

MÉTODOS DE CORREÇÃO DE AUTOVALORES E REGRESSÃO ISOTÔNICA NOS MODELOS AMMI

Lúcio Borges de ARAÚJO¹
Carlos Tadeu dos Santos DIAS¹

- RESUMO: Em experimentação agrícola, é freqüente a necessidade de análise conjunta de grupos de experimentos. Existem várias metodologias de análise e interpretação para a interação genótipo \times ambiente proveniente de um grupo de cultivares testados em vários ambientes. Entre essas metodologias destaca-se os modelos AMMI, que engloba ANOVA para os efeitos principais e para efeitos multiplicativos utiliza-se DVS. O problema de superestimação e subestimação de autovalores estimados da maneira convencional é conhecido na literatura especializada. Para superar esses problemas apresentam-se três métodos de correção dos autovalores estimados, mas nem sempre essas correções mantêm a ordem decrescente de valores. Assim, é sugerido o uso de regressão isotônica para ordenar esses dados. Verificou-se que: A regressão isotônica foi suficiente; houve redução no número de componentes significativos retidos nos modelos AMMI; o método 2 e o método 3 apresentaram, respectivamente, a menor e a maior taxa de correção da soma de quadrados da interação genótipo \times ambiente; para a medida $RMSPD_{PRESS}$, o menor valor foi obtido quando se utilizou o método de correção 2. Já o método de correção 3 apresentou o maior valor; o método 2 também se mostrou melhor quando o interesse era verificar o ganho em número de repetições relativo a uma aproximação para o número de repetições que faltam para o modelo AMMI completo apresentar uma *performance* igual ao modelo AMMI selecionado, este benefício esteve próximo de três repetições.

¹Departamento de Ciências Exatas, Universidade de São Paulo, campus de Piracicaba - ESALQ/USP, CEP 13418-900 Piracicaba, SP, Brasil. E-mail: araujob@gmail.com / ctsdias@esalq.usp.br

- PALAVRAS-CHAVE: Correção de autovalores; regressão isotônica; modelos AMMI; interação genótipo \times ambiente; eficiência da correção.

1 Introdução

Em muitos casos, o pesquisador está interessado em avaliar o desempenho de vários genótipos (tratamentos) em diversos ambientes (locais e/ou ano). Quando um conjunto de experimentos é planejado para vários locais é necessário considerar o delineamento individual em cada local e a combinação total dos delineamentos com os locais (interação genótipo \times ambiente). Logo, os dados observados podem ser organizados em uma tabela de dupla entrada, colocando, por exemplo, os genótipos nas linhas e os ambientes nas colunas.

Para investigar a existência da interação genótipo \times ambiente, é necessário usar a técnica de análise da variância (ANAVA), mas esta técnica pode ser otimizada para generalizar os resultados, embora ela não possa ser usada quando existe somente uma repetição por casela no cruzamento de genótipo por ambiente. Embora a interação represente uma das principais dificuldades encontradas pelos melhoristas durante a seleção de material genético, é importante ressaltar que com a exploração desse efeito pode-se chegar a ótimos resultados.

Várias metodologias estatísticas têm sido propostas para a interpretação da interação genótipo \times ambiente e os pesquisadores ainda continuam na busca de uma ferramenta estatística que permita extrair grande parte da informação possível desta fonte de variação. Entre tantas ferramentas disponíveis na literatura os métodos de regressão linear simples (Eberhart e Russel, 1966) e regressão linear múltipla (Silva e Barreto, 1985) têm sido as metodologias mais utilizadas, mas estas técnicas possuem limitações e têm sido alvo de várias críticas, como no caso em que a linearidade falha.

Na busca de novas ferramentas para o estudo da interação genótipo \times ambiente, o modelo AMMI (Mandel, 1961, 1969, 1971; Gollob, 1968) vem se destacando e ganhando grande aplicabilidade nos últimos anos (Duarte e Vencovsky, 1999). A metodologia AMMI engloba num único modelo a técnica ANAVA, para efeitos principais, e para efeitos multiplicativos (interação), utiliza-se análise de componentes principais (ACP) ou decomposição em valor singular (DVS), aplicado à matriz de interação.

Essa técnica multivariada é baseada no uso dos autovalores e

autovetores provenientes da matriz de interação genótipo \times ambiente. Araújo e Dias (2005), utilizando técnicas de simulação multivariada verificaram o problema de superestimação e subestimação de autovalores estimados da maneira convencional. Muirhead (1987) apresenta três métodos para corrigir autovalores estimados a partir das matrizes de covariâncias amostrais. Este autor alerta que nem sempre essas correções mantêm a ordem decrescente de valores e, dessa forma, é sugerido que se use regressão isotônica para ordenar esses dados.

O presente trabalho tem os seguintes objetivos: apresentar três técnicas analíticas para a correção de autovalores com a ordenação deles por regressão isotônica e aplicação destas correções nos modelos AMMI, em estudo de experimentos agrônomicos multiambientais; e verificar o ganho, em número de repetições, ao analisar experimentos utilizando-se as três correções para os autovalores.

2 Material e métodos

2.1 Características dos dados

Os dados a serem utilizados são os mesmos utilizados por Cornelius e Crossa (1999) e Dias e Krzanowski (2003). Foram obtidos pelo CIMMYT (Centro Internacional de Mejoramiento de Maiz y Trigo) em experimentos realizados em 34 países, caracterizando-se, assim, experimentos multiambientais. Foram utilizados 20 genótipos de trigo sendo um genótipo do tipo "durum" e os outros 19 do tipo "bread". Cada genótipo foi avaliado em 34 ambientes com 4 blocos, caracterizando assim, um delineamento aleatorizado em blocos.

2.2 Análise dos dados

Com o objetivo de verificar se existe a interação entre genótipos e ambiente, realizou-se uma análise de variância conjunta que envolve o estudo de todos os genótipos em todos os ambientes, e em cada ambiente obteve-se um delineamento aleatorizados em blocos. Hill e Rosenberger (1985) e Stroup e Muiltze (1991) mostraram que pode ser preferível, em termos de predição, assumir o efeito dos genótipos como aleatórios e, fixos, o efeito dos ambientes, obtendo o efeito da interação genótipo \times ambiente como aleatório.

Sendo a interação significativa, o próximo passo é fazer a decomposição da $SQ_{G \times E}$, para descartar um resíduo adicional presente nessa soma de quadrados. Essa decomposição é feita utilizando o fator analítico proposto por Mandel (1961, 1969, 1971) e Gollob (1968).

Antes de aplicar a decomposição, é necessário organizar os dados em uma tabela de dupla entrada (ou matriz de ordem $g \times e$) com as médias dos r blocos (ou repetições) para cada combinação de genótipos e ambientes:

$$\mathbf{Y}_{g \times e} = \begin{pmatrix} Y_{11} & Y_{12} & \dots & Y_{1e} \\ Y_{21} & Y_{22} & \dots & Y_{2e} \\ \dots & \dots & \dots & \dots \\ Y_{g1} & Y_{g2} & \dots & Y_{ge} \end{pmatrix}. \quad (1)$$

A matriz \mathbf{GE} é a matriz de interação entre os genótipos e os ambientes (matriz de resíduos dos efeitos principais), em que cada elemento $(ge)_{ij}$ de \mathbf{GE} é encontrado pela seguinte relação:

$$(ge)_{ij} = Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..} \quad (2)$$

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.

O modelo AMMI pressupõe componentes aditivos para os efeitos principais de genótipos e ambientes, e componentes multiplicativos para o efeito de interação. Então, a resposta média sobre r repetições ou blocos, do i -ésimo genótipo no j -ésimo ambiente, é representada por:

$$Y_{ij} = \mu + g_i + e_j + \sum_{k=1}^q \hat{\lambda}_k \alpha_{ik} \gamma_{jk} + \rho_{ij} + \varepsilon_{ij} \quad (3)$$

sendo que:

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

μ : é uma constante, geralmente a média;

g_i : é o efeito do i -ésimo genótipo;

e_j : é o efeito do j -ésimo ambiente;

$\hat{\lambda}_k$: é a raiz quadrada do k -ésimo autovalor da matriz $(\mathbf{GE})(\mathbf{GE})^t$ (ou $(\mathbf{GE})^t(\mathbf{GE})$), com $k = 1, 2, \dots, q$ e onde $q < p$ determina uma aproximação de mínimos quadrados para a matriz \mathbf{GE} pelos q primeiros termos da DVS, $p = \min\{g - 1, e - 1\}$, $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p$;

α_{ik} : é o i -ésimo elemento do vetor coluna $\boldsymbol{\alpha}_k$ associado a $\hat{\lambda}_k$ e α_{ik} satisfaz o contraste de ortonormalização $\sum_i \alpha_{ik} \alpha_{ik'} = 0$ para $k \neq k'$ e $\sum_i \alpha_{ik}^2 = 1$;

γ_{jk} : é o j -ésimo elemento do vetor linha $\boldsymbol{\gamma}_k$ associado a $\hat{\lambda}_k$ e γ_{jk} satisfaz o contraste de ortonormalização $\sum_j \gamma_{jk} \gamma_{jk'} = 0$ para $k \neq k'$ e $\sum_j \gamma_{jk}^2 = 1$;

ρ_{ij} : é o resíduo adicional;

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

Existem várias técnicas para atribuir os graus de liberdade associados à cada parcela da $SQ_{G \times E}$, ou seja, associada a $\hat{\lambda}_k^2$. O método Gollob (1968) destaca-se pela sua simplicidade operacional. A expressão do método para definir os graus de liberdade é:

$$GL_{IPCA_k} = g + e - 1 - 2k. \quad (4)$$

em que:

IPCA : Interaction Principal Components Analysis;

2.3 Correção dos autovalores

Para corrigir o viés dos autovalores provenientes da matriz $(\mathbf{GE})(\mathbf{GE})^t$ ou $(\mathbf{GE})^t(\mathbf{GE})$, onde $(\mathbf{GE})(\mathbf{GE})^t \sim W_g(e, \boldsymbol{\Sigma})$, usa-se quaisquer dos três métodos apresentados por Muirhead (1987):

i) A correção apresentada pelo primeiro método resulta um novo autovalor, dado por:

$$\phi_k^{(1)} = \frac{e\hat{\lambda}_k^2}{\alpha_k^{(1)}} = \frac{e\hat{\lambda}_k^2}{e - g + 1 + 2\hat{\lambda}_k^2 \sum_{k \neq k^*}^p \frac{1}{\hat{\lambda}_k^2 - \hat{\lambda}_{k^*}^2}} \quad (5)$$

ii) No segundo método, um autovalor é corrigido por:

$$\phi_k^{(2)} = \frac{e\hat{\lambda}_k^2}{\alpha_k^{(2)}} = \frac{e\hat{\lambda}_k^2}{e - g - 1 + 2\hat{\lambda}_k^2 \sum_{k \neq k^*}^p \frac{1}{\hat{\lambda}_k^2 - \hat{\lambda}_{k^*}^2}} \quad (6)$$

iii) E o terceiro método produz um autovalor, dado por:

$$\begin{aligned} \phi_k^{(3)} &= \frac{e\hat{\lambda}_k^2}{e + g + 1 - 2k} - \frac{ec\hat{\lambda}_k^2 \ln(e\hat{\lambda}_k^2)}{b + \sum_{k=1}^p [\ln(e\hat{\lambda}_k^2)]^2} \\ &= \frac{e\hat{\lambda}_k^2 (b + \sum_{k=1}^p [\ln(e\hat{\lambda}_k^2)]^2 - (e + g + 1 - 2k)c \ln(e\hat{\lambda}_k^2))}{(e + g + 1 - 2k)(b + \sum_{k=1}^p [\ln(e\hat{\lambda}_k^2)]^2)} \quad (7) \\ &= \frac{\alpha_k^{(3)}}{d_k} \end{aligned}$$

$$\text{com } b = \frac{5, 8(g-2)^2}{(e+g-1)^2}, \quad c = \frac{6(g-2)}{5(e+g-1)^2},$$

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

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

Os autovalores obtidos pelas expressões (5), (6) e (7), nem sempre se apresentam na ordem: $\phi_1 \geq \phi_2 \geq \dots \geq \phi_p$. Para colocá-los em ordem

decrecente é necessário modificar os autovalores obtidos pelos métodos de correção, já que os autovalores provenientes da DVS apresentam uma ordem decrescente. Essa modificação pode ser realizada por regressão isotônica como definido por Robertson; Wright e Dykstra (1988).

Lin e Perlman (1985) apresentam o procedimento de Stein, um algoritmo para modificar autovalores obtidos pelos métodos de correção (5), (6) e (7), ordenando-os de forma decrescente.

Lista-se os produtos $e\hat{\lambda}_k^2$ (numerador das expressões de correção 5 e 6) ou d_k (denominador da expressão de correção 7) em uma coluna e em outra coluna lista-se os valores do denominador de (5), $\alpha_k^{(1)}$, ou os valores do denominador de (6), $\alpha_k^{(2)}$, ou ainda os valores do numerador da expressão (7), $\alpha_k^{(3)}$:

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

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

Passo 2 – Reordenando as razões $\frac{e\hat{\lambda}_k^2}{\alpha_k}$ de forma que estejam em ordem decrescente:

Lista-se as razões $\frac{e\hat{\lambda}_k^2}{\alpha_k}$ à direita de cada par $(e\hat{\lambda}_k^2, \alpha_k)$ obtido 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 foi 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 este par violado com o par imediatamente acima dele, substitui-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 de razões.
- c) Reinicia-se no par imediatamente após o par trocado em **(b)** e procede-se para cima até o próximo par violado ser encontrado; então, repete-se **(b)**.
- d) Repete-se **(c)** até todas razões $\frac{e\hat{\lambda}_k^2}{\alpha_k}$ estarem em ordem decrescente.

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

2.4 Eficiência da correção dos autovalores

As correções sugeridas anteriormente devem ser avaliadas por algum procedimento estatístico. A estatística *PRESS* (“Prediction Sum of Squares”), proposta por Allen (1971), consiste em um critério para escolha das variáveis regressoras a serem incluídas no modelo. Especificamente, *PRESS* é a soma de quadrado das diferenças entre um valor observado e um valor predito pelo modelo selecionado. Para os modelos AMMI, Cornelius; Crossa e Seyedsadr (1993) definiram a estatística *PRESS* como:

$$PRESS = \sum_{i=1}^g \sum_{j=1}^e (y_{ij} - \hat{y}_{ij})^2, \quad (8)$$

em que:

y_{ij} : é a média de r repetição do i -ésimo genótipo localizado no j -ésimo ambiente.

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

Utilizando a estatística *PRESS*, Cornelius ; Crossa e Seyedsadr (1993) ajustaram um valor para a medida RMSPD que faz uma comparação aproximada com outros RMSPDs provenientes de modelos ajustados por validação cruzada. O ajuste é feito por:

$$RMSPD_{(PRESS)} = \sqrt{\frac{PRESS}{ge} + \frac{(r-1)QM_{EM}}{r}} \quad (9)$$

em que r é o número de repetição ou o número de blocos em cada ambiente do experimento.

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

$$R_{AMMI} = \frac{QM_{EM}}{QM_{EM}(modelo)} \cdot r \quad (10)$$

Então, para fazer uma estimativa do $QM_{EM}(modelo)$, Piepho (1994) sugere a seguinte expressão:

$$\hat{QM}_{EM}(modelo) = (RMSPD_{(PRESS)})^2 - \frac{QM_{EM}}{r} \quad (11)$$

Este foi o procedimento utilizado no presente estudo para avaliar o ganho em termos de número de repetições ao se fazer a correção dos autovalores nos modelos AMMI.

Todas as análises e a rotina computacional foram implementadas pelo sistema estatístico SAS (SAS Institute Inc., 2003)

3 Resultados e discussão

Na Tabela 1, apresenta-se a análise de variância conjunta bem como o desdobramento da interação genótipo \times ambiente efetuada com os dados observados. Verifica-se, ao nível de 1% de significância, que o efeito de genótipos, o efeito de ambientes e o efeito da interação genótipo \times ambiente são significativos e suas somas de quadrados (SQ) correspondem a 1,49%, 72,7% e 9,97%, respectivamente, da soma de quadrados total.

O resultado de maior interesse nessa tabela, segundo o modelo AMMI, é a soma de quadrados da interação genótipo \times ambiente,

$SQ_{G \times E} = 594.108.485,00$, que pode estar inflacionada devido a presença de ruídos na variável resposta.

Tabela 1 - Análise de variância conjunta do Experimento com 20 genótipos avaliado em 34 ambientes com 4 blocos e decomposição das somas de quadrados da interação genótipo \times ambiente

Fonte de Variação	GL	SQ	QM	F	valor p
Blocos/ambiente	102	13.961.185,00	136.874,36	0,29	1,00
Ambiente (E)	33	4.333.925.428,00	131.331.073,58	217,34	< 0,01
Genótipo (G)	19	89.066.441,00	4.687.707,42	4,95	< 0,01
Interação (G \times E)	627	594.108.485,00	947.541,44	1,97	< 0,01
IPCA 1	51	179.059.116,00	3.510.963,06	7,31	< 0,01
IPCA 2	49	91.403.060,00	1.865.368,57	3,88	< 0,01
IPCA 3	47	78.977.904,00	1.680.380,94	3,50	< 0,01
IPCA 4	45	53.975.744,00	1.199.460,98	2,50	< 0,01
IPCA 5	43	34.225.218,00	795.935,30	1,66	< 0,01
IPCA 6	41	27.621.529,60	673.695,84	1,40	0,04
IPCA 7	39	24.062.812,00	616.995,18	1,28	0,11
IPCA 8	37	23.218.126,00	627.516,92	1,31	0,10
IPCA 9	35	15.530.032,40	443.715,21	0,92	0,59
IPCA 10	33	14.201.897,60	430.360,53	0,90	0,64
IPCA 11	31	12.867.389,60	415.077,08	0,86	0,68
IPCA 12	29	10.048.801,60	346.510,40	0,72	0,86
IPCA 13	27	7.612.139,20	281.931,08	0,59	0,95
IPCA 14	25	5.691.899,60	227.675,98	0,47	0,99
IPCA 15	23	4.627.279,60	201.186,07	0,42	0,99
IPCA 16	21	4.282.728,00	203.939,43	0,42	0,99
IPCA 17	19	2.440.969,88	128.472,10	0,27	0,99
IPCA 18	17	2.308.250,76	135.779,46	0,28	0,99
IPCA 19	15	1.953.586,04	130.239,07	0,27	0,99
Resíduo	1.938	930.547.529,00	480.158,68	-	-
Total	2.719	5.961.609.068,00	-	-	-
Média (kg/ha)	3.990,80				
CV (%)	14,90				

Na mesma tabela, é feito um ajuste da interação por decomposição em valores singulares (DVS), aplicada à matriz de interação genótipo \times ambiente. Essa matriz tem posto $p = \min(19, 33) = 19$, assim a $SQ_{G \times E}$ pode ser decomposta em 19 componentes ortogonais, que são as somas de quadrados parciais.

Pelo teste F, com os graus de liberdade ajustados pelo método de Gollob (1968), verificou-se a quantidade de componentes multiplicativos não nulos a serem retidos no modelo. O teste estatístico foi aplicado ao nível de

1% em vez de 5%, pois assim reduz-se a probabilidade de ocorrer o erro tipo I , isto é, de ajustar um modelo AMMI com um maior número de componentes, quando na verdade o modelo correto é mais parcimonioso. Contudo, aumenta-se a probabilidade de ocorrer o erro tipo II , ou seja, de ajustar um modelo AMMI com menor número de componentes, sendo o modelo correto mais complexo.

A importância de controlar o erro tipo I é destacada por Cornelius ; Seyedsadr e Crossa (1992), que também consideram como bastante liberal o método de Gollob (1968), podendo resultar em falsos resultados de significância e, provavelmente, uma seleção de maior número de componentes, aumentando-se assim a ocorrência do erro tipo I .

Assim na Tabela 1 é apresentada a análise de cada componente pelo teste F, com os graus de liberdade ajustados pelo método de Gollob (1968). Nota-se, ao nível de 1% de significância que os cinco primeiros componentes são significativos para o modelo, e o primeiro componente retém 30,13% da $SQ_{G \times E}$, o segundo contém 15,38%, 13,29% é retido pelo terceiro componente, 9,08% pelo quarto e 5,76% pelo quinto componente. Esses cinco componentes juntos representam 73,64% da $SQ_{G \times E}$, que é considerada resposta padrão presente na $SQ_{G \times E}$ com 235 graus de liberdade (37% dos graus de liberdade da interação), uma vez que a técnica DVS procura maximizar a informação da variabilidade nos primeiros componentes, enquanto os outros 26,36% da $SQ_{G \times E}$, assume-se que seja devido a ruídos presente nos dados.

É possível que estejam viesados os autovalores encontrados, pois foram obtidos de maneira usual e de acordo com o que foi verificado por Araújo e Dias (2005), autovalores obtidos de maneira usual podem apresentar viés. Nos modelos AMMI, a retenção de ruído por esses componentes indica que a estimativa do modelo AMMI (como média) não é perfeita (Gauch Jr, 1992).

Utilizando as sugestões de Muirhead (1987), obtêm-se os valores dos autovalores corrigidos $\phi^{(1)}$, $\phi^{(2)}$ e $\phi^{(3)}$, obtidas pelas eq. (5), (6) e (7), respectivamente, apresentados na Tabela 2. Percebe-se que os novos autovalores não satisfazem algumas condições que deveriam ser respeitadas. A ordem decrescente dos valores não é verificada para nenhum dos métodos de correção. Os valores de $\phi^{(1)}$ e $\phi^{(2)}$ apresentam outro problema, que é o fato de assumirem valores negativos, o que não é esperado, pois são provenientes da soma de quadrados da interação genótipo \times ambiente. Para superar esses problemas utiliza-se da regressão isotônica, mais propriamente o algoritmo de Stein apresentado por Lin e Perlman (1985). Os valores $\phi^{(1)*}$, $\phi^{(2)*}$ e $\phi^{(3)*}$ também são apresentados na Tabela 2 e referem-se aos autovalores corrigidos

por regressão isotônica de $\phi^{(1)}$, $\phi^{(2)}$ e $\phi^{(3)}$, respectivamente.

A regressão isotônica mostrou-se muito eficaz para realizar os ajustes necessários nos autovalores corrigidos. Verifica-se ainda que $\sum_{k=1}^{19} \phi_k^{(1)*}$ equivale a 76% da $SQ_{G \times E}$ e o restante da $SQ_{G \times E}$ (24%) pode representar ruído detectado pela correção que estava presente na interação. A soma total dos autovalores corrigidos pelo método 2 e ajustados ($\sum_{k=1}^{19} \phi_k^{(2)*}$) é equivalente a 80% e 20% representa um suposto ruído excluído da $SQ_{G \times E}$ pela correção; e $\sum_{k=1}^{19} \phi_k^{(3)*}$ representa 28% da $SQ_{G \times E}$, uma vez que a correção considerou 72% da interação genótipos \times ambientes como supostamente ruídos, ou heterogeneidade devido a genótipos e/ou ambientes (Snee, 1982).

Tabela 2 - Correção dos autovalores da matriz $(GE)(GE)^t$ e os autovalores ajustados pela regressão isotônica

λ^2	$\phi^{(1)}$	$\phi^{(1)*}$	$\phi^{(2)}$	$\phi^{(2)*}$	$\phi^{(3)}$	$\phi^{(3)*}$
44.764.779,00	26.268.136,00	26.268.136,00	27.207.272,00	27.207.272,00	6.286.366,50	6.286.366,50
22.850.765,00	11.426.689,00	12.962.548,00	11.772.993,00	13.443.869,00	4.147.974,00	4.168.439,50
19.744.476,00	15.350.388,00	12.962.548,00	16.086.043,00	13.443.869,00	4.189.740,40	4.168.439,50
13.493.936,00	9.286.514,50	9.286.514,50	9.678.314,90	9.678.314,90	3.383.543,80	3.383.543,80
8.556.304,50	4.787.919,20	4.957.793,30	4.950.884,20	5.179.251,30	2.512.930,80	2.512.930,80
6.905.382,40	3.759.430,20	4.957.793,30	3.883.808,10	5.179.251,30	2.305.817,10	2.330.966,40
6.015.703,00	2.391.437,20	4.957.793,30	2.448.698,20	5.179.251,30	2.260.432,40	2.330.966,40
5.804.531,50	-9.118.598,00	4.957.793,30	-8.347.242,00	5.179.251,30	2.432.846,10	2.330.966,40
3.882.508,10	2.142.554,10	3.602.683,10	2.214.438,40	3.831.408,40	1.847.233,70	1.876.806,40
3.550.474,40	3.884.208,80	3.602.683,10	4.151.360,30	3.831.408,40	1.883.214,10	1.876.806,40
3.216.847,40	13.991.748,00	3.602.683,10	18.802.428,00	3.831.408,40	1.903.167,50	1.876.806,40
2.512.200,40	4.050.462,30	3.602.683,10	4.474.868,40	3.831.408,40	1.667.989,70	1.667.989,70
1.903.034,80	2.732.130,30	2.732.130,30	2.984.144,90	2.984.144,90	1.420.000,30	1.420.000,30
1.422.974,90	1.710.600,20	2.419.297,40	1.840.767,30	2.739.715,80	1.195.173,10	1.195.173,10
1.156.819,90	1.125.535,60	2.419.297,40	1.193.863,70	2.739.715,80	1.093.842,40	1.116.427,80
1.070.682,00	-3.052.528,00	2.419.297,40	-2.614.122,00	2.739.715,80	1.140.977,10	1.116.427,80
610.242,47	402.691,41	2.349.044,50	418.953,91	2.739.715,80	744.352,27	772.206,06
577.062,69	-1.219.441,00	2.349.044,50	-1.084.618,00	2.739.715,80	802.991,83	772.206,06
488.396,51	-1.485.228,00	2.349.044,50	-1.259.859,00	2.739.715,80	784.362,66	772.206,06
148.527.121,00 ^a		112.758.807,00 ^a		119.238.403,00 ^a		41.975.675,00 ^a
		0,76 ^b		0,80 ^b		0,28 ^b

a: Total da coluna; b: $\frac{\sum_{k=1}^{19} \phi_k^{(i)*}}{\sum_{k=1}^{19} \lambda_k^2}$

Os métodos de correção mostraram-se bastantes distintos quando se compara as taxas de correção. O método 3 mostrou-se bastante rigoroso e apresentou as maiores taxas de correções, ou seja, antes de aplicar o teste F para os componentes. A correção descartou grande parte da $SQ_{G \times E}$ (72%), considerando essa parte que foi descartada como ruído. O método 2 foi o que se mostrou menos rigoroso e apresentou as menores taxas de correções, e este método considerou 80% da $SQ_{G \times E}$. Já o método 1 mostrou-se intermediário aos métodos 2 e 3, mas apresentou resultados próximos do método 2.

Na Tabela 3 apresenta-se a análise de cada componente pelo teste F ,

com os graus de liberdade ajustados pelo método de Gollob (1968), para os autovalores corrigidos pelo método 1 (eq. (5)). Considerando um nível de significância de 1%, tem-se que os quatros primeiros componentes são significativos, uma vez que o primeiro, o segundo, o terceiro e o quarto componentes retêm, respectivamente, 23,30%, 11,50%, 11,50% e 8,24% da $SQ_{G \times E}$. Os quatros componentes representam 54,52% da $SQ_{G \times E}$, com 192 graus de liberdade, correspondente a 30,62% dos graus de liberdade da interação, enquanto os outros 45,48% da $SQ_{G \times E}$ são supostamente ruídos presente nos dados, descartados conjuntamente pelo teste F de Gollob e pelo método de correção 1.

Tabela 3 - Análise do Teste F para os valores singulares corrigidos pelo método 1

	GL	SQ ($\phi_k^{(1)*}$)	QM	F	valor p
1	51	26.268.136,00	515.061,49	4,29	< 0,01
2	49	12.962.548,00	264.541,80	2,20	< 0,01
3	47	12.962.548,00	275.798,89	2,30	< 0,01
4	45	9.286.514,50	206.366,99	1,72	< 0,01
5	43	4.957.793,30	115.297,52	0,96	0,54
6	41	4.957.793,30	120.921,79	1,01	0,46
7	39	4.957.793,30	127.122,91	1,06	0,37
8	37	4.957.793,30	133.994,41	1,12	0,29
9	35	3.602.683,10	102.933,80	0,86	0,70
10	33	3.602.683,10	109.172,22	0,91	0,61
11	31	3.602.683,10	116.215,58	0,97	0,51
12	29	3.602.683,10	124.230,45	1,03	0,41
13	27	2.732.130,30	101.190,01	0,84	0,69
14	25	2.419.297,40	96.771,90	0,81	0,73
15	23	2.419.297,40	105.186,84	0,88	0,63
16	21	2.419.297,40	115.204,64	0,96	0,51
17	19	2.349.044,50	123.633,92	1,03	0,42
18	17	2.349.044,50	138.179,09	1,15	0,29
19	15	2.349.044,50	156.602,97	1,30	0,19

A Tabela 4 apresenta a análise de cada valor singular pelo teste F , com os graus de liberdade ajustados pelo método de Gollob (1968), para os autovalores corrigidos pelo método 2 (eq. (6)). Considerando um nível de significância de 1%, tem-se que os quatro primeiros componentes são significativos, uma vez que o primeiro, o segundo, o terceiro e o quarto componentes retêm, respectivamente, 24,13%, 11,92%, 11,92% e 8,58% da $SQ_{G \times E}$. A união dos quatros componentes representam 56,56% da $SQ_{G \times E}$,

com 192 graus de liberdade, que representam 30,62% do graus de liberdade da interação. Os outros 43,44% da $SQ_{G \times E}$ são supostamente ruídos presente nos dados, descartados pelo teste F de Gollob e pelo método de correção 2.

Na Tabela 5 apresenta-se a análise de cada valor singular pelo teste F , com os graus de liberdade ajustados pelo método de Gollob (1968), para os autovalores corrigidos pelo método 3 (eq. (7)). Considerando um nível de significância de 1%, tem-se que todos valores singulares são não significativos, assim, o método 3 e o teste F de Gollob supõe que toda interação genótipo \times ambiente é composta por ruídos. Logo, o modelo meramente aditivo seria o melhor para analisar os dados.

Tabela 4 - Análise do Teste F para os valores singulares corrigidos pelo método 2

	GL	SQ ($\phi_k^{(2)*}$)	QM	F	valor p
1	51	27.207.272,00	533.475,92	4,44	< 0,01
2	49	13.443.869,00	274.364,67	2,29	< 0,01
3	47	13.443.869,00	286.039,77	2,38	< 0,01
4	45	9.678.314,90	215.073,66	1,79	< 0,01
5	43	5.179.251,30	120.447,70	1,00	0,46
6	41	5.179.251,30	126.323,20	1,05	0,38
7	39	5.179.251,30	132.801,32	1,11	0,30
8	37	5.179.251,30	139.979,76	1,17	0,22
9	35	3.831.408,40	109.468,81	0,91	0,61
10	33	3.831.408,40	116.103,28	0,97	0,52
11	31	3.831.408,40	123.593,82	1,03	0,42
12	29	3.831.408,40	132.117,53	1,10	0,32
13	27	2.984.144,90	110.523,89	0,92	0,58
14	25	2.739.715,80	109.588,63	0,91	0,58
15	23	2.739.715,80	119.118,08	0,99	0,47
16	21	2.739.715,80	130.462,66	1,09	0,35
17	19	2.739.715,80	144.195,57	1,20	0,24
18	17	2.739.715,80	161.159,75	1,34	0,15
19	15	2.739.715,80	182.647,72	1,52	0,08

Nota-se que, quando aplicou o método de Gollob (1968) com os métodos de correção dos autovalores, houve uma redução no número de componentes retidos no modelo.

A medida $RMSPD_{PRESS}$ reflete os desvios das predições do modelo AMMI para cada combinação de genótipos e ambientes, assim, a preferência deve recair sobre o modelo cuja diferença preditiva seja menor. Pelos valores da $RMSPD_{PRESS}$ obtidos para cada método de correção, verifica-se que a

Tabela 5 - Análise do Teste F para os valores singulares corrigidos pelo método 3

	GL	SQ ($\phi_k^{(3)*}$)	QM	F	valor p
1	51	6.286.366,50	123.262,09	1,03	0,42
2	49	4.168.439,50	85.070,19	0,71	0,93
3	47	168.439,50	3.583,82	0,03	1,00
4	45	383.543,80	8.523,20	0,07	1,00
5	43	512.930,80	11.928,62	0,10	1,00
6	41	330.966,40	8.072,35	0,07	1,00
7	39	330.966,40	8.486,32	0,07	1,00
8	37	330.966,40	8.945,04	0,07	1,00
9	35	876.806,40	25.051,61	0,21	1,00
10	33	876.806,40	26.569,89	0,22	1,00
11	31	876.806,40	28.284,08	0,24	1,00
12	29	667.989,70	23.034,13	0,19	1,00
13	27	420.000,30	15.555,57	0,13	1,00
14	25	195.173,10	7.806,92	0,07	1,00
15	23	116.427,80	5.062,08	0,04	1,00
16	21	116.427,80	5.544,18	0,05	1,00
17	19	72.206,06	3.800,32	0,03	1,00
18	17	72.206,06	4.247,42	0,04	1,00
19	15	72.206,06	4.813,74	0,04	1,00

menor medida $RMSPD_{PRESS}$ foi obtida para o modelo ajustado para os autovalores corrigidos pelo método 2, mas ficou muito próximo da medida obtida quando se utilizou o método 1 para corrigir os mesmos autovalores. Ao utilizar o método 3 e ajustar o modelo AMMI, verificou-se o maior valor para a medida $RMSPD_{PRESS}$. Nota-se, então, que existe uma tendência de relação direta entre o valor da medida $RMSPD_{PRESS}$ e a rigorosidade dos métodos de correção para autovalores, ou seja, quanto mais rigoroso for o método de correção, maior será a medida $RMSPD_{PRESS}$ (Tabela 6).

Tabela 6 - $RMSPD_{PRESS}$ e R_{AMMI} para o melhor modelo AMMI selecionado após a correção dos autovalores pelos métodos 1, 2 e 3

	$RMSPD_{PRESS}$	R_{AMMI}
método 1	409,16	2,53
método 2	408,06	2,58
método 3	555,38	0,64

De acordo com interpretação dada por Nachit et al. (1992), a medida R_{AMMI} indica uma aproximação para o número de repetições que falta para o modelo AMMI completo apresentar uma *performance* igual ao modelo AMMI selecionado, ou seja, a medida R_{AMMI} indica o número de repetições que se ganha ao analisar os dados com o modelo selecionado. Na Tabela 6 também encontra-se o ganho em termos do número de repetições ao se fazer a correção dos autovalores dos modelos AMMI. Verificou-se, assim, que os métodos 1 e 2 mostraram-se melhores, fornecendo um benefício de repetições próximo de três. Já o método de correção 3 é o que apresentou o menor ganho em número de repetições, aproximadamente uma repetição. Verifica-se, também, que existe uma tendência de relação inversa entre a medida R_{AMMI} e $RMSPD_{PRESS}$, ou seja, quanto menor foi o valor da $RMSPD_{PRESS}$, maior foi o ganho em número de repetições. Logo, esta mesma interpretação vale para a rigorosidade do método de correção, quanto menos rigoroso foi o método, maior foi o ganho quanto ao número de repetições.

Conclusões

Os resultados obtidos nessa pesquisa permitem extrair as seguintes conclusões:

a) A regressão isotônica juntamente com o algoritmo de Stein foi suficiente para transformar as correções negativas em positivas e para a ordenação dos autovalores corrigidos.

b) Os modelos AMMI selecionados são mais parcimoniosos, quanto ao número de componentes retidos, quando se utiliza qualquer um dos métodos de correção.

c) O método 2 apresenta a menor taxa de correção e o método 3, a maior taxa de correção do total da soma de quadrados da interação genótipo \times ambiente.

d) Os menores valores de $RMSPD_{PRESS}$ foram obtidos quando se utilizou o método de correção 2, mas estes ficaram muito próximos dos valores obtidos quando se utilizou o método de correção 1. Já o método de correção 3 apresentou os maiores valores de $RMSPD_{PRESS}$.

e) O método 2 também se mostrou melhor quando o interesse foi verificar o ganho em número de repetições, uma vez que este benefício esteve sempre próximo de três repetições. Já o método 3 é o que apresenta um menor ganho em número de repetições, aproximadamente uma repetição.

Agradecimentos

A Capes pela apoio financeiro.

ARAÚJO, L. B.; DIAS, C. T. S. Eigenvalue correction and isotonic regression methods in AMMI models. *Rev. Mat. Estat.*, São Paulo, v.24, n.2, p.67-85, 2006.

- **ABSTRACT:** *In agricultural research it is common to analyze groups of experiments. There are several methods of analysis and interpretation for the genotype \times environment interaction from a group of genotypes tested in several environments. These methods include AMMI models, which an ANOVA for the main effects and assume multiplicative effects for interaction using DVS. The overestimation and underestimation problem with eigenvalues estimated in the conventional way is found in the specialized literature. To correct these problems three methods are presented, but it is observed that they do not always maintain the order of values. The authors then suggest the use of isotonic regression. The results indicated that: isotonic regression was sufficient to order the values corrected by the methods; there was a reduction in the number of significant components to be kept in the AMMI models; method 2 and method 3 had, respectively, the smallest and the largest rates of correction to the sum of squares of the genotype \times environment interaction; measure $RMSPD_{PRESS}$ was smallest when method of correction 2 was used. Method of correction 3 had the largest values for $RMSPD_{PRESS}$; method 2 was also better when the interest was to verify the gain in number of replicates that lack for the full AMMI model to present a similar performance to that of the AMMI model selected. This benefit was always close to three replicates.*
- **KEYWORDS:** *Correction of eigenvalues; isotonic regression; AMMI models; genotype \times environment interaction; efficiency replicates.*

Referências

ALLEN, D. M. Mean squares error of prediction as a criterion for selecting variables. *Technometrics*, Washington, v.13, p.469-475, 1971.

ARAÚJO, L. B.; DIAS, C. T. S. Viés e distribuição de autovalores de matrizes de variâncias e covariâncias nos modelos AMMI. In: ESCOLA DE MODELOS DE REGRESSÃO, 9, 2005, São Pedro. *Anais...*, São Pedro: ABE, 2005. p.60.

- CORNELIUS, P. L.; CROSSA, J. Prediction assessment of shrinkage estimators of multiplicative models for multi-environment cultivar trials. *Crop Sci.*, Madison, v.39, p.998-1009, 1999.
- 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. *Annals...* Manhattan: Milliken, 1993. p.156-166.
- CORNELIUS, P. L.; SEYEDSADR, M.; CROSSA, J. Using the shifted multiplicative model to search for "separability" in crop cultivar trials. *Theor. Appl. Genet.*, New York, v.84, p.161-172, 1992.
- DIAS, C. T. S.; KRZANOWSKI, W. J. Model selection and cross validation in additive main effect and multiplicative interaction models. *Crop Sci.*, Madison, v.43, p.865-873, 2003.
- DUARTE, J. B.; VENCOVSKY, R. *Interação genótipo × ambiente: uma introdução à análise "AMMI"*. Ribeirão Preto: Sociedade Brasileira de Genética, 1999. 60p. (Série Monografias, 9).
- EBERHART, S. A.; RUSSELL, W. A. Stability parameters for comparing varieties. *Crop Sci.*, Madison, v.6, p.36-40, 1966.
- GAUCH JR, H. G. *Statistical analysis of regional yield trials: AMMI analysis of factorial designs*. Amsterdam: Elsevier, 1992. 277p.
- 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.
- HILL, R. R.; ROSENBERGER, J. L. Methods for combining data from germplasm evaluation trials. *Crop Sci.*, Madison, v.25, p.467-470, 1985.
- 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. cap.26, p.411-429.
- MANDEL, J. Non-additivity in two-way analysis of variance. *J. Am. Stat. Assoc.*, New York, v.56, p.878-888, 1961.
- MANDEL, J. The partitioning of interactions in analysis of variance. *J. Res. Nat. Bureau Stand., Ser. B*, Washington, v.73, p.309-328, 1969.
- MANDEL, J. A new analysis of variance model for non-additive data. *Technometrics*, Washington, v.13, n.1, p.1-18, 1971.

MUIRHEAD, R. J. Developments in eigenvalue estimation. In: GUPTA, A. K. *Advances in multivariate statistical*. Boston: Reidel, 1987. p.277-288.

NACHIT, M. M. et al. Use of AMMI and linear regression models to analyze genotype-environment interaction in durum weheat. *Theor. Appl. Genet.*, 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 interactions (AMMI) analysis. *Theor. Appl. Genet.*, New York, v.89, p.647-654, 1994.

ROBERTSON, T.; WRIGHT, F. T.; DYKSTRA, R. L. *Order restricted statistical inference*. New York: Wiley, 1988. 521p.

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

SILVA, J. G. C.; BARRETO, J. N. Aplicação de regressão liner segmentada em estudos da interação genótipo \times ambiente. In: SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA, 1., 1985, Campinas. *Resumo...* Campinas: Fundação Cargill, 1985. p.49-50.

SNEE, R. D. Nonadditivity in a two-way classification: is it interaction or nonhomogeneous variance? *J. Am. Stat. Assoc.*, New York, v.77, p.515-519, 1982.

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

Recebido em 29.11.2005.

Aprovado após revisão em 20.06.2006.