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

**Validação cruzada com correção de autovalores e regressão isotônica nos
modelos AMMI**

Pamela Piovesan

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

Piracicaba
2007

Pamela Piovesan
Licenciada em Matemática

Validação cruzada com correção de autovalores e regressão isotônica nos modelos AMMI

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

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

Piracicaba
2007

Dados Internacionais de Catalogação na Publicação (CIP)
DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - ESALQ/ USP

Piovesan, Pamela

Validação cruzada com correção de autovalores e regressão isotônica nos modelos
AMMI / Pamela Piovesan. - - Piracicaba, 2007.
77 p.

Dissertação (Mestrado) - - Escola Superior de Agricultura Luiz de Queiroz, 2007.
Bibliografia.

1. Análise de dados 2. Análise multivariada 3. Estatística aplicada 4. Genótipos 5. Meio
ambiente I. Título

CDD 519.05

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

DEDICATÓRIA

À Deus,

pela oportunidade de viver, aprimorar os meus conhecimentos e me tornar uma pessoa cada vez melhor.

*Aos meus pais **Antonio Agnelo Piovesan e Angela Maria de Almeida Piovesan**, por me proporcionarem uma base sólida em valores necessários para enfrentar todos os momentos da vida, por me encherem de amor, me tranquilizarem nas aflições e comemorarem comigo cada conquista, enfim, por sempre fazerem parte da minha vida.*

*Aos meus irmãos **Agnelo Piovesan, Vinicius Aparecido Piovesan e Gustavo Antonio Piovesan**, por cada um, à sua maneira, me incentivar e acreditar no sucesso do meu trabalho.*

Família Amo Vocês!

AGRADECIMENTOS

Ao professor Dr. Carlos Tadeu dos Santos Dias, minha eterna gratidão, a começar pelo exemplo de vida, pela orientação, pelos conselhos fundamentais, pela amizade, compreensão e incentivo constante.

Aos meus amigos de mestrado Alexandre, Ana Paula, Cristiane, Fernanda, Joseane e Juliana pelo companheirismo, amizade e união tornando a nossa caminhada mais agradável mesmo nos momentos mais difíceis de se percorrer.

Aos colegas de mestrado Edjane, Eduardo, Elton, Erick, Jalmar, Luziane, Renata, Renato, Simone, Marina e Mirian e aos colegas de doutorado Afrânio, Ana Alice, Ana Maria, Angela, Andréia, César, David, Denise, Édila, Elizabeth, Genevile, Giovana, Juliana Cespedes, Luciana, Osmar, Vanderli, e Wilson, em especial ao Lúcio pelo apoio na realização do trabalho.

À Beth pela convivência durante o ano de 2005 e por ter me ajudado em tudo o que eu precisei.

À Juliana por ter sido a minha família em Piracicaba durante o ano de 2006, por todos os momentos vividos que ficarão guardados para sempre nas minhas boas lembranças.

Aos funcionários do Departamento de Ciências Exatas da ESALQ/USP, as secretárias Solange de Assis Paes Sabadin e Luciane Brajão e aos técnicos em informática Jorge Alexandre Wiendl e Eduardo Bonilha pelos auxílios permanentes.

Aos professores do Departamento de Ciências Exatas da ESALQ/USP, Dra. Clarice Garcia Borges Demétrio, Dr. César Gonçalves de Lima, Dr. Décio Barbin, Dr. Edwin Moisés Marcos Ortega, Dr. Gerson Barreto Mourão, Dra. Roseli Aparecida Leandro, Dr. Silvio Sandoval Zocchi e Dra. Sônia Maria Stefano Piedade, pelos conhecimentos que compartilharam conosco, pela amizade e pela prontidão de nos ajudarem sempre.

Ao CNPq pelo apoio financeiro sem ao qual não seria possível a realização dessa pesquisa.

A todas as pessoas que contribuíram direta ou indiretamente para a realização deste trabalho.

SUMÁRIO

RESUMO	8
ABSTRACT	9
1 INTRODUÇÃO	10
2 REVISÃO BIBLIOGRÁFICA	13
2.1 Interação entre genótipos e ambientes	13
2.1.1 Análise de variância em cada ambiente	16
2.1.2 Análise de variância conjunta	16
2.2 Autovalor e Autovetor de uma matriz	18
2.3 Decomposição por Valores Singulares e Análise de Componentes Principais	18
2.4 Modelos AMMI	21
2.5 Procedimento para determinar o número de termos multiplicativos no modelo AMMI	23
2.6 Correção de autovalores	25
2.7 Regressão Isotônica	26
2.7.1 Regressão isotônica para um caso simples de dados ordenados	26
2.7.2 Regressão isotônica em um conjunto quase-ordenado	28
2.7.3 Regressão isotônica generalizada	29
3 METODOLOGIA	30
3.1 Material	30
3.2 Métodos	30
3.3 Análise AMMI	30
3.4 Teste de Hipóteses	33
3.5 Validação Cruzada	34
3.6 Métodos de Correção de autovalores	37

	6
3.7	Procedimento de Stein para a Isotonização 39
3.7.1	Número de blocos 40
4	RESULTADOS E DISCUSSÃO 42
4.1	Experimento 1 42
4.1.1	Análise de variância conjunta 42
4.1.2	Teste de hipóteses 42
4.1.3	Validação cruzada 43
4.1.3.1	Validação cruzada com correção de autovalores 44
4.1.4	Número de Blocos 45
4.2	Experimento 2 46
4.2.1	Análise de variância conjunta 46
4.2.2	Teste de hipóteses 46
4.2.3	Validação cruzada 47
4.2.3.1	Validação cruzada com correção de autovalores 47
4.2.4	Número de Blocos 49
4.3	Experimento 3 49
4.3.1	Análise de variância conjunta 49
4.3.2	Teste de hipóteses 50
4.3.3	Validação cruzada 50
4.3.3.1	Validação cruzada para o experimento 3 com correção de autovalores 51
4.3.4	Número de Blocos 51
4.4	Experimento 4 52
4.4.1	Análise de variância conjunta 52
4.4.2	Teste de hipóteses 52
4.4.3	Validação cruzada 53

	7
4.4.3.1 Validação cruzada com correção de autovalores	54
4.4.4 Número de Blocos	54
5 CONSIDERAÇÕES FINAIS	55
5.1 Conclusões	55
REFERÊNCIAS	57
ANEXOS	61

RESUMO

Validação cruzada com correção de autovalores e regressão isotônica nos modelos AMMI

Neste trabalho apresenta-se a aplicação dos modelos AMMI para um estudo detalhado do efeito das interações entre genótipos e ambientes em experimentos multi-ambientais. Através da decomposição da soma de quadrados dessas interações, busca-se selecionar o número de termos que explicam essa interação, descartando o ruído presente na mesma. Há duas maneiras para a escolha desses termos: validação cruzada e teste de hipóteses. O foco será na validação cruzada pela vantagem de ser um critério “preditivo” de avaliação. São apresentados dois métodos de validação cruzada, métodos esses esboçados por Eastment e Krzanowski (1982) e Gabriel (2002). Esses métodos utilizam a decomposição por valores singulares para obter os autovalores referentes à matriz de interações, cuja soma de quadrados nos dá exatamente a soma de quadrados da interação. Como esses autovalores são superestimados ou subestimados (ARAÚJO; DIAS, 2002), essas técnicas de validação serão aperfeiçoadas através da correção desses autovalores e, para reordená-los, será utilizada a regressão isotônica. Será realizado um estudo comparativo entre esses métodos, através de dados reais.

Palavras-chave: Modelos AMMI; Interação genótipo $\times$ ambiente; Validação cruzada; Correção de autovalores; Regressão isotônica

ABSTRACT

Cross-validation with eigenvalue correction and isotonic regression in models AMMI

This paper presents the application of AMMI models for a thorough study about the effect of the interaction between genotypes and environments in multi-environments experiments. Through the decomposition of the sum of squares of these interactions, one searches to select the number of terms that explains this interaction, discarding its noise in. There are two ways for choosing these terms: cross-validation and hypotheses test. The focus will be on the cross-validation for its advantage of being one prediction criterion of evaluation. Two methods of cross-validation are presented, both outlined by Eastment and Krzanowski (1982) and Gabriel (2002). These methods use the decomposition by singular values in order to obtain eigenvalues referred to the matrix of interactions, whose sum of squares accurately gives us the sum of squares of the interaction. As these eigenvalues either over- or underestimated (ARAÚJO; DIAS, 2002), these techniques of validation will be improved through the correction of these eigenvalues and, in order to rearrange them, isotonic regression will be used. A comparative study between these methods through real data will be carried out.

Keywords: AMMI models; Genotype $\times$ environment interaction; Cross-validation; Correction of eigenvalues; Isotonic regression

1 INTRODUÇÃO

Em experimentos agrícolas, os ensaios são realizados em vários ambientes para, através de uma análise estatística adequada, isolar o componente de variabilidade devido à interação entre genótipos e ambientes ($G \times E$).

Os programas de melhoramento têm como principal objetivo buscar técnicas capazes de detectar de maneira aprofundada essas interações para a seleção de genótipos consistentes e de elevada produtividade.

Os métodos tradicionais consistem em realizar o zoneamento agronômico, permitindo recomendações regionalizadas, ou em utilizar métodos de regressão linear simples e múltipla com o objetivo de identificar genótipos estáveis que podem ser recomendados de forma generalizada, desde que eles mostrem uma “performance” desejável.

A crítica feita à técnica de zoneamento agronômico é que o agrupamento é feito com base nas diferenças macro ambientais, tornando-o vulnerável às oscilações imprevisíveis que possam ocorrer no ambiente, além de não permitir o controle da interação genótipo e anos. (ROSSE, 1999).

Em relação à análise de regressão linear Crossa (1990) argumenta a possibilidade da linearidade falhar, além da técnica ser altamente dependente do tipo de genótipos e ambientes incluídos na análise e tender a simplificar padrões de resposta.

A aplicação dos métodos multivariados é tida como uma técnica que melhor explora as informações contidas nos dados. Dentre essas técnicas destacam-se a análise de componentes principais, a análise de agrupamentos e a análise AMMI.

A análise AMMI pode ajudar tanto na identificação de genótipos de alta produtividade e largamente adaptados, como na realização do zoneamento agronômico, com fins de recomendação regionalizada e seleção de locais de teste, permitindo também a análise detalhada da interação ($G \times E$), a seleção de genótipos mais produtivos, estimativas mais precisas das respostas genotípicas e fácil interpretação gráfica dos resultados da análise es-

tatística (ZOBEL et al., 1988).

O método AMMI surge, então, com a finalidade de estudar detalhadamente as interações ($G \times E$), através da decomposição da soma de quadrados dessas interações, fato que o torna vantajoso se comparado aos métodos tradicionais, além da sua boa capacidade preditiva. O AMMI é um procedimento uni-multivariado que vem ganhando destaque, pois combina num único modelo componentes aditivos para os efeitos principais e componentes multiplicativos para os efeitos da interação genótipo-ambiente, combinando sequencialmente a análise de variância (técnica univariada) e a análise de componentes principais (técnica multivariada), respectivamente, podendo a análise de componentes principais ser diretamente tratada através da decomposição por valores singulares (DVS).

A interação ($G \times E$) é representada pelos termos multiplicativos do modelo AMMI e a seleção do número ótimo desses termos pode ser feita de duas maneiras: através de testes de hipóteses como os apresentados por Gollob (1968), Mandel (1969, 1971), Cornelius (1993), Corsten e van Eijnsberger (1972), Johnson e Graybill (1972), Hegemann e Johnson (1976) e Yochmowitz e Cornell (1978) entre outros, ou através de formas de validação cruzada, apresentadas por Stone (1974), Velicer (1976), Wold (1978) e Eastment e Krzanowski (1982).

A validação cruzada é um procedimento estatístico livre de distribuição teórica que vem ganhando aplicação recente e Gauch (1988) introduziu a chamada validação preditiva, por meio da validação cruzada para selecionar o melhor modelo AMMI. Há várias formas de se fazer validação cruzada. Dias e Krzanowski (2003) descrevem duas metodologias preditivas, as mesmas apresentadas por Eastment e Krzanowski (1982) e Gabriel (2002), baseadas em dados “leave-one-out” na determinação do número dos termos multiplicativos a serem retidos no modelo AMMI, com vantagens sobre os testes de Gollob e Cornelius usados para esse mesmo fim.

A DVS, utilizada tanto para a parte multiplicativa do modelo AMMI quanto nos dois procedimentos de validação cruzada apresentados, é baseada no uso de autovalores e autovetores e como há viés na estimação dos autovetores (ARAÚJO; DIAS, 2006), serão aqui utilizados três métodos, apresentados por Muirhead (1987), para a correção das estimativas.

A regressão isotônica será utilizada para reordenar os autovalores, cuja ordem decrescente nem sempre é mantida quando se aplicam os métodos de correção de autovalores.

Assim, a presente dissertação tem os seguintes objetivos: aplicar o modelo AMMI- Additive Main effects and Multiplicative Interaction model- com o propósito de uma análise detalhada na estrutura de interação entre genótipo e ambiente em experimentos agronômicos multi-ambientais com dados balanceados; apresentar e aplicar técnicas de validação cruzada para a escolha do número de termos multiplicativos do modelo AMMI; aperfeiçoar as técnicas de validação cruzada utilizando três métodos de correção de autovalores, ordenados através da regressão isotônica; implementar uma rotina computacional para a análise dos dados segundo as técnicas de validação cruzada no modelo AMMI e avaliar os métodos de validação cruzada, aqui modificados, através do estudo de dados reais.

2 REVISÃO BIBLIOGRÁFICA

2.1 Interação entre genótipos e ambientes

Experimentos em múltiplos ambientes são importantes em agronomia porque os efeitos dos tratamentos agronômicos podem mudar diferencialmente em relação à mudança ambiental, produzindo uma interação tratamento-ambiente (CROSSA et al., 2001). À resposta diferenciada dos genótipos nos vários ambientes, chamamos de interação genótipos-ambientes ($g \times e$) (EBERHART; RUSSEL, 1996).

Em programas de melhoramento, o principal objetivo dos melhoristas é selecionar genótipos que garantam uma consistente e elevada produtividade, nos mais diversos ambientes. A baixa eficiência na análise da interação genótipo-ambiente pode reduzir a precisão de seleção de um ambiente para outro. A presença da interação genótipo-ambiente provoca um aumento do desvio padrão fenotípico, reduz a herdabilidade ao longo de ambientes e, conseqüentemente, reduz os ganhos genéticos potenciais (MATHERSON; RAYMOND, 1986).

O fenótipo é resultado da ação conjunta do genótipo, do ambiente e da interação genótipo-ambiente. O ambiente reflete as diferentes sensibilidades dos genótipos às variações do ambiente, resultando em mudanças de seus desempenhos relativos. Não basta detectar a presença de interações, deve-se também considerar a sua natureza (Vencovsky e Barriga, 1992) para a escolha consciente de qual é o melhor ambiente a receber os genótipos selecionados.

Ramalho et al (1993) relatam que a interação genótipo-ambiente pode ser simples, possibilitando o cultivar ser recomendado de forma generalizada, ou pode ser complexa, indicando a presença de material adaptado a ambientes particulares, tornando a recomendação restrita a ambientes específicos.

A significância dessas interações deve ser interpretada, a princípio, como uma indicação de que existem genótipos particularmente adaptados a ambientes específicos e outros que sejam menos influenciados pelas variações ambientais, podendo ser explorados em vários

ambientes. Nesse último caso, tem-se uma adaptação geral e o inconveniente seria o fato desses genótipos estarem associados a uma baixa produtividade, impedindo a recomendação desses cultivares de forma generalizada, acarretando maiores dificuldades e exigindo a adoção de medidas que controlem ou minimizem os efeitos dessa interação, para que seja possível proceder a uma recomendação mais segura.

Os melhoristas tendem a desenvolver uma atitude negativa contra a interação. Contudo, a presença de interação pode ser vista como um aspecto positivo, pois uma vez que ela seja positiva ao ambiente, haverá uma oportunidade de ganhos elevados. Adaptações específicas de genótipos a ambientes podem fazer a diferença entre uma boa e uma excelente variedade (GAUCH; ZOBEL, 1996).

A fase inicial de um programa de melhoramento consiste na avaliação do genótipo em apenas um ambiente, nesse caso, as estimativas da variância genética são acrescidas do efeito somente daquele ambiente nos genótipos. Assim, os ganhos genéticos previstos com a seleção não são reais. Nas etapas finais dos programas de melhoramento os experimentos são conduzidos em vários ambientes, possibilitando o isolamento da variabilidade relativa à interação genótipo-ambiente. Porém, a intensidade de seleção é baixa fazendo com que o ganho genético seja de pequena magnitude. Assim, os efeitos principais de genótipos e ambientes não podem ser interpretados de forma isolada e deve-se considerar a interação genótipo-ambiente (DUARTE; VENCovsky, 1999; KANG; MAGARI, 1996; MANDEL, 1971).

Ao longo dos últimos trinta anos, várias metodologias têm sido propostas para captar as informações relativas à essas interações e pesquisadores estatísticos continuam na busca de instrumentos mais efetivos para a análise. Uma opção seria identificar cultivares específicos para cada ambiente. Gauch e Zobel (1996) relatam que genótipos com adaptações específicas podem fazer a diferença entre uma boa e uma excelente variedade, possibilitando rendimentos diferenciados nas etapas finais de um programa de seleção. Contudo, o alto custo dessa execução inviabiliza esse procedimento (RAMALHO et al., 1993; RIBEIRO, 1999).

Um método tradicional a utilizar é o zoneamento agrônômico que consiste em

estratificar a população de ambientes em sub-regiões mais uniformes, dentro das quais a interação não se mostre significativa, permitindo, assim, recomendações regionalizadas (BRASIL, 1990; DUARTE; ZIMMERMANN, 1995).

A crítica feita a essa técnica é que o agrupamento é feito com base nas diferenças macro ambientais, tornando o genótipo vulnerável às oscilações imprevisíveis que possam ocorrer no ambiente. Além disso, não permite o controle da interação genótipo e anos (ROSSE, 1999).

A técnica que mais tem sido utilizada é a que consiste em identificar genótipos com maior estabilidade fenotípica (VENCOVSKY; BARRIGA, 1992; RAMALHO et al., 1993). Há várias metodologias propostas para medir essa estabilidade fenotípica.

Duarte e Vencovsky (1999) relatam que dentre os métodos que procuram identificar os genótipos estáveis, que menos contribuem para a interação, e que poderiam ser recomendados genericamente, desde que mostrem uma performance média desejável, estão os métodos baseados na regressão linear simples e múltipla (FINLAY; WILKINSON, 1963; EBERHART; RUSSELL, 1996; SILVA; BARRETO, 1985; CRUZ et al., 1989; BRASIL, 1990). Entretanto, Crossa (1990) argumenta que a análise de regressão linear não é informativa se a linearidade falhar, é altamente dependente do tipo de genótipos e ambientes incluídos na análise e tende a simplificar padrões de resposta, por explicar a variação da interação em apenas uma dimensão, pelo coeficiente de regressão, quando na realidade ela pode ser bastante complexa. Zobel et al (1988) reforçam que procedimentos univariados como a análise de variância são limitados em detectar a interação de fatores, mesmo em situações de elevada magnitude em termos de soma de quadrados.

Devido aos problemas citados, a tendência atual consiste em utilizar procedimentos multivariados, que podem ser úteis para melhor explorar as informações contidas nos dados, ou fazer a integração entre métodos univariados e métodos multivariados com o objetivo de se ter um melhor entendimento dos efeitos causados pela interação genótipo-ambiente, como são os procedimentos que levam aos modelos AMMI.

2.1.1 Análise de variância em cada ambiente

Em um experimento aleatorizado em blocos, para cada ambiente $j, j = 1, 2, \dots, e$, a análise de variância, sendo o efeito do genótipo considerado como fixo, é dada pelo seguinte modelo matemático:

$$Y_{ih} = \mu + g_i + b_h + \varepsilon_{ih} \quad (1)$$

em que:

Y_{ih} : é a resposta do i -ésimo genótipo no h -ésimo bloco, com $i = 1, 2, \dots, g$ e $h = 1, 2, \dots, r$;

μ : é uma constante, geralmente a média geral;

g_i : é o efeito do i -ésimo genótipo;

b_h : é o efeito do h -ésimo bloco;

ε_{ih} : é o erro experimental associado ao i -ésimo genótipo no h -ésimo bloco, assumido independente e $\varepsilon_{ih} \sim N(0, \sigma^2)$.

O esquema da análise de variância é dado por:

Tabela 1 - Esquema da análise de variância local

Fonte de Variação	Graus de Liberdade (GL)	Quadrado Médio (QM)
Blocos (B)	$(r - 1)$	QM_B
Genótipos (G)	$(g - 1)$	QM_G
Resíduos (Res)	$(g - 1)(r - 1)$	QM_{Res}
Total	$(gr - 1)$	

2.1.2 Análise de variância conjunta

Em experimentos multi-ambientais, supondo que em todos os ambientes foram delineados aleatoriamente em blocos, para analisar se há interação entre ambiente e genótipo, realiza-se a análise de variância conjunta. Em termos de predição é preferível assumir o efeito dos genótipos como aleatório e o efeito dos ambientes como fixo (HILL; ROSENBERGER, 1985; STROUP; MULITZE, 1991). O modelo matemático é dado por:

$$Y_{ijh} = \mu + g_i + e_j + (ge)_{ij} + b_{h(j)} + \varepsilon_{ijh} \quad (2)$$

em que:

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

μ : 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;

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

O esquema da análise de variância é dado por:

Tabela 2 - Esquema da análise de variância conjunta

Fonte de Variação	Graus de Liberdade (GL)	Quadrado Médio (QM)
Blocos d. ambientes (B d. E)	$e(r - 1)$	$QM_{Bd.E}$
Genótipos (G)	$(g - 1)$	QM_G
Ambientes (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	$(ger - 1)$	

Em experimentos com repetições, a análise de variância é feita com médias e para o cálculo do QM_{EM} utilizam-se os Quadrados Médios dos Resíduos obtidos das análises de variâncias individuais e calcula-se a média ponderada dada por:

$$QM_{EM} = \frac{\sum_{j=1}^j SQ_{Erro_j}}{\sum_{j=1}^j GL_{Erro_j}} \quad (3)$$

em que $j = 1, 2, \dots, e$ e os graus de liberdade das análises individuais são utilizados como pesos.

2.2 Autovalor e Autovetor de uma matriz

Para qualquer matriz quadrada $\mathbf{A}$, um escalar λ e um vetor não-nulo $\mathbf{x}$ podem ser encontrados, de tal forma que:

$$\mathbf{Ax} = \lambda\mathbf{x}$$

em que λ é chamado um *autovalor* de $\mathbf{A}$ e $\mathbf{x}$ é um *autovetor* de $\mathbf{A}$ associado a λ , também conhecidos como *valor característico* e *vetor característico* de $\mathbf{A}$ respectivamente. O vetor $\mathbf{x}$ é transformado por $\mathbf{A}$ em um múltiplo de si próprio, de tal forma que o vetor $\mathbf{Ax}$ está sobre a linha que passa por $\mathbf{x}$ e a origem.

Para encontrar λ e $\mathbf{x}$ associados à matriz $\mathbf{A}$, escreve-se

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0} \quad (4)$$

e resolve-se esse sistema homogêneo, que admite solução não trivial se e somente se

$$|\mathbf{A} - \lambda\mathbf{I}| = 0. \quad (5)$$

A equação (5) é chamada de *equação característica* e é uma equação polinomial em λ . Se $\mathbf{A}$ é $(n \times n)$ a equação característica terá n raízes, isto é, $\mathbf{A}$ terá n autovalores $\lambda_1, \lambda_2, \dots, \lambda_n$. Os λ 's não serão necessariamente distintos ou todos diferentes de zero, ou todos números reais, mas os autovalores de uma matriz simétrica serão reais (RENCHE, 2000). Depois de encontrar $\lambda_1, \lambda_2, \dots, \lambda_n$, os autovetores $\mathbf{x}$'s são encontrados usando (4) ao substituir λ por cada λ_i , $i=1, 2, \dots, n$. Claramente teremos um sistema de equações (4) indeterminado a resolver (veja restrição(5)), podendo assim atribuir valores arbitrários a alguns elementos de $\mathbf{x}$, inclusive o sinal desses.

2.3 Decomposição por Valores Singulares e Análise de Componentes Principais

Suponha que $\mathbf{X}$ é uma matriz arbitrária $m \times n$ de números reais de posto p . Então $\mathbf{X}$ pode ser expressa como a soma de p matrizes de posto 1 pela decomposição de valores singulares:

$$\mathbf{X} = \lambda_1 \mathbf{u}_1 \mathbf{v}_1' + \lambda_2 \mathbf{u}_2 \mathbf{v}_2' + \dots + \lambda_p \mathbf{u}_p \mathbf{v}_p' \quad (6)$$

em que: λ_i são números reais positivos; os vetores colunas $\mathbf{u}_i$ são ortonormais, isto é, ortogonais e de comprimento unitário, e cada um tem m elementos e os vetores linha $\mathbf{v}'_i$ são ortonormais e cada um tem n elementos. Os vetores $\mathbf{u}_1, \dots, \mathbf{u}_p$ e $\mathbf{v}'_1, \dots, \mathbf{v}'_p$ podem ser completados fazendo $\mathbf{u}_1, \dots, \mathbf{u}_m$ e $\mathbf{v}'_1, \dots, \mathbf{v}'_n$. A decomposição por valores singulares completa de $\mathbf{X}$ é então $\sum_{j=1}^t \lambda_j \mathbf{u}_j \mathbf{v}'_j$, em que $t = \min(m, n)$ e $\lambda_i = 0$ para $j = p + 1, \dots, t$ se $p < t$.

Escrevendo $\mathbf{u}_j$ como as colunas de uma matriz $\mathbf{U}$ $m \times n$, e $\mathbf{v}'_j$ como as linhas de uma matriz $\mathbf{V}'$ $n \times n$, e definindo a matriz diagonal $\mathbf{\Lambda}$ $n \times n$ com os elementos: $\lambda_{jj} = \lambda_j$ e $\lambda_{ij} = 0$ ($i \neq j$), então a decomposição de valores singulares tem a forma equivalente:

$$\mathbf{X} = \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ u_{21} & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ u_{m1} & u_{m2} & \cdots & u_{mn} \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{n1} & v_{n2} & \cdots & v_{nn} \end{pmatrix} \quad (7)$$

A ortonormalidade de $\mathbf{u}_i$ e $\mathbf{v}'_i$ assegura que $\mathbf{U}$ e $\mathbf{V}$ são matrizes ortogonais, isto é, $\mathbf{U}'\mathbf{U} = \mathbf{V}\mathbf{V}' = \mathbf{I}_n$. Os vetores $\mathbf{u}_i$ e $\mathbf{v}'_i$ são os autovetores ortonormalizados da matriz $\mathbf{X}\mathbf{X}'$ e $\mathbf{X}'\mathbf{X}$, respectivamente. Os números λ_i ($i = 1, \dots, p$) são as raízes quadradas dos autovalores não nulos de $\mathbf{X}'\mathbf{X}$ ou equivalentemente $\mathbf{X}\mathbf{X}'$.

Quando $\mathbf{X}$ é simétrica, segue que $\mathbf{U}$ é igual a $\mathbf{V}$ pois, nessa situação, $\mathbf{X}\mathbf{X}' = \mathbf{X}'\mathbf{X}$. Tem-se que $\mathbf{U} = \mathbf{V}$ é a matriz de autovetores de $\mathbf{X}$, enquanto $\mathbf{\Lambda}$ é a matriz diagonal de autovalores de $\mathbf{X}$; a decomposição de valores singulares nesse caso é a decomposição espectral de $\mathbf{X}$.

Há várias importantes aplicações da decomposição por valores singulares (DVS) como na obtenção da inversa generalizada de matrizes, na solução de equações lineares, na especificação de modelos de interação em análise de variância de um ensaio com dois fatores e nos cálculos envolvidos na técnica multivariada de componentes principais.

Uma aplicação imediata da DVS é, supondo $\mathbf{X}$ ser uma matriz $m \times n$ de posto p tendo a DVS dada por (4) ou (5), e deseja-se aproximar $\mathbf{X}$ por uma matriz $\mathbf{Y}$ de posto r , sendo $r < p$. A matriz $\mathbf{Y}$ terá p matrizes ortogonais entre si e de posto unitário (SOUZA, 1988) e

representará a melhor aproximação de posto r para $\mathbf{X}$, com a menor soma de quadrados da discrepância $S = \sum_{i=1}^m \sum_{j=1}^n (x_{ij} - y_{ij})^2 = \text{traço}\{(\mathbf{X} - \mathbf{Y})(\mathbf{X} - \mathbf{Y})'\}$ entre os elementos x_{ij} de $\mathbf{X}$ e os elementos y_{ij} de $\mathbf{Y}$. Essa aproximação chama-se *aproximação- r de quadrados mínimos* para a matriz $\mathbf{X}$ (GOOD, 1969; SOUZA, 1988).

A análise de componentes principais (ACP) tem como objetivo reduzir a dimensionalidade do conjunto de variáveis e facilitar a interpretação da interdependência 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 e (ou novas variáveis) são chamados de *componentes principais* e os valores das novas variáveis são chamados de *escores dos componentes principais* ou *coordenadas principais*.

A vantagem dessas novas variáveis é elas serem não correlacionadas e, além disso, o primeiro componente principal explicar o máximo de variabilidade dos dados, o segundo componente principal explicar o máximo da variabilidade que não foi explicada pela primeira variável e assim por diante. Dessa forma, um número pequeno de componentes principais poderá explicar uma grande porcentagem da variabilidade original dos dados. É válido notar que a variância total das novas variáveis é igual a variância total das variáveis originais, com a vantagem descrita.

Algebricamente, obtêm-se os componentes principais da seguinte maneira: Sejam $X_1, X_2, \dots, X_p$ variáveis aleatórias e o vetor aleatório $\mathbf{X}' = [X_1, X_2, \dots, X_p]$ com matriz de variâncias e covariâncias Σ . Encontram-se os pares de autovalores e autovetores ortonormalizados $(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \dots, (\lambda_p, \mathbf{e}_p)$, em que $\lambda_1 \geq \lambda_2 \geq \lambda_p$, associados a Σ e então o j -ésimo componente principal será:

$$C_j = \mathbf{e}_j' \mathbf{X} = e_{1j}X_1 + e_{2j}X_2 + \dots + e_{pj}X_p, \quad j = 1, 2, \dots, p$$

e as variâncias e covariâncias serão dadas por:

$$\begin{aligned} \text{Var}(C_j) &= \mathbf{e}_j' \Sigma \mathbf{e}_j & j &= 1, 2, \dots, p \\ \text{Cov}(C_j, C_k) &= \mathbf{e}_j' \Sigma \mathbf{e}_k = 0 & j &\neq k. \end{aligned}$$

Para relacionar a ACP e DVS, considere $\mathbf{X}$ uma matriz $n \times p$ de dados centrados na média das variáveis, isto é, $x_{ij} = x_{ij} - \bar{x}_j$, $i = 1, \dots, n$ e $j = 1, \dots, p$ e $\mathbf{S}$ a matriz de variâncias e covariâncias associada a $\mathbf{X}$. Por meio da DVS de $\mathbf{X}$, chega-se a $x_{ij} = \sum_{t=1}^p u_{it} d_t v_{tj}$. Realizando a ACP, obtêm-se $\mathbf{C}_\mathbf{X} = \mathbf{X}_{(n \times p)} \mathbf{V}_{p \times p}$, pois $\mathbf{S}_\mathbf{X} = \frac{1}{n-1} \mathbf{X}' \mathbf{X}$ e os autovetores ortonormalizados de $\mathbf{X}' \mathbf{X}$ e de $\frac{1}{n-1} \mathbf{X}' \mathbf{X}$ são os mesmos. Como, pela DVS, $\mathbf{X} = \mathbf{U} \mathbf{D}_\mathbf{X} \mathbf{V}'$, então, $\mathbf{C}_\mathbf{X} = \mathbf{U} \mathbf{D}_\mathbf{X}$, uma vez que $\mathbf{V}' \mathbf{V} = \mathbf{I}_p$. Segue também que $\mathbf{X} = \mathbf{C}_\mathbf{X} \mathbf{V}'$.

Então, escolher m componentes no modelo é equivalente a modelar os elementos de $\mathbf{X}$ por:

$$x_{ij} = \sum_{t=1}^m u_{it} d_t v_{tj} + \varepsilon_{ij}$$

em que ε_{ij} é o termo residual.

Na análise AMMI utiliza-se a ACP, tratada diretamente pela DVS, para obter uma aproximação de uma matriz de posto inferior e compreender a estrutura subjacente aos dados da matriz de interações $\mathbf{GE}$.

2.4 Modelos AMMI

A metodologia AMMI (“additive main effects and multiplicative interaction model”) é uma técnica que vem ganhando grande aceitação nos últimos anos, pois possibilita uma melhor interpretação e compreensão na interação genótipo-ambiente, combinando a técnica univariada de análise de variância com a técnica multivariada de análise de componentes principais. A grande aceitação dessa análise pode ser atribuída, entre outras coisas, à sua eficiência preditiva (ZOBEL et al., 1988).

O modelo foi inicialmente proposto por Mandel (1969, 1971) e baseia-se na decomposição das fontes de variação em efeitos aditivos de genótipos e ambientes da maneira tradicional, através da análise de variância (ANOVA) e, posteriormente, em efeitos multiplicativos para a interação genótipo-ambiente através da análise de componentes principais (ACP), permitindo um maior detalhamento da soma de quadrados da interação (ZOBEL et al., 1988).

Suponha que um conjunto de g genótipos tenha sido testado experimentalmente em e ambientes, que podem ser locais, anos, etc. As médias de cada combinação de genótipo e ambiente, obtidas de r repetições de um experimento que consiste de dados balanceados, podem ser representadas pela matriz:

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

O modelo AMMI é dado por:

$$\bar{Y}_{ij} = \mu + g_i + e_j + (ge)_{ij} + \varepsilon_{ij} \quad (8)$$

em que:

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

μ : é a performance média geral;

g_i : é o efeito fixo do genótipo i ($i=1, \dots, g$);

e_j : é o efeito fixo do ambiente j ($j=1, \dots, e$);

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

$(ge)_{ij}$: é o efeito fixo da interação específica do genótipo i e o ambiente j , no qual $(ge)_{ij}$ é representada por:

$$\sum_{k=1}^m \lambda_k \alpha_{ik} \gamma_{jk} + \rho_{ij}. \quad (9)$$

sendo que:

λ_k : é o k -ésimo valor singular da matriz de interação $(\mathbf{GE})$, com $k = 1, 2, \dots, m$ e onde $m < s$ determina uma aproximação de mínimos quadrados para a matriz $\mathbf{GE}$ pelos m primeiros termos da DVS e $s = \min\{g - 1, e - 1\}$;

α_{ik} : é o i -ésimo elemento do vetor singular coluna $\boldsymbol{\alpha}_{k(g \times 1)}$ associado a λ_k ;

γ_{jk} é o j -ésimo elemento do vetor singular linha $\boldsymbol{\gamma}'_{k(1 \times e)}$;

ρ_{ij} : é o resíduo adicional.

Tem-se as seguintes restrições:

$$\sum_i g_i = \sum_j e_j = \sum_i (ge)_{ij} = \sum_j (ge)_{ij} = 0 \quad (10)$$

uma vez que as estimativas da soma de quadrado de resíduos é feita através do método de mínimos quadrados, e para permitir solução do sistema de equações normais do modelo (8).

2.5 Procedimento para determinar o número de termos multiplicativos no modelo AMMI

Os critérios adotados para definir quantos componentes devem ser retidos no modelo, para explicar o padrão devido à interação, têm sido objeto de pesquisas de vários trabalhos (GOLLOB, 1968; MANDEL, 1971; GAUCH, 1988; GAUCH; ZOBEL, 1988; PIEPHO, 1994; PIEPHO, 1995; CORNELIUS, 1993; CORNELIUS et al., 1996). Há dois procedimentos básicos: por meio de testes de hipóteses ou validação cruzada.

Os testes de hipóteses são feitos por meio de suposições distribucionais. Gauch e Zobel (1988) comentam que avaliações feitas por esses testes podem ser ditas “posdictiva”, pois buscam um modelo que explique uma grande parte da variação dos dados observados e afirmam que tal abordagem não é eficiente para selecionar modelos parcimoniosos e capazes de desprezar ruídos de menor interesse. Por outro lado, um critério “preditivo” de avaliação prioriza a capacidade de um modelo aproximar suas predições a dados não incluídos na análise, simulando respostas futuras ainda não mensuradas.

Por outro lado, os procedimentos livres de distribuição teórica vêm ganhando aplicação crescente, pois quanto menos a escolha do modelo ou avaliação da performance de um preditor é baseada em suposições distribucionais, mais geral é o resultado. Para utilizar esses métodos são necessários procedimentos estatísticos computacionais intensivos e envolvem reamostragem de um conjunto de dados, como “jackknife”, “bootstrap” e validação cruzada.

Gauch (1988) introduziu o nome “avaliação preditiva” quando ela é baseada na validação cruzada (STONE, 1974; WOLD, 1978).

A validação cruzada é um critério “preditivo” de avaliação do modelo que prioriza sua capacidade de aproximar suas predições a dados não incluídos na análise, simulando respostas futuras ainda não mensuradas. O procedimento padrão consiste em subdividir a matriz em um número de grupos. Cada subgrupo é retirado dos dados, um de cada vez, parâmetros de interesse do preditor são avaliados através dos dados remanescentes para cada um dos modelos em estudo, e os valores retirados são então estimados (WOLD, 1976; WOLD, 1978).

Na validação, dividem-se, então, os dados em dois subconjuntos: dados para o ajuste do modelo AMMI e dados de validação para o modelo ajustado. As respostas são preditas por uma família de modelos AMMI, para diferentes valores de m , e são comparadas com os respectivos dados de validação. Obtém-se a soma de quadrados dessa diferença, dividindo-se o resultado pelo número delas. À raiz quadrada desse resultado os autores chamaram de diferença raiz quadrada média preditiva (RMSPD).

Para Crossa et al (1991), esse procedimento deve ser repetido cerca de dez vezes, obtendo-se a média dos resultados para cada membro da família de modelos. O melhor modelo é aquele que apresenta o menor valor do RMSPD e é utilizado para uma análise definitiva dos dados de todas as repetições conjuntamente. Piepho (1994) argumenta que o valor médio do RMSPD deve ser calculado a partir de 1000 diferentes aleatorizações, ao invés de 10 como sugerido por Crossa et al (1991) e ainda quando o ensaio é em blocos, o autor sugere uma modificação da partição completamente aleatória dos dados (modelagem e validação), sorteando o bloco inteiro de um ensaio e não fazer componentes para cada combinação de genótipo e ambiente, para preservar a estrutura original de blocos. Mas, estudos comprovando a sua efetividade ainda não estão disponíveis.

Krzanowski (1987) argumenta que a predição mais precisa resulta quando cada grupo deletado é tão pequeno quanto possível, ou seja, apenas um elemento da matriz e em Gauch e Zobel (1996) há também a sugestão de apenas uma observação constar no conjunto

de dados para validação, por ser mais provável a partir de $n - 1$ dados encontrar um modelo mais próximo do ideal para analisar o conjunto completo de dados.

Dias e Krzanowski (2003) propõem dois métodos baseados em um procedimento “leave-one-out” (LO) completo que otimiza o processo de validação cruzada por validar o ajuste do modelo em cada um dos dados por vez e então combinar essa validação em uma medida simples e geral de ajuste. Esses métodos foram esboçados por Krzanowski (1987) e Gabriel (2002) e serão descritos no item 4.4.

2.6 Correção de autovalores

A DVS, utilizada tanto na parte multiplicativa do modelo, quanto nos procedimentos de validação cruzada apresentados por Eastment e Krzanowski (1982) e Gabriel (2002), é baseada no uso de autovalores e autovetores. Muirhead (1987) afirma que esses autovalores estimados da maneira convencional, são viesados. Araújo e Dias (2006), utilizando técnicas de simulação multivariada, confirmaram esse problema de superestimação e subestimação. Então, para solucionar essa questão, serão utilizados três métodos de correção (item 4.5) de autovalores, tendo como principal objetivo tornar mais precisos os procedimentos de validação cruzada.

Seja Σ uma matriz de covariância populacional de ordem p , com autovalores $\lambda_1 > \lambda_2 > \dots > \lambda_p$ e $\mathbf{S}$, com autovalores $\hat{\lambda}_1 > \hat{\lambda}_2 > \dots > \hat{\lambda}_p$, a estimativa de Σ . Como os autovalores $\hat{\lambda}_1 > \hat{\lambda}_2 > \dots > \hat{\lambda}_p$ estimados tradicionalmente apresentam problema, Muirhead (1987) apresenta três métodos para corrigi-los a partir da matriz de covariância amostral.

Seja $\mathbf{X}$ uma matriz de ordem $n \times m$, em que cada linha representa um vetor normal multivariado de dimensão m com matriz de covariâncias Σ e vetor de médias μ . Segundo Johnstone (2001), se $\mu = \mathbf{0}$, a matriz $\mathbf{S} = \mathbf{X}'\mathbf{X}$ é uma estimativa de Σ e tem distribuição de Wishart com n graus de liberdade e matriz de covariâncias Σ . Para essa matriz $\mathbf{S} \sim W_m(n, \Sigma)$ Muirhead (1987) apresenta os três métodos de correção de autovalores. Porém, quando se corrigem esses autovalores, nem sempre a ordem decrescente deles é mantida e para essa ordenação utiliza-se a regressão isotônica.

2.7 Regressão Isotônica

A regressão isotônica é importante porque quando se aplicam os métodos de correção de autovalores, apresentados por Muirhead (1987), para melhorar os procedimentos de validação cruzada, os autovalores nem sempre mantêm a ordem decrescente. Então, através da regressão isotônica, os autovalores corrigidos serão reordenados, especificamente através do algoritmo para isotonização de Stein (item 4.6), e será prosseguida a análise em determinar o número de termos multiplicativos do modelo AMMI, através dos procedimentos de validação cruzada. A técnica de regressão ajusta curvas ou funções a um conjunto de dados e existem vários métodos estatísticos para verificar a qualidade do ajuste.

Em várias situações de pesquisa acredita-se, devido a informações a “priori”, que os dados devem apresentar uma característica relativa a uma ordenação. Assim, procura-se um modelo que preserve essa característica de ordem. O modelo selecionado será escolhido dentre vários modelos que considere essa ordem dos dados e esse método é denominado *regressão isotônica*. O termo isotônica (ou monotônica) refere-se ao fato de a variável resposta aumentar de acordo com o aumento da variável independente. Se a variável resposta apresentar uma característica decrescente com o aumento da variável preditora, pode-se usar o termo antitônica. Mas esses termos isotônica, monotônica e antitônica possuem o mesmo sentido.

Dentre os diversos tipos de conjuntos existentes, a regressão isotônica é feita em dois conjuntos particulares: os conjuntos com uma simples ordenação de dados e os conjuntos com uma quase ordenação de dados.

2.7.1 Regressão isotônica para um caso simples de dados ordenados

Seja $X = \{x_1 \lesssim x_2 \lesssim \dots \lesssim x_p\}$ um conjunto finito. A relação binária ($\lesssim$), entre os elementos de X , é uma relação de ordem simples se são satisfeitas as seguintes propriedades:

1. Reflexiva: $x \lesssim x$ para todo $x \in X$;
2. Transitiva: $x, y, z \in X$, $x \lesssim y$ e $y \lesssim z$ então $x \lesssim z$;

3. Anti-simétrica: $x, y \in X$, $x \lesssim y$ e $y \lesssim x$ então $x = y$;

4. Todo e qualquer elemento de X é comparável: $x, y \in X$ implica que $x \lesssim y$ ou $y \lesssim x$.

A definição de regressão isotônica para um caso de uma ordenação simples dos dados é apresentada por Robertson; Wright e Dykstra (1998):

Seja $X = \{x_1, x_2, \dots, x_p\}$ um conjunto finito com uma ordem simples $x_1 \lesssim x_2 \lesssim \dots \lesssim x_p$. Uma função f , em X , é isotônica se $f(x_1) \leq f(x_2) \leq \dots \leq f(x_p)$, sendo que $(\lesssim)$ indica a relação existente no conjunto da variável independente e $(\leq)$, a relação existente no conjunto da variável dependente. Seja também w uma função peso, positiva e definida em X .

Suponha que t seja uma função dada em X . Uma função t^* é uma regressão isotônica de t com peso w se e somente se t^* é uma função isotônica e minimiza

$$\sum_{x \in X} [t(x) - f(x)]^2 w(x) \quad (11)$$

dentro de todas classes de funções isotônicas f em X .

Assim, para uma função arbitrária t , definida em X , seja t^* a projeção de mínimos quadrados de t dentro da coleção de funções isotônicas. Especificamente, t^* é uma solução de mínimos quadrados restritos para a expressão (11).

Suponha amostras aleatórias de p populações normais com médias $\mu(x_1), \mu(x_2), \dots, \mu(x_p)$ e variância comum σ^2 . Para a i -ésima amostra, seja n_i a notação para o seu tamanho amostral, sejam $Y_{i1}, Y_{i2}, \dots, Y_{in}$ as observações e $\bar{Y}_i = \sum_{j=1}^{n_i} \frac{Y_{ij}}{n_i}$ o estimador de máxima verossimilhança (EMV), sem restrição de $\boldsymbol{\mu} = (\mu(x_1), \mu(x_2), \dots, \mu(x_p))$, o vetor de médias populacionais. Suponha também, que é conhecido que $\mu(x_1) \leq \mu(x_2) \leq \dots \leq \mu(x_p)$. Pode-se querer estimar essas médias e também satisfazer essa restrição, mas pela variabilidade amostral, a amostra de médias $\bar{Y}_1, \bar{Y}_2, \dots, \bar{Y}_p$, pode não estar ordenada, ou seja, não está obedecendo à restrição das médias populacionais.

Para obter as estimativas que fazem esta restrição, toma-se o negativo do loga-

ritmo da função de verossimilhança. Percebe-se que o EMV restrito minimiza $\sum_{i=1}^p \sum_{j=1}^{n_i} [Y_{ij} - \mu(x_i)]^2$ sujeito a $\mu(x_1) \leq \mu(x_2) \leq \dots \leq \mu(x_p)$. Somando e subtraindo $\bar{Y}_i$, esse somatório pode ser reescrito como:

$$\sum_{i=1}^p \sum_{j=1}^{n_i} [Y_{ij} - \bar{Y}_i]^2 + 2 \sum_{i=1}^p \sum_{j=1}^{n_i} [Y_{ij} - \bar{Y}_i][Y_{ij} - \mu(x_i)] + \sum_{i=1}^p [\bar{Y}_i - \mu(x_i)]^2 n_i.$$

Entretanto, a primeira soma não envolve $\mu(x_i)$ e a segunda soma é zero. Assim o EMV restrito minimiza $\sum_{i=1}^p [\bar{Y}_i - \mu(x_i)]^2 n_i$ sujeito à condição de $\mu(x_1) \leq \mu(x_2) \leq \dots \leq \mu(x_p)$. Portanto, em um conjunto $X = \{x_1, x_2, \dots, x_p\}$ com $x_1 \lesssim x_2 \lesssim \dots \lesssim x_p$, t e w são definidas em X por $t(x_i) = \bar{Y}_i$ e $w(x_i) = n_i$, então está claro que o EMV restrito de $\boldsymbol{\mu} = (\mu(x_1), \mu(x_2), \dots, \mu(x_p))$ é fornecido pela regressão isotônica t^* de t com peso w .

2.7.2 Regressão isotônica em um conjunto quase-ordenado

A relação binária $\lesssim$ em um conjunto $X = \{x_1 \lesssim x_2 \lesssim \dots \lesssim x_p\}$ é uma ordem parcial se são satisfeitas as propriedades de reflexão, transitividade e anti-simétrica. Um conjunto X , com a relação binária $\lesssim$, ainda pode ser um conjunto quase-ordenado se os elementos deste conjunto satisfizerem somente as propriedades de reflexão e transitividade, não sendo necessário que elementos sejam anti-simétricos e nem que sejam comparáveis. Assim, todo conjunto com ordem simples é um conjunto com ordem parcial e todo conjunto parcial é um conjunto com quase-ordem.

Uma função f de valor real em X é isotônica com elementos quase-ordenados ($\lesssim$) em X se $x, y \in X$ então $f(x) \leq f(y)$. Assim, seja t uma função dada em X e w uma outra função positiva em X . Uma função isotônica t^* é uma regressão isotônica de t com peso w se e somente se

$$\left\{ \sum_{x \in X} [t(x) - t^*(x)]^2 w(x) \leq \sum_{x \in X} [t(x) - f(x)]^2 w(x) \right\} \quad (12)$$

para todas funções f em X que são isotônicas.

2.7.3 Regressão isotônica generalizada

Suponha que Φ seja uma função convexa a qual é finita em um intervalo I contendo uma gama de funções t em X e tem o valor $+\infty$ fora do intervalo I . Então, é bem conhecido que Φ é contínua em I , em I^0 (conjunto dos extremos do intervalo I) e, de fato, Φ é diferenciável pela direita e pela esquerda de cada ponto em I^0 . Seja ϕ uma função não decrescente em I que coincida com alguma determinação de uma derivada de Φ em I^0 , ou seja, para cada $x \in I$, $\phi(x)$ é um número entre a derivada à esquerda de Φ em x e à derivada a direita de Φ em x . O valor de ϕ à esquerda do maior ponto de I pode ser $-\infty$ e este valor pode ser $+\infty$ à direita do menor ponto de I .

Para cada $u, v \in I$, define-se uma função $\Delta_\Phi(u, v)$ por:

$$\Delta_\Phi(u, v) = \Phi(u) - \Phi(v) - (u - v)\phi(v)$$

com $\Delta_\Phi(u, v) = 0$ e $\Delta_\Phi(u, v) = +\infty$, se u ou v não pertencer a I . Se f é uma função isotônica em X e se as possíveis f estiverem dentro de I então

$$\sum_{x \in X} \Delta_\Phi[t(x), f(x)]w(x) \geq \sum_{x \in X} \Delta_\Phi[t(x), t^*(x)]w(x) + \sum_{x \in X} \Delta_\Phi[t^*(x), f(x)]w(x).$$

Consequentemente t^* minimiza

$$\sum_{x \in X} \Delta_\Phi[t(x), f(x)]w(x)$$

na classe de todas as funções isotônicas f existentes dentro de I e maximiza

$$\sum_{x \in X} \Phi[f(x)] + [t(x) - f(x)]\phi[f(x)]w(x).$$

3 METODOLOGIA

3.1 Material

Os dados a serem utilizados são referentes a ensaios internacionais multi-ambientais de cultivares obtidos pelo CIMMYT (Centro Internacional de mejoramiento de Maiz y Trigo). São 4 experimentos, todos delineados aleatoriamente em blocos. O experimento 1 é referente a 20 genótipos de trigo avaliados em 34 ambientes com 4 blocos. O experimento 2 contém 16 genótipos de milho analisados em 24 ambientes com 4 blocos. O experimento 3 segue com 9 genótipos de milho em 4 blocos, conduzidos em 20 ambientes. O experimento 4 contém 8 cultivares de milho em 59 ambientes com 4 blocos.

3.2 Métodos

3.3 Análise AMMI

O modelo AMMI postula componentes aditivos para os efeitos principais de genótipos e ambientes e componentes multiplicativos para a interação entre genótipo e ambiente e é dado por:

$$\bar{Y}_{ij} = \mu + g_i + e_j + (ge)_{ij} + \varepsilon_{ij} \quad (13)$$

em que:

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

μ : é a performance média geral;

g_i : é o efeito fixo do genótipo i ($i=1, \dots, g$);

e_j : é o efeito fixo do ambiente j ($j=1, \dots, e$);

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

no qual $(ge)_{ij}$ é representada por:

$$\sum_{k=1}^m \lambda_k \alpha_{ik} \gamma_{jk} + \rho_{ij} \quad (14)$$

sob as restrições

$$\sum_i g_i = \sum_j e_j = \sum_i (ge)_{ij} = \sum_j (ge)_{ij} = \sum_i \sum_j (ge)_{ij} = 0 \quad (15)$$

pois as estimativas da soma de quadrado de resíduos é feita através do método de mínimos quadrados.

O ajuste do modelo (13) por mínimos quadrados é obtido, para dados balanceados, em 2 etapas sequenciais (ZOBEL et al., 1988). A primeira etapa consiste na obtenção das estimativas dos efeitos principais (g_i e e_j) e da média geral μ , estimativas essas que formam a parte aditiva do modelo, sendo ajustadas por uma análise de variância comum (ANOVA) aplicada à matriz de médias $\mathbf{Y}_{g \times e}$. Aplicada a ANOVA, obtém-se um resíduo de não aditividade ε_{ij} , e o efeito da interação ($g \times e$) recai nesse resíduo, formando a matriz de interações **GE**. Cada elemento dessa matriz de interações é obtido por:

$$\hat{\varepsilon}_{ij} = (ge)_{ij} = Y_{ij} - \bar{Y}_{i*} - \bar{Y}_{*j} + \bar{Y}_{**} \quad (16)$$

em que:

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

$\bar{Y}_{i*}$: é a média do genótipo i ;

$\bar{Y}_{*j}$: é a média do ambiente j ;

$\bar{Y}_{**}$: é a média geral do experimento.

Essa interação, constituinte da parte aditiva do modelo, é então analisada pela decomposição por valores singulares (DVS) da matriz de interações **GE**. A partir daí, a análise AMMI centra-se na decomposição da soma de quadrados dos elementos da matriz **GE**, obtida de forma direta por DVS.

Assim, λ_k é o k -ésimo valor singular de $\mathbf{GE}$, α_{ik} é o i -ésimo elemento do vetor singular coluna $\boldsymbol{\alpha}_{k(g \times 1)}$ associado a λ_k e γ_{jk} é o j -ésimo elemento do vetor singular linha $\boldsymbol{\gamma}'_{k(1 \times e)}$ associado a λ_k (GOOD, 1969; MANDEL, 1971; PIEPHO, 1995).

Os dois vetores singulares $\boldsymbol{\alpha}_k$ e $\boldsymbol{\gamma}_k$ representam os efeitos de genótipos e ambientes presentes no k -ésimo componente da interação, enquanto o valor singular λ_k é um indicador de peso desse componente. Sob o ponto de vista de análise de componentes principais, temos as seguintes correspondências:

λ_k é o k -ésimo valor singular ou a raiz quadrada do k -ésimo maior autovalor da matriz $(\mathbf{GE})(\mathbf{GE})'$ e $(\mathbf{GE})'(\mathbf{GE})$, as quais têm igual autovalores não nulos;

α_{ik} é o i -ésimo elemento do autovetor de $(\mathbf{GE})(\mathbf{GE})'$ associado com λ_k^2 ;

γ_{jk} é o j -ésimo elemento do autovetor de $(\mathbf{GE})'(\mathbf{GE})$ associado com λ_k^2 ;

A matriz $\mathbf{GE}$ nesse modelo é, então, expressa como uma soma de componentes, cada um multiplicado por λ_k , para um efeito genotípico (α_{ik}) e um efeito ambiental (γ_{jk}). O termo λ_k dá a proporção da variância devida à interação $\mathbf{G} \times \mathbf{E}$ no k -ésimo componente. Os efeitos α_{ik} e γ_{jk} representam pesos para o genótipo i e o ambiente j , naquele componente da interação (λ_k^2). O posto de $\mathbf{GE}$ é $s = \min(g - 1, e - 1)$, de tal forma que o índice k na soma de componentes multiplicativos pode variar de 1 até s . O uso de todos os s componentes recupera toda a variação da soma de quadrados da interação: $S.Q.(\mathbf{G} \times \mathbf{E}) = \sum_{k=1}^m \lambda_k^2$ e o modelo é saturado de tal forma que ele produz um ajuste exato aos dados, sem o termo residual para testar os efeitos (exceto na situação quando um resíduo independente é estimado). Quando $m < s$, o modelo é dito ser truncado. Entretanto, o modelo AMMI não tenta recuperar toda a $S.Q.(\mathbf{G} \times \mathbf{E})$, mas somente aqueles componentes mais fortes determinados por genótipos e ambientes. Conseqüentemente, o índice k é, geralmente, colocado a variar até $m < s$, de tal forma que as estimativas são obtidas dos m primeiros termos da DVS da matriz $\mathbf{GE}$ (GOOD, 1969; GABRIEL, 1978). Essa é uma análise de mínimos quadrados que conduz a um resíduo ρ_{ij} . Assim, a interação do genótipo i com o ambiente j é descrita por $\sum_{k=1}^m \lambda_k \alpha_{ik} \gamma_{jk}$, descartando o ruído dado por $\sum_{k=m+1}^s \lambda_k \alpha_{ik} \gamma_{jk}$. Aqui, como na análise de componentes principais, o componente conta, sucessivamente, para proporções decrescentes da variação presente na ma-

triz **GE** ($\lambda_1^2 \geq \lambda_2^2 \geq \dots \geq \lambda_k^2$). Dessa forma, o método AMMI é visto como um procedimento capaz de separar padrão e ruído na análise da interação, em que por padrão entende-se a lei geral de formação da matriz e que governa, nesse caso, a interação **GE**, explicada pelos m primeiros termos da DVS. O ruído presente na interação, mas não fortemente determinado por genótipos e ambientes é descartado, buscando-se melhorar a capacidade preditiva do modelo, trazendo benefícios diretos à seleção de genótipos. (ZOBEL et al., 1988).

3.4 Teste de Hipóteses

Na análise AMMI, por construção, a matriz **GE** de interações é uma matriz de desvios centrados e, portanto, com média nula, e a soma de quadrados dos seus elementos reproduz exatamente a soma de quadrados da interação. Assim, essa soma de quadrados é desdobrada em componentes relativos a cada termo da DVS, ou a cada eixo da interação da ACP ($IPCA_k$, $k=1, 2, \dots, s$). Mas, é suficiente encontrar os valores singulares de **GE** e a soma de quadrados relativa ao k -ésimo termo ou eixo da interação equivale ao quadrado do valor singular correspondente (o autovalor λ_k^2).

Para aplicar um teste F com o objetivo de selecionar o número de termos multiplicativos do modelo AMMI, é necessário determinar os graus de liberdade associados à cada um desses λ_k^2 .

O método de Gollob (1968) se destaca por ser simples e exigir pequenas operações aritméticas para a obtenção desses graus de liberdade, e utiliza a definição estrutural clássica de graus de liberdade, isto é, o número de parâmetros estimados menos o número de restrições. A expressão para determinar o número de graus de liberdade para o k -ésimo componente da interação, $k=1, 2, \dots, s$, sendo $s = \min(g-1, e-1)$, é dada por:

$$GL_{IPCA_k} = g + e - 1 + 2k.$$

Obtém-se, então, o quadro de análise de variância apresentado na Tabela 3, que é semelhante ao tradicional, com desdobramento para a fonte de variação $G \times E$.

Tabela 3 - Análise de variância conjunta computada à partir das médias pelo método de Gollob

Fonte de Variação	Graus de Liberdade (GL)	Soma de Quadrado de Gollob
Genótipos (G)	$(g - 1)$	SQ_G
Ambientes (E)	$(e - 1)$	SQ_E
Interação ($G \times E$)	$(g - 1)(e - 1)$	$SQ_{G \times E}$
IPCA ₁	$g + e - 1 - (2 \times 1)$	λ_1^2
IPCA ₂	$g + e - 1 - (2 \times 2)$	λ_2^2
IPCA ₃	$g + e - 1 - (2 \times 3)$	λ_3^2
$\vdots$	$\vdots$	$\vdots$
IPCA _s	$g + e - 1 - (2 \times s)$	λ_s^2
Erro médio/n	$ge(n - 1)$	SQ(Erro médio)
Total	$gen - 1$	SQ(Total)

3.5 Validação Cruzada

Os métodos de validação cruzada, tanto de Eastment e Krzanowski(1982) quanto de Gabriel (2002), utilizam a decomposição por valores singulares e consistem em eliminar cada elemento x_{ij} da matriz $\mathbf{X}$ com dimensão $g \times e$, prever o valor $\hat{x}_{ij}$ de x_{ij} para cada possível escolha de m (o número de componentes) e medir a discrepância entre o valor atual e o valor predito por:

$$PRESS(m) = \frac{1}{ge} \sum_{i=1}^g \sum_{j=1}^e (\hat{x}_{ij}^m - x_{ij})^2.$$

Como o objetivo é estimar cada elemento da matriz, para evitar viés, os dados de x_{ij} não devem ser utilizados nos cálculos de $\hat{x}_{ij}^m$ para cada i e j . Então, alguma forma de validação cruzada é indicada e esses dois procedimentos diferem na forma de estimar $\hat{x}_{ij}$. Através da DVS, escreve-se a matriz $\mathbf{X}$ por:

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}'$$

O procedimento de validação cruzada de Eastment e Krzanowski(1982) é descrito a seguir: Sejam $\mathbf{X}^{-i}$ a matriz resultante após deletar a i -ésima linha de $\mathbf{X}$ e centralizar

em torno das médias das colunas e $\mathbf{X}_{-j}$ a matriz resultante após deletar a j -ésima coluna de $\mathbf{X}$ e centralizar em torno das médias das colunas. Através da decomposição por valores singulares, podemos escrever:

$$\mathbf{X}^{-i} = \bar{\mathbf{U}}\bar{\mathbf{D}}\bar{\mathbf{V}}' \text{ com } \bar{\mathbf{U}} = (\bar{u}_{st}), \bar{\mathbf{V}} = (\bar{v}_{st}) \text{ e } \bar{\mathbf{D}} = \text{diag}(\bar{d}_1, \dots, \bar{d}_p)$$

$\mathbf{X}_{-j} = \tilde{\mathbf{U}}\tilde{\mathbf{D}}\tilde{\mathbf{V}}' \text{ com } \tilde{\mathbf{U}} = (\tilde{u}_{st}), \tilde{\mathbf{V}} = (\tilde{v}_{st}) \text{ e } \tilde{\mathbf{D}} = \text{diag}(\tilde{d}_1, \dots, \tilde{d}_{p-1})$ O objetivo é estimar os elementos x_{ij} da matriz de interações dado por:

$$\hat{x}_{ij} = \sum_{t=1}^m u_{it} d_t v_{tj}$$

Para a estimação, como $\mathbf{U}$ requer a informação da i -ésima linha e $\mathbf{V}$ requer a informação da j -ésima coluna, utilizam-se $\tilde{\mathbf{U}}$ e $\bar{\mathbf{V}}$ respectivamente. Para a parte central $\mathbf{D}$, utiliza-se uma composição das duas matrizes $\tilde{\mathbf{D}}$ e $\bar{\mathbf{D}}$, sendo o valor predito dado por:

$$\hat{x}_{ij}^m = \sum_{t=1}^m (\tilde{u}_{it} \sqrt{\tilde{d}_t})(\bar{v}_{tj} \sqrt{\bar{d}_t}) \quad m = 0, 1, \dots, p$$

Assim, os elementos do lado direito da equação são obtidos da DVS de $\mathbf{X}$ centrada na média após omitir a i -ésima linha e a j -ésima coluna e o valor x_{ij} não é utilizado no cálculo da sua predição e é feito o máximo uso dos elementos da matriz $\mathbf{X}$ na predição de x_{ij} .

O procedimento de Gabriel (2002) toma uma mistura de regressão e aproximação de uma matriz de posto inferior como base para sua predição. O algoritmo é descrito a seguir:

Seja a matriz $\mathbf{X}$ de interação e considere a partição

$$\mathbf{X} = \begin{pmatrix} x_{11} & \mathbf{x}'_{1*} \\ \mathbf{x}_{*1} & \mathbf{X}_{|11} \end{pmatrix}.$$

A submatriz $\mathbf{X}_{|11}$ de posto m , após aplicar a DVS, pode ser escrita como:

$$\mathbf{X}_{|11} = \sum_{k=1}^m u_k d_k v_k' = \mathbf{U} \mathbf{D} \mathbf{V}'$$

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

A estimação do elemento x_{11} , utilizando o modelo de regressão, é dada por:

$$\hat{x}_{11} = \mathbf{x}_{1*}' \hat{\beta} \quad (17)$$

em que o parâmetro $\hat{\beta}$ a ser utilizado é o referente ao modelo de regressão relativo à submatriz $\mathbf{X}_{|11}$, e utilizando $\mathbf{x}_{*1}$ na estimativa desse parâmetro, tem-se $\hat{\beta} = (\mathbf{X}_{|11}' \mathbf{X}_{|11})^{-1} \mathbf{X}_{|11}' \mathbf{x}_{*1}$, e como $\mathbf{X}_{|11} = \mathbf{U} \mathbf{D} \mathbf{V}'$, segue que $\hat{\beta} = \mathbf{V} \mathbf{D}^{-1} \mathbf{U}' \mathbf{x}_{*1}$.

Então, prediz-se x_{11} por:

$$\hat{x}_{11} = \mathbf{x}_{*1}' \mathbf{V} \mathbf{D}^{-1} \mathbf{U}' \mathbf{x}_{*1}$$

Analogamente, o valor ajustado da validação cruzada é obtido para todos os outros elementos x_{ij} , $i = 1, \dots, g$; $j = 1, \dots, m$; $(i, j) \neq (1, 1)$, e a qualidade do ajuste é dada pelas estatísticas:

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

e

$$PRECORR(m) = Corr(x_{ij}, \hat{x}_{ij} | \forall i, j)$$

e cada valor ajustado requer uma diferente partição de $\mathbf{X}$.

Tanto para o procedimento de validação cruzada de Eastment e Krzanowski (1982) quanto para o procedimento de Gabriel (2002) para a escolha do número de termos multiplicativos m , pode-se basear em alguma função apropriada de $PRESS(m)$. Porém, as características das estatísticas diferem para os métodos. O método de Gabriel (2002) produz valores que primeiro decresce e então normalmente cresce com m , sendo, então, sugerido a escolha do valor de m que minimiza a função. O método de Eastment e Krzanowski (1982) produz, geralmente, um conjunto de valores que é monotonicamente não-crescente. Devido a isso é sugerido o uso da estatística dada por:

$$W_m = \frac{PRESS(m-1) - PRESS(m)}{D_m} \div \frac{PRESS(m)}{D_r}$$

em que D_m é o número de graus de liberdade requerido para ajustar o m -ésimo componente e é dado por:

$$D_m = n + p - 2m$$

e D_r é o número de graus de liberdade remanescentes após ajustar o m -ésimo componente e é dado por:

$$D_r = D_{r-1} - [g + e - (m-1)2], \quad r = 2, 3, \dots, (g-1) \text{ e } D_1 = (g-1)e$$

A estatística W_m representa o aumento na informação preditiva média suprida pelo m -ésimo componente, dividido pela informação preditiva média em cada um dos componentes. Então, os componentes importantes são aqueles cujo valor de W_m é maior que a unidade.

Para os modelos AMMI, busca-se o efeito da interação entre genótipo e ambiente. Então, a matriz $\mathbf{X}$ a ser trabalhada é a matriz de interações $\mathbf{GE}$. Mas como se busca apenas o número de termos multiplicativos do modelo e qualquer constante aditiva pode ser absorvida no componente ε_{ij} do modelo, pode-se aplicar os procedimentos diretamente na matriz $\mathbf{Y}$ de dados, fato preferível uma vez que os valores de muitos elementos da matriz de interações são pequenos.

3.6 Métodos de Correção de autovalores

Os três métodos apresentados por Muirhead (1987) para a correção dos autovalores, no presente caso para a correção dos autovalores provenientes da matriz $(\mathbf{GE})(\mathbf{GE})'$ ou $(\mathbf{GE})'(\mathbf{GE})$ em que $(\mathbf{GE})(\mathbf{GE})' \sim W_g(e, \Sigma)$, são dados a seguir:

i)O primeiro método de correção é dado por:

$$\phi_k^{(1)} = \frac{e\hat{\lambda}_k^2}{\alpha_k^{(1)}} = \frac{e\hat{\lambda}_k^2}{e - g + 1 + 2\hat{\lambda}_k^2 \sum_{k \neq k^*}^p \frac{1}{\hat{\lambda}_k^2 - \hat{\lambda}_{k^*}^2}} \quad (18)$$

ii) O segundo método de correção é dado por:

$$\phi_k^{(2)} = \frac{e\hat{\lambda}_k^2}{\alpha_k^{(2)}} = \frac{e\hat{\lambda}_k^2}{e - g - 1 + 2\hat{\lambda}_k^2 \sum_{k \neq k^*}^p \frac{1}{\hat{\lambda}_k^2 - \hat{\lambda}_{k^*}^2}} \quad (19)$$

iii) O terceiro método de correção é dado por:

$$\phi_k^{(3)} = \frac{e\hat{\lambda}_k^2}{e + g + 1 - 2k} - \frac{ec\hat{\lambda}_k^2 \ln(e\hat{\lambda}_k^2)}{b + \sum_{k=1}^p [\ln(e\hat{\lambda}_k^2)]^2} = \frac{e\hat{\lambda}_k^2(b + \sum_{k=1}^p [\ln(e\hat{\lambda}_k^2)]^2 - (e + g + 1 - 2k)c \ln(e\hat{\lambda}_k^2))}{(e + g + 1 - 2k)(b + \sum_{k=1}^p [\ln(e\hat{\lambda}_k^2)]^2)} = \frac{\alpha_k^{(3)}}{d_k} \quad (20)$$

em que:

$$b = \frac{5,8(g-2)^2}{(e+g-1)^2},$$

$$c = \frac{6(g-2)}{5(e+g-1)^2},$$

$$\alpha_k^3 = e\hat{\lambda}_k^2(b + \sum_{k=1}^p [\ln(e\hat{\lambda}_k^2)]^2 - (e + g + 1 - 2k)c \ln(e\hat{\lambda}_k^2)) \text{ e}$$

$$d_k = (e + g + 1 - 2k)(b + \sum_{k=1}^p [\ln(e\hat{\lambda}_k^2)]^2).$$

Nos três métodos de correção, temos que:

g : é o número de genótipos do experimento;

e : é o número de ambientes do experimento;

$\hat{\lambda}_k^2$: é o k -ésimo autovalor da matriz $(\mathbf{GE})(\mathbf{GE})'$ ou $(\mathbf{GE})'(\mathbf{GE})$, com $k = 1, 2, \dots, p$, sendo

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

Porém, nem sempre os autovalores obtidos com os métodos de correção apresentados estão em ordem decrescente ($\phi_1 \geq \phi_2 \geq \dots \geq \phi_p$) e para reordená-los será utilizada a regressão isotônica.

3.7 Procedimento de Stein para a Isotonização

O algoritmo para ordenar os autovalores, obtidos pelos três métodos de correção, é apresentado por Lin e Perlman (1985) e é descrito como segue:

Listam-se os produtos $e\hat{\lambda}_k^2$ (numerador dos métodos de correção (18) e (19) ou d_k (denominador do método de correção (20)) em uma coluna e em outra coluna lista-se o denominador do método de correção (18), $\alpha_k^{(1)}$, ou o denominador do método de correção (19), $\alpha_k^{(2)}$, ou ainda o numerador do método de correção (20), $\alpha_k^{(3)}$, como segue:

$$\begin{array}{ccc} e\hat{\lambda}_1^2 \text{ ou } d_1 & \alpha_1^{(1)} \text{ ou } \alpha_1^{(2)} \text{ ou } \alpha_1^{(3)} \\ e\hat{\lambda}_2^2 \text{ ou } d_2 & \alpha_2^{(1)} \text{ ou } \alpha_2^{(2)} \text{ ou } \alpha_2^{(3)} \\ \vdots & \vdots \\ e\hat{\lambda}_p^2 \text{ ou } d_p & \alpha_p^{(1)} \text{ ou } \alpha_p^{(2)} \text{ ou } \alpha_p^{(3)} \end{array}$$

Passo 1) Fazendo todos α_k 's positivos:

a) Inicia-se pelo final da lista e procura-se para cima até que se encontre o primeiro par $(e\hat{\lambda}_k^2, \alpha_k)$ com α_k negativo.

b) Soma-se esse par com o par imediatamente acima dele, substituindo-os pelo par $(e\hat{\lambda}_k^2 + e\hat{\lambda}_{k-1}^2, \alpha_k + \alpha_{k-1})$, para que na lista um par seja menor do que o próximo.

c) Repetem-se (a) e (b), para a nova lista até que todos α_k sejam positivos.

Passo 2) Re-ordenando as razões $\frac{e\hat{\lambda}_k^2}{\alpha_k}$ de forma que estejam em ordem decrescente:

a) Listam-se as razões $\frac{e\hat{\lambda}_k^2}{\alpha_k}$ à direita de cada par $(e\hat{\lambda}_k^2, \alpha_k)$ obtidos no passo 1. Um par $(e\hat{\lambda}_k^2, \alpha_k)$ (exceto o par no final da lista) é chamado de par violado se a razão $\frac{e\hat{\lambda}_k^2}{\alpha_k}$ não é maior do que a razão $\frac{e\hat{\lambda}_{k+1}^2}{\alpha_{k+1}}$.

a) Inicia-se pelo final da lista encontrada no passo 1 e procede-se para cima até o primeiro par violado ser encontrado.

b) Soma-se esse par violado com o par imediatamente acima dele, substituem-se esses dois pares e suas razões pelo par $(e\hat{\lambda}_k^2 + e\hat{\lambda}_{k+1}^2, \alpha_k + \alpha_{k+1})$ e sua razão $\frac{e\hat{\lambda}_k^2 + e\hat{\lambda}_{k+1}^2}{\alpha_k + \alpha_{k+1}}$, formando

uma nova lista.

c) Reinicia-se no par imediatamente após o par trocado em (b) e procede-se para cima até o próximo par violado ser encontrado e repete-se (b).

d) Repete-se (c) até todas as razões $\frac{e\hat{\lambda}_k^2}{\alpha_k}$ estarem em ordem decrescente.

Passo 3) Cada razão no final da lista foi obtida por bloco acumulado de um ou mais pares consecutivos $(e\hat{\lambda}_k^2, \alpha_k)$ na lista original.

Observação: No passo 2 pode ser mais conveniente agrupar qualquer bloco em dois ou mais pares sucessivos na lista cujas razões correspondentes não estão em ordem decrescente.

3.7.1 Número de blocos

Na maioria dos estudos, existe um grande interesse na comparação do Erro Médio Quadrático do modelo ($QM_{EM}(model)$) selecionado com o Erro Médio Quadrático (QM_{EM}) do experimento. Para Nachit et al., (1992) uma aproximação para o número de blocos (ou repetições) que falta para o modelo AMMI completo apresentar uma performance igual ao modelo AMMI selecionado, ou seja, indica o número de blocos que se ganha ao analisar os dados com o modelo selecionado. Essa medida é obtida por:

$$R_{AMMI} = \frac{QM_{EM}}{QM_{EM}(model)}. \quad (21)$$

Para fazer uma estimativa do $QM_{EM}(model)$, Piepho (1994) sugere a seguinte expressão:

$$\hat{Q}M_{EM}(model) = (RMSPD)^2 - QM_{EM} \quad (22)$$

A medida RMSPD é dada por Cornélius; Crossa e Seyedsadr (1993) podendo, assim, fazer uma comparação aproximada com outros RMSPDs provenientes de modelos ajus-

tados por validação cruzada. O ajuste é feito por:

$$RMSPD_{(PRESS)} = \sqrt{\frac{PRESS}{ge} + \frac{(r-1)}{r}QM_{EM}} \quad (23)$$

sendo r o número de blocos em cada ambiente do experimento.

A medida PRESS é dada por:

$$PRESS = \sum_{i=1}^g \sum_{j=1}^e (y_{ij} - \hat{y}_{ij})^2, \quad (24)$$

em que:

y_{ij} : é a resposta média do i -ésimo genótipo no j -ésimo ambiente.

$\hat{y}_{ij}$: é a média predita pelo modelo selecionado para o i -ésimo genótipo no j -ésimo ambiente.

As análises e a rotina computacional (ver anexo) foram implementadas através do sistema de análise estatístico SAS (SAS Institute Inc., 2003).

4 RESULTADOS E DISCUSSÃO

4.1 Experimento 1

4.1.1 Análise de variância conjunta

A Tabela 4 mostra a análise de variância conjunta do experimento 1. Como podemos observar, há efeito significativo entre blocos, genótipos, ambientes e também da interação entre genótipos e ambientes.

Tabela 4 - Análise de variância conjunta do Experimento 1 com 20 genótipos avaliados em 34 ambientes com 4 blocos

Fonte de Variação	GL	SQ	QM	F	p
Blocos d. ambiente (B d. E)	102	257862519	2528064	7.14	<0.0001
Genótipo (G)	19	89066441	4687707	13.23	<0.0001
Ambiente (E)	33	4333925428	131331074	370.67	<0.0001
Interação ($G \times E$)	627	594108485	947541	2,67	<0,0001
Resíduo	1938	686646195	354307		
Total	2719	5961609068			

A análise AMMI focaliza-se no estudo dessas interações, que neste experimento corresponde a 9,97% da soma de quadrados total. Obtendo a matriz de interações, sabemos que a soma de quadrados dos autovalores dessa matriz, obtida através da DVS, nos dá exatamente a soma de quadrados da interação. Porém, busca-se nessa análise descartar um ruído, deixando no modelo somente os componentes necessários para a explicação dessa interação. A seguir há duas maneiras de escolher quantos termos reter no modelo: validação cruzada e teste de hipóteses.

4.1.2 Teste de hipóteses

A Tabela 5 apresenta o teste de hipóteses para a seleção do número de termos a serem retidos no modelo, com os graus de liberdade ajustados pelo método de Gollob (1968), nos informando que apenas um componente deve ser selecionado ao nível 1% de significância.

Tabela 5 - Decomposição de soma de quadrados de interação do Experimento 1

	GL	λ_k	SQ (λ_k^2)	QM	F	valor- p
1	51	6690,6486	44764779	877740,76	2,477348	0
2	49	4780,2473	22850765	466342,14	1,3162107	0,0711835
3	47	4443,4757	19744476	420095,24	1,1856828	0,1830846
4	45	3673,4094	13493936	299865,25	0,846344	0,7558632
5	43	2925,1162	8556304,5	198983,83	0,5616148	0,9905949
6	41	2627,8094	6905382,4	168423,96	0,4753622	0,9981421
7	39	2452,693	6015703	154248,79	0,435354	0,9991237
8	37	2409,2595	5804531,5	156879,23	0,4427782	0,9985986
9	35	1970,4081	3882508,1	110928,8	0,313087	0,9999644
10	33	1884,2702	3550474,4	107590,13	0,3036639	0,9999593
11	31	1793,5572	3216847,4	103769,27	0,2928799	0,9999544
12	29	1584,9923	2512200,4	86627,602	0,244499	0,999989
13	27	1379,5053	1903034,8	70482,77	0,1989316	0,9999976
14	25	1192,8851	1422974,9	56918,998	0,160649	0,9999994
15	23	1075,5556	1156819,9	50296,519	0,1419576	0,9999995
16	21	1034,7376	1070682	50984,857	0,1439004	0,9999983
17	19	781,18018	610242,47	32118,025	0,0906504	0,9999999
18	17	759,64643	577062,69	33944,864	0,0958065	0,9999993
19	15	698,85371	488396,51	32559,767	0,0918972	0,9999976

4.1.3 Validação cruzada

A Tabela 6 apresenta os resultados para o experimento 1 quando se aplica os procedimentos de Eastment e krzanowski (1982) e Gabriel (2002). Observamos que a estatística PRESS difere para os dois métodos. Para o método de krzanowski conforme aumenta o número de componentes, a estatística PRESS normalmente diminui. Para o método de Gabriel os valores da estatística oscilam conforme aumenta o número de componentes. Devido a essa diferença, para o método de Krzanowski utiliza-se a estatística W, cujos componentes a serem selecionados, são aqueles em que W é maior que um. Através dessa medida, observamos que o número de componentes a ser retido para esse experimento é 3. Para o método de Gabriel, seleciona-se o número de termos em que a estatística PRESS minimiza a função, portanto, para esse experimento são selecionados 5 componentes. A estatística PRECORR para os dois métodos apresenta valores próximos conforme aumenta o número de componentes.

Tabela 6 - Validação cruzada para o experimento 1 sem correção dos autovalores

m	Eastment e Krzanowski			Gabriel	
	PRESS	W	PRECORR	PRESS	PRECORR
1	193891,27	1,4403725	0,9461365	195489,97	0,9455552
2	192323,63	0,088031	0,9470625	197841,81	0,9449925
3	170883,16	1,2808254	0,9533873	171088,72	0,9525439
4	158021,88	0,7820443	0,957108	157381,94	0,9564339
5	158224,13	-0,011504	0,9571613	156528,15	0,9566904
6	158383,93	-0,008456	0,9572792	160492,88	0,9555849
7	157092,15	0,063729	0,9575781	176520,33	0,9511023
8	143497,24	0,6731519	0,9612579	161816,61	0,955242
9	143404,52	0,0041667	0,9613419	167634,14	0,9536296
10	140609,63	0,1145844	0,9623121	183000,60	0,949402
11	135448,82	0,1928891	0,9638044	193150,88	0,9466912
12	128569,66	0,2318564	0,965704	195596,89	0,9462998
13	123189,58	0,1559758	0,9672147	192330,11	0,947158
14	21481,34	0,0389401	0,9676798	196925,85	0,9461065
15	120074,97	0,0224488	0,9680899	231282,34	0,9369803
16	116127,24	0,0339949	0,9692	209015,06	0,9431825
17	115898,12	0	0,9692775	222919,78	0,939655
18	113712,99	-0,021351	0,9698622	313613,36	0,916818

4.1.3.1 Validação cruzada com correção de autovalores

Os métodos 1 e 2 apresentados por Muirhead (1987) não se mostraram eficientes nos dois procedimentos de validação cruzada, nos dando autovalores negativos em que o processo de regressão isotônica, através do algoritmo de isotonização de Stein, não conseguiu torná-lo positivo para posteriormente reordená-lo na ordem decrescente. Portanto, não foi possível utilizar esses dois métodos.

O método 3, apresentado por Muirhead (1987), foi o único que corrigiu com sucesso todos os autovalores utilizados para estimar cada valor da matriz de interações, tanto no método de validação cruzada dado por Eastment e Krzanowski (1982) quanto naquele dado por Gabriel (2002).

Os resultados dos dois procedimentos de validação cruzada com o método 3 de correção de autovalores são apresentados na Tabela 7.

Tabela 7 - Validação cruzada para o experimento 1 com correção dos autovalores pelo método 3

m	Eastment e Krzanowski			Gabriel	
	PRESS	W	PRECORR	PRESS	PRECORR
1	189228,61	1,7563843	0,9473221	207446,55	0,9424324
2	181414,08	0,4652171	0,9496187	216701,38	0,9403245
3	160751,13	1,3121797	0,9555057	192694,47	0,947129
4	147271,67	0,879463	0,9593087	183652,36	0,9499307
5	145470,82	0,1114149	0,9598278	185235,45	0,9496245
6	143470,08	0,1168754	0,960414	191413,78	0,9479854
7	142202,69	0,0690724	0,9607533	213263,53	0,9421302
8	129800,16	0,6789145	0,9642086	198737,74	0,9461683
9	129116,69	0,0341135	0,9644099	205425,77	0,9444229
10	126191,09	0,133648	0,9652759	223115,65	0,9398306
11	122076,47	0,1706329	0,9664543	236129,70	0,9365678
12	115438,74	0,2491671	0,9683118	240506,13	0,9359936
13	110533,06	0,1585069	0,9696963	235487,35	0,937208
14	109219,71	0,0332996	0,9700641	238557,41	0,9366675
15	108174,97	0,0185111	0,9703717	264484,29	0,9301109
16	103750,95	0,0426407	0,9716225	243167,35	0,9357489
17	104939,12	0	0,9713151	250590,17	0,9339717
18	102336,08	-0,028262	0,9720361	299998,59	0,9218055

Observamos que os autovalores corrigidos não alteraram o número de componentes a serem retidos no modelo quando se aplica o método de Eastment e Krzanowski. Já para o método de Gabriel o número de componentes passou a ser 4, diminuindo um componente se comparado ao mesmo método sem a correção de autovalores.

4.1.4 Número de Blocos

A Tabela 8 indica a medida R_{AMMI} tanto para a validação cruzada de Eastment e Krzanowski (1982) quanto para a validação cruzada de Gabriel (2002), quando os autovalores são corrigidos pelo método 3 de correção. Essa medida representa o número de blocos que se ganha ao analisar os dados com o modelo AMMI selecionado em vez do modelo AMMI completo. De acordo com essa medida, ao selecionar o modelo AMMI pelo método de Eastment e Krzanowski (1982) tem-se um ganho de aproximadamente 5 blocos. Para o método de Gabriel (2002), o ganho é de aproximadamente 4 blocos.

Tabela 8 - R_{AMMI} para o melhor modelo selecionado referente aos métodos de validação cruzada de Eastment e Krzanowski e de Gabriel com correção de autovalores no experimento 1

Método de validação	R_{AMMI}
Eastment e Krzanowski	4,9090291
Gabriel	3,7265732

4.2 Experimento 2

4.2.1 Análise de variância conjunta

A análise de variância conjunta do experimento 2 é apresentada na Tabela 9, informando o efeito significativo entre blocos, genótipos, ambientes e também da interação entre genótipos e ambientes.

Tabela 9 - Análise de variância conjunta do Experimento 2 com 16 genótipos avaliados em 24 ambientes com 4 blocos

Fonte de Variação	GL	SQ	QM	F	p
Blocos d. ambiente (B d. E)	72	3065470183	42575975	30,56	<0,0001
Genótipo (G)	15	304645583	20309706	14,58	<0,0001
Ambiente (E)	23	8355949394	363302148	260,81	<0,0001
Interação ($G \times E$)	345	1797966686	5211498	3,74	<0,0001
Resíduo	1080	1504417193	1392979		
Total	1535	15028449039			

4.2.2 Teste de hipóteses

O teste de hipóteses para a seleção do número de termos a serem retidos no modelo, com os graus de liberdade ajustados pelo método de Gollob (1968), para o experimento 2, é apresentado pela Tabela 10 nos informando que apenas um componente deve ser selecionado ao nível 1% de significância. Os autovalores λ_k apresentados também na Tabela 10, nos mostram a importância do primeiro componente, cujo autovalor é aproximadamente o dobro do segundo. A partir do terceiro componente, os autovalores estão mais próximos. A escolha de um componente para explicar a interação entre genótipos e ambientes para o ex-

perimento 2, está de acordo com os procedimentos de validação cruzada, tanto com o método dado por Eastment e Krzanowski (1982) quanto com o de Gabriel (2002) (quando esse utiliza o método de correção de autovalores), resultados esses que serão descritos a seguir.

Tabela 10 - Decomposição de soma de quadrados de interação do Experimento 2

	GL	λ_k	SQ (λ_k^2)	QM	F	valor- p
1	37	14493,2340	210053838	5677130,7	4,0755325	0
2	35	7376,4057	54411361	1554610,3	1,1160329	0,2963334
3	33	6250,8430	39073038	1184031,5	0,8499996	0,7103382
4	31	6003,0722	36036876	1162479,9	0,834528	0,7257644
5	29	5279,8536	27876854	961270,82	0,6900828	0,8905443
6	27	4423,4703	19567090	724707,03	0,520257	0,9800824
7	25	4063,8134	16514580	660583,19	0,4742234	0,9872297
8	23	3812,6918	14536619	632026,89	0,4537232	0,987845
9	21	3010,4929	9063067,4	431574,64	0,3098214	0,9988795
10	19	2815,4758	7926903,7	417205,46	0,2995059	0,9985119
11	17	2454,9603	6026830,1	354519,42	0,2545045	0,9990975
12	15	1948,2317	3795606,6	253040,44	0,1816542	0,9997732
13	13	1501,4923	2254479,1	173421,47	0,1244968	0,999931
14	11	1240,2846	1538305,9	139845,99	0,1003935	0,9999152
15	9	903,45138	816224,39	90691,599	0,0651062	0,9999392

4.2.3 Validação cruzada

A Tabela 11 apresenta os resultados para o experimento 2 quando se aplicam os procedimentos de Eastment e Krzanowski (1982) e Gabriel (2002) sem a correção dos autovalores. O método de Eastment e Krzanowski, através da estatística W, selecionou 1 componente para reter no modelo. Já o método de Gabriel, selecionou, através do valor dado pela estatística PRESS que minimiza a função, 5 componentes.

4.2.3.1 Validação cruzada com correção de autovalores

A Tabela 12 apresenta os resultados dos dois procedimentos de validação cruzada com o método 3 de correção de autovalores para o experimento 2.

O número de componentes a serem retidos no modelo quando se aplica o método

Tabela 11 - Validação cruzada para o experimento 2 sem correção dos autovalores

Eastment e Krzanowski				Gabriel	
m	PRESS	W	PRECORR	PRESS	PRECORR
1	788212,63	4,0848032	0,9403425	783143,67	0,9407411
2	813285,71	-0,241497	0,9385396	768854,37	0,9418528
3	839419,59	-0,225258	0,9366678	808761,62	0,9387789
4	795377,87	0,36684	0,9401871	804176,57	0,9391309
5	753792,65	0,3310079	0,9435284	765809,45	0,9421317
6	725120,46	0,2118283	0,9459519	777029,09	0,9412952
7	713449,99	0,0767558	0,9469445	864670,30	0,934798
8	669489,21	0,2626527	0,9501681	798830,64	0,9396624
9	659730,15	0,0484118	0,9509465	843449,48	0,9362608
10	637839,69	0,0857992	0,9526681	919793,21	0,931159
11	601019,7	0,1021042	0,9554972	898628,53	0,9334335
12	580445,99	0,0265835	0,9570094	859333,12	0,9363514
13	569702,93	-0,005388	0,9578346	992804,37	0,9270513
14	557784,53	-0,032051	0,958729	1162129,60	0,9154785

Tabela 12 - Validação cruzada para o experimento 2 com correção dos autovalores pelo método 3

Eastment e Krzanowski				Gabriel	
m	PRESS	W	PRECORR	PRESS	PRECORR
1	822356,14	3,5655709	0,9377543	905202,77	0,9323542
2	806278,24	0,1562036	0,9389633	934868,80	0,9309064
3	803318,04	0,0266618	0,9391764	1016292,50	0,9254147
4	753654,75	0,436565	0,9430571	1050637,70	0,9230536
5	703171,29	0,4307638	0,9469899	1026347,00	0,9253125
6	673991,9	0,2319288	0,9492286	1047447,40	0,9239643
7	661216,12	0,0906631	0,9502184	1175836,30	0,9157501
8	622364,19	0,2497054	0,9532242	1103331,00	0,9203268
9	613682,27	0,0463001	0,953888	1150784,00	0,9170835
10	592840,16	0,087891	0,9554864	1224703,60	0,9134556
11	553351,59	0,1189375	0,9585121	1190886,50	0,9168514
12	527477,73	0,036789	0,9604967	1136988,10	0,9205391
13	517749,05	-0,005369	0,9612389	1206819,00	0,9159812
14	499517,68	-0,054747	0,9626291	1252478,10	0,9127872

de Eastment e Krzanowski, com os autovalores corrigidos, não alterou, selecionando também 1 componente. Já para o método de Gabriel, o número de componentes passou a ser 1, diminuindo 4 componentes quando se compara o método sem a correção dos autovalores.

4.2.4 Número de Blocos

De acordo com a Tabela 13, através da medida R_{AMMI} , o número de blocos que se ganha ao selecionar o modelo AMMI tanto pelo método de Eastment e Krzanowski (1982) quanto para o método de Gabriel (2002), quando os autovalores são corrigidos, é de aproximadamente 3.

Tabela 13 - R_{AMMI} para o melhor modelo selecionado referente aos métodos de validação cruzada de Eastment e Krzanowski e de Gabriel com correção de autovalores no experimento 2

Método de validação	R_{AMMI}
Eastment e Krzanowski	2,9380834
Gabriel	2,5010481

4.3 Experimento 3

4.3.1 Análise de variância conjunta

A análise de variância conjunta do experimento 3 é apresentada na Tabela 14, informando o efeito significativo entre blocos, genótipos, ambientes e também da interação entre genótipos e ambientes.

Tabela 14 - Análise de variância conjunta do Experimento 3 com 9 genótipos avaliados em 20 ambientes com 4 blocos

Fonte de Variação	GL	SQ	QM	F	p
Blocos d. ambiente (B d. E)	60	118813053,8	1980217,6	3,28	<0,0001
Genótipo (G)	8	79828574,7	9978571,8	16,55	<0,0001
Ambiente (E)	19	989593771,8	52083882,7	86,40	<0,0001
Interação ($G \times E$)	152	249704161,7	1642790,5	2,73	<0,0001
Resíduo	480	289366499	602847		
Total	719	1727306061			

Tabela 15 - Decomposição de soma de quadrados de interação do Experimento 3

	GL	λ_k	SQ (λ_k^2)	QM	F	valor- p
1	26	5922,9093	35080855	1349263,6	2,2381532	0,0005193
2	24	3070,4899	9427908,4	392829,52	0,651624	0,8972384
3	22	2552,3799	6514643,1	296120,14	0,4912029	0,9760374
4	20	2320,1904	5383283,3	269164,17	0,4464884	0,9827592
5	18	1758,6717	3092926	171829,22	0,2850296	0,998536
6	16	1313,4150	1725058,9	107816,18	0,1788451	0,9998667
7	14	792,63341	628267,72	44876,265	0,0744406	0,9999986
8	12	757,03266	573098,44	47758,204	0,0792211	0,9999887

4.3.2 Teste de hipóteses

A Tabela 15 apresenta o teste de hipóteses para a seleção do número de termos a serem retidos no modelo, com os graus de liberdade ajustados pelo método de Gollob (1968) para o experimento 3, nos informando que apenas um componente deve ser selecionado ao nível 1% de significância. Observamos também na Tabela 15 os autovalores λ_k , sendo o primeiro autovalor (5922,9093) praticamente o dobro do segundo (3070,4899). A partir do segundo autovalor, eles se mostram decrescendo, porém seus valores estão mais próximos. Portanto, o primeiro componente é muito importante para explicar a interação entre genótipos e ambientes nesse experimento, o que está de acordo com os resultados quando se aplica o método de validação cruzada de Eastment e Krzanowski (1982), mas não com o método de Gabriel (2002), resultados esses que serão apresentados a seguir.

4.3.3 Validação cruzada

Os resultados de validação cruzada para o experimento 3 quando se aplicam os procedimentos de Eastment e Krzanowski (1982) e Gabriel (2002) sem a correção dos autovalores são apresentados na Tabela 16. O método de Eastment e Krzanowski, através da estatística W, selecionou 1 componente para reter no modelo. O método de Gabriel selecionou, através da estatística PRESS, 6 componentes.

Tabela 16 - Validação cruzada para o experimento 3 sem correção dos autovalores

m	Eastment e Krzanowski			Gabriel	
	PRESS	W	PRECORR	PRESS	PRECORR
1	248410,78	1,921917	0,9299879	244121,61	0,9310382
2	239768,44	0,1499454	0,9330417	245565,31	0,9307907
3	239563,82	0,0029337	0,9337207	296505,67	0,9164247
4	204237,32	0,4612477	0,9436155	305126,30	0,9148983
5	172861,54	0,3343572	0,9524079	285641,16	0,9218457
6	157752,49	0,090143	0,9567743	133751,29	0,9631414
7	155642,82	-0,000904	0,9573241	187249,25	0,9488406

4.3.3.1 Validação cruzada para o experimento 3 com correção de autovalores

Os resultados dos dois procedimentos de validação cruzada com o método 3 de correção de autovalores para o experimento 3, são apresentados na Tabela 17.

Tabela 17 - Validação cruzada para o experimento 3 com correção dos autovalores pelo método 3

m	Eastment e Krzanowski			Gabriel	
	PRESS	W	PRECORR	PRESS	PRECORR
1	245099,3	2,0134358	0,9307178	270442,78	0,9244892
2	227584,44	0,3201529	0,9358367	275004,70	0,9240014
3	224137,86	0,0528169	0,936927	352260,12	0,9036434
4	196508,65	0,3749345	0,9448762	398047,68	0,8941553
5	167930,19	0,3134906	0,9530786	372617,45	0,9035
6	153156,07	0,0907901	0,9573066	182294,39	0,9516222
7	152397,53	-0,000332	0,9575213	220879,12	0,9418806

O número de componentes a serem retidos no modelo quando se aplica tanto o método de Eastment e Krzanowski quanto o método de Gabriel não alterou quando se compararam os respectivos métodos sem a correção de autovalores.

4.3.4 Número de Blocos

O número de blocos que se ganha ao selecionar o modelo AMMI pelo método de Eastment e Krzanowski (1982) com correção de autovalores, em vez de selecionar o modelo

AMMI completo, é de aproximadamente 7. Já para o método de Gabriel (2002) com correção de autovalores, o ganho é de aproximadamente 20 blocos, como mostra a Tabela 18.

Tabela 18 - R_{AMMI} para o melhor modelo selecionado referente aos métodos de validação cruzada de Eastment e Krzanowski e de Gabriel com correção de autovalores no experimento 3

Método de validação	R_{AMMI}
Eastment e Krzanowski	6,3869298
Gabriel	19,087902

4.4 Experimento 4

4.4.1 Análise de variância conjunta

A análise de variância conjunta do experimento 4 é apresentada na Tabela 19, informando o efeito significativo entre blocos, genótipos, ambientes e também da interação entre genótipos e ambientes. Observe que a soma de quadrados da interação entre genótipos e ambientes corresponde a 5,28% da soma de quadrados total.

Tabela 19 - Análise de variância conjunta do Experimento 4 com 8 genótipos avaliados em 59 ambientes com 4 blocos

Fonte de Variação	GL	SQ	QM	F	p
Blocos d. ambiente (B d. E)	177	301817526	1705184	3,86	<0,0001
Genótipo (G)	7	82527757	11789680	26,66	<0,0001
Ambiente (E)	58	3787154892	65295774	147,67	<0,0001
Interação ($G \times E$)	406	263183718	648236	1,47	<0,0001
Resíduo	1239	547865274	602847		
Total	1887	4982549167			

4.4.2 Teste de hipóteses

O teste de hipóteses para a seleção do número de termos a serem retidos no modelo, com os graus de liberdade ajustados pelo método de Gollob (1968), para o experimento 4 é apresentado pela Tabela 20 nos informando que nenhum componente deve ser selecionado no

modelo, ou seja, não há efeito de interação entre genótipos e ambiente segundo esse critério de seleção. Podemos observar também na Tabela 20 que os autovalores λ_k estão bem próximos, o que poderia justificar a não ocorrência de efeito de interação neste experimento.

Tabela 20 - Decomposição de soma de quadrados de interação do Experimento 4

	GL	λ_k	SQ (λ_k^2)	QM	F	valor- p
1	64	4047,5314	16382511	255976,73	0,5788926	0,9967999
2	62	3465,8289	12011970	193741,45	0,4381472	0,999956
3	60	3300,1349	10890890	181514,84	0,4104967	0,9999818
4	58	3073,7283	9447805,9	162893,21	0,3683838	0,9999965
5	56	2442,0599	5963656,5	106493,87	0,2408364	1
6	54	2386,4418	5695104,4	105464,9	0,2385094	1
7	52	2324,6488	5403992,2	103922,93	0,2350222	1

4.4.3 Validação cruzada

A Tabela 18 apresenta os resultados para o experimento 4 quando se aplicam os procedimentos de Eastment e Krzanowski (1982) e Gabriel (2002) sem a correção dos autovalores. O método de Eastment e Krzanowski, através da estatística W , nos dá a informação de que nenhum termo deve ser selecionado, ou seja, para esse método só há efeito aditivo no modelo. Já o método de Gabriel, selecionou, através do valor dado pela estatística PRESS que minimiza a função, 1 componente.

Tabela 21 - Validação cruzada para o experimento 4 sem correção dos autovalores

m	Eastment e Krzanowski			Gabriel	
	PRESS	W	PRECORR	PRESS	PRECORR
1	143383,88	-0,147969	0,9667036	144160,74	0,9665111
2	149236,66	-0,174925	0,9654056	145364,10	0,9662345
3	146569,12	0,0650422	0,9661022	150514,80	0,9650403
4	136822,68	0,1895552	0,9685303	145154,12	0,9663136
5	135967,9	0,0108086	0,968724	147786,43	0,9656997
6	131599,86	0,024743	0,9698374	164380,91	0,961814

4.4.3.1 Validação cruzada com correção de autovalores

A Tabela 22 apresenta os resultados dos dois procedimentos de validação cruzada com o método 3 de correção de autovalores para o experimento 4.

Tabela 22 - Validação cruzada para o experimento 4 com correção dos autovalores pelo método 3

m	Eastment e Krzanowski			Gabriel	
	PRESS	W	PRECORR	PRESS	PRECORR
1	139572,9	-0,006664	0,9675976	145309,78	0,9662395
2	143182,3	-0,112437	0,9667961	146828,45	0,9658909
3	139400,47	0,0969537	0,9677277	152664,36	0,9645398
4	129226,55	0,2095001	0,9702123	147384,31	0,9657961
5	129068,2	0,0021093	0,9702495	149965,59	0,9651949
6	125353,64	0,0220898	0,9712052	166514,28	0,9613226

O número de componentes a serem retidos no modelo quando se aplica tanto o método de Eastment e Krzanowski quanto o método de Gabriel não alterou quando se compararam os respectivos métodos sem a correção de autovalores.

4.4.4 Número de Blocos

Na Tabela 23 é apresentada a medida R_{AMMI} e nota-se que o número de blocos que se ganha escolhendo o modelo AMMI pelo método de validação cruzada de Eastment e Krzanowski (1982) com correção de autovalores é de aproximadamente 16. Selecionando o modelo AMMI pelo método de Gabriel (2002) com correção de autovalores o ganho é de aproximadamente 13 blocos.

Tabela 23 - R_{AMMI} para o melhor modelo selecionado referente aos métodos de validação cruzada de Eastment e Krzanowski e de Gabriel com correção de autovalores no experimento 4

Método de validação	R_{AMMI}
Eastment e Krzanowski	15,325761
Gabriel	12,719609

5 Considerações Finais

A análise do experimento 2 permitiu a verificação de uma grande eficiência quando se aplica a correção dos autovalores pelo método 3 na validação cruzada dada por Gabriel (2002), pois 4 componentes foram diminuídos com essa correção. Neste caso, esse método aperfeiçoado se igualou ao método de Eastment e Krzanowski (1982) para a escolha do número de termos a ser retido no modelo. Nesse experimento podemos observar a grande importância do primeiro componente para explicar a interação entre genótipos e ambientes, pois o primeiro autovalor é muito alto se comparado aos demais.

A análise do experimento 4 é bem interessante porque embora o efeito da interação entre genótipos e ambientes na análise de variância conjunta seja significativo, observamos que a porcentagem da soma de quadrados dessa interação se comparada a soma de quadrados total é bem baixa (5,28%). O método de Eastment e Krzanowski (1982) para esse experimento nos dá a informação de que nenhum componente é importante, o que coincide com o resultado do teste F com os graus de liberdade ajustados pelo método de Gollob (1968). Podemos observar que com a análise AMMI realmente consegue-se um estudo detalhado das interações e detectar também quando não há efeito significativo dessas interações.

5.1 Conclusões

Os resultados baseados nos 4 experimentos, quando se aplicam os métodos de validação cruzada e quando se aperfeiçoam esses métodos, permitem relatar que o método de Eastment e Krzanowski (1982) seleciona modelos mais parcimoniosos. Quando esse método é aperfeiçoado através da correção de autovalores pelo método 3, e a regressão isotônica é utilizada para reordenar os autovalores corrigidos, observa-se que o número de componentes a ser retido no modelo não é alterado. Nota-se também que, na maioria dos experimentos, o resultado desse método coincidiu com o resultado obtido pelo teste F quando se decompõe a soma de quadrados e obtém-se os graus de liberdade ajustados pelo método de Gollob (1968). Já o método de Gabriel (2002) selecionou um maior número de termos para reter no modelo. Porém, quando se utiliza a correção de autovalores pelo método 3, e a regressão

isotônica para reordená-los, observa-se que na maioria dos experimentos o número de termos diminui. Essa seleção de termos multiplicativos é de grande importância, pois na prática significa um ganho do número de blocos (ou repetições) quando se seleciona o modelo AMMI em vez do modelo completo. Portanto, o aperfeiçoamento desses métodos através da correção de autovalores trouxe um grande benefício do ponto de vista prático para o analista de dados provenientes de ensaios multi-ambientais.

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.
- BRASIL, E.M. **Comparação de métodos no estudo da interação genótipo x ambiente em milho** (*Zea mays L.*). 1990. 181 p. Dissertação (Mestrado em Agronomia) - Universidade Federal de Goiás, Goiânia, 1990.
- CORNELIUS, P. L. Statistical tests and retention of terms in the additive main effects and multiplicative interaction model for cultivar trials. **Crop Science**, Madison, v. 33, p. 1186-1193, 1993.
- CORNELIUS, P.L; CROSSA, J.; SEYEDSADR, M.S. Teste and estimators of multiplicative models for variety trials. In: ANNUAL KANSAS STATE UNIVERSITY CONFERENCE ON APPLIED STATISTICS IN AGRICULTURE, 5., 1993, Manhattan. **Proceedings ...**, Manhattan: KS, 1993. p. 156-166.
- CORNELIUS, P.L.; CROSSA, J.; SEYEDSADR, M.S. Statistical tests and estimators of multiplicative models for genotype-by-environment interaction. In: KANG, M.S.; GAUCH, H.G. (Ed.) **Genotype-by-environment interaction**. Boca Raton: CRC Press, 1996. p. 199-234.
- CORSTEN, L.C.A.; EIJSBERGEN, C.A. van. Multiplicative effects in two-way analysis of variance. **Statistica Neerlandica**, Assen, v. 26, p. 61-68, 1972.
- CROSSA, J. Statistical analysis of multilocation trials. **Advances in Agronomy**, New York, v. 44, p. 55-85, 1990.
- CROSSA, J.; EEUWIJK, F. van; SAYRE, K.D.; REYNOLDS, M.P.; VARGAS, M. Interpreting treatment environment interaction in agronomy trial. **Agronomy Journal**, Madison, v. 9, p. 949-960, 2001.
- CROSSA, J.; FOX, P.N.; PFEIFER, W.H.; RAJARRAM, S.; GAUCH, H.G. AMMI adjustment for statistical analysis of an international wheat yield trial. **Theoretical Applied Genetics**, New York, v. 81, p. 27-37, 1991.
- CRUZ, C.D.; TORRES, R.A.; VENCOSKY, R. An alternative approach to the stability analysis proposed by Silva and Barreto. **Revista Brasileira de Genética**, Ribeirão Preto, v. 12, p. 567-580, 1989.
- DIAS, C.T.S.; KRZANOWSKI, W.J. Model selection and cross-validation in additive main effect and multiplicative interaction (AMMI) models. **Crop Science**, Madison, v. 43, p. 865-873, 2003.

- DUARTE, J.B.; VENCOSKY, R. **Interação genótipo x ambiente**: uma introdução à análise “AMMI”. Ribeirão Preto: Sociedade Brasileira de Genética, 1999. 60 p. (Série Monografias, 9).
- DUARTE, J.B.; ZIMMERMANN, M.J.O. Correlation among yield stability parameters in common bean. **Crop Science**, Madison, v. 35, n. 3, p. 905-912, 1995.
- EASTMENT, H.T.; KRZANOWSKI, W. J. Cross-validatory choice of the number of components from a principal components analysis. **Technometrics**, Alexandria, v. 24, p. 73-77, 1982.
- EBERHART, S.A.; RUSSELL, W.A. Stability parameters for comparing varieties. **Crop Science**, Madison, v. 6, n. 1, p. 36-40, 1996.
- FINLAY, K.W.; WILKINSON, G.N. The analysis of adaptation in a plant-breeding programme. **Australian Journal of Agricultural Research**, Melbourne, v. 14, n. 5, p. 742-754, 1963.
- GABRIEL, K.R. Least squares approximation of matrices by additive and multiplicative models. **Journal of Royal Statistical Society, Series B**, New York, v. 40, p. 186-196, 1978.
- GABRIEL, K.R. Le biplot- outil d’exploration de données multidimensionnelles. **Journal de la Societe Francaise de Statistique**, Paris, v. 143, p. 5-55, 2002.
- GAUCH, H.G. Model selection and validation for yield trials with interaction. **Biometrics**, Washington, v. 44, p. 705-715, 1988.
- 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.
- GAUCH, H.G.; ZOBEL, R.W. AMMI analysis of yield trials. 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 features of factor analytic and analysis of variance techniques. **Psychometrika**, New York, v. 33, p. 73-115, 1968.
- GOOD, I.J. Some applications of the singular decomposition of a matrix. **Technometrics**, Alexandria, v. 11, n. 4, p. 823-831, 1969.
- HEGEMANN, V.; D.E. JOHNSON. On analyzing two-way AOV with interaction. **Technometrics**, Alexandria, v. 18, p. 273-281, 1976.
- JOHNSON, D.E.; F.A. GRAYBILL. An analysis of a two-way model with interaction and no replication. **Journal of the American Statistical Association**, Alexandria, v. 67, p. 862-868, 1972.

- JOHNSTONE, I.M. On the distribution of the largest eigenvalue in principal components analysis. **The Annals of Statistics**, Hayward, v. 29, n. 2, p. 295-327, 2001.
- HILL, R.R; ROSENBERGER, J.L. Methods for combining data from germplasm evaluation trials. **Crop Science**, Madison, v. 25, p. 467-470, 1985.
- KANG, M.S; MAGARI, R. New developments in selecting for phenotypic stability in crop breeding. In: KANG, M.S.; GAUCH, H.G. **Genotype-by-environment interaction**. Boca Raton: CRC Press, 1996. chap. 1, p. 1-14.
- KRAZANOWSKI, W.J. Cross-validation in principal component analysis. **Biometrics**, Washington, v. 43, p. 575-584, 1987.
- LIN, S.P.; PERLMAN, M.D. A Monte Carlo comparison of four estimators of a covariance matrix. In: KRISHNAIAH, P.R. **Multivariate Analysis-VI**. Amsterdam: Elsevier, 1985, chap. 26, p. 411-429.
- MANDEL, J. The partitioning of interactions in analysis of variance. **Journal of Research of the National Bureau of Standards, Series B**, Washington, v. 73, p. 309-328, 1969.
- MANDEL, J. A new analysis of variance model for non-additive data. **Technometrics**, v. 13, n. 1, p. 1-18, 1971.
- MATHERSON, A.C.; RAYMOND, C.A. A review of provenance x environment interaction: its practical importance and use with particular reference to the tropics. **The Commonwealth Review**, London, v. 65, n. 4, p. 283-302, 1986.
- MUIRHEAD, R.J. Developments in eigenvalue estimation. In: GUPTA, A.K. **Advances in multivariate statistical**. Boston: Reidel, 1987. p. 277-288.
- NACHIT, M.M.; NACHIT, G.; KETATA, H.; GAUCH, H.G.; ZOBEL, R.W. Use of AMMI and linear regression models to analyze genotype-environment interaction in durum wheat. **Theoretical and Applied Genetics**, New York, v.83, p.597-601, 1992.
- PIEPHO, H.P. Best linear unbiased prediction (BLUP) for regional yield trials: a comparison to additive main effects and multiplicative interaction (AMMI) analysis. **Theoretical and Applied Genetics**, New York, v. 89, p. 647-654, 1994.
- PIEPHO, H.P. Robustness of statistical test for multiplicative terms in additive main effects and multiplicative interaction model for cultivar trial. **Theoretical and Applied Genetics**, New York, v. 90, p. 438-443, 1995.
- RAMALHO, M.A.P.; SANTOS, J.B.; ZIMMERMANN, M.J.O. **Genética quantitativa em plantas autógamas**: aplicações ao melhoramento do feijoeiro. Goiânia: UFG, 1993. 271 p.
- RENCHE, A.C. **Linear models in statistics**. New York: Wiley, 2000. 578 p.

RIBEIRO, P.H.E. **Adaptabilidade e estabilidade de cultivares de milho em diferentes épocas de semeadura, níveis de adubação e locais do estado de Minas Gerais**. 1999. 126 p. Tese (Doutorado em Agronomia) - Universidade Federal de Lavras, Lavras, 1999.

ROBERTSON, T.; WRIGHT, F.T.; DYKSTRA, R.L. **Order restricted statistical inference**. New York: Wiley, 1998. 521 p.

ROSSE, L.N. **Modelo de regressão não-linear aplicado na avaliação da estabilidade fenotípica em plantas**. 1999. 179 p. Tese (Doutorado em Genética e Melhoramento de Plantas) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 1999.

SAS INSTITUTE. **SAS 9.2**. Cary, 2003.

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. **Resumo ...** Campinas: Fundação Cargill, 1985. p. 49-50.

SOUZA, J. de. **Análise em componentes principais e suas aplicações: métodos estatísticos nas ciências psicossociais**, Brasília: Thesaurus, 1988. v. 2, 68 p.

STONE, M. Cross-validatory choice and assessment of statistical predictions (with Discussion). **Journal of the Royal Statistical Society, Series B**, London, v. 36, p. 111-148, 1974.

STROUP, W.W.; MULITZE, D.K. Nearest neighbor adjusted best linear unbiased prediction. **The American Statistician**, Washington, v. 45, p. 194-200, 1991.

VELICER, W.F. Determining the number of components from the matrix of partial correlations. **Psychometrika**, New York, v. 41, p. 321-327, 1976.

VENCOVSKY, R.; BARRIGA, P. **Genética biométrica no fitomelhoramento**. Ribeirão Preto: Sociedade Brasileira de Genética, 1992. 486 p.

WOLD, S. Pattern recognition by means of disjoint principal component models. **Pattern Recognition**, Amsterdam, v. 8, p. 127-139, 1976.

WOLD, S. Cross-validatory estimation of the number of components in factor and principal component models. **Technometrics**, Alexandria, v. 20, p. 397-405, 1978.

YOCHMOWITZ, M.G; CORNELL, R.G. Stepwise tests for multiplicative components of interaction. **Technometrics**, Alexandria, v. 20, p. 79-84, 1978.

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

ANEXOS

A seguir, serão apresentados os programas no SAS, utilizados para fazer as análises.

```
-----
Análise da variância do experimento
-----

proc glm data=TRIAL5 outstat=anavaexp;
class REP SITE GEN;
model YIELD=REP(SITE) SITE GEN SITE*GEN/ss3;
run;
data anavaexp2;
set anavaexp;
if _source_ = 'ERROR';
SQRes=ss;
gl_Res=df;
QMRes=SQRes/gl_Res;
keep gl_Res QMRes;
data anavaexp3;
set anavaexp;
if _source_ = 'REP(SITE)';
SQBloco=ss;
gl_bloco=df;
QMBloco=SQBloco/gl_bloco;
nRep=df+1;
keep gl_bloco
QMBloco nrep;
```

Obter a tabela de dupla entrada residual (GE)

```
-----

proc sort data=TRIAL5;
by site gen;
run;
proc means data=TRIAL5 noprint;
by site gen;
var YIELD;
output out = medias mean=YIELDM;
run;
proc sort data=medias;
by gen site;
run;
proc transpose data=medias out = medias2;
by gen;
id site;
var YIELDM;
run;
```

```

proc iml;
use medias2;
read all into YIELDM;
TT=YIELDM;
Y=YIELDM[,2:ncol(YIELDM)];
yyy=YIELDM[,2:ncol(YIELDM)];
g=nrow(Y);
e=ncol(Y);
Yi=j(g,e,0); do i=1 to g;
    Yi[i,]=j(1,e,sum(Y[i,])/e);
end;
Yj=j(g,e,0); do j=1 to e;
    Yj[,j]=j(g,1,sum(Y[,j])/g);
end;
Y_barra = j(g,e,sum(Y)/(g*e));
GE = Y - Yi - Yj + Y_barra;

XO=GE;
g=NROW(GE);
e=NCOL(GE);
-----

Método de validação cruzada dado por Eastment e Krzanowski (1982)
-----

start meacent(x,g);
XTOT=REPEAT(X[+,],g,1);
XBAR=XTOT/g;
X=X-XBAR;
return(X);
finish;

/* Módulo para apagar a linha r da matriz de dados */
start rowdel(y,r);
return(y[setdif(1:nrow(y),r),]);
finish;

/* Módulo para apagar a coluna J da matriz de dados */
start coldel(y,j);
return(y[,setdif(1:ncol(y),j)]);
finish;

/* Módulo para definição da função PRESS */ START
PRESS(ARG1,ARG2,ARG3,ARG4);
ARG5=(1/(ARG1*ARG2))*SUM((ARG3-ARG4)##2);
RETURN(ARG5); FINISH
PRESS;

```



```

/* Decomposição de Valor Singular de XC matriz de dados centrados
na média */
CALL SVD(U1,D1,V1, XO);
D=DIAG(D1);
U=U1[,1:e];
VL=t(V1[,1:e]);
A=U*D*VL;

/* Cria matrizes PRESS e W de zeros */
PRESS_m=SHAPE(0,g-2,1);
Wm=SHAPE(0,g-2,1);
ASSY=(1/(g*e))*SUM((XO)##2);
DR=(g-1)*e;

do m=1 to g-2; YE=SHAPE(0,g,e);
do r=1 to g;
  y0=GE;

/* Retirada da linha r */ /* Decomposição de Valor Singular com os
valores de X centrados na média */
Y0=rowdel(y0,r);
y0=meancent(y0,g-1);
CALL SVD(U1,D1,V1,y0);
DBCOR=sqrt(DIAG(D1));
UB=U1[,1:e];
VBL=t(V1[,1:e]);

/* Retirada da coluna j */
do c=1 to e;
  y0=GE;
  Y0=coldel(y0,c);

/* Decomposição de Valor Singular com os valores de X centrados na
média */
Y0=meancent(Y0,g);
  CALL SVD(U1,D1,V1,y0);
DTCOR=sqrt(DIAG(D1));
UT=U1[,1:e-1];
VTL=t(V1[,1:e-1]);
z=UT*diag(d1)*VTL;

/* Troca de sinais da matriz UT para que fiquem iguais aos da
matriz U */
SU=0; SUT=0;

```

```

DO II=1 TO g;
  DO JJ=1 TO (e-1);
    IF U[II,JJ]<0 THEN SU=-1;
    ELSE IF U[II,JJ]>0 THEN SU=1;
    IF UT[II,JJ]<0 THEN SUT=-1;
    ELSE IF UT[II,JJ]>0 THEN SUT=1;
    IF SU ^= SUT THEN UT[II,JJ]=-UT[II,JJ];
  END;
END;

/* Troca de sinais da matriz VBL para que fiquem iguais aos da
matriz VL */
SUU=0; SUUT=0;
DO III=1 TO e;
  DO JJJ=1 TO e;
    IF VL[III,JJJ]<0 THEN SUU=-1;
    ELSE IF VL[III,JJJ]>0 THEN SUU=1;
    IF VBL[III,JJJ]<0 THEN SUUT=-1;
    ELSE IF VBL[III,JJJ]>0 THEN SUUT=1;
    IF SUU ^= SUUT THEN VBL[III,JJJ]=-VBL[III,JJJ];
  END;
END;

/* Armazena os valores preditos na matriz YE */
do T=1 to m;
  YE[r,c]=YE[r,c]+UT[r,T]*DTcor[T,T]* DBcor[T,T]*VBL[T,c];
end;
end; end; YY=YIELDM[,2:ncol(YIELDM)]; g=nrow(YY); e=ncol(YY);
YYi=j(g,e,0); do i=1 to g; YYi[i,]=j(1,e,sum(YY[i,])/e); end;
YYj=j(g,e,0); do j=1 to e; YYj[,j]=j(g,1,sum(YY[,j])/g); end;
YY_barra = j(g,e,sum(YY)/(g*e)); Ychapeu=YE+YYi+YYj-YY_barra;

*Correlação;
CORR=corr(colvec(YY)||colvec(Ychapeu));
print CORR;

/* Função PRESS_m */
DM=g+e-2*m;
DR=DR-(g+e-(m-1)*2);
print dm dr;
PRESS_m[m]=PRESS(g,e,YE,GE);

/* Cálculo da Estatística W */
if m=1
  then

```

```

    Wm[m]=((ASSY-PRESS_m[m])/DM)/(PRESS_m[m]/DR);
else;
    Wm[m]=((PRESS_m[m-1]-PRESS_m[m])/DM)/(PRESS_m[m]/DR);
end;
-----
Método de validação cruzada dado por Eastment e Krzanowski (1982)
com método de correção 3 e isotonização
-----

start centlinha(x,c);
XTOT=REPEAT(X[,+],1,c);
XBAR=XTOT/c;
X=X-XBAR;
return(X);
finish;

start meancent(x,g);
XTOT=REPEAT(X[+,],g,1);
XBAR=XTOT/g;
X=X-XBAR;
return(X);
finish;

/* Módulo para apagar a linha r da matriz de dados */
start rowdel(y,r);
return(y[setdif(1:nrow(y),r),]);
finish;

/* Módulo para apagar a coluna i da matriz de dados */
start coldel(y,j);
return(y[,setdif(1:ncol(y),j)]);
finish;

/* Módulo para definição da função PRESS */
START PRESS(ARG1,ARG2,ARG3,ARG4);
ARG5=(1/(ARG1*ARG2))*SUM((ARG3-ARG4)##2);
RETURN(ARG5);
FINISH PRESS;

/* Decomposição de Valor Singular de XC matriz de dados centrados
na média */ CALL SVD(U1,D1,V1, X0); D=DIAG(D1); U=U1[,1:e];
VL=t(V1[,1:e]);
A=U*D*VL;

YE1=GE;
```

```

/* Cria matrizes PRESS e W de zeros */
PRESS_m=SHAPE(0,g-2,1);
Wm=SHAPE(0,g-2,1); ASSY=(1/(g*e))*SUM((ge)##2);
DR=(g-1)*e;

do m=1 to g-2;
YE=SHAPE(0,g,e);

do r=1 to g; y0=GE;

/* Retirada da linha r */ /* Decomposição de Valor Singular com os
valores de X centrados na média */
Y0=rowdel(y0,r);
y0=meancent(y0,g-1); y0=centlinha(y0,e);
CALL SVD(U1,D1,V1,y0);
DB=sqrt(DIAG(D1));
UB=U1[,1:e];
VBL=t(V1[,1:e]);

p=min(g-2,e-1);
lambda=D1[1:p,];
lambda2=lambda;

* Correção dos autovalores;
*Método de correção 3;
FI_3=j(p,1,0);
alpha_3=j(p,1,0);
nlambda_3=j(p,1,0);
b=5.8*((g-2)##2)/((e+g-1)##2);
cc=6*(g-2)/(5*((e+g-1)##2));
ln_li=log(e*lambda2[1:p,]);
do i=1 to p;
  alpha_3[i,]=e*lambda2[i,]*( (b + sum(ln_li##2)) -
  cc*ln_li[i,]*(e+g+1-2*i) );
  nlambda_3[i,]=(e+g+1-2*i)*(b+sum(ln_li##2));
  FI_3[i,]=alpha_3[i,]/nlambda_3[i,];
end;

* Regressão Isotonica dos autovalores corrigidos pelo método 3
*Passo 1;

do until (alpha_3>0);

do i=1 to p;
  if alpha_3[p+1-i,] < 0 then

```

```

do;
    alpha_3[p+1-i,]=(alpha_3[p+1-i,]+alpha_3[p-i,]);
    nlambda_3[p+1-i]=(nlambda_3[p+1-i,]+nlambda_3[p-i,]);
    alpha_3[p-i,]=alpha_3[p+1-i,];
    nlambda_3[p-i]=nlambda_3[p+1-i,];
end;
else
    alpha_3[p+1-i,]=alpha_3[p+1-i,];
    nlambda_3[p+1-i,]=nlambda_3[p+1-i,];
end;
Fi_3_iso=j(p,1,0);

do i=1 to p;
    Fi_3_iso[i,]=alpha_3[i,]/nlambda_3[i,];
end;
end;

* Passo 2;
ff=j(p,1,0); do until (Fi_3_iso=ff);

do i=1 to p-1;
    if Fi_3_iso[p-i+1,] > Fi_3_iso[p-i,] then
        do;
            alpha_3[p-i,]=(alpha_3[p-i+1,]+alpha_3[p-i,]);
            nlambda_3[p-i,]=(nlambda_3[p-i+1,]+nlambda_3[p-i,]);
            alpha_3[p-i+1,]=alpha_3[p-i,];
            nlambda_3[p-i+1,]=nlambda_3[p-i,];
            Fi_3_iso[p-i,]=alpha_3[p-i,]/nlambda_3[p-i,];
            Fi_3_iso[p-i+1,]=alpha_3[p-i+1,]/nlambda_3[p-i+1,];
        end;
    else
        alpha_3[p-i+1,]=alpha_3[p-i+1,];
        alpha_3[p-i,]=alpha_3[p-i,];
        nlambda_3[p-i+1,]=nlambda_3[p-i+1,];
        nlambda_3[p-i,]=nlambda_3[p-i,];
        Fi_3_iso[p-i,]=alpha_3[p-i,]/nlambda_3[p-i,];
        Fi_3_iso[p-i+1,]=alpha_3[p-i+1,]/nlambda_3[p-i+1,];
    end;
end;
A=Fi_3_iso; call sort ( A, {1}, {1}); ff=A; end;

DBc=diag(Fi_3_iso); DBcor=DBc##(.5);

/* Retirada da coluna j */
do c=1 to e;
    y0=GE;

```

```

Y0=coldel(y0,c);

/* Decomposição de Valor Singular com os valores de X centrados na
média */
Y0=meancent(Y0,g);
y0=centlinha(Y0,e-1);
CALL SVD(U1,D1,V1,y0);
    DT=sqrt(DIAG(D1));
    UT=U1[,1:e-1];
    VTL=t(V1[,1:e-1]);
    z=UT*diag(d1)*VTL;

p=min(g-1,e-2);
lambda=d1[1:p,];
lambda2=lambda;

* Correção dos autovalores;
*Método de correção 3;

FI_3=j(p,1,0);
alpha_3=j(p,1,0);
nlambda_3=j(p,1,0);
b=5.8*((g-2)##2)/((e+g-1)##2);
cc=6*(g-2)/(5*((e+g-1)##2));
ln_li=log(e*lambda2[1:p,]);
do i=1 to p;
    alpha_3[i,]=e*lambda2[i,]*( (b + sum(ln_li##2)) -
    cc*ln_li[i,]*(e+g+1-2*i) );
    nlambda_3[i,]=(e+g+1-2*i)*(b+sum(ln_li##2));
    FI_3[i,]=alpha_3[i,]/nlambda_3[i,];
end;

* Regressão Isotonica dos autovalores corrigidos pelo método 3
*Passo 1;

do until (alpha_3>0);

do i=1 to p;
    if alpha_3[p+1-i,] < 0 then
        do;
            alpha_3[p+1-i,]=(alpha_3[p+1-i,]+alpha_3[p-i,]);
            nlambda_3[p+1-i,]=(nlambda_3[p+1-i,]+nlambda_3[p-i,]);
            alpha_3[p-i,]=alpha_3[p+1-i,];
            nlambda_3[p-i,]=nlambda_3[p+1-i,];
        end;
    end;
end;

```

```

        else
            alpha_3[p+1-i,]=alpha_3[p+1-i,];
            nlambda_3[p+1-i,]=nlambda_3[p+1-i,];
        end;

    Fi_3_iso=j(p,1,0);

    do i=1 to p;
        Fi_3_iso[i,]=alpha_3[i,]/nlambda_3[i,];
    end;
end;

* Passo 2;
ff=j(p,1,0); do until (Fi_3_iso=ff);
    do i=1 to p-1;
        if Fi_3_iso[p-i+1,] > Fi_3_iso[p-i,] then
            do;
                alpha_3[p-i,]=(alpha_3[p-i+1,]+alpha_3[p-i,]);
                nlambda_3[p-i,]=(nlambda_3[p-i+1,]+nlambda_3[p-i,]);
                alpha_3[p-i+1,]=alpha_3[p-i,];
                nlambda_3[p-i+1,]=nlambda_3[p-i,];
                Fi_3_iso[p-i,]=alpha_3[p-i,]/nlambda_3[p-i,];
                Fi_3_iso[p-i+1,]=alpha_3[p-i+1,]/nlambda_3[p-i+1,];
            end;
        else
            alpha_3[p-i+1,]=alpha_3[p-i+1,];
            alpha_3[p-i,]=alpha_3[p-i,];
            nlambda_3[p-i+1,]=nlambda_3[p-i+1,];
            nlambda_3[p-i,]=nlambda_3[p-i,];
            Fi_3_iso[p-i,]=alpha_3[p-i,]/nlambda_3[p-i,];
            Fi_3_iso[p-i+1,]=alpha_3[p-i+1,]/nlambda_3[p-i+1,];
        end;
    end; A=Fi_3_iso; call sort ( A, {1}, {1}); ff=A; end;
lambcol=lambda2; alfacol=alpha_3; nlambcol=nlambda_3;
FI_1col=FI_3; Fi_3_isocol=Fi_3_iso;

DTc=diag(Fi_3_isocol); DTcor=DBc##(1/2);

/* Troca de sinais da matriz UT para que fiquem iguais aos da
matriz U */ SU=0; SUT=0; DO II=1 TO g;
    DO JJ=1 TO (e-1);
        IF U[II,JJ]<0 THEN SU=-1;
        ELSE IF U[II,JJ]>0 THEN SU=1;
        IF UT[II,JJ]<0 THEN SUT=-1;
        ELSE IF UT[II,JJ]>0 THEN SUT=1;
    end;
end;

```

```

        IF SU ^= SUT THEN UT[II,JJ]=-UT[II,JJ];
    END;
END;

/* Troca de sinais da matriz VBL para que fiquem iguais aos da
matriz VL */
    SUU=0; SUUT=0;
    DO III=1 TO e;
        DO JJJ=1 TO e;
            IF VL[III,JJJ]<0 THEN SUU=-1;
            ELSE IF VL[III,JJJ]>0 THEN SUU=1;
            IF VBL[III,JJJ]<0 THEN SUUT=-1;
            ELSE IF VBL[III,JJJ]>0 THEN SUUT=1;
            IF SUU ^= SUUT THEN VBL[III,JJJ]=-VBL[III,JJJ];
        END;
    END;

/* Armazena os valores preditos na matriz YE */
    do T=1 to m;
    YE[r,c]=YE[r,c]+UT[r,T]*DTcor[T,T]* DBcor[T,T]*VBL[T,c]; end; end;
end; YY=YIELDM[,2:ncol(YIELDM)]; g=nrow(YY); e=ncol(YY);
YYi=j(g,e,0); do i=1 to g; YYi[i,]=j(1,e,sum(YY[i,])/e); end;
YYj=j(g,e,0); do j=1 to e;
    YYj[,j]=j(g,1,sum(YY[,j])/g);
end;
YY_barra = j(g,e,sum(YY)/(g*e));
Ychapeu=YE+YYi+YYj-YY_barra;

*Correlação;
    CORR=corr(colvec(yy)||colvec(Ychapeu));
print CORR;

/* Função PRESS */
    DM=g+e-2*m;
    DR=DR-(g+e-(m-1)*2);
    PRESS_m[m]=PRESS(g,e,YE,GE);

/* Cálculo da Estatística W */
    if m=1
        then
            Wm[m]=((ASSY-PRESS_m[m])/DM)/(PRESS_m[m]/DR);
        else;
            Wm[m]=((PRESS_m[m-1]-PRESS_m[m])/DM)/(PRESS_m[m]/DR);
        end;
print PRESS_m Wm ;

```

 Validação Cruzada pelo método dado por Gabriel (2002)

```

Y=GE;
g=nrow(y);
e=ncol(y);
X0=GE;

start rowcodel(x,r,c);
return(x[setdif(1:nrow(x),r),setdif(1:ncol(x),c)]]);
finish;

start pegavrow(x,r,c);
return(x[r,setdif(1:ncol(X),c)]]);
finish;

start pegavcol(x,r,c);
return(x[setdif(1:nrow(X),r),c]);
finish;

start press_M(ARG1,ARG2,ARG3,ARG4);
ARG5=(1/(ARG1*ARG2))*sum((ARG3-ARG4)##2);
return(ARG5);
finish press_M;

/* Cria matrizes PRESS e W de zeros */ PRESSM=shape(0,g-1,1);
do M=1 to g-1;
  XE=shape(0,g,e);
  do r=1 to g;
    do c=1 to e;
      Y=X0;
      Y=rowcodel(Y,r,c);
      call svd(U,D,V,Y);
      D=diag(D);
      dd=ginv(D);
      Y=X0;
      YrT=pegavrow(Y,r,c);
      Yc=pegavcol(Y,r,c);
      XE[r,c]=YrT*V[,1:M]*ginv(D)[1:M,1:M]*U'[1:M,]*Yc;
    end;
  end;

  YY=YIELDM[,2:ncol(YIELDM)];
  g=nrow(YY);
  e=ncol(YY);

```

```

YYi=j(g,e,0);
do i=1 to g;
YYi[i,]=j(1,e,sum(YY[i,])/e);
end;
YYj=j(g,e,0); do j=1 to e;
YYj[,j]=j(g,1,sum(YY[,j])/g);
end;
YY_barra = j(g,e,sum(YY)/(g*e));

Ychapeu=XE+YYi+YYj-YY_barra;

*Correlação;
CORR=corr(colvec(YY)||colvec(Ychapeu));
print CORR;

* PRESS;
PRESSM[M]=press_M(g,e,ge,xe);
end;
PRINT PRESSM ;

-----
Validação cruzada pelo método de Gabriel com o método de correção
3 e isotonização
-----

XO=GE;
g=NROW(GE);
e=NCOL(GE);

start meancent(x,c);
XTOT=REPEAT(X[,+],1,c);
XBAR=XTOT/c;
X=X-XBAR;
return(X);
finish;

start rowcodel(x,r,c);
return(x[setdif(1:nrow(x),r),setdif(1:ncol(x),c)]); finish;

start pegavrow(x,r,c);
return(x[r,setdif(1:ncol(X),c)]);
finish;

start pegavcol(x,r,c);
return(x[setdif(1:nrow(X),r),c]);
finish;

```

```

start press_M(ARG1,ARG2,ARG3,ARG4);
ARG5=(1/(ARG1*ARG2))*sum((ARG3-ARG4)##2);
return(ARG5); finish
press_M;

/*PRESS*/
PRESSM=shape(0,g-2,1);

do M=1 to g-2;
XE=shape(0,g,e);
do r=1 to g;
do c=1 to e;
XX=X0;
X1=rowcode1(XX,r,c);
X=meancent(X1,e-1);
call svd(U,D,V,X);
p=min(g-2,e-2);
L_auto=D[1:p,];
lambda2=L_auto;

* Correção dos autovalores;
*Método de correção 3;

FI_3=j(p,1,0);
alpha_3=j(p,1,0);
nlambda_3=j(p,1,0);
b=5.8*((g-2)##2)/((e+g-1)##2);
cc=6*(g-2)/(5*((e+g-1)##2));
ln_li=log(e*lambda2[1:p,]);
do i=1 to p;
    alpha_3[i,]=e*lambda2[i,]*( (b + sum(ln_li##2)) -
    cc*ln_li[i,]*(e+g+1-2*i) );
    nlambda_3[i,]=(e+g+1-2*i)*(b+sum(ln_li##2));
    FI_3[i,]=alpha_3[i,]/nlambda_3[i,];
end;

* Regressão Isotonica dos autovalores corrigidos pelo método 3
*Passo 1;

do until (alpha_3>0);

do i=1 to p;
    if alpha_3[p+1-i,] < 0 then
        do;
            alpha_3[p+1-i,]=(alpha_3[p+1-i,]+alpha_3[p-i,]);

```

```

        nlambda_3[p+1-i]=(nlambda_3[p+1-i,]+nlambda_3[p-i,]);
        alpha_3[p-i,]=alpha_3[p+1-i,];
        nlambda_3[p-i]=nlambda_3[p+1-i,];
    end;
else
    alpha_3[p+1-i,]=alpha_3[p+1-i,];
    nlambda_3[p+1-i,]=nlambda_3[p+1-i,];
end;

Fi_3_iso=j(p,1,0);

do i=1 to p;
    Fi_3_iso[i,]=alpha_3[i,]/nlambda_3[i,];
end;
end;

* Passo 2;
ff=j(p,1,0);
do until (Fi_3_iso=ff);
do i=1 to p-1;
    if Fi_3_iso[p-i+1,] > Fi_3_iso[p-i,] then
        do;
            alpha_3[p-i,]=(alpha_3[p-i+1,]+alpha_3[p-i,]);
            nlambda_3[p-i,]=(nlambda_3[p-i+1,]+nlambda_3[p-i,]);
            alpha_3[p-i+1,]=alpha_3[p-i,];
            nlambda_3[p-i+1,]=nlambda_3[p-i,];
            Fi_3_iso[p-i,]=alpha_3[p-i,]/nlambda_3[p-i,];
            Fi_3_iso[p-i+1,]=alpha_3[p-i+1,]/nlambda_3[p-i+1,];
        end;
    else
        alpha_3[p-i+1,]=alpha_3[p-i+1,];
        alpha_3[p-i,]=alpha_3[p-i,];
        nlambda_3[p-i+1,]=nlambda_3[p-i+1,];
        nlambda_3[p-i,]=nlambda_3[p-i,];
        Fi_3_iso[p-i,]=alpha_3[p-i,]/nlambda_3[p-i,];
        Fi_3_iso[p-i+1,]=alpha_3[p-i+1,]/nlambda_3[p-i+1,];
    end;
end;

A=Fi_3_iso;
call sort ( A, {1}, {1});
ff=A;
end;

Dcor=diag(Fi_3_iso);
X=GE;

```

```

YrT=pegavrow(X,r,c);
Yc=pegavcol(X,r,c);
XE[r,c]=YrT*V[,1:M]*inv(Dcor)[1:M,1:M]*U'[1:M,]*Yc; end; end;

YY=YIELDM[,2:ncol(YIELDM)];

g=nrow(YY); e=ncol(YY); YYi=j(g,e,0); do i=1 to g;
YYi[i,]=j(1,e,sum(YY[i,])/e); end; YYj=j(g,e,0); do j=1 to e;
YYj[,j]=j(g,1,sum(YY[,j])/g); end; YY_barra =j(g,e,sum(YY)/(g*e));

Ychapeu=XE+YYi+YYj-YY_barra;

*CORRELAÇÃO;
CORR=corr(colvec(yy)||colvec(ychapeu));
print corr;

/* PRESS_m */
PRESSM[M]=press_M(g,e,XE,GE);
end;
PRINT PRESSM ;

-----
Análise AMMI método de Gollob
-----

proc iml;
use anavaexp2;
read var {QMRes} into QMRes;
read var {gl_Res} into gl_Res;
use anavaexp3;
read var {gl_bloco} into gl_bloco;
read var {QMBloco} into QMBloco;
read var {nRep} into nRep;
use medias2;
read all into YIELDM;
Y=YIELDM[,2:ncol(YIELDM)];
g=nrow(Y);
e=ncol(Y);
Yi=j(g,e,0);
do i=1 to g;
    Yi[i,]=j(1,e,sum(Y[i,])/e);
end;
Yj=j(g,e,0);
do j=1 to e;
    Yj[,j]=j(g,1,sum(Y[,j])/g);
end;
Y_barra = j(g,e,sum(Y)/(g*e));

```

```

GE = Y - Yi - Yj + Y_barra;
SQGen=e*sum((Yi[,1] - Y_barra[,1])##2);
SQSite=g*sum((Yj[1,] - Y_barra[1,])##2); SQInt=sum(GE##2);
p=min(g-1,e-1);
call SVD (U,L,V,GE);
L_auto=L[1:p,];

SQ_gollob=j(p,1,0);
gl_gollob=j(p,1,0);
QM_gollob=j(p,1,0);
F_gollob=j(p,1,0);
p_valor_gollob=j(p,1,0);

do i=1 to p;
    SQ_gollob[i,]=(L_auto[i,]##2);
    gl_gollob[i,]=g + e - 1 - 2*i;
    QM_gollob[i,]=SQ_gollob[i,]/gl_gollob[i,];
    F_gollob[i,]=QM_gollob[i,]/(QMRes);
p_valor_gollob[i,]=1-probf(F_gollob[i,],gl_gollob[i,],gl_Res);
    if p_valor_gollob[i,] < 0.0001 then
        do;
            p_valor_gollob[i,]= 0;
        end;
    else
        p_valor_gollob[i,]=p_valor_gollob[i,];
    end;

print "Análise AMMI pelo método de Golllob";
print gl_gollob SQ_gollob QM_gollob F_gollob p_valor_gollob;
quit;

*Rammi;
RM=(PRESS_m +(3/4)*(QMRes))##(1/2);
QM=(RM##2)-(QMRES);
R=(QMRes)/QM;

```