### Universidade de São Paulo Escola Superior de Agricultura "Luiz de Queiroz"

### Distribuição empírica dos autovalores associados à matriz de interação dos modelos AMMI pelo método bootstrap não-paramétrico

### Kuang Hongyu

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

Piracicaba 2012 Kuang Hongyu Bacharel em Matemática

### Distribuição empírica dos autovalores associados à matriz de interação dos modelos AMMI pelo método bootstrap não-paramétrico

### versão revisada de acordo com a resolução CoPGr 5890 de 2010

Orientador: Prof. Dr. CARLOS TADEU DOS SANTOS DIAS

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

Piracicaba 2012

# Dados Internacionais de Catalogação na Publicação DIVISÃO DE BIBLIOTECA - ESALQ/USP

Hongyu, Kuang Distribuição empírica dos autovalores associados à matriz de interação dos modelos AMMI pelo método bootstrap não-paramétrico / Kuang Hongyu. - - versão revisada de acordo com a resolução CoPGr 5890 de 2010. - - Piracicaba, 2012. 102 p. : il.

Dissertação (Mestrado) - - Escola Superior de Agricultura "Luiz de Queiroz", 2012.

1. Análise de variância 2. Estimação não-paramétrica 3. Genótipos 4. Interação genótipo-ambiente 5. Modelos matemáticos 6. Probabilidade I. Título

> CDD 519.5 H772d

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

### Dedicatória

Aos meus pais,

Kuang Songzhi e Zhen Yaci, pelo exemplo de vida, por sempre acreditarem em mim, por todo apoio e confiança em mim depositada, pois sem a participação de vocês não teria chegado até aqui.

A minha namorada,

Fabiane de Lima Silva, por toda ajuda, incentivo, ensinamentos, conselhos, companheirismo, o amor sempre constante e por me fazer tão feliz.

A minha irmã,

Kuang Qixia, pelo amizade, incentivo, apoio e por fazer parte da minha vida.

#### AGRADECIMENTOS

Ao Prof. Dr. Carlos Tadeu dos Santos dias pela orientação, pela confiança em mim depositada, por todo incentivo e pelas críticas e sugestões que foram fundamentais para o desenvolvimento desta pesquisa e para meu crescimento profissional.

Ao Prof. Dr. Lúcio Borge de Araújo, pela confiança, amizade, ensinamentos e apoio na elaboração da dissertação.

Ao Prof. Dr. Gerson Barreto Mourão, pelos ensinamentos e amizade.

Ao Prof. Dr. Ronald Vencovsky, pela paciência, dedicação de seu tempo e discussões que enriqueceram este trabalho.

Aos professores do Departamento de Matemática da Universidade Federal de Uberlândia por todos os ensinamentos e amizade.

Aos professores do programa de Pós-graduação em Estatística e Experimentação Agronômica Dr. César Gonçalves de Lima, Dra. Clarice Garcia Borges Demétrio, Dr. Décio Barbin, Dr. Edwin Ortega, Dr. Gabriel Adrián Sarriés, Dra. Roseli Aparecida Leandro, Dr. Sílvio Zocchi, Dra. Sônia Maria de Stefano Piedade, Dra Taciana Villela Savian e Dr. Vitor Augusto Ozaki por todos os ensinamentos e amizade.

Aos amigos de turma do mestrado: Ana Julia, Daniel, Gabriel, Ítalo, João, Josy, Luiz, Patricia Gonzalez e Veronika, pelo companheirismo durante o curso.

Aos amigos Daniel e Veronika pela ajuda e ensinamento no desenvolvimento deste trabalho.

Aos amigos Diego Wyllyam do Vale, Iuri Ferreira, Gilvano Ebling Brondani e Vinícius Sandri Diaz pela agradável convivência e amizade.

As colegas da Pós-Graduação em Estatística e Experimentação Agronômica, Cristiane Rodrigues, Josiane Rodrigues, Izabela Oliveira e Luzia pela ajuda e incentivo.

Aos amigos dos cursos de mestrado e doutorado do Departamento de Ciência Exatas da ESALQ/USP, em especial a Ana Patrícia Bastos Peixoto, Elizabeth Mie Hashimoto, Gláucia Tatiana Ferrari, Marcelino Alves Rosa de Páscoa, Ricardo Alves de Olinda, Tiago Almeida de Oliveira, Tiago Viana Flor de Santana, pelo auxílio em momentos de dúvidas, e principalmente pela sincera amizade. A todos os alunos do curso de Pós-Graduação em Estatística e Experimentação Agronômica da ESALQ/USP, com os quais tive o prazer de conviver.

À secretária da Pós-Graduação em Estatística e Experimentação Agronômica, Luciane Brajão, pelo apoio, amizade, ajuda e confiança depositada em mim, que me ajudou muito durante esses dois anos de mestrado.

À secretária do LCE, Solange Paes de Assis Sabadin, pelo amizade, apoio e ajuda sempre que preciso.

Aos funcionários do Departamento de Ciências Exatas da ESALQ/USP Eduardo Bonilha, Jorge Alexandre Wiendl e Rosni Honofre Aparecido Pinto pela amizade, ajuda e apoio técnico.

Às bibliotecárias da ESALQ/USP, Eliana Maria Garcia e Beatriz Helena Giongo pela ajuda e revisão da dissertação.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES, pela concessão de bolsa de estudos.

Àqueles que contribuiram direta ou indiretamente para a realização deste trabalho. "No meio de qualquer dificuldade encontra-se a oportunidade."

Albert Einstein

"Não deixe o barulho da opinião dos outros abafar sua voz interior. E mais importante, tenha a coragem de seguir seu coração e sua intuição. Eles de alguma forma já sabem o que você realmente quer se tornar. Tudo o mais é secundário." 

## SUMÁRIO

RESUMO	11
ABSTRACT	13
LISTA DE FIGURAS	15
LISTA DE TABELAS	17
1 INTRODUÇÃO	19
REFERÊNCIAS	21
2 REVISÃO BIBLIOGRÁFICA	23
2.1 Interação genótipo x ambiente	23
2.2 Análise de variância	25
2.3 Análise de variância conjunta	28
2.4 Autovalores e Autovetores de uma matriz	30
2.5 Decomposição em valores singulares	32
2.6 Análise de Componentes Principais	33
2.7 Histórico dos modelos AMMI	35
2.8 Modelo AMMI	36
2.9 Estimação de autovalores e distribuição amostral	41
2.10 O método Bootstrap e estimador do erro padrão	43
2.110 número de repetições bootstrap $B$	47
REFERÊNCIAS	49
3 ADAPTABILIDADE E ESTABILIDADE DE CULTIVARES DE MILHOS	
ATRAVÉS DO MODELO AMMI	53
Resumo	53
Abstract	53
3.1 Introdução	54
3.2 Material e Métodos	55
3.3 Resultados e Discussão	57
3.4 Conclusão	64
REFERÊNCIAS	65

4 DISTRIBUIÇÃO EMPÍRICA	DOS A	UTOVALOR	LES ASSOCIAD	DOS À
MATRIZ DE INTERAÇÃO	PELO	MÉTODO	BOOTSTRAP	NÃO-
PARAMÉTRICO				67
Resumo				67
Abstract				67
4.1 Introdução				68
4.2 Material e Métodos				69
4.3 Resultados e Discussão				75
4.4 Conclusão				
REFERÊNCIAS				89
APÊNDICES				93

#### RESUMO

# Distribuição empírica dos autovalores associados à matriz de interação dos modelos AMMI pelo método bootstrap não-paramétrico

A interação genótipos  $\times$  ambientes  $(G \times E)$  foi definido por Shelbourne (1972) como sendo a variação entre genótipos em resposta a diferentes condições ambientais. Sua magnitude na expressão fenotípica do caráter pode reduzir a correlação entre fenótipo e genótipo, inflacionando a variância genética e, por sua vez, parâmetros dependentes desta, como herdabilidade e ganho genético com a selecão. Estudos sobre a adaptabilidade e a estabilidade fenotípica permitem particularizar os efeitos da interação  $G \times E$  ao nível de genótipo e ambiente, identificando a contribuição relativa de cada um para a interação total. Várias metodologias estatísticas têm sido propostas para a interpretação da interação  $G \times E$  proveniente de um grupo de cultivares testados em vários ambientes. Entre essas metodologias destaca-se os modelos AMMI (Additive Main Effects and Multiplicative Interaction Model), que vem ganhando grande aplicabilidade nos últimos anos. O modelo AMMI é um método uni-multivariado, que engloba uma análise de variância para os efeitos principais, que são os efeitos dos genótipos (G) e os ambientes (E) e para os efeitos multiplicativos (interação genótipo  $\times$  ambiente), para a qual utiliza-se a decomposição em valor singular (DVS). Essa técnica multivariada baseia-se no uso dos autovalores e autovetores provenientes da matriz de interação  $G \times E$ . Araújo e Dias (2005) verificaram o problema de superestimação e subestimação de autovalores estimados da maneira convencional. Efron(1979) propôs uma técnica de simulação numérica chamada Bootstrap para avaliar tais incertezas. O método Bootstrap consiste em uma técnica de reamostragem que permite aproximar a distribuição de uma função das observações a partir da distribuição empírica dos dados. Por meio desse método, podem ser estimados o erro-padrão da referida estimativa e os intervalos de confiança, com o intuito de fazer inferência sobre os parâmetros em questão. O objetivo deste trabalho será estudar o efeito da interação  $G \times E$ , avaliar a adaptabilidade e estabilidade de genótipos em diferentes ambientes através do modelo AMMI, com as análises através dos gráficos Biplot, encontrar a distribuição empírica dos autovalores e calcular o intervalo de confiança através o método Bootstrap não-paramétrico. Com o estudo da distribuição empírica dos autovalores poder-se-á validar os testes de hipóteses propostos na literatura para identificar o número de IPCA (Incremental Principal Component Analysis) para seleção dos modelos AMMI, e propor um teste para seleção dos modelos.

Palavras-chave: Interação genótipos × ambientes; Modelos AMMI; *Biplot*; Método *Bootstrap*; Intervalo de confiança

### ABSTRACT

#### Empirical distribution of eigenvalues associated with the interaction matrix of the AMMI models for non-parametric bootstrap method

The genotype  $\times$  environment interaction  $(G \times E)$  was defined by Shelbourne (1972) as the variation among genotypes in response to different environmental conditions. Its magnitude in phenotypic expression of the character can reduce the correlation between genotype and phenotype, inflating the genetic variance and, in turn, dependent on the parameters, as heritability and genetic gain with selection. Studies on the phenotypic adaptability and stability allow particularize the effects of interaction  $G \times E$  at the level of genotype and environment, identifying the relative contribution of each to the total interaction. There are several methods of analysis and interpretation for the genotype  $\times$  environment interaction from a group of genotype tested in several environments. These methods include AMMI models (Additive Main Effects and Multiplicative Interaction Model), coming gaining great applicability past years. The AMMI model is a uni-multivariate method, that includes an analysis of variance for the main effects (the effects of the genotypes (G) and environments (E)) and assumes multiplicative effects for the genotype  $\times$  environment interaction, using a singular value decomposition (DVS). This method estimates the eigenvalues and eigenvectors deriving from the matrix of genotype  $\times$  environment interaction. Araújo and Dias (2005) found an overestimation and underestimation problem with the eigenvalues in the conventional way. Efron (1979) proposed a numerical resampling technique called Bootstrap for evaluate such uncertainties. The bootstrap method consists of a resampling technique that allows to approximate the distribution of a function of the observations from the empirical distribution of the data. Through this method, can be estimated by the standard error of that estimate and confidence intervals, in order to make inferences about the parameters in question. The aim of this work was to study the effect of genotype  $\times$  environment interaction ( $G \times E$ ), evaluate the adaptability and stability of genotypes in different environments through the AMMI model, with the analysis through the Biplot graphs, find the empirical distribution of eigenvalues and calculate the confidence interval using the nonparametric bootstrap, the study of the empirical distribution of eigenvalues serve to validate the hypothesis tests proposed in the literature to identify the number of IPCA (Incremental Principal Component Analysis) for selecting the AMMI model, and propose a test for selection of models.

Keywords: Genotype  $\times$  environment interaction; AMMI models; Biplot; Bootstrap method; Confidence interval

### LISTA DE FIGURAS

Figura 1 -	Biplot AMMI1 (Médias vs PC1) para os dados de produtividade de milho	
	$(kgha^{-1})$ , com nove genótipos (G) e vinte ambientes (E)	63
Figura 2 -	Biplot AMMI2 (PC1 vs PC2) para os dados de produtividade de milho	
	$(kgha^{-1})$ , com nove genótipos (G) e vinte ambientes (E)	64
Figura 3 -	Gráfico de Histograma (esquerda) e Q-Q plot (direita) do primeiro autovalor	
	${\rm com}\;100\;{\rm reamostragens\;utilizando\;a\;metodologia\;\textit{bootstrap}, aplicada\;na\;matriz}$	
	de interação genótipos $\times$ ambientes $\hfill\hf$	76
Figura 4 -	Gráfico de Envelope simulado sob o primeiro autovalor com 100 reamostra-	
	gens utilizando a metodologia <i>bootstrap</i> , aplicada à matriz de interação	
	genótipos $\times$ ambientes $\ldots \ldots \ldots$	77
Figura 5 -	Gráfico de Histograma (esquerda) e Q-Q plot (direita) do segundo autovalor	
	${\rm com}\;100\;{\rm reamostragens\;utilizando\;a\;metodologia\;\textit{bootstrap}, aplicada na matriz}$	
	de interação genótipos $\times$ ambientes $\hfill\hf$	78
Figura 6 -	Gráfico de Envelope simulado sob o segundo autovalor com 100 reamostragens	
	utilizando a metodologia <i>bootstrap</i> , aplicada à matriz de interação genótipos	
	$\times$ ambientes $\ldots$ $\ldots$ $\ldots$ $\ldots$	78
Figura 7 -	Gráfico de Histograma (esquerda) e Q-Q plot (direita) do terceiro autovalor	
	$\operatorname{com} 100$ reamostragens utilizando a metodologi a $\mathit{bootstrap},$ aplicada na matriz	
	de interação genótipos $\times$ ambientes $\hfill\hf$	79
Figura 8 -	Gráfico de Envelope simulado sob o terceiro autovalor com 100 reamostragens	
	utilizando a metodologia <i>bootstrap</i> , aplicada à matriz de interação genótipos	
	$\times$ ambientes $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$	80
Figura 9 -	Gráfico de Histograma (esquerda) e Q-Q plot (direita) do quarto autovalor	
	$\operatorname{com} 100$ reamostragens utilizando a metodologi a $\mathit{bootstrap},$ aplicada na matriz	
	de interação genótipos $\times$ ambientes $\ .$	81
Figura 10	- Gráfico de Envelope simulado sob o quarto autovalor com 100 reamostragens	
	utilizando a metodologia <i>bootstrap</i> , aplicada à matriz de interação genótipos	
	$\times$ ambientes	81

Figura 11 - Gráfico de Histograma (esquerda) e Q-Q plot (direita) do quinto autovalor	
com 100 reamostragens utilizando a metodologia <i>bootstrap</i> , aplicada na matriz	
de interação genótipos $\times$ ambientes $\hfill\hf$	82
Figura 12 - Gráfico de Envelope simulado sob o quinto autovalor com 100 reamostragens	
utilizando a metodologia <i>bootstrap</i> , aplicada à matriz de interação genótipos	
$\times$ ambientes $\ldots$ $\ldots$ $\ldots$ $\ldots$	83
Figura 13 - Gráfico de Histograma (esquerda) e Q-Q plot (direita) do sexto autovalor com	
100 reamostragens utilizando a metodologia <i>bootstrap</i> , aplicada na matriz de	
interação genótipos $\times$ ambientes $\hdots$	84
Figura 14 - Gráfico de Envelope simulado sob o sexto autovalor com 100 reamostragens	
utilizando a metodologia <i>bootstrap</i> , aplicada à matriz de interação genótipos	
$\times$ ambientes $\ldots$ $\ldots$ $\ldots$ $\ldots$	85
Figura 15 - Gráfico de Histograma (esquerda) e Q-Q plot (direita) do sétimo autovalor	
${\rm com}\;100\;{\rm reamostragens}\;{\rm utilizando}\;{\rm a}\;{\rm metodologia}\;bootstrap, {\rm aplicada}\;{\rm na}\;{\rm matriz}$	
de interação genótipos $\times$ ambientes $\hfill\hf$	86
Figura 16 - Gráfico de Envelope simulado sob o sétimo autovalor com 100 reamostragens	
utilizando a metodologia <i>bootstrap</i> , aplicada à matriz de interação genótipos	
$\times$ ambientes $\ldots$ $\ldots$ $\ldots$ $\ldots$	86
Figura 17 - Gráfico de Histograma (esquerda) e Q-Q plot (direita) do oitavo autovalor	
com 500 reamostragens utilizando a metodologia $bootstrap,$ aplicada na matriz	
de interação genótipos $\times$ ambientes $\hfill\hf$	87
Figura 18 - Gráfico de Envelope simulado sob o oitavo autovalor com 500 reamostragens	
utilizando a metodologia <i>bootstrap</i> , aplicada à matriz de interação genótipos	
$\times$ ambientes	88

### LISTA DE TABELAS

Tabela 1 -	Esquema da análise de variância para um experimento inteiramente aleatorizado	27
Tabela 2 -	Esquema da análise de variância para experimentos de um mesmo grupo de	
	genótipos avaliado em $e$ locais com $r$ blocos	29
Tabela 3 -	Esquema da análise de variância para experimentos de um mesmo grupo de	
	genótipos avaliado em $e$ locais com $r$ repetições	30
Tabela 4 -	O coeficiente de variação de $\hat{ep}_{boot}$ em função do coeficiente de variação da	
	ideal estimativa bootstrap $\widehat{ep}_\infty$ e o número de amostras bootstrap $B,$ a partir	
	da equação (50) assumindo $\Delta=0$	48
Tabela 5 -	Por centagem da soma de quadrados da interação $(G\times E)$ captada por com-	
	ponente principal (PC)	59
Tabela 6 -	Análise de variância conjunta dos experimentos para à produtividade de	
	milho e decomposição das somas de quadrados da interação $(G\times E)$	60
Tabela 7 -	Coordenadas dos eixos das abscissas (x) e de coordenadas (y), para os dois	
	gráficos Biplot	62
Tabela 8 -	Por centagem da soma de quadrados da interação $(G\times E)$ captada por com-	
	ponente principal (PC)	75

### 1 INTRODUÇÃO

Em alguns estudos, é comum a análise de experimentos com dois fatores, cada um com diferentes níveis. Estes experimentos proporcionam uma tabela de dados de dupla entrada e geralmente, a análise destes dados é realizada através da análise de variância (ANOVA) porém, cumprindo algumas pressuposições, tais como: número de repetições de cada combinação do fator deve ser maior do que um, homogeneidade de variâncias, normalidade dos resíduos, erros independentes e identicamente distribuídos e efeitos aditivos no modelo (GARCÍA-PEÑA; DIAS, 2009). Todavia, existem estudos nos quais é importante a análise da interação ainda que as pressuposições mencionadas não sejam satisfeitas. No melhoramento genético vegetal, por exemplo, o pesquisador está interessado em selecionar os genótipos com desempenhos superiores em diferentes ambientes, porém, a baixa eficiência na análise da interação dos genótipos com os ambientes ( $G \times E$ ) pela ANOVA pode representar um problema para os geneticistas, que aproveitam o efeito da interação para a seleção dos genótipos com produções elevadas.

A interação  $G \times E$  é um componente da variação fenotípica resultante do comportamento diferente dos genótipos quando repetidos a mais de um ambiente. Sua magnitude na expressão fenotípica do caráter pode reduzir a correlação entre fenótipo e genótipo, inflacionando a variância genética e por sua vez, parâmetros dependentes desta, como herdabilidade e ganho genético com a seleção (LAVORANTI; DIAS; VENCOVSKY, 2002).

Estudos sobre a adaptabilidade e a estabilidade fenotípica permitem particularizar os efeitos da interação  $G \times E$  ao nível de genótipo e ambiente, identificando a contribuição relativa de cada um para a interação total. Várias metodologias estatísticas têm sido propostas para a interpretação da interação  $G \times E$  e os pesquisadores ainda continuam na busca de uma ferramenta estatística que permita extrair grande parte da informação possível desta fonte de variação (ARAÚJO; DIAS, 2006). Entre tantas ferramentas disponíveis na literatura os métodos de regressão linear simples (EBERHART; RUSSEL, 1966) e regressão linear múltipla (SILVA; BARRETO, 1985) têm sido as metodologias mais utilizadas, mas estas técnicas possuem limitações e têm sido alvo de várias críticas, como no caso em que a linearidade falha.

O modelo AMMI (Additive Main Effects and Multiplicative Interaction Model)

(MANDEL, 1961, 1969,1971; GOLLOB, 1968), que vem ganhando grande aplicabilidade nos últimos anos (DUARTE; VENCOVSKY, 1999). A metodologia AMMI engloba num único modelo a técnica ANOVA, para efeitos principais, e para efeitos multiplicativos (interação), utiliza-se análise de componentes principais (ACP) ou decomposição em valor singular (DVS), aplicado à matriz de interação (ARAÚJO; DIAS, 2006).

Esta análise pode ajudar tanto na identificação de genótipos de alta produtividade e largamente adaptados, como na localização do chamado zoneamento agronômico, com fins de recomendação regionalizada e seleção de locais de teste (GAUCH; ZOBEL, 1996).

Essa técnica multivariada é baseada no uso dos autovalores e autovetores provenientes da matriz de interação  $G \times E$ . Araújo e Dias (2005), utilizando técnicas de simulação multivariada verificaram o problema de superestimação e subestimação de autovalores estimados da maneira convencional.

Efron (1979) propôs uma técnica de simulação numérica chamada *Bootstrap* para avaliar tais incertezas. O *Bootstrap* é um procedimento robusto de simulação estatística (SSP) para atribuir medidas de precisão das estimativas dos estimadores de parâmetros estatísticos. Um dos atrativos deste método é que ele consegue responder muitas questões estatísticas reais sem termos que lançar mão de fórmulas matemáticas complexas.

O método *Bootstrap* consiste em uma técnica de reamostragem que permite aproximar a distribuição de uma função das observações a partir da distribuição empírica dos dados. Por meio desse método, podem ser estimado o erro-padrão da referida estimativa e os intervalos de confiança, com o intuito de fazer inferência sobre os parâmetros em questão. Assim, a presente dissertação tem os seguintes objetivos:

- a) Estudar o efeito da interação  $G \times E$ , avaliar a adaptabilidade e estabilidade de genótipos em diferentes ambientes utilizando o modelo AMMI.
- b) Encontrar a distribuição empírica dos autovalores, e calcular o intervalo de confiança usando o método *Bootstrap* não-paramétrico.

### Referências

ARAÚJO, L.B.; DIAS, C.T.S. Métodos de correção de autovalores e regressão isotônica nos modelos AMMI. **Revista de Matemática e Estatística**, Jaboticabal, v.24, n.2, p.71-89, 2006.

DUARTE, J.B.; VENCOVSKY, R. Interação genótipo x ambiente: uma introdução à análise "AMMI". Ribeirão Preto: Sociedade Brasileira de Genética, 1999. 60p. (Série Monografias, 9).

EBERHART, S.A.; RUSSELL, W.A. Stability parameters for comparing varieties. Crop Science, Madison, v.6, n.1, p.36-40, 1966.

EFRON, B. Bootstrap methods: another look at jakknife. Annals of Statistics, Hayward, v.7, n.1, p.1-26, 1979.

GARCÍA-PEÑA, M.; DIAS, C.T.S. Análise dos modelos aditivos com interação multiplicativa (AMMI) bivariados. **Revista Brasileira de Biometria**, São Paulo, v.27, n.4, p.586-602, 2009.

GAUCH, H.G.; ZOBEL, R.W. AMMI analysis of yield trails. In: KANG, M.S.; GAUCH, H.G. (Ed.). Genotype by environment interaction. New York: CRC Press, 1996. p.416-428.

GOLLOB, H.F. A statistical model which combines feature of factor analitic and analysis of variance techniques. **Psychometrika**, New York, v.33, p.73-115, 1968.

LAVORANTI, O.J.; DIAS, C.T.S.; VENCOVSKY, R. Estudo da estabilidade e adaptabilidade fenotípica de progênies de *Eucalyptus grandis*, via metodologia AMMI. **Boletim de Pesquisa Florestal**, Curitiba, v.44, n.1, p.107-124, jan./jun. 2002.

MANDEL, J. Non-additivity in two-way analysis of variance. Journal of the American Statistical Association, New York, v.56, p.878-888, 1961.

SILVA, J.G.; BARRETO, J.N. Aplicação da regressão linear segmentada em estudos da interação genótipo x ambiente. In: SIMPÓSIO DE ESTATÍSTICA APLICADA A EXPERIMENTAÇÃO AGRONÔMICA, 1., 1985. Campinas. **Resumos...** Campinas: Fundação Cargill, 1985. p.49-50.

### 2 REVISÃO BIBLIOGRÁFICA

#### 2.1 Interação genótipo x ambiente

A interação genótipo × ambiente  $(G \times E)$  foi definida por Shelbourne (1972) como sendo a variação entre genótipos em resposta a diferentes condições ambientais. Neste sentido, na presença da interação, os resultados das avaliações podem mudar de um ambiente para outro, ocasionando mudanças na posição relativa dos genótipos ou mesmo na magnitude das suas diferenças (FALCONER; MACKAY, 1996). A interação  $(G \times E)$  deve ser encarada como um fenômeno biológico com suas implicações no melhoramento de plantas, e não como um fenômeno de simples efeito estatístico, cumprindo buscar a explicação evolutiva do evento se quiser tirar proveito de seus efeitos benéficos indesejáveis sobre a avaliação de genótipos e recomendação de cultivares (CHAVES, 2001).

A seleção e recomendação de genótipos mais produtivos é objetivo básico dos programas de melhoramento genético de qualquer espécie cultivada (CARVALHO et al., 2002). O processo de seleção é freqüentemente realizado avaliando-se o desempenho dos genótipos em diferentes ambientes (ano, local, época de semeadura). Contudo, a decisão de lançamento de novas cultivares normalmente é dificultada pela ocorrência da interação ( $G \times E$ ).

A interação  $(G \times E)$  ocorre quando há respostas diferenciadas dos genótipos testados em diferentes ambientes(CARVALHO et al., 2002). Ela pode ser simples, quando é proporcionada pela diferença de variabilidade entre genótipos nos ambientes, e complexa, quando denota a falta de correlação entre medidas de um mesmo genótipo em ambientes distintos e indica haver inconsistência na superioridade de genótipos com a variação ambiental (ROBERTSON, 1959). Somente quando ocorre a interação complexa haverá dificuldades no melhoramento (CRUZ; REGAZZI, 1994).

As características relacionadas à produção vegetal estão condicionadas ao controle genético do organismo, ao ambiente em que é cultivado e à interação entre esses dois fatores. As diferentes respostas fenotípicas frente a mudanças nas condições ambientais resultam em comportamentos distintos dos genótipos, caracterizando a interação. Se a expressão de um determinado genótipo depende dos genes e do ambiente em que é avaliado, a interação  $(G \times E)$  deve ser mais um fator a ser considerado na análise. Assim, no melhoramento, o processo de seleção depende também da estimação dessa interação, para que não ocorra a queda inesperada de desempenho de um material testado.

Em um determinado ambiente, a manifestação fenotípica é o resultado da ação do genótipo sob influência do meio. Entretanto, quando se considerar uma série de ambiente, detecta-se, além dos efeitos genéticos e ambientais, um efeito adicional, proporcionado pela interação destes.

A avaliação da interação  $(G \times E)$  torna-se de grande importância no melhoramento, pois, no caso de sua existência, há possibilidades de o melhor genótipo em um ambiente não o ser em outro. Este fato influencia o ganho de seleção e dificulta a recomendação de cultivares com ampla adaptabilidade. Pela importância desta interação, cabe ao melhorista avaliar sua magnitude e tecnologia e fornecer subsídio que possibilitem adotar procedimentos para sua minimização e, ou, seu aproveitamento.

O potencial genético de determinado cultivar depende das condições ambientais o qual está sujeito, tais como, condições climáticas, tipo de solo, mudanças e melhorias no ambiente. A interação  $(G \times E)$  reflete as diferentes respostas dos genótipos às condições ambientais, sendo um dos aspectos mais importantes na seleção e recomendação dos melhores genótipos com alta produtividade e qualidade pelos melhoristas (WARZECHA et al., 2011; FALCONER; MACKAY, 1996).

A interação  $(G \times E)$  surge quando um genótipo responde de maneira diferente às variações ambientais e nesta situação, os melhores genótipos em um local podem não o ser em necessariamente os melhores em outros locais.

No melhoramento genético de plantas, os ensaios multiambientais são importantes para testar a adaptação geral e específica das cultivares. Uma cultivar desenvolvendo-se em diferentes ambientes mostrará uma flutuação significativa na produtividade em relação a outras cultivares. Essas mudanças são influenciadas por diferentes condições ambientais e são referidas como interação ( $G \times E$ ) (ARCINIEGAS-ALARCÓN; DIAS, 2009).

Cockerham (1963) atribuiu o aparecimento de interações  $(G \times E)$  como sendo devido a respostas diferenciais do mesmo conjunto gênico em ambientes distintos ou pela expressão de diferentes conjuntos gênicos em diferentes ambientes. Quando um mesmo conjunto de genes se expressa em diferentes ambientes, as diferenças nas respostas podem ser explicadas pela heterogeneidade das variâncias genéticas e experimentais ou por ambas e quando diferentes conjuntos de genes se expressam em ambientes distintos, as diferenças nas respostas explicam-se pela inconsistência das correlações genéticas entre os valores de um mesmo caráter em dois ambientes (FALCONER, 1989). A interação ( $G \times E$ ) também pode surgir em função de fatores fisiológicos e bioquímicos próprios de cada genótipo cultivado (CRUZ; REGAZZI, 1994). Ainda que a falta de ajuste do modelo estatístico adotado ao conjunto de dados pode ser uma das causas da interação ( $G \times E$ ) significativa (CHAVES; VENCOVSKY; GERALDI, 1989).

A interação  $(G \times E)$  pode ser reduzida, utilizando-se cultivares com ampla adaptabilidade e boa estabilidade, ou estratificando-se a região considerada em sub-regiões com características ambientais semelhantes, dentro das quais a interação passa a ser não-significativa (RAMALHO; SANTOS; ZIMMERMANN, 1993).

#### 2.2 Análise de variância

A Análise de Variância (ANOVA) é um procedimento utilizado para comparar dois ou mais tratamentos. Existem muitas variações da ANOVA devido aos diferentes tipos de experimentos que podem ser realizados.

Em um experimento, cada observação  $Y_{ij}$  pode ser decomposta conforme o modelo a seguir, para um experimento com delineamento inteiramente aleatorizado:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \qquad i = 1, \dots, I \ e \ j = 1, \dots, J \tag{1}$$

em que:  $Y_{ij}$  é a observação do *i*-ésimo tratamento na *j*-ésima repetição;  $\mu$  é o efeito constante (média geral);  $\tau_i$  é o efeito do *i*-ésimo tratamento;  $\varepsilon_{ij}$  é o erro associado ao *i*-ésimo tratamento na *j*-ésima repetição assumido como:  $\varepsilon_{ij} \stackrel{IID}{\sim} N(0, \sigma^2)$ . Aqui, *IID* significa que os erros devem ser independentes e identicamente distribuídos.

Em um experimento, existe o interesse em testar se há diferenças entre as médias dos tratamentos, o que equivale a testar as hipóteses:

$$\begin{cases} H_0: \tau_1 = \tau_2 = \dots = \tau_I = 0\\ H_1: Pelo menos um contraste nos efeitos \tau_i é significativo. \end{cases}$$

De forma equivalente, podemos escrever tais hipóteses da seguinte forma:

$$\begin{cases} H_0: \mu_1 = \mu_2 = \dots = \mu_I \\ H_1: Pelo menos um contraste nas médias \mu_i é significativo$$

 $\operatorname{em}\operatorname{que}$ 

$$\mu_i = \mu + \tau_i \qquad i = 1, 2, \dots, I.$$

Note que, se a hipótese nula for verdadeira, todos os tratamentos terão uma média comum  $\mu$ . A análise de variância, baseia-se na decomposição da variação total da variável resposta em partes que podem ser atribuídas aos tratamentos (variação entre) e ao erro experimental (variação dentro). Essas variações podem ser medidas por meio das somas de quadrados definidas para cada um dos seguintes componentes:

$$SQTotal = \sum_{i=1}^{I} \sum_{j=1}^{J} y_{ij}^2 - C, \ em \ que \ C = \frac{\left(\sum_{i=1}^{I} \sum_{j=1}^{J} y_{ij}\right)^2}{IJ};$$
(2)

$$SQTrat = \frac{\sum_{i=1}^{I} y_{i.}^2}{J} - C,$$
(3)

em que  $y_{i.} = \sum_{j=1}^{J} y_{ij}$  o total do *i*-ésimo tratamento. E a soma de quadrados dos resíduos pode ser obtida por diferença:

$$SQRes = SQTotal - SQTrat$$

A *SQTrat* também é chamada de "variação Entre", que é a variação existente entre os diferentes tratamentos e a *SQRes* é chamada de "variação Dentro" que é função das diferenças existentes entre as repetições de um mesmo tratamento.

Essas soma de quadrados podem ser organizadas em uma tabela, denominada tabela da análise de variância, como apresentado na tabela 1.

Para testar a hipótese  $H_0$ , utiliza-se o teste F apresentado na tabela da Análise de Variância (Tabela 1). Convém lembrar que esse teste é válido se os pressupostos assumidos para os erros do modelo estiverem satisfeitos.

Fonte de	Graus de	Somas de	Quadrados	F calculado
Variação	Liberdade	Quadrados	Médios	
Tratamentos	I-1	SQTrat	QMTrat	QMTrat/QMRes
Resíduo	I(J-1)	SQRes	QMRes	
Total	IJ-1	SQTotal		

Tabela 1 - Esquema da análise de variância para um experimento inteiramente aleatorizado

em que QMTrat = SQTrat/(I-1) e QMRes = SQRes/(I(J-1)).

Pode-se mostrar que o quociente QMTrat/QMRes tem distribuição F com (I-1) e I(J-1) graus de liberdade, supondo que,  $y_{ij}$  são variáveis aleatórias independentes, todos os tratamentos têm variância iguais a  $\sigma^2$  e  $Y_{ij} \sim N(\mu_i, \sigma^2)$ . Por esses motivos, os pressupostos da ANOVA devem ser testados ou avaliados em qualquer análise.

Se Fcalculado para tratamentos for maior que Ftabelado, ao nível  $\alpha$  de significância:  $F_{(\alpha, (I-1), I(J-1))}$ , logo rejeitamos a hipótese de nulidade  $H_0$ , ou seja, existem evidências de diferenças significativas em pelo menos um contraste nas médias de tratamentos, ao nível  $\alpha$  de significância escolhido. Caso contrário, não rejeitamos a hipótese de nulidade  $H_0$ , ou seja, não há evidências de diferença significativa entre os níveis de tratamentos, ao nível  $\alpha$ de significância escolhido.

Em um experimento, cada observação  $Y_{ij}$  pode ser decomposta conforme o modelo a seguir, para um experimento com delineamento em blocos aleatorizados:

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \qquad i = 1, \dots, I \ e \ j = 1, \dots, J$$

em que:  $Y_{ij}$  é a observação do *i*-ésimo tratamento no *j*-ésimo bloco;  $\mu$  é a média geral;  $\tau_i$  é o efeito do *i*-ésimo tratamento;  $\beta_j$  é o efeito do *j*-ésimo bloco;  $\varepsilon_{ij}$  é o erro associado ao *i*-ésimo tratamento no *j*-ésimo bloco assumido como:  $\varepsilon_{ij} \stackrel{IID}{\sim} N(0, \sigma^2)$ .

#### 2.3 Análise de variância conjunta

Em muitas áreas da pesquisa agronômica, o pesquisador dificilmente teria condições de conduzir um experimento num determinado local com o objetivo de tirar conclusões mais amplas dos seus resultados para toda uma região, tendo em vista que as condições climáticas variam muito de um local para outro, afetando, sobremaneira, o comportamento dos seres vivos. Tal fato ocorre, por exemplo, nos experimentos com vegetais, nos quais o solo e as condições climáticas têm influência muito grande (FERREIRA, 2000).

Nestes casos, só uma experimentação apropriada permitiria ao pesquisador dar uma solução definitiva ao problema de generalização dos resultados de um experimento para toda uma região. Temos, então, a análise conjunta de experimentos.

O modelo para a análise estatística de um conjunto de observações provenientes de grupos de experimentos em blocos é:

$$Y_{ijh} = \mu + g_i + e_j + (ge)_{ij} + b_{h(j)} + \varepsilon_{ijh}$$

$$\tag{4}$$

#### em que

 $Y_{ijh}$  é a resposta do *i*-ésimo genótipo no *j*-ésimo ambiente e no *h*-ésimo bloco, com  $i = 1, 2, \ldots, g, j = 1, 2, \ldots, e$  e  $h = 1, 2, \ldots, r$ ;

 $\mu$ : é uma constante, geralmente a média geral;

 $g_i$ : é o efeito do *i*-ésimo genótipo;

 $e_j$ : é o efeito do *j*-ésimo ambiente;

 $(ge)_{ij}$ : é o efeito da interação do *i*-ésimo genótipo com o *j*-ésimo ambiente;

 $b_{h(j)}$ : é o efeito do h-ésimo bloco dentro do j-ésimo ambiente;

 $\varepsilon_{ijh}$ : é o erro experimental associado ao *i*-ésimo genótipo, no *j*-ésimo ambiente e no *h*-ésimo bloco, assumido ser independente e  $\varepsilon_{ijh} \stackrel{IID}{\sim} N(0, \sigma^2)$ .

O quadro para análise de variância para este modelo é como mostrado na Tabela 2 a seguir:

Tabela 2 - Esquema da análise de variância para experimentos de um mesmo grupo de genótipos avaliado em e locais com r blocos

Fonte de Variação	Graus de Liberdade $(GL)$	Quadrado Médio $(QM)$
Blocos d. ambiente $(Bd.E)$	e(r-1)	$QM_{Bd.E}$
$\operatorname{Gen{\acute{o}tipos}}(G)$	(g - 1)	$QM_G$
Ambiente $(E)$	(e - 1)	$QM_E$
Interação (G $\times E)$	(g-1)(e-1)	$QM_{G \times E}$
Resíduos $(Res)$	e(g-1)(r-1)	$QM_{EM}$
Total	(ger - 1)	

O modelo mais simples e comum para a análise estatística de um conjunto de dados, com essa estrutura, é dado por (DUARTE; VENCOVSKY, 1999):

$$Y_{ij} = \mu + g_i + e_j + (ge)_{ij} + \varepsilon_{ij} \tag{5}$$

em que

 $Y_{ij}$  é a resposta média do genótipo i no ambiente j; com  $i = 1, 2, \ldots, g$  e  $j = 1, 2, \ldots, e$ ;

 $\mu$ : é uma constante, geralmente a média geral;

 $g_i$ : é o efeito fixo do genótipo i;

 $e_j$ : é o efeito fixo do ambiente j;

 $(ge)_{ij}$ : é o efeito da interação específica do genótipo *i* com o ambiente *j*; e

 $\varepsilon_{ij}$ : é o erro experimental médio, assumido independente e  $\varepsilon_{ij} \stackrel{IID}{\sim} N(0, \sigma^2)$ .

O quadro de ANOVA correspondente tem, portanto, a estrutura mostrada na Tabela 3, assumindo um delineamento experimental inteiramente aleatorizado:

Tabela 3 - Esquema da análise de variância para experimentos de um mesmo grupo de genótipos avaliado em e locais com r repetições

Fonte de Variação	Graus de Liberdade $(GL)$	Quadrado Média $\left( QM\right)$
$\operatorname{Gen{\acute{o}tipos}}(G)$	(g - 1)	$QM_G$
Ambiente $(E)$	(e - 1)	$QM_E$
Interação (G $\times E)$	(g-1)(e-1)	$QM_{G \times E}$
Erro Médio/r	ge(r-1)	$QM_{EM}$
Total	(ger - 1)	

A análise efetuada com média exige cálculo a parte da estimativa da variância residual média  $(QM_{EM})$ , assumida comum a toda a série de experimentos. Isto é feito obtendose a média ponderada dos Quadrados Médios Residuais  $(QM_{Res})$  obtida de todas as ANOVA's individuais de experimentos, onde os graus de liberdade são os pesos, ou seja:

$$QM_{EM} = \frac{\sum_{j} SQ_{Res_j}}{\sum_{j=1} GL_{Res_j}};$$
(6)

 $\operatorname{com} \, j = 1, 2, \dots, e.$ 

### 2.4 Autovalores e Autovetores de uma matriz

Dado um operador linear T: V  $\rightarrow$  V, estaremos interessados, em saber quais vetores são levados em um múltiplo de si mesmo, isto é, procuraremos um vetor  $v \in$  V e um escalar  $\lambda \in R$  tais que

$$T(\boldsymbol{v}) = \lambda \boldsymbol{v}.\tag{7}$$

Neste caso  $T(\boldsymbol{v})$  será um vetor de mesma direção que  $\boldsymbol{v}$ , ou melhor,  $T(\boldsymbol{v})$  e  $\boldsymbol{v}$  estão sobre a mesma reta suporte.

Como  $\boldsymbol{v} = 0$  satisfaz a equação (7) para todo  $\lambda$ , estaremos interessados em determinar  $\boldsymbol{v} \neq 0$  que satisfaça a condição acima.

Seja T:V $\rightarrow$ V um operador linear. Um vetor  $v \in$ V,  $v \neq 0$ , é um autovetor de T se existe  $\lambda \in R$  tal que

$$T(\boldsymbol{v}) = \lambda \boldsymbol{v}.$$

O número real  $\lambda$  é denominado *autovalor* de T associado ao autovetor  $\boldsymbol{v}$ .

Sempre que um vetor  $\boldsymbol{v}$  é autovetor de um operador linear T associado ao autovalor  $\lambda$ , isto é,  $T(\boldsymbol{v}) = \lambda \boldsymbol{v}$ , o vetor  $k\boldsymbol{v}$ , para qualquer real  $k \neq 0$ , é também um autovalor de T associado ao mesmo  $\lambda$ . De fato:

$$T(k\boldsymbol{v}) = kT(\boldsymbol{v}) = k(\lambda\boldsymbol{v}) = \lambda(k\boldsymbol{v}).$$
(8)

Seja  $\boldsymbol{A}$  a matriz canônica de T, de forma que  $T_A(\boldsymbol{v}) = \boldsymbol{A}\boldsymbol{v}$ . Então, autovalores  $\lambda$  de  $\boldsymbol{A}$  e autovetores  $\boldsymbol{v}_{\lambda}$  de V são soluções da equação  $\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{v}$ . Ou seja:

$$Av = \lambda v \Leftrightarrow Av - \lambda Iv = 0 \Leftrightarrow (A - \lambda I)v = 0$$
(9)

em que I é a matriz identidade de mesma ordem que A. Observamos que a última equação,  $(A - \lambda I)v = 0$ , representa um sistema homogêneo, que é sempre compatível, cuja solução desejada, conforme definição, é a não-trivial, ou seja,  $v \neq 0$ . Assim, devemos ter o sistema em questão como compatível e indeterminado, de forma que

$$\boldsymbol{s}(\lambda) = det(\boldsymbol{A} - \lambda \boldsymbol{I}) = 0. \tag{10}$$

A função  $\mathbf{s}(\lambda)$  é um polinômio de grau p conhecido como polinômio característico da transformação e as raízes da equação  $det(\mathbf{A} - \lambda \mathbf{I}) = 0$ , chamada equação característica de  $\mathbf{A}$ (ou de T), são os autovalores de  $\mathbf{A}$  (ou de T). Para cada autovalor  $\lambda$ , corresponde um conjunto  $W = \{\mathbf{v} \in V/\mathbf{A}\mathbf{v} = \lambda \mathbf{v}\}$  de autovetores, que é obtido pela solução da equação  $\mathbf{A}\mathbf{v} = \lambda \mathbf{v}$  ou, equivalentemente, da equação  $(\mathbf{A} - \lambda \mathbf{I}) = 0$ .

As principais propriedades dos autovalores e autovetores de matrizes simétricas e de covariâncias são (ARAÚJO, 2005):

- 1. Se A é uma matriz simétrica, seus autovalores são reais;
- 2. Se A é uma matriz simétrica e seus autovalores são distintos, então os autovalores associados são ortogonais entre si e formam uma base do espaço vetorial;
- Se A é uma matriz de covariâncias e, portanto, positiva definida, então seus autovalores são sempre positivos;

- 4. Se  $\lambda_i$ , com i = 1, 2, ..., p são os autovalores de  $\boldsymbol{A}$ , então o  $Tr(\boldsymbol{A}) = \sum_{i=1}^p \lambda_i$  e  $det(\boldsymbol{A}) = \prod_{i=1}^p \lambda_i$ ;
- 5. Seja X uma matriz retangular de ordem  $n \times m$ , cujo posto seja p < min(n, m). Supondo n < m, então, os p autovalores não nulos de XX' são iguais aos p autovalores não nulos de X'X. Os demais autovalores são iguais a zero.

#### 2.5 Decomposição em valores singulares

Dada uma matriz  $\mathbf{A}$ ,  $m \times n$ , a matriz  $\mathbf{A}' \mathbf{A}$ ,  $n \times n$ , é uma matriz simétrica com *n* valores próprios (ou autovalores) reais não negativos, iguais ou distintos,  $\lambda_1^2, \lambda_2^2, \ldots, \lambda_n^2$ .

Designam-se por valores singulares da matriz A as n raízes quadradas dos valores próprios da matriz A'A,  $\lambda_i^* = \sqrt{\lambda_i^2}$ .

A decomposição em valores singulares ("singular value decomposition", DVS) é um dos resultados mais importantes da álgebra linear, tanto computacional quanto teórico, que é aplicada em diversas áreas de conhecimento. A DVS é a base dos métodos mais precisos para a resolução de problemas de mínimos quadrados, para a determinação de posto de matrizes, do espaço-imagem(R(.)) e do espaço nulo (N(.)) de matrizes e de solução de vários problemas envolvendo normas euclidianas $(\parallel . \parallel)$  (ARAÚJO, 2005).

Qualquer matriz X de ordem  $n \times m$ , com posto p < min(n,m), é possível expressar cada elemento de X da seguinte maneira (MANDEL, 1982):

$$x_{ij} = \lambda_1 \gamma_{1i} \alpha_{1j} + \lambda_2 \gamma_{2i} \alpha_{2j} + \ldots + \lambda_p \gamma_{pi} \alpha_{pj} = \sum_{k=1}^p \lambda_k \gamma_{ki} \alpha_{kj},$$
(11)

em que  $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k \ldots \geq \lambda_p$ . Esta é a definição de DVS da matriz  $\boldsymbol{X}$ .

Os p vetores  $\gamma_i$  são ortogonais a cada vetor  $\gamma_j$ , para  $i \neq j$  e i, j = 1, 2, ..., p, assim como os p vetores  $\alpha_i$ . Além disso, considera-se que cada um desses vetores tem tamanho unitário (vetores normalizados), ou seja

$$\sum_{i} \boldsymbol{\gamma}_{ki} = \sum_{j} \boldsymbol{\alpha}_{kj} = 1, \qquad \forall k.$$
(12)

Então  $\boldsymbol{X}$  pode ser expressa na forma:

$${}_{n}\boldsymbol{X}_{m} = {}_{n}\boldsymbol{U}_{p}\boldsymbol{S}_{p}\boldsymbol{V}_{m}^{'}$$

$$\tag{13}$$

A matriz S é diagonal cujos elementos na diagonal são os valores singulares e todos os  $\lambda_k$ são positivos. As colunas da matriz U são os vetores  $\gamma$  e as colunas de V são os vetores  $\alpha$ da equação (11) e (12). Com a ortogonalidade e a normalidade dos vetores  $\gamma$  e  $\alpha$  pode ser expressa na forma matricial:

$$\boldsymbol{U}'\boldsymbol{U} = \boldsymbol{V}\boldsymbol{V}' = \boldsymbol{I}_p. \tag{14}$$

Sendo que I indica uma matriz identidade de ordem p. Os valores singulares  $\lambda_k$  são obtidos, extraindo-se a raiz quadrada dos autovalores diferentes de zero da matriz X'X ou a raiz quadrada dos autovalores diferentes de zero da matriz XX'. As colunas de U são os autovetores da matriz XX', ordenadas de acordo com a magnitude dos autovalores e as linhas de V' são os autovetores da matriz X'X, ordenadas de acordo com a magnitude dos autovalores (ARAÚJO, 2005).

### 2.6 Análise de Componentes Principais

A análise de componentes principais é uma técnica multivariada de modelagem da estrutura de covariância. A técnica foi inicialmente descrita por Pearson (1901) e uma descrição de métodos computacionais práticos veio muito mais tarde com Hotelling (1933, 1936) quem a usou com o propósito determinado de analisar as estruturas de correlação.

A análise de componentes principais (ACP) tem como objetivo reduzir a dimensionalidade do conjunto de variáveis e facilitar a interpretação da independência entre elas. Para isso, são obtidas combinações lineares das variáveis originais que, geometricamente representam a seleção de novos sistemas de coordenadas obtidos pela rotação do sistema original que tem p variáveis aleatórias como eixos das coordenadas. Esses novos eixos ortogonais (novas variáveis) são chamados de *componentes principais* e os valores das novas variáveis são chamadas de *escores dos componentes principais* ou *coordenadas principais*. Essas novas variáveis são não correlacionadas (PIOVESAN 2007).

Em geral escolhemos o componente principal de maior importância (o primeiro componente principal) como sendo aquele de maior variância  $(\max_{i} \lambda_{i})$ , que explique o máximo de variabilidade dos dados, o segundo componente de maior importância, o que apresenta a segunda maior variância e assim sucessivamente, até o componente principal de menor importância.

A obtenção dos componentes principais é realizada por meio da diagonalização de matrizes simétricas positivas semi-definidas. Então, podemos calcular os componentes principais facilmente e utilizá-los em diferentes aplicações nas mais variadas áreas científicas. Esta facilidade é função da existência de inúmeros programas capazes de realizar cálculos matriciais para diagonalizar uma matriz simétrica positiva semi-definida. Muitos pesquisadores têm utilizado a análise de componentes principais para resolver problemas como da multicolinearidade em regressão linear, para estimar fatores, que representam outra técnica multivariada de modelagem da matriz de covariâncias, para realizar a modelagem da interação entre fatores em experimentos sem repetição, estudos de divergência e agrupamento entre genótipos em estudo de genética e melhoramento de plantas e animais, entre outras possibilidades (JOHNSON; WICHERN, 1998).

Para obter os componentes principais de uma forma geral, seja um conjunto de pvariáveis  $X_1, X_2, \dots, X_p$  com médias  $\mu_1, \mu_2, \dots, \mu_p$  e variância  $\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2$ , respectivamente. Estas variáveis não são independentes e portanto, possuem covariância entre a *i*-ésima e *k*-ésima variável definida por  $\sigma_{ik}$ , para  $i \neq k = 1, 2, \dots, p$ . Então as p variáveis podem ser expressas na forma vetorial por:  $\mathbf{X} = [X_1, X_2, \dots, X_p]'$ , com vetor de médias  $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_p]'$  e matriz de covariância  $\boldsymbol{\Sigma}$ . Encontram-se os pares de autovalores e autovetores  $(\lambda_1, e_1), (\lambda_2, e_2), \dots,$  $(\lambda_p, e_p)$ , em que  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ , associados a  $\boldsymbol{\Sigma}$  e então o *i*-ésimo componente principal é definido por:

$$\mathbf{Z}_{i} = \mathbf{e}_{i}^{'} \mathbf{X} = e_{i1} X_{1} + e_{i2} X_{2} + \dots + e_{ip} X_{p}$$
 (15)

em que  $i = 1, 2, \dots, p$ .

A variável  $Z_i$ , é uma variável latente, ou seja, não é mensurada a partir do experimento ou levantamento amostral (JOHNSON; WICHERN, 1998). O objetivo é determiná-la a partir das p variáveis contidas no vetor X. A idéia é projetar os pontos coordenados originais em um plano maximizando a distância entre eles, o que equivale a maximizar a variabilidade da variável latente  $Z_i$ . A variância de  $Z_i$  é dada por

$$Var(\boldsymbol{Z}_{i}) = Var(\boldsymbol{e}_{i}'\boldsymbol{X}) = \boldsymbol{e}_{i}'Var(\boldsymbol{X})\boldsymbol{e}_{i} = \boldsymbol{e}_{i}'\boldsymbol{\Sigma}\boldsymbol{e}_{i}$$
(16)

com  $i = 1, 2, \cdots, p$ , e também que

$$Cov(\boldsymbol{Z}_{i}, \boldsymbol{Z}_{k}) = \boldsymbol{e}_{i}^{'} \boldsymbol{\Sigma} \boldsymbol{e}_{k} = \boldsymbol{e}_{i}^{'} \lambda_{k} \boldsymbol{e}_{k} = \lambda_{k} \boldsymbol{e}_{i}^{'} \boldsymbol{e}_{k} = 0$$
(17)
com  $i \neq k$ , uma vez que  $e_i$  e  $e_k$  são ortogonais.

Se utilizamos a decomposição espectral da matriz  $\Sigma$ , dada por  $\Sigma = P\Lambda P'$ , em que P é a matriz composta pelos autovetores de  $\Sigma$  em suas colunas e  $\Lambda$ , a matriz diagonal de autovalores de  $\Sigma$ , então, podemos observar que

$$tr(\boldsymbol{\Sigma}) = tr(\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}') = tr(\boldsymbol{\Lambda}\boldsymbol{P}'\boldsymbol{P}) = tr(\boldsymbol{\Lambda}\boldsymbol{I}) = tr(\boldsymbol{\Lambda}) = \sum_{i=1}^{p} \lambda_{i}.$$
 (18)

Mas,  $tr(\Sigma)$  é obviamente dada pela soma dos elementos da diagonal, ou seja,

$$tr(\mathbf{\Sigma}) = \sum_{i=1}^{p} \sigma_{ii},\tag{19}$$

portanto,

$$\sum_{i=1}^{p} \sigma_{ii} = \sum_{i=1}^{p} \lambda_i, \tag{20}$$

que significa que a variabilidade total contida nas variáveis originais é igual a variabilidade total contida nos componentes principais (JOHNSON; WICHERN, 1998).

## 2.7 Histórico dos modelos AMMI

Em melhoramento de plantas, ensaios multi-ambientes (MET-multienvironment trials) são importantes para testar adaptação geral e específica de cultivar. Um cultivar crescendo em diferentes ambientes irá freqüentemente mostrar flutuações significantes na performance de produção relativa a outras cultivares. Essas mudanças são influenciadas por diferentes condições experimentais e são referidas como GEI (DIAS, 2005).

O termo GEI como sendo linearmente relacionados ao efeito de ambiente, isto é colocar  $(ge)_{ij} = \xi_i e_j + d_{ij}$  em que  $\xi_i$  é o coeficiente de regressão linear do *i*-ésimo genótipo na média ambiental e  $d_{ij}$  é um desvio (YATES; COCHRAN, 1938). Este procedimento foi posteriormente usado por Finlay e Wilkinson (1963) e modificado por Eberhart e Russell (1966). Tukey (1949) propôs um teste para GEI usando  $(ge)_{ij} = Kg_ie_j$  (K é uma constante). Mandel (1961) generalizou o modelo de Tukey colocando  $(ge)_{ij} = \lambda \gamma_i e_j$  para genótipos e  $(ge)_{ij} = \lambda g_i \alpha_j$  para ambientes e assim obtendo um feixe de linhas retas que podem ser testadas para concorrência (isto é, se os  $\gamma_i$  ou os  $\alpha_j$  são todos os mesmos) ou não concorrência (DIAS, 2005). Gollob (1968) e Mandel (1969, 1971) propuseram um termo bilinear GEI  $(ge)_{ij} = \sum_{k=1}^{n} \lambda_k \gamma_{ik} \alpha_{jk}$  no qual  $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n$  e  $\gamma_{ik}, \alpha_{jk}$  satisfazem a restrição de ortonormalidade  $\sum_i \gamma_{ik} \gamma_{ik'} = \sum_j \alpha_{jk} \alpha_{jk'} = 0$  para  $k \ne k'$  e  $\sum_i \gamma_{ik}^2 = \sum_j \alpha_{jk}^2 = 1$ .

Isto conduz ao modelo linear-bilinear  $\overline{Y}_{ij} = \mu + g_i + e_j + \sum_{k=1}^n \lambda_k \gamma_{ik} \alpha_{jk} + \overline{\varepsilon}_{ij}$ , o qual é uma generalização da regressão no modelo de média, com mais flexibilidade para descrever GEI porque mais que uma dimensão genótipo ambiente é considerada (DIAS, 2005). Zobel, Wright e Gauch (1988) e Gauch (1988) chamaram esse modelo de "efeitos principais aditivos e interação multiplicativa" AMMI-Additive Main Effects and Multiplicative Interaction Model.

Uma família de modelos multiplicativos pode então ser gerada retirando-se o efeito principal de genótipo (modelo de regressão para ambientes), (SREG – Site Regression Model), o efeito principal de ambientes (GREG – Genotype Regression Model), ou ambos efeitos (COMM – Complete Multiplicative Model) (DIAS, 2005). Um outro modelo multiplicativo (SHMM – Shifted Multiplicative Model), é útil para estudar crossover GEI (CROSSA; CORNELIUS; YAN, 2002).

#### 2.8 Modelo AMMI

A interação de genótipos com ambientes  $(G \times E)$  representa uma das principais dificuldades encontradas pelo melhorista durante sua atividade seletiva (DUARTE; VEN-COVSKY, 1999). Embora várias metodologias tenham sido proposta ao longo dos últimos trinta anos, visando mitigar os desafios impostos pela interação  $G \times E$ , os pesquisadores atuais continuam buscando instrumentos estatísticos ainda mais efetivos nessa tarefa.

A aplicação de métodos multivariados pode ser útil para melhorar e explorar as informações contidas nos dados (CROSSA, 1990). Recomenda, então, técnicas como a *análise de componentes principais* (ACP), a análise de agrupamentos e o procedimento *AMMI* (*additive main effects and multiplicative interaction analysis*), que vem ganhando grande aplicabilidade nos últimos anos (DUARTE; VENCOVSKY, 1999).

Adota-se a análise AMMI tendo em mente vários aspectos de interesse para o melhorista. Um deles é o de procurar entender melhor o fenômeno da interação  $G \times E$ . Vale lembrar que, tradicionalmente, a manifestação dessas interações era vista, muitas vezes, como tendo uma natureza errática, sem uma explicação simples. Conforme será exposto, o método AMMI permite contornar esta dificuldade, pois fornece uma lógica ou padrão subjacente na manifestação das interações.

Numa análise conjunta, com dados de avaliação de n genótipos em m ambientes, existem m.n efeitos de interação que são estimáveis. Cada um desses efeitos, portanto, é desdobrado nas mencionadas componentes. Uma dada componente, por sua vez, é formada por uma contribuição do genótipo e outra do ambiente, contribuições essas que são multiplicadas entre si, além de uma constante, para gerar a parte da interação explicada pela componente em questão. Essa idéia de quantificar um efeito de interação pela ação conjunta do genótipo e do ambiente, pela multiplicação das respectivas contribuições, é o fundamento básico deste modelo.

Vale atentar para o significado da sigla AMMI. Ela especifica que a média de uma variável, obtida num genótipo e num certo ambiente, é resultado dos efeitos principais, genéticos e ambientais, que se somam, ou são aditivos (*AM: Additive Main effects*) no modelo geral. Os efeitos de interação, por outro lado, são resultado da multiplicação entre uma contribuição do genótipo e outra do ambiente (*Multiplicative Interaction effects*). Gauch e Zobel (1996) informam que essa análise pode ajudar tanto na identificação de genótipos de alta produtividade e largamente adaptados, como na realização do chamado *zoneamento agronômico*, com fins de recomendação regionalizada e seleção de locais de teste.

Suponha que um conjunto de g genótipos tenha sido testado experimentalmente em e ambientes (locais, anos etc). Assim, as médias de cada combinação de genótipo e ambiente obtidas de r repetições por experimento (um conjunto balanceado de dados), podem ser representadas pela matriz:

$$\boldsymbol{Y}_{g \times e} = \begin{pmatrix} \bar{Y}_{11} & \bar{Y}_{12} & \cdots & \bar{Y}_{1e} \\ \bar{Y}_{21} & \bar{Y}_{22} & \cdots & \bar{Y}_{2e} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{Y}_{g1} & \bar{Y}_{g2} & \cdots & \bar{Y}_{ge} \end{pmatrix}$$
(21)

Sendo a interação significativa, o próximo passo é fazer a decomposição da  $SQ_{G\times E}$ , para descartar um resíduo adicional presente nessa soma de quadrados (ARAÚJO,

2005). Essa decomposição é feita utilizando o fator analítico proposto por Gollob (1968) e Mandel (1969, 1971) e tem a seguinte expressão:

$$(ge)_{ij} = \sum_{k=1}^{p} \lambda_k \gamma_{ik} \alpha_{jk}, \qquad (22)$$

em que  $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ , e  $\gamma_{ik}$ ,  $\alpha_{jk}$  satisfazem o contraste de orto-normalização  $\sum_i \gamma_{ik} \gamma'_{ik} = \sum_j \alpha_{jk} \alpha'_{jk} = 0$  para  $k \neq k'$  e  $\sum_i \gamma^2_{ik} = \sum_j \alpha^2_{jk} = 1$ .

O modelo AMMI pressupõe componentes aditivos para os efeitos principais de genótipos  $(g_i)$  e de ambientes  $(e_j)$  e componentes multiplicativos para o efeito da interação  $(ge)_{ij}$ . Assim, a resposta média sobre r repetições ou blocos do *i*-ésimo genótipo no *j*-ésimo ambiente é representada por:

$$Y_{ij} = \mu + g_i + e_j + \sum_{k=1}^n \lambda_k \gamma_{ik} \alpha_{jk} + \rho_{ij} + \varepsilon_{ij}$$
(23)

Os componentes aditivos e multiplicativos do modelo podem ser escritos da seguinte forma:

parte aditiva  

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

parte multiplicativa

$$\downarrow$$

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

com: n

né número de eixos ou componentes principais selecionados para descrever o padrão da interação G $\times$ E;

sendo que:

 $Y_{ij}$ : é a resposta média do *i*-ésimo genótipo no *j*-ésimo ambiente, com i = 1, 2, ..., g e j = 1, 2, ..., e;

 $g_i$ : é o efeito do *i*-ésimo genótipo;

 $e_j$ : é o efeito do *j*-ésimo ambiente;

 $\lambda_k$ : é a raiz quadrada do k-ésimo autovalor da matriz  $(\mathbf{GE})(\mathbf{GE}^T)$  ou  $(\mathbf{GE}^T)(\mathbf{GE})$ , com  $k = 1, 2, \dots, p \in p = \min\{g - 1, e - 1\};$ 

 $\gamma_{ik}$ : é o *i*-ésimo elemento do vetor coluna  $\alpha_k$  associado a  $\lambda_k$ ;

 $\alpha_{jk}$ : é o j-ésimo elemento do vetor linha  $\gamma_k$  associado a  $\lambda_k$ ;

 $\rho_{ij}$ : é o resíduo adicional;

 $\varepsilon_{ij}$ : é erro experimental associado ao *i*-ésimo genótipo no *j*-ésimo ambiente, assumido ser independente e  $\varepsilon_{ij} \stackrel{IID}{\sim} N(0, \frac{\sigma^2}{r}).$ 

O termo  $(ge)_{ij}$  (interação no modelo tradicional), na metodologia AMMI é representado pela soma de p parcelas, cada uma da multiplicação de  $\lambda_k$ , expresso na mesma unidade de  $\boldsymbol{Y}_{ij}$ , por um efeito de genótipo  $(\gamma_{ik})$  e um efeito ambiental  $(\alpha_{jk})$ , ambos adimensionais, ou seja,  $\sum_{k=1}^{n} \lambda_k \gamma_{ik} \alpha_{jk}$  (n: termos da interação). O termo  $\lambda_k$  traz uma informação relativa à variação devida à interação  $G \times E$ , na k-ésima parcela. De forma que a soma das p parcelas recompõe toda a variação  $(SQ_{G\times E} = \sum_{k=1}^{p} \lambda_k^2)$ , e os efeitos  $\gamma_{ik}$  e  $\alpha_{jk}$  representam os pesos do genótipo i e do ambiente j, naquela parcela da interação (DUARTE; VENCOVSKY, 1999).

Pela abordagem AMMI não se busca recuperar toda a  $SQ_{G\times E}$ , mas apenas a parcela mais fortemente determinada por genótipo e ambiente (linhas e colunas da matriz GE), ou seja, o padrão (parte determinística ou sistemática). Assim, a interação do genótipo i com o ambiente j $(\sum_{k=1}^{n} \lambda_k \gamma_{ik} \alpha_{jk})$ , descartando-se o resíduo adicional  $\rho_{ij}$  dado por:  $\sum_{k=n+1}^{p} \lambda_k \gamma_{ik} \alpha_{jk}$ . Aqui, como em ACP, esses eixos captam, sucessivamente, porções cada vez menores da variação presente na matriz GE  $(\lambda_1^2 \ge \lambda_2^2 \ge \ldots \ge \lambda_p^2)$ . Por isso, o método AMMI é visto como um procedimento capaz de separar padrão e ruído na análise da  $SQ_{G\times E}$ :  $\sum_{k=1}^{n} \lambda_k \gamma_{ik} \alpha_{jk}$  e  $\sum_{k=n+1}^{p} \lambda_k \gamma_{ik} \alpha_{jk}$ , respectivamente (DUARTE; VENCOVSKY, 1999; WEBER; WRICKE; WESTERMANN, 1996).

A matriz GE é matriz de interação entre genótipos × ambiente (matriz de resíduos) obtida do modelo (21), ou seja, cada elemento  $(ge)_{ij}$  da matriz GE é encontrado pela seguinte relação:

$$(ge)_{ij} = Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{.}$$

em que:

 $Y_{ij}$ : é a média das repetições do genótipo *i* no ambiente j, com i = 1, 2, ..., g e j = 1, 2, ..., e;  $\overline{Y}_{i}$ : é a média do genótipo *i*;

 $\overline{Y}_{.j}$ : é a média do amiente j;

 $\overline{Y}_{..}$  é a média geral do experimento.

Uma propriedade da DVS que determina a partição da  $SQ_{G\times E}$  na análise AMMI é como segue considere-se, a princípio, a decomposição exata da matriz  $\mathbf{G}E_{(g\times e)}$  por DVS:

$$\underbrace{\boldsymbol{GE}}_{\boldsymbol{B}} = \underbrace{\boldsymbol{U}}_{(\boldsymbol{P}_{BB'})} \underbrace{\boldsymbol{S}}_{(\boldsymbol{\Lambda}^{1/2})} \underbrace{\boldsymbol{V}'}_{(\boldsymbol{P}_{B'B})} = \sum_{k=1}^{p} \lambda_k \gamma_k \boldsymbol{\alpha}'_k;$$
(24)

com  $p = min\{(g-1), (e-1)\}.$ 

Portanto, a DVS de GE pode ser escrita como uma soma de p matrizes de posto unitário, podemos obter a partição da soma de quadrados dos elementos da matriz GE:

$$\underbrace{\mathbf{GE}}_{\substack{[\mathbf{B}=[b_{ij}]\\(g\times e)}} = \underbrace{\lambda_1 \gamma_1 \alpha'_1}_{\substack{[\mathbf{M}=[b_{1ij}]\\(g\times e)}} + \underbrace{\lambda_2 \gamma_2 \alpha'_2}_{\substack{[\mathbf{M}=[b_{2ij}]\\(g\times e)}} + \dots + \underbrace{\lambda_p \gamma_p \alpha'_p}_{\substack{[\mathbf{M}=[b_{1ij}]\\(g\times e)}} \right) \tag{25}$$

$$\sum_{i,j} b_{ij}^2 = \sum_{i,j} (g\hat{e})_{ij}^2 = \sum_{i,j} b 1_{ij}^2 + \sum_{i,j} b 2_{ij}^2 + \dots + \sum_{i,j} b p_{ij}^2 = SQ_{G \times E}$$
(26)

$$\sum_{k} \lambda_{k}^{2} = \sum_{i,j} \left(g\hat{e}\right)_{ij}^{2} = \lambda_{1}^{2} + \lambda_{2}^{2} + \dots + \lambda_{p}^{2} = \sum_{k} SQ_{IPCA_{k}} = SQ_{G\times E}$$
(27)

em que o termo IPCA vem de "Incremental Principal Component Analysis", ou seja,  $IPCA_k$ : o efeito de interação do componente principal associado ao k-ésimo eixo.

O modelo AMMI procura explicar a  $SQ_{G\times E}$  por meio de uma aproximação de posto n para a matriz GE (com n preferivelmente bem menor do que p). Ou seja, o objetivo da análise é descrever a interação  $G \times E$  por um número reduzido de eixos ( $n = 1 \ a \ 3$ ), resultando num modelo informativo, mas que consuma poucos graus de liberdade. Com isso, o modelo é capaz de captar o padrão subjacente à interação e descartar os ruídos presentes nos dados e sem interesse agronômico, os quais só prejudicam a capacidade preditiva do modelo. Essa separação em padrão e ruídos pode ser expressa em termos de soma de quadrados da seguinte forma (DUARTE; VENCOVSKY, 1999):

$$SQ_{G\times E} = \sum_{k=1}^{p} \lambda_{k}^{2} = (\sum_{k=1}^{n} \lambda_{k}^{2}) + (\sum_{k=n+1}^{p} \lambda_{k}^{2}) = (SQ_{G\times E[padr\tilde{a}o]}) + (SQ_{G\times E[ruidos]})$$
(28)

O número de eixos ou componentes da ACP, retidos para a maioria das aplicações, é o menor possível (dois ou três no máximo), com o qual se pretende reduzir a dimensão do sistema e fornecer uma descrição mais parcimoniosa à cerca da estrutura da interação (LAVORANTI, 2003). Um dos procedimentos usuais adotados para a definição do número de eixos a serem retidos, de maneira a explicar o padrão relacionado à interação, consiste em determinar graus de liberdade associados à parcela da  $SQ_{G\times E}$  relacionada a cada membro da família AMMI (DUARTE; VENCOVSKY, 1999). Obtém-se, então, o quadrado médio (QM) correspondente a cada parcela (ou modelo). Em seguida é obtido um teste F, avaliando-se a significância de cada componente em relação ao  $QM_{erro\ médio}$ . Dessa forma o ponto de parada que determina a seleção do modelo ( $AMMI_0, AMMI_1, \dots$  ou  $AMMI_n$ ) baseia-se na significância do teste F para os sucessivos eixos da interação. O resíduo AMMI, reunindo os eixos descartados da interação, também pode ser testado de maneira a assegurar o seu caráter desprezível.

Existem várias técnicas para atribuir os graus de liberdade associados à cada parcela da  $SQ_{G\times E}$ , ou seja, associada a  $\hat{\lambda}_k^2$ . O método de Gollob (1968) destaca-se por ser muito fácil de ser aplicado e exigir pequenas operações aritméticas:

$$GL_{IPCA_K} = g + e - 1 - 2k.$$
 (29)

A representação gráfica em biplot, é uma técnica bastante útil na análise de componentes principais pois, o gráfico usado para representar simultaneamente as linhas e colunas, referentes a uma matriz de dados, pode indicar a existência de agrupamentos entre as observações, assim como, mostrar as variâncias e correlações entre as variáveis (LAVORANTI, 2003). Essa técnica foi introduzida por Gabriel (1971), e é baseada no fato de que, uma matriz de ordem maior do que dois pode ser aproximada por uma matriz de ordem dois, sendo que cada elemento da matriz é representado pelo produto interno dos vetores correspondentes à sua linha e coluna. Dessa forma, qualquer matriz de n observações e p variáveis pode ser representada por um biplot.

# 2.9 Estimação de autovalores e distribuição amostral

Seja  $\Sigma$  uma matriz simétrica e positiva definida de ordem p, pode-se encontrar um conjunto de autovalores  $\lambda_1 > \lambda_2 > \ldots > \lambda_p > 0$  e autovetores  $e_1, e_2, \ldots, e_p$ . De forma análoga, é possível encontrar os autovalores e os autovetores de uma matriz S que estima  $\Sigma$ , com os autovalores  $\hat{\lambda}_1 > \hat{\lambda}_1 > \ldots > \hat{\lambda}_p > 0$  e os autovetores  $\hat{e}_1, \hat{e}_2, \ldots, \hat{e}_p$ , com decomposições espectrais  $S = \hat{P} \hat{\Lambda} \hat{P}^T$  e  $\Sigma = P \Lambda P^T$ . Tomando uma amostra aleatória  $X_1, X_2, \ldots, X_m$ , proveniente de uma população normal multivariada de dimensão m com matriz de covariância populacional  $\Sigma$  e vetor de média  $\mu$  conhecido é possível obter a matriz S. Para uma amostra grande, que cada  $\hat{\lambda}_i$  têm uma distribuição independente e também tem uma distribuição assintoticamente normal, ou seja,  $\hat{\lambda}_i \stackrel{a}{\sim} N(\lambda_i, \frac{2\lambda_i^2}{n})$  (JOHNSON; WICHERN, 1998; ANDERSON, 1958). Araújo e Dias (2005) confirmam essa declaração, simulando amostras multivariadas e depois verificando a aderência da distribuição dos autovalores à distribuição normal especificada.

Podemos utilizar este resultado para obter um intervalo de aproximadamente 100(1- $\alpha$ )% de confiança para estimarmos  $\lambda_i$ . Assim, a partir da afirmativa probabilística

$$P(-Z_{\alpha/2} \le \frac{\hat{\lambda}_i - \lambda_i}{\lambda_i \sqrt{2/(n-1)}} \le Z_{\alpha/2}) \approx 1 - \alpha, \tag{30}$$

Desta forma o intervalo de aproximadamente  $100(1-\alpha)\%$  de confiança para  $\lambda_i$  é

$$\frac{\hat{\lambda}_i}{1 + Z_{\alpha/2}\sqrt{2/(n-1)}} \le \lambda_i \le \frac{\hat{\lambda}_i}{1 - Z_{\alpha/2}\sqrt{2/(n-1)}},\tag{31}$$

em que  $Z_{\alpha/2}$  é o quantil superior  $100\alpha/2$  da normal padrão e assumindo que  $Z_{\alpha/2}\sqrt{2/(n-1)} < 1$  (JOHNSON; WICHERN, 1998; ANDERSON, 1958).

Para evitarmos o inconveniente de ter de satisfazer a condição  $Z_{\alpha/2}\sqrt{2/(n-1)} < 1$ , temos duas alternativas. A primeira é utilizamos o estimador da variância de  $\hat{\lambda}_i$  dada por  $\hat{\lambda}_i^2/(n-1)$ na afirmativa de probabilidade anterior, que resulta no seguinte intervalo de aproximadamente 100(1- $\alpha$ )% de confiança:

$$\hat{\lambda}_i - Z_{\alpha/2}\hat{\lambda}_i \sqrt{\frac{2}{n-1}} \le \lambda_i \le \hat{\lambda}_i + Z_{\alpha/2}\hat{\lambda}_i \sqrt{\frac{2}{n-1}}.$$
(32)

A segunda alternativa é obter o intervalo de confiança, considerando a distribuição de  $\ln(\hat{\lambda}_i)$ , que é aproximadamente dada por

$$\ln(\hat{\lambda}_i) \sim N(\ln(\lambda_i), \frac{2}{n-1}).$$

Esta distribuição possui variância que independe de  $\lambda_i$ , de onde facilmente podemos obter que o intervalo de aproximadamente  $100(1-\alpha)\%$  de confiança para  $\ln(\lambda_i)$  é dado por  $\ln(\hat{\lambda}_i) \pm Z_{\alpha/2}\sqrt{2/(n-1)}$ . Se utilizamos a transformação inversa para obtermos o intervalo de aproximadamente  $100(1-\alpha)\%$  de confiança para  $\lambda_i$  teremos

$$\hat{\lambda}_{i} \exp\left\{-Z_{\alpha/2}\sqrt{\frac{2}{n-1}}\right\} \le \lambda_{i} \le \hat{\lambda}_{i} \exp\left\{Z_{\alpha/2}\sqrt{\frac{2}{n-1}}\right\}$$
(33)

Os p autovalores, variâncias dos componentes principais, são independentes assintoticamente e a região de confiança para todos os  $\lambda_i$  é obtida combinando simplesmente os intervalos individuais dados pelas equações (31), (32) e (33), escolhendo um coeficiente de confiança individual para obtermos um coeficiente de confiança global desejado. Isso pode ser feito substituindo  $Z_{\alpha/2}$ em todos eles por  $Z_{\alpha^*/2}$ , em que  $\alpha^*$  pode ser determinada pelo procedimento de Bonferroni dado por  $\alpha^* = \alpha/m$ , em que m = p se todos autovalores forem considerados ou m < p, se apenas um subconjunto de intervalos for obtidos (JOHNSON; WICHERN, 1998; ANDERSON, 1958).

### 2.10 O método Bootstrap e estimador do erro padrão

O método *bootstrap* foi proposto por Efron (1979), como um procedimento de reamostragem amplamente utilizado na obtenção de estimativas pontuais e intervalares, bem como na avaliação da acurácia de estimativas e testes. Segundo Martinez e Louzada-Neto (2001) se comparado a outras técnicas estatísticas, o método teve uma gênese um tanto tardia, devido à sua dependência do uso de computadores. Os progressos da informática experimentados nas últimas décadas do século XX possibilitaram a popularização do uso do computador e incrementaram o surgimento e acesso a *softwares* matemáticos e estatísticos. Conseqüentemente, as aplicações de métodos *bootstrap* nas mais diferentes áreas da estatística se intensificaram.

Procedimentos *bootstrap* para a identificação de valores críticos para ajustar estatísticas são fáceis de implementar e de fato, tem sido implementado em três programas distintos que fazem interface com o software comercial para identificar os ajustes estatísticos razoáveis, dado um conjunto de itens estimados e parâmetros personalizado (STONE, 2007; SU; SHEU; WANG, 2007; WOLFE, 2008).

É sabido que as distribuições nulas de estatísticas comumente empregadas a ajustes estatísticos não seguem uma distribuição com uma forma paramétrica conhecida (KARABATSOS, 2000; MOLENAAR; HOIJTINK, 1990; SMITH, 1988, 1991; WANG; CHEN, 2005). Nesse caso o *bootstrap* é muito útil, pois é uma técnica que não exige diferentes fórmulas para cada problema e pode ser utilizada em casos gerais, não dependendo da distribuição original da estatística do parâmetro estudado.

O *bootstrap* pode ser implementado tanto na estatística não-paramétrica quanto na paramétrica, dependendo apenas do conhecimento do problema. No caso não-paramétrico, o método *bootstrap* reamostra os dados com reposição, de acordo com uma distribuição empírica estimada, tendo em vista que, em geral, não se conhece a distribuição subjacente aos dados. No caso paramétrico, quando se tem informação suficiente sobre a forma da distribuição dos dados, a amostra *bootstrap* é formada realizando-se a amostragem diretamente nessa distribuição com parâmetros desconhecidos substituídos por estimativas paramétricas. A distribuição da estatística de interesse aplicada aos valores da amostra *bootstrap*, condicional aos dados observados, é definida como a distribuição *bootstrap* dessa estatística (LAVORANTI, 2003).

Seja uma amostra aleatória baseada em n observações independentes  $x_1, x_2, \ldots, x_n$ . O erro padrão de uma média  $\bar{\mathbf{x}}$  baseada nesta amostra é estimado pela expressão:

$$\widehat{ep}(\bar{x}) = \sqrt{\frac{s^2}{n}},\tag{34}$$

em que

$$s^{2} = \frac{\sum_{i=1}^{n} (x_{i} - \bar{x})^{2}}{n - 1}$$
(35)

é o estimador não viciado da variância. Note-se que o erro padrão não é uma estimativa de uma quantidade pertinente a uma população, mas uma medida da incerteza da média amostral vista como uma estimativa da média populacional (ALTMAN, 1991). A expressão (34) deixa claro que a magnitude desta incerteza diminui conforme o tamanho da amostra n aumenta (MARTINEZ; LOUZADA-NETO, 2001).

Seja X uma variável aleatória com distribuição F, sendo sua esperança denotada por  $\mu_F$  e sua variância denotada por  $\sigma_F^2$ . Usaremos a notação  $X \sim (\mu_F, \sigma_F^2)$  e escrevemos  $\hat{F}_n \rightarrow$  $(x_1, x_2, \ldots, x_n)$  para indicar que  $\mathbf{x} = (x_1, x_2, \ldots, x_n)$  é uma amostra aleatória de tamanho n obtida de uma população com função de probabilidade F. A média  $\bar{X}$  é também uma variável aleatória e tem esperança  $\mu_F$  e variância  $\sigma_F^2/n$ , ou seja,  $\bar{X} \sim (\mu_F, \sigma_F^2/n)$ . Note-se então que X e  $\bar{X}$  têm a mesma esperança, entretanto, o desvio padrão de  $\bar{X}$  é definido como a raiz quadrada da variância de  $\bar{X}$ (MARTINEZ; LOUZADA-NETO, 2001), ou seja,

$$ep_F(\bar{X}) = \sqrt{\operatorname{var}_F(\bar{X})} = \frac{\sigma_F}{\sqrt{n}}$$
(36)

A expressão (34) fornece um estimador para (36). Entretanto, nem todos os estimadores têm expressões de fácil manejo para seu respectivo erro padrão, como (34). Isto significa que o trabalho de encontrar medidas de precisão para outros estimadores  $\hat{\theta} = s(X)$ , que não a média, pode ser algo bastante complicado. Sob certas condições, o teorema central do limite diz que a distribuição de  $\bar{X}$  é aproximadamente normal quando n é grande (LEITE; SINGER, 1990), ou seja,

$$\bar{X} \sim N(\mu_F; \sigma_F^2/n), \tag{37}$$

e conseqüentemente,

$$\frac{\bar{X} - \mu_F}{\sigma_F / \sqrt{n}} \sim N(0; 1). \tag{38}$$

Usando uma tabela da distribuição normal padrão, de (38) podemos escrever, por exemplo, a probabilidade

$$P(\left|\bar{X} - \mu_F\right| < \frac{2\sigma_F}{\sqrt{n}}) \doteq 0,954,$$
(39)

em que o erro padrão mostra-se uma útil medida da precisão da estimativa da média. A probabilidade da distância entre a média amostral e a populacional ser de pelo menos dois erros padrão é aproximadamente 0,954. Verifica-se assim, que quanto menor o erro padrão da média, a expressão (39) sugere uma melhor aproximação entre as médias amostral e populacional (MARTINEZ; LOUZADA-NETO, 2001).

Uma vantagem do método *bootstrap* é que esta técnica não depende inteiramente do teorema central do limite, já que, em suas aplicações, medidas de precisão são obtidas diretamente dos dados (EFRON; TIBSHIRANI, 1993).

Observada uma amostra aleatória de tamanho n, oriunda de uma distribuição F, define-se uma função distribuição empírica  $\hat{F}_n$  como uma distribuição discreta, que atribui probabilidade  $n^{-1}$  a cada valor  $x_i$ , i = 1, ..., n. Uma amostra bootstrap  $\boldsymbol{x}^* = (x_1^*, x_2^*, ..., x_n^*)$  é obtida reamostrando aleatoriamente n vezes, com reposição, as observações  $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ , onde verifica-se que  $\hat{F}_n \to (x_1^*, x_2^*, ..., x_n^*)$ .

Selecionadas *B* amostras *bootstrap*,  $\boldsymbol{x}^{*1}$ ,  $\boldsymbol{x}^{*2}$ , ...,  $\boldsymbol{x}^{*B}$ , de forma independente, estimase  $\theta$  em cada uma destas amostras através de  $\hat{\theta}^{*b} = s(\boldsymbol{x}^{*b})$ , b = 1, 2, ..., B. Uma expressão para o estimador *bootstrap* do erro padrão da estatística  $\hat{\theta}$  é dada por

$$\widehat{ep}_{boot} = \sqrt{\frac{\sum_{b=1}^{B} (s(\boldsymbol{x}^{*b}) - s(.))^2}{B - 1}}$$
(40)

46

em que

$$s(.) = \frac{\sum_{b=1}^{B} s(\boldsymbol{x}^{*b})}{B}$$
(41)

O estimador *bootstrap* ideal de  $ep_F(\hat{\theta})$  o limite de  $\hat{ep}_{boot}$  quando B vai para o infinito (EFRON; TIBSHIRANI, 1993), ou seja,

$$\lim_{B \to \infty} \hat{ep}_{boot} = ep_{\hat{F}}(\hat{\theta}^*) \tag{42}$$

O estimador *bootstrap* ideal e sua aproximação (40) são chamados estimadores *bootstrap* não paramétricos, já que se baseiam em  $\hat{F}$ , um estimador não paramétrico de F. Um estimador bootstrap paramétrico do erro padrão é baseado em um estimador  $\hat{F}$  de F derivado de um modelo paramétrico. Por exemplo, ao invés de estimarmos F pela função distribuição empírica  $\hat{F}$ , podemos assumir que a população tem distribuição normal (MARTINEZ; LOUZADA-NETO, 2001).

Baseado em seu erro padrão estimado  $\hat{ep}$ , usualmente escrevemos intervalos de confiança para um parâmetro  $\theta$  de interesse, na forma

$$\hat{\theta} \pm z_{\alpha} \, \hat{ep},$$
(43)

em que  $z_{\alpha}$ é o 100  $\alpha$ -ésimo percentil de uma distribuição normal padrão.

Dado que, por construção,  $L_I(x) \leq \hat{\theta} \leq L_S(x)$ , intervalos de confiança como (43) são ditos simétricos, pois neste caso,  $\hat{\theta} - L_I(x)$  é igual a  $L_S(x) - \hat{\theta}$ , ou seja, a estimativa dada por  $\hat{\theta}$  é o ponto médio entre  $L_I(x)$  e  $L_S(x)$ . Assim, diz-se que um intervalo de confiança é simétrico quando a quantidade

$$S(x) = \frac{\left|L_I(x) - \hat{\theta}\right|}{\left|L_S(x) - \hat{\theta}\right|}$$
(44)

é igual a 1. Denominaremos S(x) de índice de simetria de um intervalo de confiança.

A motivação de (43) está na probabilidade de  $\hat{\theta}$  ter, assintoticamente, distribuição normal, ou seja,  $\hat{\theta} \sim N(\theta; \hat{e}p^2)$ . Desta propriedade, segue que

$$Z = \frac{\hat{\theta} - \theta}{\hat{ep}} \dot{\sim} N(0; 1), \tag{45}$$

e daí podemos escrever a probabilidade

$$P_{F}[\theta \in (\hat{\theta} - z_{1-\alpha/2} \,\widehat{ep}; \,\hat{\theta} + z_{1-\alpha/2} \,\widehat{ep})] = 1 - \alpha \tag{46}$$

sendo os dados amostrados procedentes de uma distribuição  $F, F \rightarrow \mathbf{X} = (X_1, X_2, \dots, X_n).$ 

De (46), definimos a expressão

$$(\hat{\theta} - z_{1-\alpha/2} \,\widehat{ep}; \,\hat{\theta} + z_{1-\alpha/2} \,\widehat{ep}) \tag{47}$$

como o intervalo de confiança padrão com probabilidade de cobertura  $1 - \alpha$ , ou nível de confiança  $100(1-\alpha)\%$ .

Quando  $\hat{\theta}$  é a média amostral, Gosset mostrou que a expressão (45) tem uma melhor aproximação em

$$Z = \frac{\hat{\theta} - \theta}{\hat{ep}} \dot{\sim} t_{n-1},\tag{48}$$

em que  $t_{n-1}$  se refere à distribuição t de Student com n-1 graus de liberdade. Baseado neste fato, o intervalo de confiança descrito em (47) seria melhor escrito

$$(\hat{\theta} - t_{(1-\alpha/2), n-1} \,\hat{e}p; \,\hat{\theta} + t_{(1-\alpha/2), n-1} \,\hat{e}p),$$
(49)

sendo  $t_{\alpha,n-1}$  o  $\alpha$ -ésimo percentil da distribuição t com n-1 graus de liberdade.

Intervalos de confiança da forma (47) ou (49) são aproximados, dado que a probabilidade de cobertura não será exatamente igual a  $1 - \alpha$ , conforme seria o desejado. Intervalos *bootstrap* também são aproximados, entretanto, oferecem melhor aproximação que os intervalos de confiança padrão.

#### 2.11 O número de repetições bootstrap B

As técnicas de reamostragem com reposição têm adquirido espaço no meio científico, em especial a técnica de *bootstrap*. Esta consiste basicamente de retirar amostras com reposição em uma amostra original. Geralmente são retiradas de 100 a 2000 amostras. Com essas amostras é possível estimar erros-padrão e CV's (Coeficiente de variação) dos estimadores, determinando sua precisão (EFRON; TIBSHIRANI, 1993). Além disso, pode-se utilizá-la em diversas situações para estimação de parâmetros, obtenção de intervalos de confiança para os parâmetros analisados, obtenção de distribuição empírica dos estimadores e determinação do tamanho da amostra (MANLY, 1997).

O número de repetições *bootstrap* utilizado para avaliar  $\hat{e}p_{boot}$  é o número tomada para B, quando  $B = \infty$ ,  $\hat{e}p_{boot}$  se torna a estimativa ideal do *bootstrap* para  $\hat{e}p_{\infty}$ . Uma resposta aproximada, mas bastante satisfatória podem ser redigidas em termos do coeficiente de variação de  $\hat{e}p_{boot}$ , a proporção de  $\hat{e}p_{boot}$ 's o desvio padrão para a sua expectativa. A variabilidade aumentou devido a parar após repetições *bootstrap* B, ao invés de ir até o infinito, se reflete em um aumento do coeficiente de variação (EFRON; TIBSHIRANI, 1993),

$$cv(\hat{ep}_{boot}) = \left\{ cv(\hat{ep}_{\infty})^2 + \frac{E(\Delta) + 2}{4B} \right\}^{1/2}$$
(50)

em que  $\Delta$  é um parâmetro que mede a longa cauda da distribuição de  $\hat{\theta}^*$ :  $\Delta$  é zero para a distribuição normal, que varia de -2 para as distribuições de cauda mais curta para valores arbitrariamente grande quando F é de cauda longa, na prática, geralmente  $\Delta$  é menor do que 10. O coeficiente de variação na equação (50) refere-se a variação tanto na reamostragem (*bootstrap*) nível e ao nível de amostragem da população. A estimativa ideal  $\hat{e}p_{\infty} = ep_{\hat{F}}(\hat{\theta}^*)$  não é perfeito. Ele ainda pode ter considerável variabilidade como uma estimativa do  $ep_F(\hat{\theta})$ , devido à variabilidade de  $\hat{F}$  como uma estimativa de F, Por exemplo, se  $x_1, x_2, \ldots, x_n$  é uma amostra aleatória de uma distribuição normal e  $\hat{\theta} = \bar{x}$ , então  $cv(\hat{e}p_{\infty}) = 1/\sqrt{2n}$ , igualando 0,22 para n = 10. Na equação (50) tem uma importante consequência prática: para os valores de  $cv(\hat{e}p_{\infty})$  e  $\Delta$  que podem surgir na prática,  $cv(\hat{e}p_B)$  não é muito maior do que  $cv(\hat{e}p_{\infty})$  de  $B \geq 200$ .

Na Tabela 4 compara-se o  $cv(\hat{ep}_{boot})$  com  $cv(\hat{ep}_{\infty})$  para várias opções de B, assumindo  $\Delta = 0$ . Muitas vezes podemos esperar ter  $cv(\hat{ep}_{\infty})$  não inferior a 0,10, caso em que B = 100 dá resultados bastante satisfatórios. Aqui estão duas regras definidas por Efron e Tibshirani (1993): 1) Mesmo um pequeno número de repetições *bootstrap*, seja B = 25, geralmente é informativo e B = 50 muitas vezes é suficiente para dar uma boa estimativa de  $ep_F(\hat{\theta})$ .

2) Raramente são mais do que com B = 200 repetições necessárias para estimar um erro padrão

Tabela 4 - O coeficiente de variação de  $\hat{ep}_{boot}$  em função do coeficiente de variação da ideal estimativa bootstrap  $\hat{ep}_{\infty}$  e o número de amostras bootstrap B, a partir da equação (50) assumindo  $\Delta = 0$ 

		E	$3 \rightarrow$			
		25	50	100	200	$\infty$
$cv(\widehat{ep}_\infty)$	$0,\!25$	$0,\!29$	$0,\!27$	$0,\!26$	$0,\!25$	$0,\!25$
$\downarrow$	0,20	$0,\!24$	$0,\!22$	$0,\!21$	$0,\!21$	0,20
	$0,\!15$	$0,\!21$	$0,\!18$	$0,\!17$	$0,\!16$	$0,\!15$
	$0,\!10$	$0,\!17$	$0,\!14$	$0,\!12$	$0,\!11$	$0,\!10$
	$0,\!05$	$0,\!15$	$0,\!11$	$0,\!09$	$0,\!07$	$0,\!05$
	0,00	$0,\!14$	$0,\!10$	$0,\!07$	$0,\!05$	0,00

# Referências

ALTMAN, D.G. Practical statistics for medical research. London: Chapman; Hall, 1991. 611p.

ANDERSON, T.W. An introduction to multivariate statistical analysis. New York: Wiley, 1958. 374p.

ARAÚJO, L.B. Métodos de correção de autovalores e regressão isotônica nos modelos **AMMI**. 2005. 75p. Dissertação (Mestrado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2005.

ARAÚJO, L.B.; DIAS, C.T.S. Viés e distribuição de autovalores de matrizes de variância e covariâncias nos modelos AMMI. In: ESCOLA DE MODELOS DE REGRESSÃO, 9., 2005, São Pedro. **Anais...** São Pedro: ABE, 2005.

ARCINIEGAS-ALARCÓN, S.; DIAS, C.T.S. Análise AMMI com dados imputados em experimentos de interação genótipo x ambiente de algodão. **Pesquisa Agropecuária Brasileira**, Brasília, v.44, n.11, p.1391-1397, 2009.

CARVALHO, C.G.P.; ARIAS, C.A.A.; TOLEDO, J.F.F.; ALMEIDA, L.A.; KIIHL, R.A.S.; OLIVEIRA, M.F. Interação genótipo x ambiente no desempenho produtivo da soja no Paraná. **Pesquisa Agropecuária Brasileira**, Brasília, v.37, n.7, p.989-1000, 2002.

CHAVES, J.L. Interação de cultivares com ambientes. In: NASS,L.L.; VALOIS, A.C.C.; MELO, I.S.; VALADARES, M.C. **Recursos genéticos e melhoramento - plantas**. Rondonópolis: Fundação MT, 2001, p.673-713.

CHAVES, L.J.; VENCOVSKY, L.; 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.259-269, 1989.

COCKERHAM, C.C. Estimation of genetics variance. In: HANSON, W.D.; ROBINSON, H.F. (Ed) **Statistical genetics and plant breeding**. Madison: National Academy of Sciences, 1963. chap.2, p.53-94.

CROSSA, J. Statistical Analyses of multilocation trials. Advances in Agronomy, New York, v.44, p.55-85, 1990.

CROSSA, J.; CORNELIUS, P.L.; YAN, W. Biplot of linear-bilinear models for studying crossover genotype x environment interaction. **Crop Science**, Madison, v.42, p.619-633, 2002.

CRUZ, C.D.; REGAZZI, A.J. Modelos biométricos aplicados ao melhoramento genético. Viçosa: UFV, 1994. 390p.

DIAS, C.T.S. Método para a escolha de componentes em modelo de efeito principal aditivo e interação multiplicativa. 2005. Tese (Livre-Docência)-Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba. 2005.

DUARTE, J.B.; VENCOVSKY, R. Interação genótipo x ambiente: uma introdução à análise "AMMI". Ribeirão Preto: Sociedade Brasileira de Genética, 1999. 60p. (Série Monografias, 9).

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

EFRON, B. Bootstrap methods: another look at jakknife. Annals of Statistics, Hayward, v.7, n.1, p.1-26, 1979.

EFRON, B.; TIBSHIRANI, R. An introduction to the bootstrap. New York: Chapman; Hall, 1993. 436p.

FALCONER, D.S. Introduction to quantitative genetics. Harlow: Longman, 1989, 438p.

FALCONER, D.S.; MACKAY, T.F.C. Introduction to quantitative genetics. Edinburgh: Longman, 1996.

FERREIRA, P.V. Estatística experimental aplicada à Agronomia. Maceió: ADUFAL, 2000. 422p.

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

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

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

GAUCH, H.G.; ZOBEL, R.W. AMMI analysis of yield trails. In: KANG, M.S.; GAUCH, H.G. (Ed) **Genotype by environment interaction**. New York: CRC Press, 1996. p.416-428.

GOLLOB, H.F. A statistical model which combines feature of factor analitic and analysis of variance techniques. **Psychometrika**, New York, v.33, p.73-115, 1968.

HOTELLING, H. Analysis of a complex of statistical variables into principal components. **The Journal Educational Psychology**, Cambridge, v.24, p.498-520, 1933.

-----. Simplifield calculation of principal components. **Psychometrika**, Williamsburg, v.1, p.27-35, 1936.

JOHNSON, R.A.; WICHERN, D.W. Applied multivariate statistical analysis. Madison: Prentice Hall International, 1998. 816p.

KARABATSOS, G. A critique of Rasch residual fit statistics. Journal of Applied Measurement, Maple Grove, v.1, p.152-176, 2000.

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

LEITE, J.G.; SINGER, J.M. Métodos assintóticos em estatística: fundamentos e aplicações. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 9., 1990, São Paulo. **Anais...** São Paulo: IME/USP, 1990. 130p.

MANDEL, J. Non-additivity in two-way analysis of variance. Journal of the American Statistical Association, New York, v.56, p.878-888, 1961.

...... A new analysis os variance model for non-adittive data. **Technometrics**, Alexandria, v.13, n.1, p.1-18, 1971.

MANLY, B.F.J. Randomization, bootstrap and Monte Carlo methods in biology. 2nd ed. New York: Chapman; Hall, 1997. 399p

MARTINEZ, E.Z.; LOUZADA-NETO, F. Estimação interval via bootstrap. **Revista de** Matemática e Estatística. São Paulo, v.19, p.217-251, 2001.

MOLENAAR, I.W.; HOIJTINK, H. The many null distributions of person fit indices. **Psychometrika**, Williamsburg, v.55, p.75-106, 1990.

PEARSON, K. On lines and planes of closest fitto systems of points in space. **Philosophical Maga**zine, London, v.2, p.559-72, 1901.

PIOVESAN, P. Validação cruzada com correção de autovalores e regressão isotônica nos modelos AMMI. 2007. 77p. Dissertação (Mestrado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2007.

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

ROBERTSON, A. Experimental design on the mesurement of herdabilities and genetic correlations. **Biometrical genetics**. New York: Pergamon, 1959. 186p.

SHELBOURNE, C. Genotype environment interaction: its implications in forest tree improvement. In: IUFRO GENETIC SABRAO JOINT SYMPOSIA, Tokyo, 1972, **Proceedings**. 28p.

SMITH, R.M. The distributional properties of Rasch standardized residuals. Educational and Psychological Measurement, Durham, v.48, p.657-667, 1988.

STONE, C.A. A computer program for assessing goodnes of fit of item response theory models based on posterior expectations. **Applied Psychological Measurement**, St. Paul, v.28, p.143-144, 2007.

SU, Y.H.; SHEU, C.F.; WANG, W.C. Computing confidence intervals of item fit in the family of rasch models using the bootstrap method. **Journal of Applied Measurement**, Maple Grove, v.8, p.190-203, 2007.

TUKEY, J.W. One degree of freedom for non-additivity. **Biometrics**, Washington, v.5, p.232-242, 1949.

WANG, W.C.; CHEN, C.T. Item parameter recovery, standard error estimates, and fit statistics of the Winsteps program for the family of Rasch models. Educational and Psychological Measurement, Durham, v.65, p.376-404, 2005.

WARZECHA, T.; ADAMSKI, T.; KACZMAREK, Z.; SURMA, M.; CHELKOWSKI, J.; WIŚNIEWSKA, H.; KRYSTKOWIAK, K.; KUCZYŃSKA, A. Genotype-by-environment interaction of barley DH lines infected with Fusarium culmorum (W.G.Sm.)Sacc. **Field Crops Research**, Amsterdam, v.120, p.21-30, 2011.

WEBER, W.E.; WRICKE, G.; WESTERMANN, T. Selection of genotypes and prediction of performance by analysis genotype-by-environment interactions. In: KANG, M.S; GAUCH, H.G. **Genotype by environment interactions**. Boca Raton: CRC Press, 1996, chap.13, p.353-371.

WOLFE, E.W. RBF. sas (Rasch Bootstrap Fit): A SAS macro for estimating critical values for Rasch model fit statistics. **Applied Psychological Measurement**, St. Paul, v.32, p.585-586, 2008.

YATES, F., AND W.G. COCHRAN. The analysis of groups of experiments. Journal of Agricultural Science, Cambridge, v.28, p.556-580, 1938.

ZOBEL, R.W.; WRIGHT, M.J.; GAUCH, JR.H.G. Statistical analysis of a yield trial. **Agronomy Journal**, Madison, v.80, p.388-393, 1988.

# 3 ADAPTABILIDADE E ESTABILIDADE DE CULTIVARES DE MILHOS ATRAVÉS DO MODELO AMMI

### Resumo

O objetivo deste trabalho foi estudar o efeito da interação genótipo × ambiente  $(G \times E)$ , avaliar a adaptabilidade e a estabilidade da produtividade de milho  $(kgha^{-1})$  de noves genótipos utilizando o modelo AMMI (modelos de efeitos principais aditivos e interação multiplicativa). Os ensaios foram realizados em vinte ambientes, o delineamento experimental foi o aleatorizado em blocos completos, com quatro repetições. O modelo AMMI permitiu identificar as melhores combinações entre genótipos e ambientes em relação à variável resposta, os efeitos de ambientes, interação  $(G \times E)$  e os três dos oitos eixos da análise de componentes principais da interação foram significativos (p < 0,01), o que levaria à seleção do modelo AMMI3. Os componentes principais explicaram 81,7% da soma de quadrados da interação  $(G \times E)$ . Os genótipos G6, G3 e G7 apresentaram maior adaptabilidade e estabilidade, sendo recomendado para ser cultivados em todos os ambientes em estudo. Os genótipos G5 e G4 apresentaram produtividade média alta, porém foram instáveis e com adaptações específicas aos ambientes de alta qualidade. O ambiente A12 apresentou a maior produtividade média (7516,500  $kgha^{-1}$ ) e o ambiente A19 apresentou a menor média (4837,614  $kgha^{-1}$ ),podem ser considerados exemplos de ambiente favorável e desfavorável, respectivamente.

Palavras-chave: Interação genótipo  $\times$  ambiente; Adaptabilidade e estabilidade; Modelos AMMI; Análise de componentes principais

#### Abstract

The aim of this work was to study the effect of genotype × environment interaction  $(G \times E)$ , evaluate the adaptability and stability in the productivity of corn  $(kgha^{-1})$  of nines genotypes using AMMI model (additive main effects and multiplicative interaction models). The experiment were conducted at twenty environments. The experimental design was a randomized complete block design with four replications. The AMMI model identified the best combinations of genotypes and environments in relation to the response variable. The effects of environments, interaction  $(G \times E)$  and the floirst three axes of principal component analysis of the interaction were significant (p < 0,01). The principal components explained 81.7% of the sum of squares of the interaction  $(G \times E)$ 

*E*). The genotypes G6, G3, G7 showed greater adaptability and stability and is recommended to be grown in all environments under study. The genotypes G5 e G4 presented the highest yield averages, but it was unstable and specific adaptation to high-quality environment. The environment A12 presented the highest yield averages (7516,500  $kgha^{-1}$ ) e the environment A19 presented the lowest yield averages (4837,614  $kgha^{-1}$ ), can be considered examples of environment favorable and unfavorable, respectively.

Keywords: genotype  $\times$  environment interaction; adaptability and stability; AMMI models; Analysis of principal components

### 3.1 Introdução

A interação genótipo × ambiente  $(G \times E)$  reflete as diferentes respostas dos genótipos às condições ambientais, ou seja, os melhores genótipos em determinado ambiente podem não ser em outros ambientes, pois dependendo das condições ambientais que forem submetidos podem não expressar todo o potencial genético dificultando, assim, a recomendação dos melhores genótipos pelos melhoristas (LAVORANTI et al., 2002; FALCONER; MACKAY, 1996; WARZECHA et al., 2011; DUARTE; VENCOVSKY, 1999).

O interesse em estudos sobre interação  $G \times E$  têm sido crescente, todavia, às mudanças climáticas bruscas que tem ocorrido recentemente tendem a desafiar os objetivos de seleção para os programas de melhoramento genético. Estudos como estes permitem investigações que auxiliam a deduzir a estabilidade e adaptabilidade genotípica dos cultivares ou animais de diferentes composições genéticas em diversos ambientes, e otimizar os ganhos genéticos de acordo com os diferentes ambientes.

Como no melhoramento genético, o pesquisador está interessado em selecionar os genótipos com desempenhos superiores em diferentes ambientes, a baixa eficiência na análise da interação dos genótipos com os ambientes pela ANOVA (Análise de Variância) pode representar um problema para os geneticistas, que aproveitam o efeito da interação para a seleção dos genótipos com produções elevadas.

Em virtude disto, diversos métodos estatísticos destinados à avaliação da interação  $G \times E$  estão disponíveis no sentido de entender melhor este efeito e a escolha do método mais adequado depende dos dados experimentais, especialmente do número de ambientes disponíveis, da precisão requerida e do tipo de informação desejada (COSTA et al., 2010), sendo assim, o modelo AMMI nos últimos anos vem se destacando e ganhando grande aplicabilidade (MANDEL, 1961; GOLLOB, 1968; FREIRE FILHO et al., 2003; WARZECHA et al., 2011). Esta análise pode ajudar tanto

na identificação de genótipos de alta produtividade e largamente adaptados, como na localização do chamado zoneamento agronômico, com fins de recomendação regionalizada e seleção de locais de teste (GAUCH; ZOBEL, 1996).

O objetivo deste trabalho foi estudar o efeito da interação  $G \times E$ , avaliar a adaptabilidade e a estabilidade da produção de milho de noves genótipos em vintes ambientes utilizando o modelo AMMI.

# 3.2 Material e Métodos

Os dados utilizados neste trabalho são os mesmos utilizados por Cornelius e Crossa (1999) e Dias e Krzanowski (2003). Foram obtidos pelo CIMMYT (CENTRO INTERNACIONAL DE MEJORAMIENTO DE MAIZ Y TRIGO) em experimentos realizados em 34 países, caracterizandose, assim, experimentos multiambientais. Foram utilizados 9 genótipos de milho, Cada genótipo foi avaliado em 20 ambientes com 4 blocos, caracterizando assim, um delineamento aleatorizado em blocos.

Os dados referentes à produção  $(kgha^{-1})$  foram submetidos inicialmente a análise de variância individual por local e posteriormente uma análise conjunta, considerando todos os locais para se avaliar a variabilidade genética entre os tratamentos e a presença de interação  $G \times E$ . Após a detecção da presença de interação foi realizada a análise de adaptabilidade e estabilidade fenotípica pelo modelo AMMI.

O modelo AMMI combina dois métodos na sua análise: análise de variância e a decomposição por valor singular em um único modelo, componentes aditivos para os efeitos principais de genótipos  $(g_i)$  e de ambientes  $(e_j)$ , e componentes multiplicativos para os efeitos da interação  $(g_e)_{ij}$ (GARCÍA-PEÑA; DIAS, 2009; GAUCH JUNIOR, 2006; PIOVESAN et al., 2008;). Assim, para a resposta média do *i*-ésimo genótipo no *j*-ésimo ambiente sobre *r* blocos (repetições) foi adotado o modelo conforme a equação:

$$Y_{ij} = \mu + g_i + e_j + \sum_{k=1}^n \lambda_k \alpha_{ik} \gamma_{jk} + \rho_{ij} + \varepsilon_{ij};$$
(51)

em que:  $y_{ij}$  é a média do *i*-ésimo genótipo no *j*-ésimo ambiente, com i=1, 2, ..., g e j=1, 2, ..., e;  $\mu$  é uma constante e geralmente é a média geral;  $g_i$  e  $e_j$  são os efeitos do *i*-ésimo genótipo e *j*-ésimo ambiente, respectivamente;  $\lambda_k$  é *k*-ésimo valor singular de **GE** com  $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ ;  $\gamma_{ik}$  e  $\alpha_{jk}$ são elementos dos *k*-ésimos valores singulares correspondentes ao *i*-ésimo genótipo e *j*-ésimo ambiente respectivamente;  $\rho_{ij}$  é o resíduo da interação  $G \times E$ ; *n* é o número de eixos ou componentes principais (PC) retidos pelo modelo; r é o número de blocos e  $\varepsilon_{ij}$  é o erro médio experimental associado ao *i*-ésimo genótipo no *j*-ésimo ambiente, assumidos independentes e  $\varepsilon_{ij} \stackrel{IID}{\sim} N(0, \frac{\sigma^2}{r})$ .

Na primeira fase os efeitos principais, na parte aditiva (média geral, efeitos de genótipos e ambientes), são ajustados por uma análise de variância comum aplicada à matriz de média  $(\mathbf{Y}_{(g\times e)})$ sobre os r blocos, resultando em um resíduo de não aditividade, isto é, na interação  $G \times E$ , dada por  $(ge)_{ij}$ , essa interação constitui a parte multiplicativa do modelo. Na segunda fase a interação é analisada pela decomposição por valores singulares (DVS) da matriz de interações ( $\mathbf{GE}(g\times e) = [(ge)_{ij}]$ ), ou por análise de componentes principais (DUARTE; VENCOVSKY, 1999).

A matriz **GE** é a matriz de interação entre os genótipos e os ambientes, (matriz de resíduo dos efeitos principais), em que cada elemento  $(ge)_{ij}$  de **GE** são dados por:

$$(ge)_{ij} = Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..}$$

$$(52)$$

em que  $Y_{ij}$ : é a média das repetições do genótipo *i* no ambiente *j*, com i=1, 2, ..., g e j=1, 2, ..., e;  $\overline{Y}_{i}$ : é a média do genótipo *i*,  $\overline{Y}_{.j}$ : é a média do ambiente j,  $\overline{Y}_{..}$ : é a média geral do experimento.

Existem várias técnicas para atribuir os graus de liberdade a um modelo AMMI, um dos procedimentos usuais consiste em determinar os grau de liberdade associados à cada parcela da  $SQ_{G\times E}$ , ou seja, associada a  $\lambda_k^2$ , relacionada a cada membro da família de modelos AMMI, obtém-se o quadrado médio (QM) correspondente a cada parcela (ou modelo), em seguida, é obtido um teste F avaliando-se a significância de cada componente em relação ao  $QM_{erro\ médio}$ . Isso resulta num quadro de análise de variância semelhante ao tradicional, com desdobramento para fonte de variação da interação  $G \times E$  (DUARTE; VENCOVSKY, 1999).

Dessa forma, a definição do número de eixos a serem retidos para explicar o padrão relacionado à interação leva em consideração a proporção da  $SQ_{G\times E}$  acumulada até o *n*-ésimo eixo  $(\sum_{k=1}^{n} \lambda_k^2/SQ_{G\times E})$  (GAUCH; ZOBEL, 1988). O ponto de parada que determina a seleção do modelo (AMMI0, AMMI1, ..., ou AMMIF) baseia-se na significância do teste *F* para os sucessivos termos da interação. O método de Gollob (1968) é um dos mais utilizados para atribuir graus de liberdades a um modelo AMMI, a expressão do método é:  $GL_{PC_k} = g + e - 1 - 2k$ , com k = 1, 2, ..., p, p=min  $\{g - 1, e - 1\}$ , em que  $PC_k$ : o k-ésimo eixo de componente principal.

Piepho (1995) mostrou o teste  $F_{R,n}$  proposto por Cornelius et al. (1992) é mais robusto que o proposto por (GOLLOB, 1968), a estatística é definida como  $F_{R,n} = \frac{(SQ_{G\times E} - \sum_{k=1}^{n} \lambda_k^2)}{f_2 QM_{Erro} médio}$ , em que  $f_2 = (g - 1 - n)(e - 1 - n)$  sendo que n é o número de termos multiplicativos incluídos no modelo. A estatística  $F_{R,n}$  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  $\operatorname{GL}_{Erro médio}$ .

Os resultados foram analisados pela representação gráfica em biplot, que é uma técnica bastante útil na análise de componentes principais, pois o gráfico usado para representar, simultaneamente, as linhas e colunas referentes a uma matriz de dados, pode indicar a existência de agrupamentos entre as observações, assim como mostrar as dispersões e correlações entre as variáveis (LAVORANTI et al., 2002).

O gráfico biplot foi introduzida por Gabriel (1971), baseia-se na aproximação de uma matriz de posto 8 pela decomposição por valores singulares (DVS) de uma matriz, por outra de posto inferior. Na análise AMMI, a DVS é aplicada à matriz de interações **GE** (matriz de posto conhecido p):

$$\widehat{\boldsymbol{GE}}_{(g \times e)} = \sum_{k=1}^{n} \lambda_k \boldsymbol{\gamma}_k \boldsymbol{\alpha}'_k = \underbrace{\widetilde{\boldsymbol{U}}}_{(\boldsymbol{P}_{BB'})} \underbrace{\widetilde{\boldsymbol{S}}}_{(\boldsymbol{\Lambda}^{1/2})} \underbrace{\widetilde{\boldsymbol{V}}'}_{(\boldsymbol{P}_{B'B})}$$
(53)

em que,  $\widetilde{U}_{(g \times e)}$  tem em suas colunas apenas os p primeiros vetores  $\gamma_{(g \times 1)}$ ;  $\widetilde{S}_{(n)}$  é a matriz diagonal com os primeiros valores singulares:  $\lambda_1, \lambda_2, \ldots, \lambda_p$  (com n<p) e  $\widetilde{V}'_{(g \times e)}$  tem em suas primeiras linhas os n primeiros vetores  $\alpha'_{(1 \times e)}$ .

Assim, na análise AMMI tem-se a decomposição exata dada por  $\boldsymbol{G}\boldsymbol{E}_{(g\times e)} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}' = \boldsymbol{U}\boldsymbol{S}^{\frac{1}{2}}\boldsymbol{S}^{\frac{1}{2}}\boldsymbol{V}' = \boldsymbol{G}\boldsymbol{H}'$  e a decomposição aproximada por *n* componentes  $\boldsymbol{G}\widehat{\boldsymbol{E}}_{(g\times e)} = \widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{S}}\widetilde{\boldsymbol{V}}' = \widetilde{\boldsymbol{G}}\widetilde{\boldsymbol{H}}' = \sum_{k=1}^{n} (\lambda_{k}^{\frac{1}{2}}\boldsymbol{\gamma}_{k}^{\frac{1}{2}})(\boldsymbol{\gamma}_{k}^{\frac{1}{2}}\boldsymbol{\alpha}_{k}')$  sendo  $\boldsymbol{G}_{(g\times e)}$  efeitos de genótipos e  $\widetilde{\boldsymbol{H}}'_{(n\times e)}$  efeitos de locais (LAVORANTI et al., 2002).

Todas as análises e a rotina computacional (no apêndice A) foram implementadas através do sistema estatístico SAS (SAS Institute Inc., 2003).

#### 3.3 Resultados e Discussão

Os resultados da análise de variância conjunta relativa à produtividade de milho  $(kgha^{-1})$  mostraram que o efeito de genótipo, ambiente e o efeito da interação foram significativos (p < 0, 01). Esse resultado indica que os genótipos apresentaram comportamento diferenciado nos ambientes, o que justifica estudo mais aprofundado sobre o comportamento dos genótipos para identificar suas magnitudes de interação com os ambientes (FREIRE FILHO et al., 2003).

O coeficiente de variação (CV%) do experimento mostrou-se de baixa magnitude (15,982%), permitindo inferir que os experimentos foram bem conduzidos, e a produtividade da média geral é 4858,139  $kgha^{-1}$  (Tabela 5).

Na análise AMMI a soma de quadrados da interação (SQ<sub>G×E</sub> = 249704162,00) corresponde exatamente à soma dos autovalores ( $\sum_{k=1}^{n} \lambda_k^2$ ) (Tabela 5 e Tabela 6), que pode estar inflacionado devido à presença de ruídos (variação não explicável) na variável resposta, em virtude disto, faz-se necessário realizar um ajuste da interação por decomposição por valores singulares, aplicada à matriz de interação  $G \times E$ . Esta matriz por sua vez tem posto p = 8, ou seja, pode ser decomposta em oito componentes principais, que são a soma de quadrados parciais na análise de variância.

Observando à decomposição da interação genótipo × ambiente dos experimentos através do modelo AMMI (Tabela 5), constatou-se que o primeiro, segundo e terceira componente principal ( $PC_s$ ) foi significativo (p < 0,01) pelo teste F de Gollob (1968) e explicam respectivamente 56,2%, 15,1% e 10,4% da variação da  $SQ_{G\times E}$ . Estes três componentes principais juntos somam 81,7%da  $SQ_{G\times E}$ , que é considerada uma resposta padrão presente na  $SQ_{G\times E}$  com 72 graus de liberdade (47,37% dos graus de liberdade da interação).

A não-significância para os  $PC_4$ ,  $PC_5$ ,  $PC_6$ ,  $PC_7$  e  $PC_8$  são desprezíveis e contém apenas ruído (variação aleatória não relacionada com o fenômeno da interação), que pode diminuir a eficiência da interpretação da estabilidade dos genótipos e ambientes na análise gráfica. Assim, a interpretação gráfica, considerando apenas a variação contida nos três primeiros eixos da ACP, é suficiente para avaliar a estabilidade dos genótipos e ambientes (ROCHA, 2002; FREIRE FILHO et al., 2003).

Os resultados para o desdobramento da interação  $G \times E$ , correspondente a cada membro da família de modelos AMMI, ou seja, os modelos AMMIO, sendo toda a interação  $(G \times E)$ , com 152 GL; modelo AMMI1 é o restante da interação, depois de retirados os 26 GL e a soma de quadrados atribuída ao primeiro eixo do componente principal (PC1); modelo AMMI2 é o restante da interação, após retirados os 24 GL e a soma de quadrados atribuída ao segundo eixo do componente principal (PC2); modelo AMMI3, com 22 GL e a soma de quadrados atribuída ao terceiro eixo do componente principal (PC3); modelo AMMI4, com 20 GL e a soma de quadrados atribuída ao quatro eixo do componente principal (PC4) e assim por diante, até modelo AMMI8, com 12 GL e a soma de quadrados atribuída ao oitavo eixo do componente principal (PC8)(Tabela 5 e Tabela 6).

Os resultados pela análise de variância conjunta efetuada com os dados observados, verifica-se, ao nível de 1% de significância que o efeito de ambientes, o efeito de blocos dentro de ambientes, o efeito de genótipos e o efeito da interação genótipo  $\times$  ambiente foram significativos e suas somas de quadrados (SQ) correspondem a 57,29%, 6,88%, 4,62% e 14,46%, respectivamente, da soma de quadrados total (Tabela 5).

#### Avaliando a família dos modelos AMMI (AMMI0, AMMI1, AMMI2, AMMI3, AMMI4,

AMMI5, AMMI6, AMMI7 e AMMI8) verificou-se pelo teste F, com os graus de liberdade ajustados pelo método de Gollob (1968) foram significativos para os três dos oitos componentes principais da interação com p < 0,01, o que levaria à seleção do modelo AMMI3, e seriam necessários três componentes principais para explicar a interação de forma significativa, pois somados (PC1, PC2 e PC3) explicaram grande porcentagem da variabilidade da interação dos dados (81,7%) e apresentaram o melhor padrão de resposta para os genótipos aos diferentes ambientes (Tabela 5).

 $\mathrm{F_{Gollob}}^{5}$ Proporção AC(%)<sup>1</sup>  $\mathrm{GL}^2$  $SQ^3$  $QM^4$  $\mathbf{PC}$ Proporção Valor-pPC1 0,5628,95 56,226140323418 5397054,5 < 0.01PC20,15171,3 2437711634 1571318,1 2,61< 0.01PC30,10481,7 2226058572 1184480,6 1,96< 0.01PC40,086 90,3 201,7921533133 1076656,9 0,0192 PC50,050 95,31812371704 687316,9 1,140,3093 PC60,7744 0,028 98,1166900236 431264,7 0,72PC70,010 99,1142513071 179505,1 0,300,9939 PC8 0,009 100,0122292394 191032,8 0,320,9858 Total 1.00249704162 \_ \_ \_

Tabela 5 - Porcentagem da soma de quadrados da interação  $(G \times E)$  captada por componente principal (PC)

 $AC(\%)^1$ : proporção acumulada;  $GL^2$ : graus de liberdade;  $SQ^3$ : soma de quadrados;  $QM^4$ : quadrado médio;  $F_{Gollob}^5$ : teste F de Gollob (1968).

Uma outra alternativa para fazer a seleção dos modelos AMMI é utilizar o teste  $F_R$ de Cornelius, definido por Piepho (1995) como um dos mais robustos, por este critério o modelo selecionado foi o AMMI1, ao nível de 1% de significância. Esse eixo explicou 56,2% da  $SQ_{G\times E}$ , os demais eixos representam variações da  $SQ_{G\times E}$  mais predominantemente ricas em ruído. Assim, o modelo AMMI1 foi aquele que melhor descreve o padrão inerente à interação  $G \times E$  (Tabela 6).

produtividade	Î
á	i
entos para	
rime	,
expe	
$\operatorname{dos}$	
conjunta	9
variância	,
de	
Análise	,
Tabela 6 -	

de milho e

$\widehat{E}$
X
$\mathcal{O}$
$\sim$
nteração
•—
da
ladrados
Ę
de c
somas
das
$\overline{\mathbf{O}}$
composição
de

F.V.	$\mathrm{GL}_{\mathrm{Gollob}}$	SQ	QM	FGollob	Valor-p	GL <sub>Cornelius</sub>	FCornelius	Valor-p
Ambiente (E)	19	989593772	52083883	26, 30	< 0,01	I	I	
Blocos/Ambiente	09	118813054	1980218	3,28	< 0,01	I	Ι	I
Genótipo (G)	×	79828575	9978572	16,55	< 0,01	I	I	I
Interação (G $\times{\rm E})$	152	249704162	1642791	2,73	< 0,01	152	2,73	< 0,01
PC1	26	140323418	5397054, 5	8,95	< 0,01	126	1,44	< 0,01
PC2	24	37711634	1571318,1	2,61	< 0,01	102	1,17	0,1487
PC3	22	26058572	1184480, 6	1,96	< 0,01	80	0,9457	0,6116
PC4	20	21533133	1076656,9	1,79	0,0192	09	0,6657	0,9739
PC5	18	12371704	687316, 9	1,14	0,3093	42	0,4623	0,9986
PC6	16	6900236	431264, 7	0,72	0,7744	26	0,307	0,9997
PC7	14	2513071	179505,1	0,30	0,9939	0	Ι	I
PC8	12	2292394	191032,8	0,32	0,9858	0	Ι	I
Resíduo	480	289366499	602847	I	Ι	I	Ι	I
Total	719	1727306061	I	I	I	I	Ι	I
Média Geral	4858, 139							
CV (%)	15,982							

A última etapa da análise AMMI consiste na representação gráfica dos genótipos e ambientes no Biplot (GABRIEL, 1971). para isso, faz-se necessária a determinação de suas coordenadas para os eixos singulares de interação. A partir das mesmas matrizes U,  $S \in V$ , resultantes da decomposição por valores singulares - DVS da matriz **GE**, obtém-se novamente os resultados de interesse (GARCÍA-PEÑA; DIAS, 2009).

Nesse caso, seria possível a construção de dois gráficos biplot: i) AMMI1 - Médias vs PC1 (Figura 1); ii) AMM2 - PC1 vs PC2 (Figura 2). Os gráficos na forma de biplot (AMMI1 e AMMI2) foram utilizados para analisar a dispersão da variação dos genótipos e ambientes e da interação entre estes. O biplot AMMI1 contém a variação dos efeitos aditivos principais de genótipos e ambientes, visualizada na horizontal do gráfico (eixo das abscissas), e a variação dos efeitos multiplicativos da interação  $G \times E$ , visualizada na vertical (eixo das ordenadas). O biplot AMMI2 são visualizados apenas os efeitos multiplicativos da interação  $G \times E$  contidos nos dois primeiros  $PC_s$  e dispersos em ambas as ordenadas do gráfico (FREIRE FILHO et al., 2003).

O primeiro eixo singular da análise AMMI captura a maior porcentagem de "padrão" e um incremento de "ruídos" (GAUCH, 1988). Com isso, uma alta proporção da  $SQ_{G\times E}$  explicada pelos dois primeiros eixos (71,3%), espera-se capturar a maior parte do "padrão" devido a interação  $G \times E$ . Dessa forma, os escores de genótipos e de ambientes são plotados só até o segundo eixo, para cada uma das variáveis respostas.

A interpretação do gráfico biplot permite avaliar a estabilidade dos genótipos nos ambientes através da observação da magnitude e do sinal dos escores dos genótipos e ambientes para o eixo da interação (Tabela 7). Portanto, os escores baixos (próximos de zero) representam os genótipos e ambientes mais estáveis (DUARTE; VENCOVSKY, 1999; CARBONELL et al., 2004; MAIA et al., 2006).

Os genótipos mais produtivos foram G6, G5, G4, G9, G3, G7 e G1 em ordem decrescente (Figura 1; Tabela 7). Os genótipos mais estáveis foram G7, G2, G3 e G6 (Figura 1 e Figura 2), estes podem ser cultivados em todos os locais onde foram feitos os estudos. Entre estes, apenas os genótipos G6, G3 e G7 apresentaram produtividade acima da média geral, e estão entre os dois grupos (produtivos e estáveis), indicando que esses genótipos associaram adaptabilidade com estabilidade. Porém nem sempre os genótipos com as produtividades médias altas são estáveis, é o caso do genótipo G5 que mostrou adaptabilidade específica ao ambiente A15, e também do genótipo G4 ao ambiente A11, associado com altas médias (Figura 1, Figura 2 e Tabela 7).

Genótipos	Biplot A	MMI1	Biplot .	AMMI2	Ordem para
Ambientes	MÉDIA(x)	PC1(y)	PC(1)	PC(2)	MÉDIAS
G1	4617,029	-29,339	-29,339	12,543	7
G2	4603,259	-19,051	-19,051	17,630	8
G3	4822,914	$-12,\!672$	$-12,\!672$	8,227	5
G4	5217,849	39,639	39,639	-0,181	3
G5	5246,874	21,784	21,784	-18,983	2
G6	$5330,\!171$	$20,\!865$	20,865	-12,166	1
G7	4672,328	-9,142	-9,142	20,996	6
G8	4283,195	-36,872	-36,872	-38,442	9
G9	4929,634	24,787	24,787	$10,\!376$	4
A1	$3612,\!531$	$-17,\!300$	-17,300	-14,893	17
A2	4209,444	$1,\!353$	$1,\!353$	9,847	15
A3	$5104,\!919$	-19,826	-19,826	-9,723	8
A4	$5229,\!997$	$11,\!386$	11,386	-4,902	7
A5	$4955,\!231$	$6,\!611$	6,611	-4,862	11
A6	$6305,\!306$	-0,482	-0,482	11,234	3
A7	3229,342	-6,243	-6,243	-5,799	18
A8	4027,708	-51,791	-51,791	$23,\!155$	16
A9	4970,644	$2,\!922$	$2,\!922$	$2,\!479$	10
A10	$2936,\!425$	-16,785	-16,785	-3,760	19
A11	$5307,\!192$	$29,\!962$	29,962	3,804	6
A12	$7516,\!500$	$12,\!311$	12,311	17,482	1
A13	$6332,\!100$	19,200	19,200	13,493	2
A14	6052,297	-0,512	-0,512	$1,\!307$	4
A15	5049,028	$15,\!976$	$15,\!976$	$-15,\!667$	9
A16	5406,164	6,079	6,079	$0,\!123$	5
A17	4876,689	0,008	0,008	-23,808	12
A18	4551,547	16,714	16,714	18,909	14
A19	$2652,\!103$	-6,783	-6,783	-4,210	20
A20	4837,614	-2,802	-2,802	-14,210	13

Tabela 7 - Coordenadas dos eixos das abscissas (x) e de coordenadas (y), para os dois gráficos Biplot

A adaptação específica indica maior média de produtividade de genótipo em um determinado ambiente, verifica-se que os genótipos G4 e G9 foram mais adaptados aos ambientes A11, A13 e A18; dos genótipos G5 e G6 ao ambiente A15; do genótipo G8 aos ambientes A1, A3, A10, A17, A20, A7 e A19; do genótipo G1 aos ambientes A8 e A3 e do genótipo G7 aos ambientes A2 e A6. Além disso, constata-se falta de adaptação dos genótipos G4 e G9 aos ambientes A1, A3 e A10; dos genótipos G5 e G6 aos ambientes A18, A2 e A6; dos genótipos G1, G2 e G3 aos ambientes A11 e A15 e do genótipo G8 aos ambientes A11, A18, A13 e A12 (Figura 2).

Em relação aos ambientes, destacam-se, com menor contribuição, os ambientes A9, A16, A14, A4, A5, A19 e A7 (Figura 2). Os ambientes mais discrepantes foram A19, A10, A8, A13 e A11. O A12 apresentou a maior produtividade média (7516,500  $kgha^{-1}$ ), e o A19 a menor média (4837,614  $kgha^{-1}$ ), e podem ser considerados exemplos de ambiente favorável e desfavorável, respectivamente (Figura 1 e Figura 2). Os ambientes A20 e A17, A9 e A5 foram muito semelhantes pela Figura 1, essa informação é importante, uma vez que se pode substituir um desses locais por outro mais representativo da região, em que as cultivares serão recomendadas. A análise AMMI pode ser utilizada com eficiência na identificação de condições ambientais superiores para a exploração agrícola (seleção de locais de cultivo) e genótipos de superior desempenho médio (GAUCH; PIEPHO; ANNICCHIARICO, 2008; YAN, 2010).



Figura 1 - Biplot AMMI1 (Médias vs PC1) para os dados de produtividade de milho (kgha<sup>-1</sup>), com nove genótipos (G) e vinte ambientes (E)



Figura 2 - Biplot AMMI2 (PC1 vs PC2) para os dados de produtividade de milho (kgha<sup>-1</sup>), com nove genótipos (G) e vinte ambientes (E)

# 3.4 Conclusão

Os genótipos G6, G3 e G7 apresentaram maior adaptabilidade e estabilidade, sendo recomendado para ser cultivados em todos os ambientes em estudo. Os genótipo G5 e G4 apresentaram produtividade média alta, porém foram instáveis e com adaptações específicas aos ambientes de alta qualidade. O ambiente A12 apresentou a maior produtividade média (7516,500  $kgha^{-1}$ ) e o ambiente A19 apresentou a menor média (4837,614  $kgha^{-1}$ ), podem ser considerados exemplos de ambiente favorável e desfavorável, respectivamente.

# Referências

CARBONELL, S. A. M.; AZEVEDO FILHO, J. A.; DIAS, L. A. S.; GARCIA, A. A. F.; MORAIS, L. K. Common bean cultivars and lines interactions with environments. **Scientia Agricola**, Piracicaba, v.61, p.169-177, 2004.

CORNELIUS, P. L.; CROSSA, J. Prediction assessment of shrinkage estimators of multiplicative models for multi-environment trials. **Crop Science**, Madison, v.39, p.998-1009, 1999.

CORNELIUS, P. L.; SEYEDSADR, M.; CROSSA, J. Using the shifted multiplicative model to search for "separability" in crop cultivar trials. **Theoretical and Applied Genetics**, Berlin, v.84, p.161-172, 1992.

COSTA, E. F. N.; SOUZA, J.C.; LIMA, J. L.; CARDOSO, G. A. Interação entre genótipos e ambientes em diferentes tipos de híbridos de milho. **Pesquisa Agropecuária Brasileira**, Brasília, v.45, n.12, p.1433-1440, 2010.

DIAS, C.T.S.; KRZANOWSKI, W.J. Model selection and cross validation in additive main effect and multiplicative interaction models. **Crop Science**, Madison, v.43, p.865-873, 2003.

DUARTE, J.B.; VENCOVSKY, R. Interação genótipo x ambiente: uma introdução à análise "AMMI". Ribeirão Preto: Sociedade Brasileira de Genética, 1999. 60p. (Série Monografias, 9).

FALCONER, D.S; MACKAY, T.F.C. Introduction to quantitative genetics. 4nd ed. Edinburgh: Longman Group Limited, 1996. 464p.

FREIRE FILHO, F. R.; RIBEIRO, V. Q.; ROCHA, M. M.; LOPES, Â. C. A. Adaptabilidade e estabilidade da produtividade de grãos de genótipos de caupi enramador de tegumento mulato. **Pesquisa Agropecuária Brasileira**, Brasília, v.38, n.5, p.591-598, 2003.

GABRIEL, K.R. The biplot graphic display of matrices with application to principal components analysis. **Biometrika**, London, v.58, p.453-467, 1971.

GARCÍA-PEÑA, M.; DIAS, C.T.S. Análise dos modelos aditivos com interação multiplicativa (AMMI) bivariados. **Revista Brasileira de Biometria**, São Paulo, v.27, n.4, p.586-602, 2009.

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

GAUCH, H.G.; PIEPHO, H.G.; ANNICCHIARICO, P. Statistical Analysis of Yield Trials by AMMI and GGE: Further Considerations. **Crop Science**, Madison, v.48, n3, p.866-889, 2008.

GAUCH, H.G.; ZOBEL, R.W. Predictive and postdictive success of statistical analysis of yield trials. **Theoretical and Applied Genetics**, New York, v.76, p.1-10, 1988.

-----. AMMI analysis of yield trails. In: KANG, M.S.; GAUCH, H.G. (Ed.). Genotype by environment interaction. New York: CRC Press, 1996. p.416-428.

GAUCH JUNIOR, H.G. Statistical analysis of yield trials by AMMI and GGE. Crop Science, Madison, v.46, p.1488-1500, 2006.

GOLLOB, H.F. A statistical model which combines feature of factor analitic and analysis of variance techniques. **Psychometrika**, New York, v.33, p.73-115, 1968.

LAVORANTI, O.J.; DIAS, C.T.S.; VENCOVSKY, R. Estudo da estabilidade e adaptabilidade fenotípica de progênies de *Eucalyptus grandis*, via metodologia AMMI. **Boletim de Pesquisa Florestal**, Curitiba, v.44, n.1, p.107-124, 2002.

MAIA, M. C. M.; VELLO, N. A.; ROCHA, M. de M.; PINHEIRO, J. B.; SILVA JÚNIOR, N. F. da. Adaptabilidade e estabilidade de linhagens experimentais de soja selecionadas para caracteres agronômicos através de método uni-multivariado. **Bragantia**, Campinas, v.65, p.215-226, 2006.

MANDEL, J. Non-additivity in two-way analysis of variance. Journal of the American Statistical Association, New York, v.56, p.878-888, 1961.

PIEPHO, H. P. Robustness of statistical test for multiplicative terms in the additive main effects and multiplicative interaction model for additive main effects and multiplicative interaction model for cultivar trial. **Theoretical and Applied Genetics**, Berlin, v.90, n.3/4, p.438-443, 1995.

PIOVESAN, P.; ARAÚJO, L. B.; DIAS, C. T. S. Validação cruzada com correção de autovalores e regressão isotônica nos modelos de efeitos principais aditivos e interação multiplicativa. **Ciência Rural**, Santa Maria, Online, ISSN0103-8478, 06.11.2008.

ROCHA, M. M. Seleção de linhagens experimentais de soja para adaptabilidade e estabilidade fenotípica. 2002. 174 f. Tese (Doutorado em Genética e Melhora mento de Plantas) - Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, 2002.

SAS Institute inc. SAS 9.2. Cary, 2003.

WARZECHA, T.; ADAMSKI, T.; KACZMAREK, Z.; SURMA, M.; CHELKOWSKI, J.; WIŚNIEWSKA, H.; KRYSTKOWIAK, K.; KUCZYŃSKA, A. Genotype-by-environment interaction of barley DH lines infected with Fusarium culmorum (W.G.Sm.)Sacc. **Field Crops Research**, Amsterdam, v.120, p.21-30, 2011.

YAN, W. Optimal Use of Biplots in Analysis of Multi-Location Variety Test Data. Acta Agronomica Sinica, Beijing, v.36, n.11, p.1805-1819, 2010.

# 4 DISTRIBUIÇÃO EMPÍRICA DOS AUTOVALORES ASSOCIADOS À MATRIZ DE INTERAÇÃO PELO MÉTODO BOOTSTRAP NÃO-PARAMÉTRICO

### Resumo

Esta pesquisa visou estudar a distribuição empírica dos autovalores e calcular o intervalo de confiança, com o auxílio da reamostragem *Bootstrap* não-paramétrico no modelo biométrico AMMI (modelos de efeitos principais aditivos e interação multiplicativa). Os ensaios foram realizados em vinte ambientes, o delineamento experimental foi o aleatorizado em blocos completos, com quatro repetições. O modelo AMMI permitiu identificar as melhores combinações entre genótipos e ambientes em relação à variável resposta, os efeitos de ambientes, interação ( $G \times E$ ) e os três dos oitos eixos da análise de componentes principais da interação foram significativos (p < 0,01), o que levaria à seleção do modelo AMMI3. Os componentes principais explicaram 81,7% da soma de quadrados da interação ( $G \times E$ ). Com a utilização do método *bootstrap* não-paramétrico aplicado na matriz de resíduo, constitui uma eficiente alternativa para encontrar a distribuição empírica dos autovalores e com a aplicação do teste de normalidade de Shapiro-Wilk para cada autovalor, concluímos que o primeiro autovalor até o quinto autovalor apresentaram uma distribuição normal, a partir do sexto autovalor até o oitavo autovalor não apresentaram o mesmo resultado anterior.

Palavras-chave: Distribuição empírica; Autovalores; Intervalo de confiança; *Bootstrap*; Modelos AMMI

#### Abstract

This study aimed to investigate the empirical distribution of eigenvalues and calculate confidence interval, by applying AMMI (Additive Main Effects and Multiplicative Interaction Model) and nonparametric Bootstrap methods. The study was done in twenty environments, the experimental design used was randomized blocks, with four replications. The AMMI model identified the best combinations of genotypes and environments in relation to the response variable, the effects of environments, interaction ( $G \times E$ ) and the first three axes of principal component analysis of the interaction were significant (p < 0.01). The principal components explained 81.7% of the sum of squares of the interaction ( $G \times E$ ). By using nonparametric Bootstrap method applied to the waste matrix, is an efficient alternative to find the empirical distribution of eigenvalues, with the application of the normality test of Shapiro-Wilk for each eigenvalue, conclude that the first eigenvalue to the fifth eigenvalue showed a normal distribution, from the sixth eigenvalue to the eighth eigenvalue did not show the same result as before.

Keywords: Empirical distribution; Eigenvalues; Confidence interval; Bootstrap; AMMI models

# 4.1 Introdução

Muitos dados coletados em experimentos agronômicos são multivariados, e recebem influência de vários fatores como por exemplo, o ambiente onde foi realizado o experimento, genótipos, tratamentos agronômicos etc (DIAS; KRZANOWSKI, 2003). A expressão das características das plantas cultivadas está ligada ao controle genético, ao ambiente em que são cultivadas e à interação entre esses dois fatores (YAN; KANG, 2003; MOHAMMADI; AMRI, 2009). A resposta distinta dos genótipos em diferentes condições ambientais é denominada de interação genótipo × ambiente  $(G \times E)$ , a qual reduz a correlação entre os valores fenotípicos e genotípicos e dificulta a seleção e recomendação de genótipos adaptados e estáveis (YAN; HOLLAND, 2010; WARZECHA et al., 2011).

Em virtude disto, diversos métodos estatísticos destinados à avaliação da interação  $G \times E$  estão disponíveis no sentido de entender melhor este efeito e a escolha do método mais adequado depende dos dados experimentais, especialmente dos números de ambientes disponíveis, da precisão requerida e do tipo de informação desejada (COSTA et al., 2010). Sendo assim, o modelo AMMI (modelos de efeitos principais aditivos e interação multiplicativa) nos últimos anos vem se destacando e ganhando grande aplicabilidade (MANDEL, 1961; GOLLOB, 1968; FREIRE FILHO et al., 2003; WARZECHA et al., 2011). Em seguida, com a metodologia AMMI-*bootstrap* (LAVORANTI, 2003) e a técnica de reamostragem *bootstrap* não-paramétrico aplicada à matriz de resíduo, para encontrar a distribuição empírica dos autovalores.

O método *bootstrap* foi proposto por Efron (1979), como um procedimento de reamostragem amplamente utilizado na obtenção de estimativas pontuais e intervalares, bem como na avaliação da acurácia de estimativas e testes. Segundo Martinez e Louzada-Neto (2001) se comparado a outras técnicas estatísticas, o método teve uma gênese um tanto tardia, devido à sua dependência do uso de computadores. Os progresso da informática experimentados nas últimas décadas do século XX possibilitaram a popularização do uso do computador e incrementaram o surgimento e acesso a *softwares* matemáticos e estatísticos. Conseqüentemente, as aplicações de métodos *bootstrap* nas mais diferentes áreas da estatística se intensificaram.

Procedimentos bootstrap para a identificação de valores críticos para ajustar estatísticas

são fáceis de implementar e, de fato, tem sido implementado em três programas distintos que fazem interface com o software comercial para identificar os ajustes estatísticos razoáveis, dado um conjunto de itens estimados e parâmetros personalizado (STONE, 2007; SU; SHEU; WANG, 2007; WOLFE, 2008).

É sabido que as distribuições nulas de estatísticas comumente empregadas a ajustes estatísticos não seguem uma distribuição com uma forma paramétrica conhecida (KARABATSOS, 2000; MOLENAAR; HOIJTINK, 1990; SMITH, 1988, 1991; WANG; CHEN, 2005). Nesse caso o *bootstrap* é muito útil, pois é uma técnica que não exige diferentes fórmulas para cada problema e pode ser utilizada em casos gerais, não dependendo da distribuição original da estatística do parâmetro estudado.

O *bootstrap* pode ser implementado tanto na estatística não-paramétrica quanto na paramétrica. No caso não-paramétrico, o método *bootstrap* reamostra os dados com reposição, de acordo com uma distribuição empírica estimada, tendo em vista que, em geral, não se conhece a distribuição subjacente aos dados. No caso paramétrico, quando se tem informação suficiente sobre a forma da distribuição dos dados, a amostra *bootstrap* é formada realizando-se a amostragem diretamente nessa distribuição com parâmetros desconhecidos substituídos por estimativas paramétricas. A distribuição da estatística de interesse aplicada aos valores da amostra *bootstrap*, condicional aos dados observados, é definida como a distribuição *bootstrap* dessa estatística (LAVORANTI, 2003).

O objetivo deste trabalho foi propor procedimento empírico de *Bootstrap* nãoparamétrico aplicada à matriz de resíduo, estimada via metodologia AMMI, para encontrar a distribuição empírica dos autovalores, com a aplicação do teste de normalidade de Shapiro-Wilk para cada autovalor e calcular o intervalo de confiança *Bootstrap*. Com o estudo da distribuição empírica dos autovalores servirá para validar os testes de hipóteses propostos na literatura para identificar o número de *IPCA* (*Incremental Principal Component Analysis*) para seleção dos modelos AMMI, e propor um teste para seleção dos modelos.

# 4.2 Material e Métodos

Os dados utilizados neste trabalho são os mesmos utilizados por Cornelius e Crossa (1999), Dias e Krzanowski (2003) e Araújo (2005). Foram obtidos pelo CIMMYT (CENTRO IN-TERNACIONAL DE MEJORAMIENTO DE MAIZ Y TRIGO) em experimentos realizados em 34 países, caracterizando-se, experimentos multiambientais. Foram utilizados 9 genótipos de milho, Cada genótipo foi avaliado em 20 ambientes com 4 blocos, caracterizando assim, um delineamento aleatorizado em blocos.

O modelo AMMI pressupõe componentes aditivos para os efeitos principais de genótipos  $(g_i)$  e de ambientes  $(e_j)$  e componentes multiplicativos para o efeito da interação  $(ge)_{ij}$ . Assim, a resposta média do *i*-ésimo genótipo no *j*-ésimo ambiente é representada pelo seguinte modelo matemática:

$$Y_{ij} = \mu + g_i + e_j + \sum_{k=1}^n \lambda_k \alpha_{ik} \gamma_{jk} + \rho_{ij} + \varepsilon_{ij};$$
(54)

em que:  $y_{ij}$  é a média do *i*-ésimo genótipo no *j*-ésimo ambiente, com i=1, 2, ..., g e j=1, 2, ..., e;  $\mu$  é uma constante e geralmente é a média geral;  $g_i$  e  $e_j$  são os efeitos do *i*-ésimo genótipo e *j*-ésimo ambiente, respectivamente;  $\lambda_k$  é *k*-ésimo valor singular da matriz de interação **GE** com  $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ ;  $\gamma_{ik}$  e  $\alpha_{jk}$  são elementos dos *k*-ésimos valores singulares correspondentes ao *i*-ésimo genótipo e *j*-ésimo ambiente respectivamente;  $\rho_{ij}$  é o resíduo da interação  $G \times E$ ; *n* é o número de eixos ou componentes principais (PC) retidos pelo modelo; *r* é o número de blocos e  $\varepsilon_{ij}$  é o erro médio experimental associado ao *i*-ésimo genótipo no *j*-ésimo ambiente, assumidos independentes e  $\varepsilon_{ij} \stackrel{IID}{\sim} N(0, \frac{\sigma^2}{r})$ ; O qual sugere que os dados observados, possam ser organizados em uma matriz de ordem  $g \times e$ .

O desenvolvimento da metodologia AMMI-bootstrap (LAVORANTI, 2003) consiste em se executar o método AMMI seguindo a proposta de Gollob (1968), a qual garante pelo menor rigor dos graus de liberdade, um número maior de fatores multiplicativos e, conseqüentemente, capta todo o padrão devido à interação  $G \times E$  (MAIA et al., 2006). A partir desse modelo, foram determinadas as estimativas das médias dos genótipos (*i*) nos ambientes (*j*) ( $\hat{Y}_{ij}$ ) livres de interferência de ruídos, e, a partir desse novo conjunto de dados, foram obtidas as matrizes de resíduos ou dos efeitos (ge)<sub>*ij*</sub>,  $\left|\widehat{GE}_{G \times E} = (\widehat{ge}_{ij})\right|$  representadas por:

$$\widehat{\boldsymbol{GE}}_{G \times E} = \begin{vmatrix} \widehat{ge}_{11} & \widehat{ge}_{12} & \cdots & \widehat{ge}_{1e} \\ \widehat{ge}_{21} & \widehat{ge}_{22} & \cdots & \widehat{ge}_{2e} \\ \vdots & \vdots & \cdots & \vdots \\ \widehat{ge}_{g1} & \widehat{ge}_{g2} & \cdots & \widehat{ge}_{ge} \end{vmatrix},$$

com:

$$\widehat{g}\widehat{e}_{ij} = Y_{ij} - \widehat{\overline{Y}}_{i.} - \widehat{\overline{Y}}_{.j} + \widehat{\overline{Y}}_{..} \quad , \tag{55}$$

em que  $\hat{\overline{Y}}_{i,:}$  é a média dos valores estimados do genótipo  $i; \hat{\overline{Y}}_{j,:}$  é a média dos valores estimados do ambiente  $j; \hat{\overline{Y}}_{..}$ : é a média geral dos valores estimados.
Das matrizes  $\widehat{GE}_{G\times E}$ , podemos reamostrar *B* matrizes *bootstrap* para genótipos  $(GE_g^*)$ , sendo o sorteio com reposição executado nas linhas; ou também pode reamostrar *B* matrizes *bootstrap* para ambientes  $(GE_e^*)$ , sendo o sorteio com reposição executado nas colunas, *B* é o número de repetição *bootstrap*. Assim, obteve-se (MAIA et al., 2006):

$$oldsymbol{GE}_{kl}^{*} = egin{bmatrix} ge_{11}^{*} & ge_{12}^{*} & \cdots & ge_{1e}^{*} \ ge_{21}^{*} & ge_{22}^{*} & \cdots & ge_{2e}^{*} \ dots & dots & \ddots & dots \ ge_{g1}^{*} & ge_{g2}^{*} & \cdots & ge_{ge}^{*} \ \end{bmatrix}$$

com: k = g ou k = e, para genótipos ou ambientes, respectivamente e  $l = (1, 2, \dots, B)$ .

A partir da matriz de interação  $\widehat{GE}_{G\times E}$ , utilizando a técnica de decomposição por valores singulares (DVS) obtendo-se os valores singulares que é dados por  $\lambda_s$ , com  $s = 1, 2, \dots, p$ ,  $\lambda_1 > \lambda_2 > \dots > \lambda_p > 0$  e  $p = min\{g - 1, e - 1\}$ , ou pela técnica da análise de componentes principais (ACP) obtendo-se um conjunto de p autovalores a partir de  $(\widehat{GE}_{G\times E})(\widehat{GE}_{G\times E}^T)$  ou  $(\widehat{GE}_{G\times E}^T)(\widehat{GE}_{G\times E}), \operatorname{com} \lambda_1^2 > \lambda_2^2 > \dots > \lambda_p^2 > 0.$ 

Da mesma forma podemos encontrar os autovalores das matrizes  $GE_{kl}^*$  gerada a partir da  $\widehat{GE}_{G\times E}$  e estimar outros conjuntos de autovalores  $\lambda_{1i}^2 > \lambda_{2i}^2 > \cdots > \lambda_{pi}^2 > 0$ , com k = g ou k = e, para genótipos ou ambientes, respectivamente,  $l = 1, 2, \cdots, B$  e  $i = 1, 2, \cdots, B$ . Como as matrizes  $GE_{kl}^*$  são as estimativas de  $\widehat{GE}_{G\times E}$ , temos que os  $\lambda_{ki}^2$  são estimativas de  $\lambda_s^2$  com  $s = 1, 2, \cdots, p$ (MUIRHEAD, 1987). Segundo esse autor  $\lambda_{11}^2$  superestimar  $\lambda_1^2$ , enquanto que  $\lambda_{p1}^2$  subestimar  $\lambda_p^2$ , e esses  $\lambda_{ps}^2$  tendem a algum valor central.

As técnicas de reamostragem com reposição têm adquirido espaço no meio científico, em especial a técnica de *bootstrap*. Esta consiste basicamente de retirar amostras com reposição em uma amostra original. Consiste, basicamente, na replicação do processo de estimação via reamostragem pela amostra ou distribuição da variável, caso seja conhecida, com parâmetros estimados via amostra. Utilizando esta técnica, é possível realizar comparações estatísticas em diversas áreas do conhecimento científico, tais como: economia, ecologia, genética, biologia, agronomia, etc, e suas subáreas (EFRON; TIBSHIRANI, 1993; DAVISON; HINKLEY, 1997; JHUN; JEONG, 2000). Aplicações da técnica *bootstrap* ao estudo da variabilidade de parâmetros hídricos do solo são encontradas nos trabalhos de Schaap e Leij (1998) e Hendrickx e Wierenga (1990).

Seja uma amostra aleatória baseada em n observações independentes  $x_1, x_2, \ldots, x_n$ . O erro padrão de uma média  $\bar{\mathbf{x}}$  baseada nesta amostra é estimado pela expressão:

$$\widehat{ep}(\bar{x}) = \sqrt{\frac{s^2}{n}} , \qquad (56)$$

em que

$$s^{2} = \frac{\sum_{i=1}^{n} (x_{i} - \bar{x})^{2}}{n - 1}$$
(57)

é o estimador não viciado da variância. Note-se que o erro padrão não é uma estimativa de uma quantidade pertinente a uma população, mas uma medida da incerteza da média amostral vista como uma estimativa da média populacional (ALTMAN, 1991). A expressão (56) deixa claro que a magnitude desta incerteza diminui conforme o tamanho da amostra n aumenta (MARTINEZ; LOUZADA-NETO, 2001).

Seja X uma variável aleatória com distribuição F, sendo sua esperança denotada por  $\mu_F$  e sua variância denotada por  $\sigma_F^2$ . Usaremos a notação  $X \sim (\mu_F, \sigma_F^2)$  e escrevemos  $\hat{F}_n \rightarrow$  $(x_1, x_2, \ldots, x_n)$  para indicar que  $\mathbf{x} = (x_1, x_2, \ldots, x_n)$  é uma amostra aleatória de tamanho n obtida de uma população com função de probabilidade F. A média  $\bar{X}$  é também uma variável aleatória e tem esperança  $\mu_F$  e variância  $\sigma_F^2/n$ , ou seja,  $\bar{X} \sim (\mu_F, \sigma_F^2/n)$ . Note-se então que X e  $\bar{X}$  têm a mesma esperança, entretanto, o desvio padrão de  $\bar{X}$  é definido como a raiz quadrada da variância de  $\bar{X}$ (MARTINEZ; LOUZADA-NETO, 2001), ou seja,

$$ep_F(\bar{X}) = \sqrt{\operatorname{var}_F(\bar{X})} = \frac{\sigma_F}{\sqrt{n}}$$
(58)

A expressão (56) fornece um estimador para (58). Entretanto, nem todos os estimadores têm expressões de fácil manejo para seu respectivo erro padrão, como (56). Isto significa que o trabalho de encontrar medidas de precisão para outros estimadores  $\hat{\theta} = s(X)$ , que não a média, pode ser algo bastante complicado.

Sob certas condições, o teorema central do limite diz que a distribuição de  $\bar{X}$  é aproximadamente normal quando n é grande (LEITE; SINGER, 1990), ou seja,

$$\bar{X} \sim N(\mu_F; \sigma_F^2/n), \tag{59}$$

e conseqüentemente,

$$\frac{X - \mu_F}{\sigma_F / \sqrt{n}} \sim N(0; 1). \tag{60}$$

Usando uma tabela da distribuição normal padrão, de (60) podemos escrever, por exemplo, a probabilidade

$$P(\left|\bar{X} - \mu_F\right| < \frac{2\sigma_F}{\sqrt{n}}) \doteq 0,954,\tag{61}$$

em que o erro padrão mostra-se uma útil medida da precisão da estimativa da média. A probabilidade da distância entre a média amostral e a populacional ser de pelo menos dois erros padrão é aproximadamente 0,954. Verifica-se assim, que quanto menor o erro padrão da média, a expressão (61) sugere uma melhor aproximação entre as médias amostral e populacional (MARTINEZ; LOUZADA-NETO, 2001).

Uma vantagem do método *bootstrap* é que esta técnica não depende inteiramente do teorema central do limite, já que, em suas aplicações, medidas de precisão são obtidas diretamente dos dados (EFRON; TIBSHIRANI, 1993).

Observada uma amostra aleatória de tamanho n oriunda de uma distribuição F, definese uma função distribuição empírica  $\hat{F}_n$  como uma distribuição discreta, que atribui probabilidade  $n^{-1}$ a cada valor  $x_i, i = 1, ..., n$ . Uma amostra bootstrap  $\boldsymbol{x}^* = (x_1^*, x_2^*, ..., x_n^*)$  é obtida reamostrando aleatoriamente n vezes, com reposição, as observações  $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ , onde verifica-se que  $\hat{F}_n \to (x_1^*, x_2^*, ..., x_n^*)$ .

Selecionadas B amostras *bootstrap*,  $\boldsymbol{x}^{*1}$ ,  $\boldsymbol{x}^{*2}$ , ...,  $\boldsymbol{x}^{*B}$ , de forma independente, estimase  $\theta$  em cada uma destas amostras através de  $\hat{\theta}^{*b} = s(\boldsymbol{x}^{*b})$ , b = 1, 2, ..., B. Uma expressão para o estimador *bootstrap* do erro padrão da estatística  $\hat{\theta}$  é dada por

$$\hat{ep}_{boot} = \sqrt{\frac{\sum_{b=1}^{B} \left( s(\boldsymbol{x}^{*b}) - s(.) \right)^2}{B - 1}}$$
(62)

em que

$$s(.) = \frac{\sum_{b=1}^{B} s(\boldsymbol{x}^{*b})}{B}$$
(63)

O estimador *bootstrap* ideal de  $ep_F(\hat{\theta})$  o limite de  $\hat{ep}_{boot}$  quando *B* vai para o infinito (EFRON; TIBSHIRANI, 1993), ou seja,

$$\lim_{B \to \infty} \hat{e} \hat{p}_{boot} = e p_{\hat{F}}(\hat{\theta}^*) \tag{64}$$

O estimador *bootstrap* ideal e sua aproximação (64) são chamados estimadores *bootstrap* não paramétricos, já que se baseiam em  $\hat{F}$ , um estimador não paramétrico de F. Um estimador *bootstrap* paramétrico do erro padrão é baseado em um estimador  $\hat{F}$  de F derivado de um modelo paramétrico. Por exemplo, ao invés de estimarmos F pela função distribuição empírica  $\hat{F}$ , podemos assumir que a população tem distribuição normal (MARTINEZ; LOUZADA-NETO, 2001).

Sendo  $\hat{ep}_{boot}(\hat{\theta}), \hat{\theta} = s(x)$ , a estimativa *bootstrap* do erro padrão de  $\hat{\theta}$ , definida em (62), o intervalo de confiança *bootstrap* padrão para  $\theta$ , com probabilidade de cobertura de aproximadamente  $1 - \alpha$ , é dado por

$$\hat{\theta} \pm z_{1-\alpha/2} \widehat{ep}_{boot}(\hat{\theta}). \tag{65}$$

A maior vantagem deste método reside na sua simplicidade algébrica de encontrar um intervalo de confiança para  $\theta$ . Entretanto, sendo a equação (65) uma seqüência de

$$Z = \frac{\hat{\theta} - \theta}{\hat{ep}_{boot}(\hat{\theta})} \sim N(0; 1)$$
(66)

Os números de replicações *bootstrap* necessárias para uma boa estimativa do erropadrão e do intervalo de confiança é muito importante, foram discutidas por Efron e Tibshirani (1993), Kendall e Stuart (1977) e Efron (1987). Para obter uma boa estimativa do erro-padrão através do *bootstrap* são necessárias entre 25 e 200 replicações e que para uma boa estimativa dos limites de confiança seriam necessárias mais de 500 replicações (EFRON; TIBSHIRANI, 1993). Além disso, pode-se utilizá-la em diversas situações para estimação de parâmetros, obtenção de intervalos de confiança para os parâmetros analisados, obtenção de distribuição empírica dos estimadores e determinação do tamanho da amostra (MANLY, 1997). Utilizamos neste trabalho 1000 replicações para a construção de intervalos de confiança e 100 replicações para o erro-padrão.

Com a técnica de reamostragem *bootstrap* não-paramétrico no modelo AMMI, foram reamostradas 100 matrizes de genótipos × ambientes, obtém-se  $B_1=100$  conjuntos de autovalores,  $(\lambda_{1i}^2 > \lambda_{2i}^2 > \cdots > \lambda_{pi}^2 > 0)$ , em que  $i = 1, 2, \cdots, 100$  e  $p = min\{(g-1), (e-1)\} = 8$ . Para a construção de intervalos de confiança foram reamostradas 1000 matrizes de genótipos × ambientes, obtém-se  $B_2=1000$  conjuntos de autovalores,  $(\lambda_{1i'}^2 > \lambda_{2i'}^2 > \cdots > \lambda_{pi'}^2 > 0)$ ,  $i' = 1, 2, \cdots, 1000$  e  $p = min\{(g-1), (e-1)\} = 8$ .

Para cada autovalor foram feitas as análises separadas, como histograma, Q-Q plot, Envelope simulado e com a aplicação do teste de normalidade de Shapiro-Wilk através da técnica de reamostragem *bootstrap*. Portanto, utilizando a metodologia AMMI-*bootstrap* (LAVORANTI, 2003) e a técnica de reamostragem *bootstrap* não-paramétrico podemos encontrar a distribuição empírica dos autovalores e construir os intervalos de confiança. Todas as análises, gráficos e a rotina computacional (no apêndice B) foram implementadas através do sistema estatístico software R (2010).

#### 4.3 Resultados e Discussão

Os resultados utilizando a análise AMMI relativa à produtividade de milho  $(kgha^{-1})$ , obtém-se os elementos do desdobramento da  $SQ_{G\times E}$  por decomposição por valores singulares (DVS), aplicada à matriz de interações, correspondentes aos quadrados dos oito valores singulares  $(\lambda_s^2)$ , com  $s = 1, 2, \dots, 8$ , ou equivalentemente, aos oito autovalores de  $(\widehat{GE}_{G\times E})(\widehat{GE}_{G\times E}^T)$  ou  $(\widehat{GE}_{G\times E}^T)(\widehat{GE}_{G\times E})(Tabela 8)$ :

$\mathbf{PC}$	Proporção	Proporção $\mathrm{AC}(\%)^1$	$\mathrm{GL}^2$	$\mathrm{SQ}^3$	$QM^4$	${\rm F_{Gollob}}^5$	Valor- $p$
PC1	$0,\!562$	56,2	26	140323418	5397054,5	$8,\!95$	$< 0,\!01$
PC2	$0,\!151$	$71,\!3$	24	37711634	1571318,1	$2,\!61$	$< 0,\!01$
PC3	0,104	81,7	22	26058572	1184480,6	$1,\!96$	$< 0,\!01$
PC4	0,086	$90,\!3$	20	21533133	1076656, 9	$1,\!79$	0,0192
PC5	$0,\!050$	$95,\!3$	18	12371704	687316, 9	$1,\!14$	0,3093
PC6	0,028	98,1	16	6900236	431264,7	0,72	0,7744
PC7	0,010	99,1	14	2513071	$179505,\!1$	0,30	0,9939
PC8	0,009	100,0	12	2292394	$191032,\!8$	$0,\!32$	$0,\!9858$
Total	1,00	_	_	249704162	_	_	_

Tabela 8 - Porcentagem da soma de quadrados da interação  $(G \times E)$  captada por componente principal (PC)

 $AC(\%)^1$ : proporção acumulada;  $GL^2$ : graus de liberdade;  $SQ^3$ : soma de quadrados;  $QM^4$ : quadrado médio;  $F_{Gollob}^5$ : teste F de Gollob (1968).

Observando à decomposição da interação genótipo × ambiente dos experimentos através do modelo AMMI (Tabela 8), constatou-se que o primeiro, segundo e terceira componente principal  $(PC_s)$  foi significativo (p < 0, 01) pelo teste F de Gollob (1968), e explicam respectivamente 56,2%, 15,1% e 10,4% da variação da SQ<sub>G×E</sub>. Estes três componentes principais juntos somam 81,7% da  $SQ_{G\times E}$ , que é considerada uma resposta padrão presente na  $SQ_{G\times E}$  com 72 graus de liberdade (47,37% dos graus de liberdade da interação).

A não-significância para os  $PC_4$ ,  $PC_5$ ,  $PC_6$ ,  $PC_7$  e  $PC_8$  são desprezíveis e contém apenas ruído (variação aleatória não relacionada com o fenômeno da interação), que pode diminuir a eficiência da interpretação da estabilidade dos genótipos e ambientes na análise gráfica. Assim, a interpretação gráfica, considerando apenas a variação contida nos três primeiros eixos da ACP, é suficiente para avaliar a estabilidade dos genótipos e ambientes (ROCHA, 2002; FREIRE FILHO et al., 2003).

O histograma é uma forma de descrição gráfica de dados quantitativos, agrupados em classes de frequência e a grande vantagem desta ferramenta é permitir facilmente obter informações sobre a distribuição de freqüencias de um determinado grupo de dados. Observando o gráfico da histograma e Q-Q plot dos valores estimados do primeiro autovalor com  $B_1 = 100$  amostras por meio da técnica *bootstrap*, apresentou uma forma simétrica e estão bem próximos à uma distribuição normal. Também podem ser verificadas pelo Q-Q plot, nota-se que os pontos estão dispostos próximo de linha reta, o que leva a concluir que o primeiro autovalor apresentou um bom ajuste à distribuição normal (Figura 3).



Figura 3 - Gráfico de Histograma (esquerda) e Q-Q plot (direita) do primeiro autovalor com 100 reamostragens utilizando a metodologia *bootstrap*, aplicada na matriz de interação genótipos × ambientes

Os dados do primeiro autovalor apresentaram uma média amostral  $\hat{\lambda}_1^2 = 35080855$  e cujo erro padrão é estimado por  $\hat{e}p_{boot}(\hat{\lambda}_1^2) = 796031, 3$  com  $B_1 = 100$  amostras. Os intervalos de confiança (99%) bootstrap com  $B_2 = 1000$  reamostragens para a média foram obtidos pelo método padrão : [9856041 ; 53858254].

O gráfico de Envelope simulado sob o primeiro autovalor com 100 reamostragens utilizando a metodologia *bootstrap*, aplicada na matriz de interação genótipos × ambientes, nota-se que todos os pontos se encontra contidos dentro do envelope (Figura 4). A distribuição do primeiro autovalor apresentou uma distribuição normal, pois pelo teste de normalidade de Shapiro-Wilk, reportando o valor da estatística W e o valor-p foram: W = 0,9905 e valor-p = 0,7074.



Figura 4 - Gráfico de Envelope simulado sob o primeiro autovalor com 100 reamostragens utilizando a metodologia *bootstrap*, aplicada à matriz de interação genótipos × ambientes

Observa-se que o histograma do segundo autovalor com  $B_1 = 100$  amostras, apresentou uma forma simétrica e estão bem próxima de uma distribuição normal, ou também podem ser verificadas pelo Q-Q plot, nota-se que os pontos estão dispostos próximo de linha reta, o que leva a concluir que o segundo autovalor apresentou um bom ajuste à distribuição normal (Figura 5).



Figura 5 - Gráfico de Histograma (esquerda) e Q-Q plot (direita) do segundo autovalor com 100 reamostragens utilizando a metodologia *bootstrap*, aplicada na matriz de interação genótipos × ambientes



Figura 6 - Gráfico de Envelope simulado sob o segundo autovalor com 100 reamostragens utilizando a metodologia *bootstrap*, aplicada à matriz de interação genótipos  $\times$  ambientes

Os dados do segundo autovalor apresentaram uma média amostral  $\hat{\lambda}_2^2 = 11365486$  e cujo erro padrão é estimado por  $\hat{ep}_{boot}(\hat{\lambda}_2^2) = 290021$  com  $B_1 = 100$  amostras. Os intervalos de confiança (99%) bootstrap com  $B_2 = 1000$  reamostragens para a média foram obtidos pelo método padrão : [39997 ; 14816570].

O gráfico de Envelope simulado sob o segundo autovalor com 100 reamostragens utilizando a metodologia *bootstrap*, nota-se que todos os pontos se encontra contidos dentro do envelope (Figura 6). A distribuição do segundo autovalor apresentou uma distribuição normal, pois por meio do teste de normalidade de Shapiro-Wilk, reportando o valor da estatística W e o valor-p foram: W = 0,981 e valor-p = 0,1591.

O histograma do terceiro autovalor com  $B_1 = 100$  amostras, apresentou uma forma simétrica e estão bem próxima de uma distribuição normal, ou verificando pelo gráfico Q-Q plot, nota-se que os pontos estão dispostos próximo de linha reta, o que leva a concluir que o terceiro autovalor apresentou um bom ajuste à distribuição normal (Figura 7).



Figura 7 - Gráfico de Histograma (esquerda) e Q-Q plot (direita) do terceiro autovalor com 100 reamostragens utilizando a metodologia *bootstrap*, aplicada na matriz de interação genótipos × ambientes

Os dados do terceiro autovalor apresentaram uma média amostral  $\hat{\lambda}_3^2 = 6234229$  e cujo erro padrão é estimado por  $\hat{ep}_{boot}(\hat{\lambda}_3^2) = 156486, 3$  com  $B_1 = 100$  amostras. Os intervalos de confiança (99%) bootstrap com  $B_2 = 1000$  reamostragens para a média foram obtidos pelo método padrão : [2419947 ; 10913305].

O gráfico de Envelope simulado sob o terceiro autovalor com 100 reamostragens utilizando a metodologia *bootstrap*, nota-se que todos os pontos se encontra contidos dentro do envelope (Figura 8). A distribuição do terceiro autovalor apresentou uma distribuição normal, pois pelo teste de normalidade de Shapiro-Wilk, reportando o valor da estatística W e o valor-p foram: W = 0,9819e valor-p = 0,1865.

O histograma do quarto autovalor com  $B_1 = 100$  amostras, apresentou uma forma simétrica e estão bem próxima de uma distribuição normal, ou verificando pelo gráfico Q-Q plot, nota-se que os pontos estão dispostos próximo de linha reta, o que leva a concluir que o quarto autovalor apresentou um bom ajuste à distribuição normal (Figura 9).

Os dados do quarto autovalor apresentaram uma média amostral  $\hat{\lambda}_4^2 = 3390191$  e cujo erro padrão é estimado por  $\hat{ep}_{boot}(\hat{\lambda}_4^2) = 103357, 3$  com  $B_1 = 100$  amostras. Os intervalos de confiança (99%) bootstrap com  $B_2 = 1000$  reamostragens para a média foram obtidos pelo método padrão : [4253955 ; 10349045].



Figura 8 - Gráfico de Envelope simulado sob o terceiro autovalor com 100 reamostragens utilizando a metodologia *bootstrap*, aplicada à matriz de interação genótipos  $\times$  ambientes



Figura 9 - Gráfico de Histograma (esquerda) e Q-Q plot (direita) do quarto autovalor com 100 reamostragens utilizando a metodologia *bootstrap*, aplicada na matriz de interação genótipos  $\times$  ambientes



Figura 10 - Gráfico de Envelope simulado sob o quarto autovalor com 100 reamostragens utilizando a metodologia *bootstrap*, aplicada à matriz de interação genótipos  $\times$  ambientes

O gráfico de Envelope simulado sob o quarto autovalor com 100 reamostragens utilizando a metodologia *bootstrap*, nota-se que todos os pontos se encontra contidos dentro do envelope (Figura 10). A distribuição do quarto autovalor apresentou uma distribuição normal, pois pelo teste de normalidade de Shapiro-Wilk, reportando o valor da estatística W e o valor-p foram: W = 0,9858e valor-p = 0,3635.

O histograma do quinto autovalor com  $B_1 = 100$  amostras, apresentou uma forma simétrica e estão bem próxima de uma distribuição normal, ou verificando por meio do gráfico Q-Q plot, nota-se que grande parte dos pontos estão dispostos próximo de linha reta, apenas dois pontos ficou fora da reta, mas não influenciam os resultados e a distribuição dos dados, o que leva a concluir que o quinto autovalor apresentou um bom ajuste à distribuição normal (Figura 11).

Os dados do quinto autovalor apresentaram uma média amostral  $\hat{\lambda}_5^2 = 1693611$  e cujo erro padrão é estimado por  $\hat{ep}_{boot}(\hat{\lambda}_5^2) = 70691, 93$  com  $B_1 = 100$  amostras. Os intervalos de confiança (95%) bootstrap com  $B_2 = 1000$  reamostragens para a média foram obtidos pelo método padrão : [2640740 ; 6538943].



Figura 11 - Gráfico de Histograma (esquerda) e Q-Q plot (direita) do quinto autovalor com 100 reamostragens utilizando a metodologia *bootstrap*, aplicada na matriz de interação genótipos × ambientes

O gráfico de Envelope simulado sob o quinto autovalor com 100 reamostragens utilizando a metodologia *bootstrap*, nota-se que a maioria dos pontos se encontra contidos dentro do envelope, apenas dois pontos encontraram fora dos limites de envelope, mas como esses dois pontos que têm freqüência baixa, que não influenciar os resultados e a distribuição dos dados (Figura 12). A distribuição do quinto autovalor apresentou uma distribuição normal, pois pelo teste de normalidade de Shapiro-Wilk, reportando o valor da estatística W e o valor-p foram: W = 0,9804 e valor-p= 0,1416.



Figura 12 - Gráfico de Envelope simulado sob o quinto autovalor com 100 reamostragens utilizando a metodologia *bootstrap*, aplicada à matriz de interação genótipos × ambientes

O histograma do sexto autovalor apresentou uma forma com "dois picos", o lado direito do histograma tem a forma de um histograma "despenhadeiro", o lado esquerdo do histograma tem um pico e com o valor próxima de zero, isso acontece pois o valor do sexto autovalor foi menor em relação aos autovalores anteriores ( $\lambda_1^2 > \lambda_2^2 > \cdots > \lambda_8^2 > 0$ ). O valor de cada autovalor  $\lambda_s^2$  é igual a variância do componente principal  $PC_s$ , em que  $s = 1, 2, \cdots, p$ ;  $p = min\{(g-1), (e-1)\} = 8$ . O primeiro componente é o que apresenta a maior variância e assim sucessivamente. Observando pelo gráfico Q-Q plot, nota-se que uma parte dos pontos estão dispostos próximo à linha reta, mas existem vários pontos que não seguiram a linha reta, logo influenciaram os resultados e a distribuição dos dados, o que leva a concluir que os autovalores 6 não apresentaram um ajuste à distribuição normal (Figura 13).

Os dados do sexto autovalor apresentaram uma média amostral  $\hat{\lambda}_6^2 = 601648, 7$  e cujo erro padrão é estimado por  $\hat{ep}_{boot}(\hat{\lambda}_6^2) = 49659, 09$  com  $B_1 = 100$  amostras. Os intervalos de confiança (99%) bootstrap com  $B_2 = 1000$  reamostragens para a média foram obtidos pelo método padrão : [1545877 ; 4197074].



Figura 13 - Gráfico de Histograma (esquerda) e Q-Q plot (direita) do sexto autovalor com 100 reamostragens utilizando a metodologia *bootstrap*, aplicada na matriz de interação genótipos × ambientes

O gráfico de Envelope simulado sob o sexto autovalor com 100 reamostragens utilizando a metodologia *bootstrap* apresentou vários pontos que estão fora dos limites de envelope, embora haja bastante pontos que ficaram dentro dos limites, mas esses pontos que estão fora são suficientes para influenciar os resultados e a distribuição dos dados (Figura 14). Pelo teste de normalidade de Shapiro-Wilk, reportando o valor da estatística W e o valor-p foram: W = 0,9017 e valor-p = 0,0042, portanto a distribuição do sexto autovalor não apresentou uma distribuição normal.



Figura 14 - Gráfico de Envelope simulado sob o sexto autovalor com 100 reamostragens utilizando a metodologia *bootstrap*, aplicada à matriz de interação genótipos × ambientes

O histograma do sétimo autovalor apresentou uma forma como "histograma ilha isolada" ou " retângulos isolados", que tem algumas faixas de valores que ficaram isoladas da grande maioria dos dados, gerando uma barra separada. Isso acontece por causa do valor do sétimo autovalor foi menor do que os outros autovalores anteriores, quase próxima de zero, que foi mostrado no lado esquerdo do histograma e o lado direito tem a forma de um histograma "despenhadeiro". Observando por meio do gráfico Q-Q plot, nota-se que poucos pontos estão dispostos próximo à linha reta, existem grande parte dos pontos ficaram fora da linha reta, com isso influenciaram os resultados e a distribuição dos dados, o que leva a concluir que o sétimo autovalor não apresentou um ajuste à distribuição normal (Figura 15).

O gráfico de Envelope simulado sob o sétimo autovalor com 100 reamostragens utilizando a metodologia *bootstrap*, nota-se que quase todos os pontos estão fora dos limites de envelope (Figura 16). Pelo teste de normalidade de Shapiro-Wilk, reportando o valor da estatística W e o valor-p foram: W = 0,5899 e valor-p = 0,0077, portanto a distribuição do sétimo autovalor não apresentou uma distribuição normal.



Figura 15 - Gráfico de Histograma (esquerda) e Q-Q plot (direita) do sétimo autovalor com 100 reamostragens utilizando a metodologia *bootstrap*, aplicada na matriz de interação genótipos  $\times$  ambientes



Figura 16 - Gráfico de Envelope simulado sob o sétimo autovalor com 100 reamostragens utilizando a metodologia *bootstrap*, aplicada à matriz de interação genótipos  $\times$  ambientes

Os dados do sétimo autovalor apresentaram uma média amostral  $\hat{\lambda}_7^2 = 129060, 4$  e cujo erro padrão é estimado por  $\hat{e}p_{boot}(\hat{\lambda}_7^2) = 22281, 94$  com  $B_1 = 100$  amostras. Os intervalos de confiança (99%) bootstrap com  $B_2 = 1000$  reamostragens para a média foram obtidos pelo método padrão : [617398 ; 1680491].

O histograma do oitavo autovalor apresentou apenas um pico, essas faixas de valores bem próximas a zero, por causa do valor do oitavo autovalor que foi o menor entre todos os autovalores. Observando pelo gráfico Q-Q plot, nota-se que quase todos os pontos estão fora da linha reta, o que leva a concluir que o oitavo autovalor não apresentou um ajuste à distribuição normal (Figura 17).

O gráfico de Envelope simulado sob o oitavo autovalor com 100 reamostragens utilizando a metodologia *bootstrap*, nota-se que quase todos os pontos estão fora dos limites de envelope (Figura 18). Pelo teste de normalidade de Shapiro-Wilk, reportando o valor da estatística W e o valor-p foram: W = 0,2441 e valor-p < 0,0022, portanto a distribuição do oitavo autovalor não apresentou uma distribuição normal.

Os dados do oitavo autovalor apresentaram uma média amostral  $\hat{\lambda}_8 = 7350, 7$  e cujo erro padrão é estimado por  $\hat{ep}_{boot}(\hat{\lambda}_8^2) = 3824, 227 \text{ com } B_1 = 100 \text{ amostras.}$  Os intervalos de confiança (99%) bootstrap com  $B_2 = 1000$  reamostragens para a média foram obtidos pelo método padrão : [1016415 ; 1260182].



Figura 17 - Gráfico de Histograma (esquerda) e Q-Q plot (direita) do oitavo autovalor com 500 reamostragens utilizando a metodologia *bootstrap*, aplicada na matriz de interação genótipos × ambientes



Figura 18 - Gráfico de Envelope simulado sob o oitavo autovalor com 500 reamostragens utilizando a metodologia *bootstrap*, aplicada à matriz de interação genótipos  $\times$  ambientes

## 4.4 Conclusão

O método *bootstrap* não-paramétrico aplicado a matriz de resíduo, estimado via metodologia AMMI, constitui uma eficiente alternativa para encontrar a distribuição empírica dos autovalores e calcular o intervalo de confiança. Portanto, com a aplicação do teste de normalidade de Shapiro-Wilk para cada autovalor, concluímos que o primeiro autovalor até o quinto autovalor apresentaram uma distribuição normal, a partir do sexto autovalor até o oitavo autovalor não apresentaram o mesmo resultado anterior.

#### Referências

ALTMAN, D.G. Practical statistics for medical research. London: Chapman; Hall, 1991. 611p.

ARAÚJO, L.B. Métodos de correção de autovalores e regressão isotônica nos modelos **AMMI**. 2005. 75p. Dissertação (Mestrado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2005.

CORNELIUS, P. L.; CROSSA, J. Prediction assessment of shrinkage estimators of multiplicative models for multi-environment trials. **Crop Science**, Madison, v.39, p.998-1009, 1999.

COSTA, E. F. N.; SOUZA, J.C.; LIMA, J. L.; CARDOSO, G. A. Interação entre genótipos e ambientes em diferentes tipos de híbridos de milho. **Pesquisa Agropecuária Brasileira**, Brasília, v.45, n.12, p.1433-1440, 2010.

DAVISON, A.C.; HINKLEY, D.V. Bootstrap methods and their application. New York: Cambridge University Press. 1997.

DIAS, C.T.S.; KRZANOWSKI, W.J. Model selection and cross validation in additive main effect and multiplicative interaction models. **Crop Science**, Madison, v.43, p.865-873, 2003.

EFRON, B. Bootstrap methods: another look at jakknife. Annals of Statistics, Hayward, v.7, n.1, p.1-26, 1979.

EFRON, B.; TIBSHIRANI, R. An introduction to the bootstrap. New York: Chapman; Hall, 1993. 436p.

FREIRE FILHO, F. R.; RIBEIRO, V. Q.; ROCHA, M. M.; LOPES, Â. C. A. Adaptabilidade e estabilidade da produtividade de grãos de genótipos de caupi enramador de tegumento mulato. **Pesquisa** Agropecuária Brasileira, Brasília, v.38, n.5, p.591-598, 2003.

GOLLOB, H.F. A statistical model which combines feature of factor analitic and analysis of variance techniques. **Psychometrika**, New York, v.33, p.73-115, 1968.

HENDRICKX, J.M.H.; WIERENGA, P.J. Variability of soil water tension in a trickle irrigated Chile pepper field. Irrigation science, New York, v.11, p.23-30, 1990.

JHUN, M.; JEONG, H. Applications of bootstrap methods for categorical data analysis. Computational Statistics and Data Analysis, Amsterdam, v.35, p.83-91, 2000.

KARABATSOS, G. A critique of Rasch residual fit statistics. Journal of Applied Measurement, v.1, p.152-176, 2000.

KENDALL, M. G.; STUART, M. A. **The advanced theory of statistics**. 4th ed. London: Charles Griffin, 1977. v.1.

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

LEITE, J.G.; SINGER, J.M. Métodos assintóticos em estatística: fundamentos e aplicações. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 9., 1990, São Paulo. Anais... São Paulo: IME/USP, 1990. 130p.

MAIA, M. C. M.; VELLO, N. A.; ROCHA, M. de M.; PINHEIRO, J. B.; SILVA JÚNIOR, N. F. da. Adaptabilidade e estabilidade de linhagens experimentais de soja selecionadas para caracteres agronômicos através de método uni-multivariado. **Bragantia**, Campinas, v.65, p.215-226, 2006.

MANDEL, J. Non-additivity in two-way analysis of variance. Journal of the American Statistical Association, New York, v.56, p.878-888, 1961.

MANLY, B.F.J. Randomization, bootstrap and Monte Carlo methods in biology. 2nd ed. New York: Chapman; Hall, 1997. 399p

MARTINEZ, E.Z.; LOUZADA-NETO, F. Estimação interval via bootstrap. **Revista de** Matemática e Estatística, São Paulo, v.19, p.217-251, 2001.

MOHAMMADI, R.; AMRI, A. Analysis of genotype x environment interactions for grain yield in durum wheat. **Crop Science**, Madison, v.49, n.4, p.1177-1186, 2009.

MOLENAAR, I.W.; HOIJTINK, H. The many null distributions of person fit indices. **Psychometrika**, Williamsburg, v.55, p.75-106, 1990.

MUIRHEAD, R.J. Developments in eigenvalue estimation. In: GUPTA, A.K. (Ed.) Advances in multivariate statistical analysis. Dordrecht: Reidel, 1987. p.277-288.

R DEVELOPMENT CORE TEAM. R: A language and environment for statistical computing. **R** Foundation for Statistical Computing, Vienna, 2010. ISBN 3-900051-07-0.

ROCHA, M. M. Seleção de linhagens experimentais de soja para adaptabilidade e estabilidade fenotípica. 2002. 174 f. Tese (Doutorado em Genética e Melhora mento de Plantas) - Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, 2002.

SCHAAP, M.G.; LEIJ, F.J. Using neural networks to predict soil water retention and soil conductivity. Soil and Tillage Research, Amsterdam, v.47, p.37-42, 1998.

SMITH, R.M. The distributional properties of Rasch standardized residuals. Educational and Psychological Measurement, Durham, v.48, p.657-667, 1988.

STONE, C.A. A computer program for assessing goodnes of fit of item response theory models based on posterior expectations. **Applied Psychological Measurement**, St. Paul, v. 28, p. 143-144, 2007.

SU, Y.H.; SHEU, C.F.; WANG, W.C. Computing confidence intervals of item fit in the family of rasch models using the bootstrap method. **Journal of Applied Measurement**, Maple Grove, v.8, p.190-203, 2007.

TACONELI, C.A.; BARRETO, M.C.M. Avaliação de uma proposta de intervalos de confiança bootstrap em amostragem por conjuntos ordenados perfeitamente. **Revista de Matemática e Estatística**. São Paulo, v.23, n.3, p.33-53, 2005.

WANG, W.C.; CHEN, C.T. Item parameter recovery, standard error estimates, and fit statistics of the Winsteps program for the family of Rasch models. Educational and Psychological Measurement, v.65, p.376-404, 2005.

WARZECHA, T.; ADAMSKI, T.; KACZMAREK, Z.; SURMA, M.; CHELKOWSKI, J.; WIŚNIEWSKA, H.; KRYSTKOWIAK, K.; KUCZYŃSKA, A. Genotype-by-environment interaction of barley DH lines infected with Fusarium culmorum (W.G.Sm.)Sacc. **Field Crops Research**, Amsterdam, v.120, p.21-30, 2011.

WOLFE, E.W. RBF. sas (Rasch Bootstrap Fit): A SAS macro for estimating critical values for Rasch model fit statistics. **Applied Psychological Measurement**, St. Paul, v. 32, p. 585-586, 2008.

YAN, W.; HOLLAND, J.B. A Heritability-adjusted GGE biplot for test environment evaluation. **Euphytica**, Wageningen, v.171, n.3, p.355-369, 2010.

YAN, W.; KANG, M.S. **GGE Biplot analysis:** a graphical tool for breeders, geneticists, and agronomists. Flórida:Boca Raton, 2003.

# APÊNDICES

**APÊNDICE A** 

```
a) *Análise de variância conjunta;
proc anova data = AMMIMED;
class AMB GENOT;
model YIELD = AMB GENOT AMB*GENOT;
test H = AMB GENOT E = AMB*GENOT;
run;
*Matriz de Médias;
proc sort data=AMMIMED;BY GENOT AMB;run;
proc means data=AMMIMED mean;
var YIELD;
by GENOT AMB;
OUTPUT OUT=MEDIAS_1;
run;
PROC PRINT DATA=MEDIAS_1;RUN;
DATA AMMI;
SET MEDIAS_1;
IF _STAT_ = ''MEAN";
DROP OBS _TYPE_
                  _FREQ_ _STAT_;
PROC PRINT; RUN;
/*
b) Ajuste dos efeitos principais e obtenção da matriz de interações GxE:
*/
title ''ESTUDO DE INTERAÇÃO GxE { MÉTODO AMMI";
title2 ('Análise com dados de Média (QM'originais/r)";
proc glm data=AMMI;
   class GENOT AMB;
   model YIELD=GENOT AMB;
  run; title2 'Médias de Ambientes e Genótipos';
   Lsmeans GENOT AMB;
   Output out=GEAM r=GE;
run;
proc sort data=AMMI;
by GENOT;
proc transpose data=AMMI out=MEDIAS (drop=_name_) prefix=A;
  var YIELD;
  by GENOT;
run;
proc print data=MEDIAS (drop=GENOT);
  title2 ''A matriz X de médias (gxa)";
```

run; proc sort data=GEAM; by GENOT; proc transpose data=GEAM out=GXE (drop=\_name\_) prefix=A; var GE; by GENOT; run; proc print data=GXE (drop=GENOT); title2 ''Matriz de interações GxE (gxa)"; run; /\* II) A decomposição da integração GxE a) DVS da matriz de interações GxE: \*/ proc iml; GLERR= 540; QMERR= 755888; use work.MEDIAS; setin work.MEDIAS; read all var {A1 A2 A3 A4 A5 A6 A7 A8 A9 A10 A11 A12 A13 A14 A15 A16 A17 A18 A19 A20} into X; ng=nrow(X); na=ncol(X); GM=X[,:]; AM=X[:,]; GG=X[:,:]; CORR=(ng\*na)\*(GG##2); SQTR=ssq(X)-CORR; SQGEN=na\*ssq(GM)-CORR; SQAMB=ng\*ssq(AM)-CORR; SQINT=SQTR-SQGEN-SQAMB; GLMOD=(ng-1)+(na-1)+(ng-1)\*(na-1); XG=GM\*J(1,na,1);XA=J(ng,1,1)\*AM; MG=J(ng,1,1)\*GG\*J(1,na,1); GE=X-XG-XA+MG;EIGV=EIGVAL(GE\*GE'); \*print EIGV ; \*QUIT; NNUL1=EIGV<-1E-8; NNUL2=EIGV>1E-8; NNUL=NNUL1+NNUL2; r=(NNUL^=0)[+,]; title2 'Decomposição por Valores Singulares (A=USV')da matriz GE'; SQGE=ssq(GE); print 'SQ dos elementos da matriz GE = SQ(GxE):' SQGE; print 'O posto da matriz GE:' r; CALL SVD(eG,VS,eA,GE);S=DIAG(VS[1:r,]); U=eG[,1:r];VT=eA[,1:r]'; print'Matriz de valores singulares(S):' S[format=12.4]/ 'Matriz de autovalores (U) associados a genótipos (AA'):' U[format=12.6]/ 'Matriz de autovetores (V') associados a ambientes (A'A): VT[format=12.6]/;

```
b) Desdobramento da SQ(GxE) por AMMI (ACP ou DVS):
*/
Li=VS##2; AUTV=Li[1:r,];
SAUTV=SUM(Li);
PRP=INV(SAUTV)*Li;PROP=PRP[1:r,];
RESD1=SAUTV-Li[1:1,];
PACUM=CUSUM(PROP)*100;
LINH=(1:r)';
RESUMO=LINH||AUTV||PROP||PACUM;
colu={'CP/Eixo''Autovalor''Prop./CP''% Acumulada'};
title2 'Resultados do desdobramento da SQ(GxE) por AMMI (DVS ou ACP)';
print 'Notas:',
           $'_____',
           'Autovalores da ACP<=>Desdobramento da SQ(GxE) por AMMI',
           'Proporção da SQ(GxE) atribuída a cada CP (ou eixo singular de interação)',
 'Percentagem da SQ(GxE) acumulada até o n-ésimo CP(ou eixo singular)'
 '______';$
print RESUMO [colname=colu format=12.4];
print, 'Soma dos Autovalores<=>SQ(GxE):' SAUTV/;
title2;
/*
c) Escores de genótipos e ambientes e aproximações sucessivas de GE e X:
*/
n=0;
 GXEI=j(ng,na,0); GLACU=0;
 SQGEI=ssq(GXEI);
 PROPI=INV(SAUTV)*SQGEI;
 PSQTR=inv(SQTR)*(SQGEN+SQAMB);
 XPO=XG+XA-MG;
         , _____, ____,
 print
          'Resultados da avaliação do modelo: AMMI' n;
          ' SQmod/SQtrat:' PSQTR' SQge(mod)/SQge:' PROPI,
 print
          ·_____, /;
 print 'Matriz de respostas preditas (gxa): Modelo AMMI' n;
 print XP0[format=12.2]/;
do n=1 to r by 1; i=n;
 VG=eG[,i]; ES=S[i,i]; VA=eA[,i];
 IN=VG*ES*VA';
 G=VG*(sqrt(ES));ESGE=G';
 H=sqrt(ES)*VA;ESAM=H';
 if i=1 then ECP1=G//H;
 if i=2 then ECP2=G//H;
```

```
if i=3 then ECP3=G//H;
  if i=4 then ECP4=G//H;
  if i=5 then ECP5=G//H;
  if i=6 then ECP6=G//H;
  if i=7 then ECP7=G//H;
  if i=8 then ECP8=G//H;
 GEI=G*H'; ZE=IN-GEI;
  SQGEI=SQGEI+ssq(GEI);
  PROPI=INV(SAUTV)*SQGEI;
  GLAMI=ng+na-1-(2*i); QMAMI=inv(GLAMI)*(ssq(GEI));
 GLACU=GLACU+GLAMI;
  GLRAMI=(ng-1)*(na-1)-GLACU;
  if i<r then QMRAMI=inv(GLRAMI)*(SAUTV-SQGEI); else QMRAMI=0;
  FMODI=inv(QMERR)*QMAMI; FRESI=inv(QMERR)*QMRAMI;
  GXEI=GXEI+GEI;
  XPI=XG+XA-MG+GXEI;
  PSQTR=inv(SQTR)*(SQGEN+SQAMB+SQGEI);
  prob1=1-probf(FMODI,GLAMI,GLERR);
  if i<r then prob2=1-probf(FRESI,GLRAMI,GLERR); else prob2={.};</pre>
  FV1=GLAMI||QMAMI||FMODI||PROB1;FV2=GLRAMI||QMRAMI||FRESI||PROB2;
  FV3=GLERR||QMERR||{.}||{.};
  FV=FV1//FV2//FV3;
  lin={'AMMIn(GxE)' 'Res.GxE/AMMIn' 'Erro médio'};
 col={'GL''QM''teste F''Pr>F'};
  print 'Resultados da avaliação do modelo: AMMI' n;
  print FV[rowname=lin colname=col format=12.4];
  print 'SQmod/SQtrat:' PSQTR 'SQge(mod)/SQge:' PROPI/;
  print 'Escores para genótipos e ambientes / Modelo ajustado: AMMI' n,
  'Escores de genótipos' ESGE [format=12.4],
  'Escores de ambientes' ESAM [format=12.4]/;
  print 'Aproximação DVS para a Matriz de interações (gxa): Modelo AMMI' n;
 print GXEI[format=12.4]/;
  print 'Matriz de respostas preditas (gxa): Modelo AMMI' n;
  print XPI[format=12.2]/;
end;
```

```
/*
```

d)Resultados para a representação gráfica (Biplot):

\*/

```
GEN=1:ng;AMB=1:na;
OBJ=(GEN||AMB)';
PROD=GM//AM';
RES=OBJ||PROD||ECP1||ECP2||ECP3||ECP4||ECP5||ECP6||ECP7||ECP8;
RESGEN=RES[1:9,1:4];
```

```
RESAMB=RES[10:29,1:4];
VAR1={'GENOTIPOS' 'MEDIAS' 'IPCA1' 'IPCA2' 'IPCA3'};
VAR2={'AMBIENTES' 'MEDIAS' 'IPCA1' 'IPCA2' 'IPCA3'};
create RESULT1 from RESGEN [colname=VAR1];
append from RESGEN;
close RESULT1;
create RESULT2 from RESAMB [colname=VAR2];
append from RESAMB;
close RESULT2;
quit;
proc print data=RESULT1;
title2 "Resultados úteis para a representação gráfica em Biplot";
title3 ('Médias e escores (3 primeiros IPCA's) para genótipos";
run;
proc print data=RESULT2;
title2 ''Resultados úteis para a representação gráfica em Biplot";
title3 ('Médias e escores (3 primeiros IPCA's) para ambientes";
run;
/* e) Representação gráfica (Biplot): */
data BIPLOT1GEN(keep=xsys ysys x y color function position size text style);
length text \ 10;
set RESULT1;
text=GENOTIPOS;
style = 'ZAPFB';
xsys='2'; ysys='2'; color='blue'; position='5'; function='label';
size=1.2:
x=IPCA1;
y=IPCA2;
data BIPLOT1AMB(keep=xsys ysys x y color function position size text style);
length text \ 10;
set RESULT2;
text=AMBIENTES;
style = 'SWISSB';
xsys='2'; ysys='2'; color='red'; position='5'; function='label';
size=1.4;
x=IPCA1;
y=IPCA2;
data GRAFICO1;
set BIPLOT1GEN BIPLOT1AMB;
data BIPLOT2GEN(keep=xsys ysys x y color function position size text style);
length text \ 10;
set RESULT1;
text=GENOTIPOS;
style = 'ZAPFB';
xsys='2'; ysys='2'; color='blue'; position='5'; function='label';
size=1.2;
x=MEDIAS;
```

```
y=IPCA1;
data BIPLOT2AMB(keep=xsys ysys x y color function position size text style);
length text \ 10;
set RESULT2;
text=AMBIENTES;
style = 'SWISSB';
xsys='2'; ysys='2'; color='red'; position='5'; function='label';
size=1.4;
x=MEDIAS;
y=IPCA1;
data GRAFICO2;
set BIPLOT2GEN BIPLOT2AMB;
proc gplot data=GRAFIC01;
symbol1 v=none i=none color=white;
plot y*x=1 y*x=1/anno=GRAFICO1 overlay vref=0 href=0 ;
title1 'Biplot IPCA2(y) vs IPCA1(x)';
title2 'Obs.: N genótipos maiores e ambientes menores';
run;
proc gplot data=GRAFICO2;
symbol1 v=none i=none color=white;
plot y*x=1 y*x=1/anno=GRAFICO2 overlay vref=0 href=0;
title1 'Biplot IPCA1(y) vs MEDIAS(x)';
title2 'Obs.: N genótipos maiores e ambientes menores';
run;
```

### **APÊNDICE B**

```
dados<-read.csv2("dados.csv",head=TRUE);</pre>
library(agricolae)
library(klaR)
library(boot)
AMB
      <- dados$AMB
GENOT <- dados$GENOT
bloco <- dados$bloco
PROD <- dados$PROD
data.ammi <- as.data.frame(cbind(AMB,GENOT,bloco,PROD))</pre>
model<- AMMI(AMB,GENOT,bloco,PROD,ylim=c(-30,40),xlim=c(-55,50))</pre>
envelope.kuang <- function(valores){</pre>
#valores <- res.linha$t[,1]</pre>
res<-sort(valores)
env<-quantile(res,c(.05,0.5,.95))</pre>
for (i in 1:999){
 y.env<-rnorm(rep(1,100), mean(valores),sd(valores))</pre>
 res<-cbind(res,sort(y.env))</pre>
}
for (i in 1:nrow(res)){
 env<-cbind(env,quantile(res[i,],c(.05,0.5,.95)))</pre>
}
matplot(qnorm((1:nrow(res))/(nrow(res)+1)),cbind(t(env[,-1]),res[,1]),
col=c(1,1,1,1),pch=c(NULL,NULL,1),cex=.5,type=c("l","l","l","p"),lty=c(1,2,1,1),
xlab="quantile", ylab="valores")
}
geno <- 9
enva <- 20
genoamb <- matrix(NaN,nrow=20 , ncol= 9)</pre>
for (ncol in 1:9){
  for (nrow in 1:20){
    genoamb[nrow,ncol] <- model$genXenv[ncol,nrow]</pre>
  }
}
boot.linha <- function(matrix, indices)</pre>
{
  newmatrix
                  <- matrix[indices,]
  newmatrixcentr <- newmatrix</pre>
  decomp <- svd(newmatrixcentr)</pre>
vprop <- (decomp$d)^2</pre>
return((decomp$d)^2)
}
res.linha = boot(data=t(genoamb),statistic=boot.linha,R=100)
###Análises feitas para o primeiro autovalor com B1=100 amostras.
```

```
plot(res.linha,index=1)
shapiro.test(res.linha$t[,1])
envelope.kuang(res.linha$t[,1])
hist(res.linha$t[,1], prob=TRUE)
lines(density(res.linha$t[,1]))
hist(res.linha$t[,1], prob = T, main = "")
mean(res.linha$t[,1])
sd(res.linha$t[,1])
curve(dnorm(x, mean=38490473, sd=7209990), lwd=2, add=T)
####Análises feitas para o segundo autovalor com B1=100 amostras.
plot(res.linha,index=2)
shapiro.test(res.linha$t[,2])
mean(res.linha$t[,2])
sd(res.linha$t[,2])
hist(res.linha$t[,2], prob = T, main = "")
curve(dnorm(x, mean=11632954, sd=2728419),lwd=2,add=T)
envelope.kuang(res.linha$t[,2])
####Análises feitas para o terceiro autovalor com B1=100 amostras.
plot(res.linha,index=3)
shapiro.test(res.linha$t[,3])
mean(res.linha$t[,3])
sd(res.linha$t[,3])
hist(res.linha$t[,3], prob = T, main = "")
curve(dnorm(x, mean=6481337, sd=1431400),lwd=2,add=T)
envelope.kuang(res.linha$t[,3])
##### Análises feitas para o quarto autovalor com B1=100 amostras.
plot(res.linha,index=4)
shapiro.test(res.linha$t[,4])
mean(res.linha$t[,4])
sd(res.linha$t[,4])
hist(res.linha$t[,4], prob = T, main = "")
curve(dnorm(x, mean=3559668, sd=1050636),lwd=2,add=T)
envelope.kuang(res.linha$t[,4])
########Análises feitas para o quinto autovalor com B1=100 amostras.
plot(res.linha,index=5)
shapiro.test(res.linha$t[,5])
mean(res.linha$t[,5])
sd(res.linha$t[,5])
hist(res.linha$t[,5], prob = T, main = "")
curve(dnorm(x, mean=1612974, sd=684341),lwd=2,add=T)
envelope.kuang(res.linha$t[,5])
####Análises feitas para o sexto autovalor com B1=100 amostras.
plot(res.linha,index=6)
shapiro.test(res.linha$t[,6])
mean(res.linha$t[,6])
sd(res.linha$t[,6])
hist(res.linha$t[,6], prob = T, main = "")
curve(dnorm(x, mean=612416.9, sd=521416.9), lwd=2, add=T)
```

```
envelope.kuang(res.linha$t[,6])
####Análises feitas para o sétimo autovalor com B1=100 amostras.
plot(res.linha,index=7)
shapiro.test(res.linha$t[,7])
mean(res.linha$t[,7])
sd(res.linha$t[,7])
hist(res.linha$t[,7], prob = T, main = "")
curve(dnorm(x, mean=123195.5, sd=223897.5), lwd=2, add=T)
envelope.kuang(res.linha$t[,7])
####Análises feitas para o oitavo autovalor com B1=100 amostras.
plot(res.linha,index=8)
shapiro.test(res.linha$t[,8])
mean(res.linha$t[,8])
sd(res.linha$t[,8])
hist(res.linha$t[,8], prob = T, main = "")
curve(dnorm(x, mean=9906.516, sd=43387.15), lwd=2, add=T)
envelope.kuang(res.linha$t[,8])
#####0s intervalos de confiança bootstrap com B2=1000 amostras.
res.linha = boot(data=t(genoamb),statistic=boot.linha,R=1000)
ic.kuang <- boot.ci(res.linha,type='norm',index=1)</pre>
ic.kuang
res.linha$t0[1]
ic.kuang <- boot.ci(res.linha,type='norm',index=2)</pre>
ic.kuang
res.linha$t0[2]
ic.kuang <- boot.ci(res.linha,type='norm',index=3)</pre>
ic.kuang
res.linha$t0[3]
ic.kuang <- boot.ci(res.linha,type='norm',index=4)</pre>
ic.kuang
res.linha$t0[4]
ic.kuang <- boot.ci(res.linha,type='norm',index=5)</pre>
ic.kuang
res.linha$t0[5]
ic.kuang <- boot.ci(res.linha,type='norm',index=6)</pre>
ic.kuang
res.linha$t0[6]
ic.kuang <- boot.ci(res.linha,type='norm',index=7)</pre>
ic.kuang
res.linha$t0[7]
ic.kuang <- boot.ci(res.linha,type='norm',index=8)</pre>
ic.kuang
res.linha$t0[8]
```