2069956	0.62343042	0.3127979	0.61180412	0.9645030	0.9515130	0.6230791

	A8	A9	A10	A11	A12
A1	0.9046036	0.2948124	0.5254919	0.6697576	0.2069956


```

A2  0.3284755 0.7450558 0.4760301 0.1945190 0.6234304
A3  0.5186570 0.4614297 0.3296578 0.2665173 0.3127979
A4  0.2618379 0.7483567 0.5142789 0.1032340 0.6118041
A5  1.1515692 1.0774870 1.2479226 1.0527106 0.9645030
A6  0.3357768 1.0704085 0.7789404 0.4386837 0.9515130
A7  0.4467220 0.7214474 0.4151223 0.3099996 0.6230791
A8  0.0000000 0.9799371 0.7739543 0.2524039 0.8285784
A9  0.9799371 0.0000000 0.3415987 0.7275784 0.1680504
A10 0.7739543 0.3415987 0.0000000 0.5377946 0.3191754
A11 0.2524039 0.7275784 0.5377946 0.0000000 0.5778910
A12 0.8285784 0.1680504 0.3191754 0.5778910 0.0000000

```

*

*Stress- Vetor de stress para interação

```

[1] 701.8241 411.4849 385.9026 377.3501 373.7053 371.6068 370.0556
    368.7449 367.5783 366.5254 365.5758 364.7247 363.9681 363.3011
[15] 362.7181 362.2122 361.7763 361.4027 361.0842 360.8137 360.5847
    360.3913 360.2282 360.0909 359.9753 359.8780 359.7962 359.7273
[29] 359.6693 359.6205 359.5793 359.5445 359.5152 359.4904 359.4694
    359.4516 359.4365 359.4237 359.4129 359.4037 359.3959 359.3892
[43] 359.3835 359.3787 359.3746 359.3711 359.3681 359.3655 359.3633
    359.3614 359.3599 359.3585 359.3573 359.3563 359.3555 359.3547
[57] 359.3541 359.3536 359.3531 359.3527 359.3524

```

*

*valores preditos

\$G1

	Prediction	Actual	RelAbsErr%
A1	26.97705	30.65	30.3549192
A2	23.34532	21.47	15.4985031
A3	18.51202	18.87	2.9585498
A4	24.04104	24.02	0.1738965
A5	17.70626	18.68	8.0474778
A6	19.37624	21.16	14.7418466
A7	26.50474	28.96	20.2913685
A8	29.65326	28.85	6.6385070
A9	24.66161	21.44	26.6248974
A10	22.00130	21.45	4.5562030
A11	20.95604	18.55	19.8846248
A12	29.91212	30.59	5.6022882

\$G2

	Prediction	Actual	RelAbsErr%
A1	26.70814	29.34	19.043859
A2	23.48383	23.19	2.126135
A3	19.11946	17.86	9.113337
A4	24.11094	22.96	8.328041
A5	18.36247	20.45	15.105115
A6	19.90179	16.63	23.674340
A7	26.28447	25.65	4.590987
A8	29.04057	30.45	10.198454
A9	24.64093	27.77	22.641642
A10	22.27164	23.04	5.559782
A11	21.35231	23.06	12.356686
A12	29.24725	25.28	28.706556

\$G3

	Prediction	Actual	RelAbsErr%
A1	32.05915	30.99	5.062259
A2	25.63751	26.69	4.983375
A3	19.49823	20.31	3.843618
A4	27.46646	24.01	16.365826
A5	16.73309	14.52	10.478643
A6	19.85930	20.52	3.128321
A7	30.31111	26.41	18.471188
A8	35.44244	35.64	0.935419
A9	25.89760	31.11	24.679910
A10	22.81456	23.11	1.398868
A11	23.89773	26.80	13.741787
A12	33.57449	34.18	2.866988

\$G4

	Prediction	Actual	RelAbsErr%
A1	28.85979	32.26	19.631695
A2	25.30261	26.12	4.719337
A3	17.41014	16.03	7.968487
A4	24.84180	28.31	20.024222
A5	19.64742	17.91	10.031297
A6	20.05864	22.71	15.308081
A7	30.01050	32.59	14.893178
A8	34.34739	31.33	17.421401
A9	30.69469	33.35	15.330898
A10	25.15103	26.94	10.328932
A11	19.05489	16.10	17.060567
A12	39.41134	32.38	40.596629

\$G5

	Prediction	Actual	RelAbsErr%
A1	32.65515	28.99	19.049654
A2	27.20410	30.55	17.390340
A3	21.89261	22.34	2.325312
A4	28.83109	32.27	17.873767
A5	19.19806	18.51	3.576177
A6	22.13550	22.73	3.089915
A7	31.12495	31.32	1.013776
A8	35.32162	37.75	12.621508
A9	27.27440	26.92	1.842006
A10	24.68400	24.30	1.995831
A11	25.87509	23.31	13.332051
A12	33.62032	31.68	10.084816

\$G6

	Prediction	Actual	RelAbsErr%
A1	30.70046	26.75	18.5380584
A2	26.54923	26.17	1.7795823
A3	20.02934	20.14	0.5192746
A4	26.97770	28.24	5.9235043
A5	20.10121	23.67	16.7470237
A6	21.60658	18.83	13.0294811
A7	30.66250	33.42	12.9399418
A8	34.66240	35.96	6.0891413
A9	29.28254	26.96	10.8988296
A10	25.37513	21.83	16.6359774

```

A11  22.53450  24.56  9.5049496
A12  36.34946  40.14 17.7875959
$G7

```

```

      Prediction Actual RelAbsErr%
A1    31.71463  33.74  6.9648224
A2    28.51009  28.87  1.2376409
A3    19.02606  18.30  2.4967784
A4    27.30866  27.49  0.6235971
A5    23.20852  24.56  4.6474664
A6    22.71466  22.11  2.0793095
A7    33.88260  34.32  1.5041430
A8    38.57992  37.79  2.7163706
A9    36.45942  32.49 13.6499989
A10   29.23414  31.07  6.3131365
A11   20.18270  19.73  1.5567240
A12   46.76205  47.38  2.1249908
$G8

```

```

      Prediction Actual RelAbsErr%
A1    33.50024  31.97  5.7854019
A2    29.22598  27.29  7.3193998
A3    20.97692  20.83  0.5554601
A4    29.12448  29.26  0.5123683
A5    22.48993  23.33  3.1760524
A6    23.47066  24.02  2.0769075
A7    34.22813  34.56  1.2547215
A8    38.92927  39.71  2.9517132
A9    33.98678  30.43 13.4471981
A10   28.56690  28.69  0.4654187
A11   23.21707  24.12  3.4137176
A12   42.93269  47.28 16.4359551
$G9

```

```

      Prediction Actual RelAbsErr%
A1    32.83521  31.56  4.5076484
A2    29.05648  28.88  0.6238210
A3    19.97022  21.71  6.1497968
A4    28.34740  25.37 10.5245536
A5    22.96306  20.97  7.0450933
A6    23.16817  22.76  1.4427955
A7    34.36941  30.66 13.1120925
A8    39.18039  39.03  0.5316096
A9    35.62382  40.33 16.6354977
A10   29.12976  28.74  1.3777215
A11   21.63543  24.15  8.8885444
A12   45.50911  49.26 13.2587056

```

*

* Erro Médio Absoluto dos genótipos

```

      G1      G2      G3      G4      G5      G6      G7      G8      G9
12.948176 13.453617  8.830739 16.110400  8.682341 10.866599  3.826655  4.782696  7.008049

```

*

* Vetor do Stress por interação

```

[1]  0.3728397  0.3728394

```

Resultados referente a variável SST

* Matriz de dados originais da variável SST

	G1	G2	G3	G4	G5	G6	G7	G8	G9
A1	8.69	8.83	9.39	10.42	11.00	9.48	9.89	9.17	10.08
A2	11.31	11.19	11.76	10.32	13.34	12.32	12.98	12.95	13.05
A3	9.66	10.32	10.20	11.57	12.06	11.09	11.41	12.51	10.88
A4	10.89	11.20	12.27	11.87	12.90	12.47	12.70	12.17	13.38
A5	12.73	13.71	11.65	12.68	8.73	9.71	7.65	6.51	7.04
A6	12.47	13.13	11.50	12.85	14.18	13.34	14.59	14.66	15.37
A7	10.23	11.27	12.08	11.04	12.46	11.45	13.22	14.51	13.69
A8	14.35	11.66	12.52	11.11	13.32	11.63	14.74	12.81	12.02
A9	9.43	8.52	9.16	9.13	9.42	9.82	9.49	11.44	10.45
A10	9.21	9.59	10.56	9.26	10.35	9.34	11.49	13.16	12.68
A11	11.42	11.41	11.33	12.69	12.58	12.25	12.13	12.10	12.66
A12	8.57	9.14	10.14	10.31	8.91	11.28	11.02	10.45	11.09

* Matriz de dados transformados da variável SST calculados dos dados originais

	G1	G2	G3	G4	G5	G6
A1	-0.20015043	-0.19391120	-0.15754533	-0.0574545698	-0.04087773	-0.158529301
A2	0.06336392	0.04295431	0.06751332	-0.0670978461	0.15199404	0.103510340
A3	-0.09432972	-0.03798245	-0.07480291	0.0472339350	0.05112119	-0.001669816
A4	0.02552157	0.04384757	0.10996663	0.0728326024	0.11845431	0.115612142
A5	0.18163805	0.24605928	0.05811555	0.1388443428	-0.27200763	-0.134557335
A6	0.16100239	0.20283348	0.04515641	0.1521622052	0.21305952	0.183053423
A7	-0.03699879	0.05007812	0.09436057	0.0003434347	0.08375051	0.030276112
A8	0.30142658	0.08409797	0.13013674	0.0066639975	0.15049367	0.045874349
A9	-0.11842727	-0.22964987	-0.18234445	-0.1896159116	-0.19593791	-0.123292495
A10	-0.14203352	-0.11134532	-0.04011735	-0.1754775575	-0.10178648	-0.173407365
A11	0.07304284	0.06242395	0.03026345	0.1396326756	0.09333525	0.097812319
A12	-0.21405563	-0.15940583	-0.08070263	-0.0680673081	-0.25159876	0.015317628

	G7	G8	G9
A1	-0.15893194	-0.23740996	-0.14598100
A2	0.11295363	0.10774854	0.11225387
A3	-0.01596592	0.07318108	-0.06960802
A4	0.09114591	0.04562666	0.13722679
A5	-0.41575043	-0.58000779	-0.50492609
A6	0.22988028	0.23177545	0.27588330
A7	0.13127475	0.22149082	0.16013138
A8	0.24010881	0.09687887	0.03003767
A9	-0.20021747	-0.01623126	-0.10993228
A10	-0.00897899	0.12383468	0.08349169
A11	0.04522564	0.03985820	0.08191316
A12	-0.05074428	-0.10674527	-0.05049046

*

*Matriz Delta de inter-distancias

	G1	G2	G3	G4	G5	G6	G7	G8	G9
G1	0.000000	3.068763	3.090933	5.374979	5.979678	6.128535	8.009787	10.677423	9.980168
G2	3.068763	0.000000	2.750242	2.666903	6.924625	4.902778	8.567529	10.742068	9.452092
G3	3.090933	2.750242	0.000000	3.303833	4.188170	3.037961	5.834682	8.145793	7.110292

```
G4  5.374979  2.666903  3.303833  0.000000  6.956855  3.387289  8.047188  9.619855  7.928095
G5  5.979678  6.924625  4.188170  6.956855  0.000000  4.333307  2.102792  4.957730  5.055329
G6  6.128535  4.902778  3.037961  3.387289  4.333307  0.000000  4.868361  6.232638  4.624738
G7  8.009787  8.567529  5.834682  8.047188  2.102792  4.868361  0.000000  2.886868  3.489956
G8  10.677423  10.742068  8.145793  9.619855  4.957730  6.232638  2.886868  0.000000  2.216679
G9  9.980168  9.452092  7.110292  7.928095  5.055329  4.624738  3.489956  2.216679  0.000000
```

*

*Stress- Vetor de stress para interação

```
[1] 29.62323 16.54381 15.11519 14.44827 14.12738 13.96643 13.87742 13.82022 13.77700
[10] 13.73984 13.70522 13.67152 13.63798 13.60423 13.57010 13.53549 13.50040 13.46484
[19] 13.42888 13.39263 13.35620 13.31977 13.28353 13.24771 13.21256 13.17835 13.14532
[28] 13.11375 13.08387 13.05587 13.02992 13.00613 12.98456 12.96520 12.94801 12.93291
[37] 12.91976 12.90841 12.89871 12.89047 12.88353 12.87771 12.87286 12.86885 12.86553
[46] 12.86281 12.86058 12.85875 12.85726 12.85604 12.85506 12.85425 12.85360 12.85307
[55] 12.85264 12.85229 12.85201 12.85178 12.85159 12.85144 12.85132 12.85122 12.85114
[64] 12.85107 12.85102 12.85097 12.85094 12.85091 12.85088 12.85086 12.85085 12.85084
[73] 12.85082
```

*

*valores preditos

\$G1

	Prediction	Actual	RelAbsErr%
A1	9.310660	8.69	10.7380627
A2	11.130667	11.31	3.1026436
A3	10.328315	9.66	11.5625466
A4	11.472993	10.89	10.0863769
A5	12.455333	12.73	4.7520317
A6	12.422643	12.47	0.8193306
A7	10.556609	10.23	5.6506740
A8	13.119933	14.35	21.2814442
A9	8.319251	9.43	19.2171165
A10	8.850137	9.21	6.2259971
A11	11.657184	11.42	4.1035237
A12	9.067070	8.57	8.5998190

\$G2

	Prediction	Actual	RelAbsErr%
A1	9.567397	8.83	14.2080309
A2	11.138341	11.19	0.9953544
A3	10.453413	10.32	2.5705710
A4	11.474915	11.20	5.2970079
A5	12.689019	13.71	19.6720723
A6	12.273667	13.13	16.4996820
A7	10.586740	11.27	13.1649415
A8	13.000965	11.66	25.8374673
A9	8.579330	8.52	1.1431666
A10	9.027300	9.59	10.8420093
A11	11.670562	11.41	5.0204583
A12	9.294759	9.14	2.9818717

\$G3

	Prediction	Actual	RelAbsErr%
A1	9.948107	9.39	16.6103354
A2	11.497766	11.76	7.8045860
A3	10.812565	10.20	18.2310924

A4	11.632705	12.27	18.9671121
A5	11.258738	11.65	11.6446955
A6	12.424574	11.50	27.5170875
A7	11.220097	12.08	25.5923383
A8	12.529396	12.52	0.2796452
A9	9.475279	9.16	9.3833076
A10	10.020267	10.56	16.0634868
A11	11.656801	11.33	9.7262174
A12	9.938661	10.14	5.9922185

\$G4

	Prediction	Actual	RelAbsErr%
A1	10.28374	10.42	3.662787
A2	11.26326	10.32	25.356582
A3	10.84684	11.57	19.439846
A4	11.52397	11.87	9.301981
A5	12.72582	12.68	1.231674
A6	11.99104	12.85	23.090331
A7	10.84697	11.04	5.189053
A8	12.58757	11.11	39.719637
A9	9.47440	9.13	9.258048
A10	9.74532	9.26	13.046250
A11	11.69087	12.69	26.858435
A12	10.02608	10.31	7.632323

\$G5

	Prediction	Actual	RelAbsErr%
A1	9.489707	11.00	27.711789
A2	12.782010	13.34	10.238340
A3	11.241713	12.06	15.014438
A4	12.667076	12.90	4.273844
A5	9.183316	8.73	8.317733
A6	14.556080	14.18	6.900550
A7	12.744750	12.46	5.224762
A8	13.509074	13.32	3.469247
A9	9.575054	9.42	2.845019
A10	10.902367	10.35	10.135174
A11	12.372508	12.58	3.807201
A12	9.983123	8.91	19.690328

\$G6

	Prediction	Actual	RelAbsErr%
A1	9.787028	9.48	7.675703
A2	11.918064	12.32	10.048397
A3	10.950136	11.09	3.496610
A4	11.892236	12.47	14.444108
A5	9.899211	9.71	4.730283
A6	13.032187	13.34	7.695317
A7	11.830347	11.45	9.508677
A8	12.536393	11.63	22.659830
A9	9.722184	9.82	2.445405
A10	10.575887	9.34	30.897165
A11	11.746118	12.25	12.597039
A12	10.059544	11.28	30.511393

\$G7

	Prediction	Actual	RelAbsErr%
A1	9.420093	9.89	6.62773737

A2	13.227958	12.98	3.49729411
A3	11.416852	11.41	0.09664543
A4	12.904779	12.70	2.88828389
A5	8.002948	7.65	4.97810445
A6	15.143640	14.59	7.80874694
A7	13.459832	13.22	3.38268083
A8	13.360643	14.74	19.45496477
A9	9.981656	9.49	6.93450468
A10	11.648534	11.49	2.23602100
A11	12.411620	12.13	3.97207828
A12	10.212688	11.02	11.38663443

\$G8

	Prediction	Actual	RelAbsErr%
A1	9.587382	9.17	5.1212502
A2	13.431303	12.95	5.9055599
A3	11.585791	12.51	11.3399822
A4	12.767026	12.17	7.3254670
A5	6.494461	6.51	0.1906678
A6	15.007192	14.66	4.2600228
A7	14.187416	14.51	3.9580900
A8	12.321389	12.81	5.9952258
A9	11.071396	11.44	4.5227467
A10	13.058285	13.16	1.2480316
A11	12.015160	12.10	1.0409866
A12	10.830410	10.45	4.6676017

\$G9

	Prediction	Actual	RelAbsErr%
A1	9.671591	10.08	4.9028632
A2	13.336295	13.05	3.4369176
A3	11.588874	10.88	8.5098971
A4	12.795084	13.38	7.0217970
A5	7.086012	7.04	0.5523679
A6	14.915356	15.37	5.4579162
A7	13.912303	13.69	2.6687016
A8	12.599979	12.02	6.9625382
A9	10.842985	10.45	4.7177020
A10	12.650921	12.68	0.3490883
A11	12.146355	12.66	6.1662049
A12	10.744159	11.09	4.1517576

*

*Erro Médio Absoluto

	G1	G2	G3	G4	G5	G6	G7	G8	G9
	8.844964	9.852719	13.984344	15.315579	9.802369	13.059160	6.105308	4.631303	4.574813

*

* Vetor do Stress por interação

[1] 0.1722107 0.1722106