

INTERAÇÃO GENÓTIPOS X AMBIENTES

UMA INTRODUÇÃO À ANÁLISE “AMMI”

*João Batista Duarte
Roland Vencovsky*

(Série Monografias, Nº 9)
Sociedade Brasileira de Genética
Ribeirão Preto, SP – Brasil
1999

Apresentação

O tratamento estatístico dado à interação de genótipos com ambientes (GxE), nesta última década, tem sofrido uma forte reorientação. Seja pelos grandes avanços obtidos nos procedimentos estatístico-computacionais, seja pela pouca eficácia dos métodos usuais de análise da interação. Esta mudança de rumos, nítida em publicações recentes, pode ser percebida numa rápida consulta ao livro *Genotype by Environment Interaction*, editado em 1996, por Manjit S. Kang (Univ. Louisiana, EUA) e Hugh G. Gauch, Jr (Univ. Cornell, EUA). Nos seus quatorze capítulos, escritos por autoridades internacionais no assunto, o enfoque tradicional univariado é substituído por uma abordagem que combina técnicas univariadas e multivariadas. Neste sentido, entre os métodos já disponíveis, a análise *AMMI* (*Additive Main effects and Multiplicative Interaction analysis*) ocupa lugar de destaque pelo seu maior volume de aplicações.

Dada a escassez completa de publicações nacionais com esta orientação, os autores entenderam que seria útil a elaboração e divulgação deste texto. Assim, apresenta-se aqui uma introdução à análise *AMMI*, aplicada ao estudo da interação GxE. Procurou-se, inicialmente, situar as limitações do enfoque tradicional, de maneira a justificar a adoção dessa nova abordagem. Buscou-se simplicidade e objetividade no tratamento do tema, incluindo exemplos ilustrativos, de maneira a facilitar a iniciação daqueles ainda não familiarizados com o assunto. Dessa forma, acredita-se que a publicação possa ser útil para professores, pesquisadores e estudantes que estejam interessados no entendimento e aplicação dessa técnica de grande potencialidade.

Os autores desejam externar seus agradecimentos à Universidade Federal de Goiás, à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e aos Departamentos de Genética e de Matemática e Estatística da Escola Superior de Agricultura “Luiz de Queiroz”, da Universidade de São Paulo, pelo apoio durante a realização deste trabalho. Devemos também um agradecimento especial ao professor Dr. Carlos Tadeu dos Santos Dias (DME-ESALQ/USP), pelo incentivo e assessoria na preparação do programa computacional de análise. É preciso ressaltar ainda que a publicação deste texto só foi possível graças à colaboração decisiva da Sociedade Brasileira de Genética.

João Batista Duarte
Roland Vencovsky

Sumário

1. INTRODUÇÃO	4
2. A INTERAÇÃO $G \times E$ E O ENFOQUE ANALÍTICO TRADICIONAL ...	7
3. O MÉTODO “AMMI”	10
3.1. O modelo matemático	10
3.2. A análise AMMI	11
3.2.1. Considerações iniciais	11
3.2.2. Ajuste dos efeitos principais por ANOVA	13
3.2.3. Ajuste da interação por DVS (ou por ACP)	14
3.2.4. A seleção do modelo AMMI	17
3.2.5. Predição de respostas genotípicas por ambiente	23
3.3. A representação gráfica em <i>biplot</i>	24
3.3.1. Conceituação e origem do <i>biplot</i> AMMI	25
3.3.2. Os tipos mais usados de <i>biplot</i> AMMI	26
3.3.3. Interpretando um <i>biplot</i> AMMI	28
3.3.4. O <i>biplot</i> para diagnose do modelo de análise	32
4. EXEMPLO NUMÉRICO	33
5. CONSIDERAÇÕES FINAIS	43
6. REFERÊNCIAS BIBLIOGRÁFICAS	45
7. APÊNDICES	49
7.1. Decomposição por valores singulares (DVS)	50
7.2. Associação entre a DVS e a representação <i>biplot</i>	54
7.3. Um programa SAS® para análise AMMI	57

INTERAÇÃO GENÓTIPOS X AMBIENTES

UMA INTRODUÇÃO À ANÁLISE “AMMI”

*João Batista Duarte*¹

*Roland Vencovsky*²

1. INTRODUÇÃO

A interação de genótipos com ambientes (*GxE*) representa uma das principais dificuldades encontradas pelo melhorista durante sua atividade seletiva. Nas etapas preliminares deste processo (com avaliações normalmente em uma só localidade), a interação *GxE* pode inflacionar as estimativas da variância genética, resultando em superestimativas dos ganhos genéticos esperados com a seleção (ganhos reais inferiores aos previstos). Nas fases finais, via de regra, os ensaios são conduzidos em vários ambientes (locais, anos e/ou épocas), o que possibilita o isolamento daquele componente da variabilidade; muito embora, neste momento a intensidade de seleção seja baixa, o que já minimizaria seus efeitos sobre previsões de ganho genético. Por outro lado, a presença dessa interação, na maioria das vezes, faz com que os melhores genótipos num determinado local não o sejam em outros. Isto dificulta a recomendação de genótipos (cultivares) para toda a população de ambientes amostrada pelos testes. Estatisticamente isso decorre da impossibilidade de interpretar, de forma aditiva, os efeitos principais de genótipos e ambientes (Kang & Magari, 1996; Mandel, 1971). Neste texto, o enfoque recairá sobre implicações desta interação em etapas finais de um programa de seleção.

Paradoxalmente, se não houvesse interação *GxE*, uma variedade adaptaria-se à maioria dos ambientes de cultivo (locais, anos e estações), de forma que um único ensaio, conduzido num só local (ainda com repetições), seria suficiente para fornecer resultados universais. Contudo, na realidade, variedades de ampla adaptação normalmente apresentam rendimentos sub-ótimos, o que desperta para a possibilidade de explorar vantajosamente os efeitos dessa interação. Assim, embora os melhoristas tendam a desenvolver uma atitude negativa contra a interação *GxE* (barreira às elevadas herdabilidades e ganhos com seleção), deve ser lembrado que interações positivas, associadas com características previsíveis do ambiente, oferecem a oportunidade dos rendimentos mais elevados. Logo, interação não é apenas um problema, mas também uma oportunidade a ser aproveitada. Adaptações específicas de genótipos a ambientes podem fazer a diferença entre uma boa e uma excelente variedade (Gauch & Zobel, 1996). Mas, para que seja possível explorar estes efeitos positivos é preciso dispor de métodos estatísticos capazes de captar esse tipo de informação.

^{1/} - Professor adjunto na Escola de Agronomia e Engenharia de Alimentos, Universidade Federal de Goiás. Caixa Postal 131, Campus II / UFG, CEP 74001-970 Goiânia, GO, Brasil. *E-mail*: jbduarte@agro.ufg.br.

^{2/} - Professor titular do Departamento de Genética, Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo. Piracicaba, SP, Brasil.

Embora várias metodologias tenham sido propostas ao longo dos últimos trinta anos, visando mitigar os desafios impostos pela interação $G \times E$, os pesquisadores atuais continuam buscando instrumentos estatísticos ainda mais efetivos nesta tarefa. Entre os métodos tradicionais, alguns permitem estratificar a população de ambientes em sub-regiões, dentro das quais a interação não se mostre significativa (*zoneamento agrônomo*), permitindo, assim, recomendações regionalizadas (Brasil, 1990; Duarte & Zimmermann, 1991). Outros métodos procuram identificar os genótipos que menos contribuem para a interação (genótipos estáveis) e que poderiam ser recomendados para toda a população de ambientes, desde que mostrem também uma performance média desejável. Nesta última categoria são de maior uso os métodos de regressão linear simples e múltipla (Finlay & Wilkinson, 1963; Eberhart & Russell, 1966; Silva & Barreto, 1985; Cruz *et al.*, 1989; Brasil, 1990; entre outros).

Apesar do emprego generalizado, os métodos baseados em regressão possuem limitações que têm sido freqüentemente relatadas na literatura. Crossa (1990) argumenta que a análise de regressão linear não é informativa se a linearidade falha, é altamente dependente do grupo de genótipos e ambientes incluídos, e tende a simplificar modelos de resposta, explicando a variação devida à interação em uma única dimensão, quando na realidade ela pode ser bastante complexa. Acrescenta-se aqui o fato de esses procedimentos, em geral, não informarem sobre interações específicas de genótipos com ambientes (se positivas ou negativas), dificultando explorar vantajosamente os efeitos da interação. Crossa (1990) sugere que a aplicação de métodos multivariados pode ser útil para melhor explorar as informações contidas nos dados. Recomenda, então, técnicas como a *análise de componentes principais* (ACP), a *análise de agrupamentos* e o procedimento *AMMI* (*additive main effects and multiplicative interaction analysis*), que vem ganhando grande aplicabilidade nos últimos anos.

Zobel *et al.* (1988) também reforçam a tese de que procedimentos univariados como a *análise de variância* (ANOVA) tenham limitações em detectar interação de fatores; mesmo em situações de elevada magnitude em termos de soma de quadrados (SQ). Exemplificam casos em que a $SQ_{G \times E}$ atinge 20 a 50% da SQ_{TOTAL} (de tratamentos: combinações de genótipos e ambientes) e, mesmo assim, o quadrado médio da interação $G \times E$ não atinge significância estatística (5% de probabilidade). Acrescentam que a regressão linear também explica, na maioria das vezes, apenas uma pequena porção da $SQ_{G \times E}$. Ademais, enquanto o método *AMMI* captou 71% desta soma de quadrados, a análise pela regressão linear de Finlay & Wilkinson (1963) só o fez para 7,9%. Concluíram, então, que só uma análise apropriada permitiria captar *padrões* agrônomo e estatisticamente importantes presentes na interação $G \times E$, e que o método *AMMI* mostrou-se bastante promissor neste sentido.

Os trabalhos precursores da análise *AMMI* são tributados, conforme citações de Gauch & Zobel (1996) e De Cauwer & Ortiz (1998), especialmente a Fisher & Mackenzie (1923), Gilbert (1963), Gollob (1968) e Mandel (1969; 1971). Entre as denominações iniciais recebidas pela metodologia estão: *FANOVA*, *MI*, *ACP* duplamente centrada e análise “biplot”. Esta última sendo mais inadequada,

tendo em vista que *biplot* refere-se apenas a um tipo de gráfico e não a um modelo particular. Aplicações *AMMI* em agricultura foram estimuladas, um pouco mais tarde, por Kempton (1984), Gauch (1985; 1986) e Zobel *et al.* (1988), na modelagem e busca de um melhor entendimento das interações *GxE* (Gauch & Zobel, 1996). Gauch (1988) introduziu outra motivação ao aplicá-la na análise de dados de experimentos, para garantir maior precisão e eficiência à análise. Desde então, o uso da análise *AMMI* só tem crescido a cada dia. A partir de 1994 tem-se publicado mais sobre aplicações desse tipo de análise na agricultura, do que em todos os anos anteriores (Gauch & Zobel, 1996).

A análise *AMMI* combina, num único modelo, componentes aditivos para os efeitos principais (genótipos e ambientes) e componentes multiplicativos para os efeitos da interação *GxE*. Gauch & Zobel (1996) informam que esta análise pode ajudar tanto na identificação de genótipos de alta produtividade e largamente adaptados, como na realização do chamado *zoneamento agrônomo*, com fins de recomendação regionalizada e seleção de locais de teste. Zobel *et al.* (1988) sustentam que o método *AMMI* permite uma análise mais detalhada da interação *GxE*, garante a seleção de genótipos mais produtivos (capazes de capitalizar interações positivas com ambientes), propicia estimativas mais precisas das respostas genotípicas e possibilita uma fácil interpretação gráfica dos resultados da análise estatística.

Este texto tem o objetivo de apresentar os princípios da análise *AMMI*, como uma alternativa para a interpretação e melhor compreensão do fenômeno da *interação de fatores*, com ênfase na interação *GxE*. Como será visto adiante, o procedimento modela efeitos principais e interação, seqüencialmente, combinando *análise de variância* (técnica univariada) e *análise de componentes principais* (técnica multivariada), respectivamente. Neste tipo de aplicação, seguindo a abordagem aqui adotada, a *ACP* pode ser tratada diretamente através da técnica matemática denominada *decomposição por valores singulares (DVS)* (Eckart & Young, 1936; Mandel, 1971), aplicada à matriz de interações. Dessa forma, como salientam Zobel *et al.* (1988), o modelo incorpora componentes aditivos e multiplicativos em uma análise de quadrados mínimos integrada e poderosa.

2. A INTERAÇÃO $G \times E$ E O ENFOQUE ANALÍTICO TRADICIONAL

Para ilustrar o significado e as implicações da interação $G \times E$, será admitido, inicialmente, que um conjunto de g genótipos tenha sido testado experimentalmente em a ambientes (locais, anos, etc). Assim, as médias de cada combinação de genótipos e ambientes (tratamentos), obtidas a partir de r repetições por experimento (um conjunto balanceado de dados), podem ser representadas pela tabela ou matriz de dados ($Y_{(g \times a)}$) a seguir:

$gen_i \backslash amb_j$	1	2	...	a	
1	Y_{11}	Y_{12}	...	Y_{1a}	= $Y_{(g \times a)}$
2	Y_{21}	Y_{22}	...	Y_{2a}	
...	
g	Y_{g1}	Y_{g2}	...	Y_{ga}	

O modelo mais simples e comum para a análise estatística de um conjunto de dados, com essa estrutura, é dado por (análise conjunta de variância para grupos de experimentos):

$$Y_{ij} = \mu + g_i + e_j + (ge)_{ij} + \varepsilon_{ij} \quad (1)$$

em que:

Y_{ij} : é a resposta média do genótipo i no ambiente j ;

μ : é uma constante comum às respostas (normalmente a média geral);

g_i : é o efeito fixo do genótipo i ($i = 1, 2, \dots, g$);

e_j : é o efeito fixo do ambiente j ($j = 1, 2, \dots, a$);

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

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

O quadro de ANOVA correspondente tem, portanto, a estrutura mostrada a seguir (assumindo um delineamento experimental inteiramente casualizado):

Fontes de Variação	Graus de Liberdade (GL)	
Genótipos (G)	$g-1$	} $(ga-1) \Rightarrow$ TRATAMENTOS
Ambientes (E)	$a-1$	
Interação $G \times E$	$(g-1)(a-1)$	
Erro médio/ r	$ga(r-1)$	
Total	$gar-1$	

A análise efetuada com médias exige cálculo à parte da estimativa da variância residual média ($QM_{\text{Erro médio}}$), assumida comum a toda a série de experimentos. Isto é feito obtendo-se a média ponderada dos QM 's do *Erro*

(obtidos nas ANOVA's individuais de experimentos), homogêneos, usando seus respectivos graus de liberdade como pesos; ou seja:

$$QM_{\text{Erro médio}} = (\sum_j SQ_{\text{Erro}_j}) / (\sum_j GL_{\text{Erro}_j}) ; \text{ onde: } j = 1, 2, \dots, a \text{ ambientes (experimentos).}$$

Gauch & Zobel (1996) apontam que, em geral, os componentes G , E e $G \times E$ representam, proporcionalmente, cerca de 10:70:20 da SQ_{TRAT} (relativa à matriz Y). Assim, um estudo mais pormenorizado da fonte de variação $G \times E$, que em certos casos chega a atingir 60% daquela soma de quadrados, é justificável mesmo em situações em que um teste F aponta $QM_{G \times E}$ não significativo.

A implicação estatística básica para uma interação de fatores significativa, está diretamente relacionada à impossibilidade de uma interpretação aditiva para os efeitos principais (Mandel, 1971; Kang & Magari, 1996). Isto pode resultar simplesmente da desigualdade das variâncias nas repostas genóticas de um ambiente para outro, sem modificação no posicionamento relativo dos genótipos (interação simples), até a completa falta de correlação nos comportamentos genotípicos entre os ambientes (interação complexa ou cruzada). Neste último caso, a presença da interação determina uma inconsistência na ordenação (*ranking*) dos genótipos de um ambiente para outro, como ilustra a Figura 1.

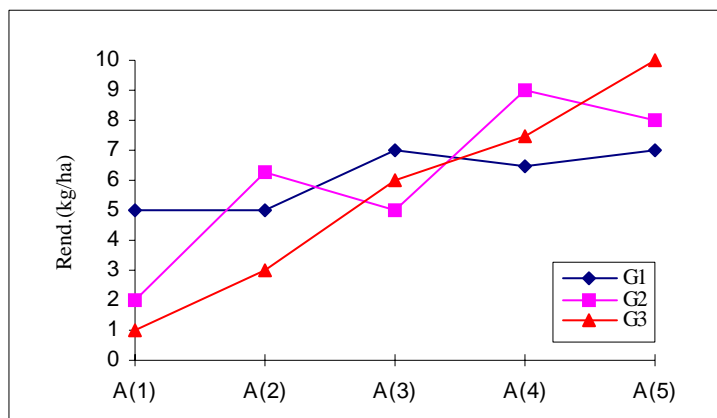


Figura 1. Comportamento das médias de três genótipos em cinco ambientes.

Diante desse quadro, vários métodos estatísticos têm sido empregados buscando um melhor entendimento dos efeitos da interação $G \times E$. Entre eles estão aqueles que se baseiam no ajuste de uma regressão linear das respostas genotípicas em cada ambiente, sobre um índice da qualidade ambiental I_j (Figura 2). Normalmente toma-se como índice a média do ambiente para o caráter em estudo ($I_j = \bar{Y}_{.j} = \sum_i Y_{ij} / g$), ou simplesmente o desvio de ambiente em relação à média geral ($I_j = \bar{Y}_{.j} - \bar{Y}_{..}$).

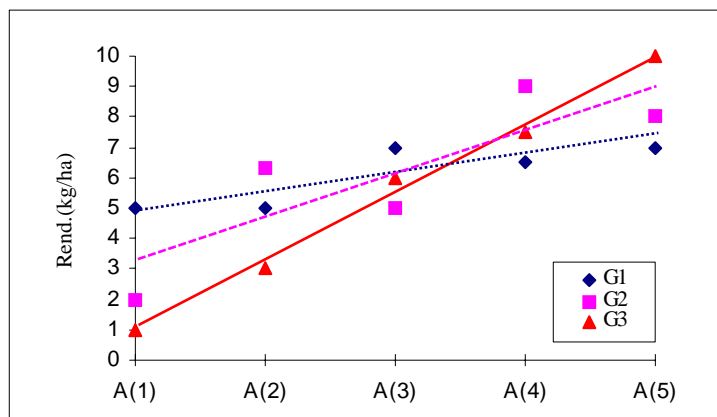


Figura 2. Respostas lineares ajustados para três genótipos em cinco ambientes.

Críticas têm sido feitas com relação à pressuposição de independência da variável explanatória (o índice ambiental é obtido a partir dos próprios dados). Contudo, isso não parece relevante quando o número de genótipos é elevado. Neste momento, é merecedor de maior destaque o fato de que, para se ajustar regressões individuais dos genótipos às variações ambientais, o desdo-bramento tradicionalmente feito não toma apenas a variação devida à interação $G \times E$, mas aquela decorrente das fontes de variação: *Ambientes + Interação $G \times E$* (efeitos ambientais dentro de genótipos). Esse procedimento usual, sem dúvida, auxilia no processo de recomendação de cultivares pois fornece uma descrição do padrão médio de resposta de cada genótipo frente à melhoria da qualidade ambiental. No entanto, obscurece a identificação de relações importantes no discernimento das verdadeiras causas da interação $G \times E$ em si.

Embora existam outros enfoques no estudo de interações $G \times E$, como por exemplo os propostos por Wricke (1965), Shukla (1972) e Magari & Kang (1997), que procuram avaliar a contribuição de cada genótipo para a interação propriamente dita, este tipo de procedimento, em geral, é preterido em relação aos métodos de regressão. Assim, a crítica feita anteriormente recai, especialmente, sobre estes últimos, que têm tido aplicação muito maior. Por outro lado, a pressuposição de que toda a $SQ_{G \times E}$ é importante para descrever o padrão de resposta diferencial dos genótipos aos ambientes, assumida nos métodos de Wricke (1965) e Shukla (1972), torna-os também sujeitos a questionamentos. A abordagem estatística sob a qual se assenta a análise *AMMI*, por exemplo, procura descartar um resíduo adicional presente nesta soma de quadrados.

3. O MÉTODO “AMMI”

3.1. O modelo matemático

Como comentado anteriormente, a análise AMMI combina, num único modelo, componentes aditivos para os efeitos principais de genótipos (g_i) e ambientes (e_j), e componentes multiplicativos para o efeito da interação $(ge)_{ij}$. Assim, a resposta média de um genótipo i num ambiente j é dada por:

$$Y_{ij} = \mu + g_i + e_j + \sum_{k=1}^n \lambda_k \gamma_{ik} \alpha_{jk} + \rho_{ij} + \varepsilon_{ij} ; \quad (2)$$

com $(ge)_{ij}$ modelado por: $\sum_{k=1}^n \lambda_k \gamma_{ik} \alpha_{jk} + \rho_{ij}$.

Sob as restrições de identificabilidade: $\sum_i g_i = \sum_j e_j = \sum_i (ge)_{ij} = \sum_j (ge)_{ij} = 0$, além da média geral (μ) e do erro experimental médio (ε_{ij}), os demais termos do modelo resultam da chamada *decomposição por valores singulares (DVS)* da matriz de interações: $\mathbf{GE}_{(g \times a)} = [(g\hat{e})_{ij}]$. O procedimento para obter essa decomposição será tratado mais adiante. A matriz de interações é obtida como resíduo do ajuste aos efeitos principais, por ANOVA, aplicada à matriz de médias $\mathbf{Y}_{(g \times a)} = [Y_{ij}]$. Assim, λ_k é o k -ésimo valor singular de \mathbf{GE} (escalar) e, $\gamma_{k(g \times 1)}$ e $\alpha'_{k(1 \times a)}$ são os respectivos vetores singulares (vetor coluna e vetor linha) associados a λ_k (Good, 1969; Mandel, 1971; Piepho, 1995). Logo, γ_{ik} e α_{jk} são os elementos relacionados ao genótipo i e ao ambiente j dos vetores γ_k e α'_k , respectivamente. O índice k ($k = 1, 2, \dots, p$; onde: $p = \min \{g-1, a-1\}$, é o posto de \mathbf{GE}), tomado até n no somatório ($n < p$), determina uma aproximação de mínimos quadrados para a matriz \mathbf{GE} pelos n primeiros termos da DVS (Good, 1969; Gabriel, 1978); deixando-se um resíduo adicional denotado por ρ_{ij} . Para $n=p$ não se tem mais a aproximação e sim uma decomposição exata da matriz, implicando em ρ_{ij} nulo.

Para ilustrar os componentes aditivos e multiplicativos, no modelo, pode-se escrevê-lo ainda da seguinte forma:

$$Y_{ij} = \underbrace{\mu + g_i + e_j}_{\text{parte aditiva.}} + \underbrace{\sum_{k=1}^p \lambda_k \gamma_{ik} \alpha_{jk}}_{(ge)_{ij}: \text{multiplicativa.}} + \varepsilon_{ij}$$

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

com: $n < p$ e $\sum_{k=n+1}^p \lambda_k \gamma_{ik} \alpha_{jk} = \rho_{ij}$.

Sob o ponto de vista da *análise de componentes principais*, além dos termos já definidos anteriormente, tem-se ainda as seguintes correspondências:

λ_k : é a raiz quadrada do k -ésimo autovalor das matrizes $(\mathbf{GE})(\mathbf{GE})'$ e $(\mathbf{GE})'(\mathbf{GE})$ (de iguais autovalores não nulos) $\Rightarrow \lambda_k^2$ é o k -ésimo autovalor;

γ_{ik} : é o i -ésimo elemento (relacionado ao genótipo i) do k -ésimo autovetor de $(\mathbf{GE})(\mathbf{GE})'$ associado a λ_k^2 ; e

α_{jk} : é o j -ésimo elemento (relacionado ao ambiente j) do k -ésimo autovetor de $(\mathbf{GE})'(\mathbf{GE})$ associado a λ_k^2 .

Note-se que o termo $(ge)_{ij}$ (interação no modelo tradicional) é agora descrito como uma soma de p parcelas, cada uma resultante da multiplicação de λ_k , expresso na mesma unidade de Y_{ij} , por um efeito genotípico (γ_{ik}) e um efeito ambiental (α_{jk}), ambos adimensionais. O termo λ_k traz uma informação relativa à variação devida à interação $G \times E$, na k -ésima parcela. De forma que a soma das p parcelas recompõem toda a variação ($SQ_{G \times E} = \sum_{k=1}^p \lambda_k^2$), como será visto mais adiante. Os efeitos γ_{ik} e α_{jk} representam pesos para o genótipo i e para o ambiente j , naquela parcela da interação (λ_k^2).

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

Por *padrão* entende-se a lei geral de formação da matriz e que governa o fenômeno em estudo (a interação $G \times E$), aqui explicado pelos n primeiros termos da *DVS* ou eixos da *ACP*. O exame destes eixos possibilita a identificação de fatores ambientais e genéticos mais diretamente relacionados à interação. A porção *ruído* (*noise*, em inglês), presente em \mathbf{GE} mas não fortemente determinada por genótipos e ambientes (linhas e colunas da matriz), é descartada buscando melhorar a capacidade preditiva do modelo. Isso traz benefícios diretos à seleção de genótipos (Zobel *et al.*, 1988). Tais aspectos são tidos como vantagens do método *AMMI* em relação àqueles mais tradicionais. Acrescenta-se o fato de a análise *AMMI* considerar, isoladamente e de forma multiplicativa, os efeitos de interação (causa da não aditividade nos efeitos principais).

3.2. A análise AMMI

3.2.1. Considerações iniciais

O ajuste do modelo por *quadrados mínimos* é obtido, para dados balanceados, em duas etapas sequenciais (Zobel *et al.*, 1988): (i) os efeitos principais, na parte aditiva (média geral, efeitos genotípicos e ambientais), são

ajustados por uma *análise de variância* comum (*univariada*), aplicada à matriz de médias ($\mathbf{Y}_{(g \times a)}$), resultando em um resíduo de não aditividade, isto é na *interação* $G \times E$, dada por $(g\hat{e})_{ij}$; (ii) esta interação, constituinte da parte multiplicativa do modelo, é, em seguida, analisada pela *decomposição por valores singulares* (DVS) da matriz de interações ($\mathbf{GE}_{(g \times a)} = [(g\hat{e})_{ij}]$), ou por *análise de componentes principais* (ACP) como prefere referir um grande número de autores.

Embora as abordagens por DVS e por ACP sejam equivalentes, neste texto, como em Piepho (1994; 1995) e Gauch & Zobel (1996), optou-se por aquela em DVS. Neste enfoque é possível estimar simultaneamente os parâmetros multiplicativos do modelo a partir da matriz \mathbf{GE} , enquanto por ACP são necessárias duas análises, uma a partir de $(\mathbf{GE})(\mathbf{GE})'$ e outra com $(\mathbf{GE})'(\mathbf{GE})$, aumentando o trabalho e as chances de equívocos na interpretação dos resultados. Isto ocorre porque a extração separada dos componentes principais para genótipos e ambientes pode resultar em escores com sinais trocados, o que compromete seriamente a interpretação das relações adaptativas dos genótipos aos ambientes. Ademais, a análise AMMI centra-se na decomposição da soma de quadrados dos elementos da matriz \mathbf{GE} ($SQ_{G \times E}$), o que é obtido de forma direta por DVS. Por ACP, voltada mais para a identificação de variáveis complexas de alto poder discriminatório (as componentes principais), isto seria obtido indiretamente, pelas duas análises, através do desdobramento das variâncias totais associadas às *variáveis* genótipos e ambientes; ou seja: $\sum_{i=1}^g s_i^2 = \sum_{k=1}^p \lambda_k^2$ e $\sum_{j=1}^a s_j^2 = \sum_{k=1}^p \lambda_k^2$, respectivamente. Logo, faz-se necessária a aplicação de uma ACP integrada para linhas e colunas, o que é perfeitamente possível³. Apesar disso, a utilização de *softwares* estatísticos para a realização de uma ACP é, sem dúvida, de maior domínio público do que a técnica DVS. Embora a análise integrada anteriormente referida, normalmente não faça parte dos procedimentos padrões (*default*) dos *softwares* que realizam ACP.

É oportuno ressaltar que o caráter multivariado da análise AMMI difere um pouco do inerente à maioria das aplicações de técnicas multivariadas. Aqui, as medidas experimentais referem-se a um único caráter biológico (ex: produtividade de grãos), que recebe, num determinado momento, um tratamento estatístico multivariado. Na maior parte dos casos, os dados já são provenientes de vários caracteres (ex: altura da planta, produtividade de grãos, percentagem de óleo, ciclo, etc), buscando-se, por exemplo, através de técnicas como a ACP, encontrar combinações lineares destas variáveis que melhor descrevem a variabilidade dos indivíduos (ex: acessos de um banco de germoplasma). Contudo, um conjunto univariado de dados, arranjados numa matriz de genótipos por ambientes pode

^{3/} - Numa ACP dupla aplicada a uma dada matriz \mathbf{A} , para evitar erros de interpretação em decorrência de sinais invertidos nos escores de suas linhas e colunas, as duas análises devem ser realizadas de forma integrada, justificando a antiga denominação ACP duplamente centrada. Isso é feito obtendo-se os autovetores de $\mathbf{A}'\mathbf{A}$ a partir dos autovetores de $\mathbf{A}\mathbf{A}'$ (ou vice-versa), ao invés de obtê-los independentemente. Seja, então, \mathbf{u}_k o k -ésimo autovetor de $\mathbf{A}\mathbf{A}'$, associado ao autovalor λ_k^2 . O k -ésimo autovetor de $\mathbf{A}'\mathbf{A}$ é dado por: $\mathbf{v}_k = (1/\lambda_k)\mathbf{A}'\mathbf{u}_k$. No caso inverso, tendo-se \mathbf{v}_k , obtém-se \mathbf{u}_k por: $\mathbf{u}_k = (1/\lambda_k)\mathbf{A}\mathbf{v}_k$. Neste procedimento é conveniente obter primeiramente os autovetores da matriz de menor dimensão (entre $\mathbf{A}\mathbf{A}'$ e $\mathbf{A}'\mathbf{A}$), o que facilita, inclusive, encontrar os autovetores da outra matriz (aquela de maior dimensão).

também ser tratado estatisticamente da mesma forma que a *matriz de dados* (indivíduos por variáveis), própria daquelas aplicações.

3.2.2. Ajuste dos efeitos principais por ANOVA

A primeira etapa da análise *AMMI* consiste numa análise de variância comum, aplicada à matriz de médias ($\mathbf{Y}_{(g \times a)}$). O modelo subjacente à análise é:

$$Y_{ij} = \mu + g_i + e_j + \varepsilon_{ij} \quad (3)$$

em que ε_{ij} , agora, representa o resíduo da não aditividade dos efeitos principais, equivalente ao termo $(ge)_{ij}$ do modelo (1).

As soluções de *quadrados mínimos ordinários* para o sistema de equações normais correspondente, sob as restrições de identificabilidade $\sum_i g_i = \sum_j e_j = 0$, são dadas por:

$$\hat{\mu} = \bar{Y}_{..} ; \quad \hat{g}_i = \bar{Y}_{i.} - \bar{Y}_{..} ; \quad \hat{e}_j = \bar{Y}_{.j} - \bar{Y}_{..} ; \quad \text{com:}$$

$$\bar{Y}_{..} = \sum_{i,j} Y_{ij} / ga ; \quad \bar{Y}_{i.} = \sum_j Y_{ij} / a ; \quad \text{e} \quad \bar{Y}_{.j} = \sum_i Y_{ij} / g.$$

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

$$\hat{Y}_{ij} = \bar{Y}_{i.} + \bar{Y}_{.j} - \bar{Y}_{..} ; \quad \text{e}$$

$$\hat{\varepsilon}_{ij} = Y_{ij} - \hat{Y}_{ij} \Leftrightarrow (g\hat{e})_{ij} = Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..}.$$

A partir deste último resultado pode-se construir a matriz de interações mostrada a seguir ($\mathbf{GE}_{(g \times a)}$), objeto de estudo da segunda etapa da análise *AMMI*:

gen _i \ amb _j	1	2	...	a	
1	$(g\hat{e})_{11}$	$(g\hat{e})_{12}$...	$(g\hat{e})_{1a}$	= $\mathbf{GE}_{(g \times a)}$
2	$(g\hat{e})_{21}$	$(g\hat{e})_{22}$...	$(g\hat{e})_{2a}$	
...	
g	$(g\hat{e})_{g1}$	$(g\hat{e})_{g2}$...	$(g\hat{e})_{ga}$	

Outros resultados de interesse são as somas de quadrados para cada fonte de variação, conforme o modelo de análise de variância. Elas podem ser obtidas a partir das médias de cada genótipo em cada ambiente, já ajustadas para a constante μ , pelas seguintes expressões:

$$SQ_G = a \sum_i (\hat{g}_i)^2 ; \text{ com } (g-1) \text{ graus de liberdade;}$$

$$SQ_E = g \sum_j (\hat{e}_j)^2 ; \text{ com } (a-1) \text{ graus de liberdade; e}$$

$$SQ_{Res} = \sum_{i,j} \hat{e}_{ij}^2 \Leftrightarrow SQ_{G \times E} = \sum_{i,j} (g \hat{e}_{ij})^2 ; \text{ com } (g-1)(a-1) \text{ graus de liberdade.}$$

É notório que o desdobramento da soma de quadrados e dos graus de liberdade associados à interação $G \times E$ (sob ênfase neste trabalho) pode ser feito de várias formas. Sendo esta SQ , em geral, de elevada magnitude (pelo menos 20% da SQ_{TRAT}), como citam Gauch & Zobel (1996), a busca de partições mais adequadas torna-se ainda mais interessante. O objetivo final sempre é identificar algum padrão importante que, uma vez incorporado no modelo, possa trazer uma melhoria na sua capacidade preditiva (Gauch & Zobel, 1988). O que o método *AMMI* faz representa apenas uma dessas maneiras alternativas de promover esse desdobramento. A análise procura capturar *padrões* presentes na estrutura dos dados, que possam contribuir para melhor explicar a resposta diferencial dos genótipos quando cultivados em diversos ambientes.

3.2.3. Ajuste da interação por DVS (ou por ACP)

a) A decomposição por valores singulares (DVS):

A *decomposição por valores singulares* é uma técnica da álgebra matricial, de propriedades matemáticas muito fortes. Sua proposição é atribuída a Eckart & Young (1936), do que vem a sinonímia *teorema de Eckart-Young* (Souza, 1988). Embora a *DVS* represente a base teórica de aplicações como a *ACP*, uma vez que é a própria generalização da chamada *decomposição espectral* (técnica restrita a matrizes quadradas e simétricas⁴), boa parte dos livros textos de análise multivariada parece não reconhecer este fato.

Good (1969) ilustra algumas das possibilidades de uso da *DVS* em matemática e estatística. Um exemplo é a sua utilização para a obtenção de *inversas generalizadas* de matrizes, como a *inversa de Moore-Penrose*, algebricamente a mais forte delas. As inversas generalizadas têm aplicação geral em modelos lineares (lemma, 1988). Em estatística, a técnica tem sido aplicada também em análise de regressão, com ênfase para a identificação e solução de problemas de colinearidade, através do que se denomina *regressão por componentes principais* (Mandel, 1982). Em modelos de posto incompleto, a exemplo dos que ocorrem em delineamentos, o enfoque através da *DVS* permite trabalhar com a melhor solução aproximada para o sistema de equações normais, além de garantir uma abordagem rica, tanto teórica como didaticamente (Carvalho Filho, 1987).

⁴/ A decomposição espectral de uma matriz $\mathbf{A}_{(m)}$, quadrada e simétrica, é dada por: $\mathbf{A}_{(m)} = \sum_k \lambda_k^2 \mathbf{e}_k \mathbf{e}_k'$, onde λ_k^2 é o seu k -ésimo autovalor e \mathbf{e}_k é o correspondente autovetor associado a λ_k^2 (com $k=1, 2, \dots, p$; onde p é o posto de \mathbf{A}); ou ainda: $\mathbf{A}_{(m)} = \mathbf{P}_A \mathbf{\Lambda}_{(p)} \mathbf{P}_A'$, onde $\mathbf{\Lambda}_{(p)}$ é a matriz diagonal com os referidos p autovalores e \mathbf{P}_A ($m \times p$) é a matriz ortogonal com os respectivos p autovetores dispostos em suas colunas.

Por outro lado, uma aplicação imediata da *DVS* reside na construção de uma aproximação para uma matriz real qualquer, de posto p , por p parcelas (matrizes) ortogonais entre si e de posto unitário (Souza, 1988). A decomposição exata para uma matriz real $\mathbf{B}_{(l \times c)}$, por *DVS*, é dada por: $\mathbf{B} = \sum_k \lambda_k \mathbf{u}_k \mathbf{v}_k'$; em que λ_k é o k -ésimo valor singular de \mathbf{B} , com \mathbf{u}_k e \mathbf{v}_k' vetores singulares (coluna e linha, respectivamente) associados ($k=1,2, \dots, p$; com $p \leq \min\{l, c\}$). Assim, fazendo-se o somatório de um número n de termos, inferior a p , obtém-se uma aproximação de posto n para a matriz \mathbf{B} . Ademais, a matriz resultante representa a melhor aproximação de posto n para \mathbf{B} (com a menor soma de quadrados dos desvios dos elementos de \mathbf{B} para aqueles correspondentes na matriz aproximada). Daí o nome *aproximação- n de quadrados mínimos* para a matriz \mathbf{B} (Good, 1969; Gabriel, 1971; Gabriel, 1978; Souza, 1988).

Arranjando-se os p valores singulares, em ordem decrescente, numa matriz diagonal ($\mathbf{S}_{(p)}$) e os respectivos vetores singulares \mathbf{u}_k e \mathbf{v}_k' , em duas matrizes $\mathbf{U}_{(l \times p)}$ e $\mathbf{V}'_{(p \times c)}$, respectivamente, a expressão anterior pode ainda ser escrita da seguinte maneira: $\mathbf{B} = \mathbf{U} \mathbf{S} \mathbf{V}'$. Logo, \mathbf{U} terá, seqüencialmente, os p vetores \mathbf{u}_k em suas colunas e \mathbf{V}' , da mesma forma, os p vetores \mathbf{v}_k' em suas linhas⁵.

O uso da *DVS*, para obter a aproximação de uma matriz por outra de posto inferior, representa a aplicação adotada pela análise *AMMI* para compreender a estrutura subjacente aos dados da matriz de interações (\mathbf{GE}). Ou seja, \mathbf{GE} assumirá o papel da matriz \mathbf{B} anteriormente definida. Aos menos iniciados em álgebra matricial apresentam-se, ao final deste texto, no Apêndice 7.1, exemplos ilustrativos do procedimento para a obtenção da *DVS* de uma matriz, bem como algumas de suas propriedades.

b) O desdobramento da $\mathbf{SQ}_{\mathbf{GxE}}$ por *DVS*:

A seguir será ilustrada uma propriedade da *DVS* que determina a partição da $\mathbf{SQ}_{\mathbf{GxE}}$ na análise *AMMI*. Considere-se, a princípio, a decomposição exata da matriz $\mathbf{GE}_{(g \times a)}$ por *DVS* (com as correspondências ao item anterior):

$$\underbrace{\mathbf{GE}}_{\mathbf{B}} = \underbrace{\mathbf{U}}_{(\mathbf{P}_{\mathbf{B} \cdot \mathbf{B}'})} \underbrace{\mathbf{S}}_{(\mathbf{A}^{1/2})} \underbrace{\mathbf{V}'}_{(\mathbf{P}_{\mathbf{B}' \cdot \mathbf{B}})} = \sum_{k=1}^p \lambda_k \gamma_k \alpha'_k; \text{ com } p = \min \{(g-1), (a-1)\}.$$

Portanto, a *DVS* de \mathbf{GE} pode ser escrita como uma soma de p matrizes de posto unitário, o que permite ilustrar a propriedade de interesse (partição da soma de quadrados dos elementos da matriz \mathbf{GE}):

⁵/ A inversa generalizada de Moore-Penrose de uma matriz $\mathbf{B}_{(l \times c)}$ é dada por: $\mathbf{B}^+_{(l \times c)} = \mathbf{V} \mathbf{S}^{-1} \mathbf{U}'$; sendo $\mathbf{S}_{(p)}$ uma matriz diagonal, sua inversa é facilmente obtida por: $\mathbf{S}^{-1} = \text{diag}\{1/\lambda_1, 1/\lambda_2, \dots, 1/\lambda_p\}$.

$$\begin{aligned}
 \underbrace{\mathbf{GE}}_{\substack{\mathbf{B}=[b_{ij}] \\ (\mathbf{g} \times \mathbf{a})}} &= \underbrace{\lambda_1 \gamma_1 \alpha'_1}_{\substack{\mathbf{B1}=[b1_{ij}] \\ (\mathbf{g} \times \mathbf{a})}} + \underbrace{\lambda_2 \gamma_2 \alpha'_2}_{\substack{\mathbf{B2}=[b2_{ij}] \\ (\mathbf{g} \times \mathbf{a})}} + \dots + \underbrace{\lambda_p \gamma_p \alpha'_p}_{\substack{\mathbf{Bp}=[bp_{ij}] \\ (\mathbf{g} \times \mathbf{a})}} . \\
 \sum_{i,j} b_{ij}^2 &= \sum_{i,j} (g\hat{e})_{ij}^2 = \sum_{i,j} b1_{ij}^2 + \sum_{i,j} b2_{ij}^2 + \dots + \sum_{i,j} bp_{ij}^2 = SQ_{\mathbf{GxE}} \\
 \sum_k \lambda_k^2 &= \sum_{i,j} (g\hat{e})_{ij}^2 = \lambda_1^2 + \lambda_2^2 + \dots + \lambda_p^2 = \sum_k SQ_{IPCA_k} = SQ_{\mathbf{GxE}}
 \end{aligned}$$

Como \mathbf{GE} é, por construção, uma matriz de desvios centrados e, portanto, com média nula, a soma de quadrados de seus elementos reproduz exatamente a $SQ_{\mathbf{GxE}}$. Assim, essa soma de quadrados é desdobrada em componentes relativos a cada termo da DVS , ou a cada eixo de interação da ACP ($IPCA_k$, conforme notação internacional)⁶. Observa-se ainda que, para obter o desdobramento dessa soma de quadrados não é necessário encontrar as p matrizes parciais. Mas é suficiente encontrar os valores singulares de \mathbf{GE} ou, equivalentemente, os autovalores de $(\mathbf{GE})'(\mathbf{GE})$ ou de $(\mathbf{GE})(\mathbf{GE})'$. A soma de quadrados relativa ao k -ésimo termo ou eixo de interação equivale ao quadrado do valor singular correspondente (o autovalor λ_k^2): $SQ_{\mathbf{GxE}} [IPCA_k] = \lambda_k^2$. Ademais, dada as propriedades desse tipo de decomposição, as p parcelas obtidas para a $SQ_{\mathbf{GxE}}$, são ortogonais e, portanto, independentes.

Tomando-se de forma cumulativa os sucessivos termos do desdobramento, obtém-se aproximações cada vez melhores para a $SQ_{\mathbf{GxE}}$ original. Porém, sendo $\lambda_1^2 \geq \lambda_2^2 \geq \dots \geq \lambda_p^2$, é possível que alguns poucos primeiros termos já sejam suficientes para descrever uma alta proporção dessa soma de quadrados. Isso pode permitir a seleção de um modelo parcimonioso (consumindo poucos graus de liberdade) e com boa capacidade para descrever a variabilidade devida à interação original (clássica). Logo, o modelo $AMMI$ procura explicar a $SQ_{\mathbf{GxE}}$ por meio de uma aproximação de posto n para a matriz \mathbf{GE} (com n preferivelmente bem menor do que p). Ou seja, o objetivo da análise é descrever a interação \mathbf{GxE} por um número reduzido de eixos ($n=1$ a 3), resultando num modelo informativo mas que consuma poucos graus de liberdade. Com isso, espera-se que o modelo seja capaz de: (i) captar o **padrão** subjacente à interação, ou seja, a possível lei geral que governa o fenômeno; e (ii) descartar os **ruídos** presentes nos dados e sem interesse agrônômico, os quais só prejudicam a capacidade preditiva do modelo. Esta separação em **padrão** e **ruído**, pode também ser expressa em termos de soma de quadrados da seguinte forma:

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

^{6/} - A equivalência entre as abordagens por DVS e ACP pode ser vista através das matrizes: $\mathbf{\Lambda}$, diagonal com os autovalores de \mathbf{BB}' ou de $\mathbf{B}'\mathbf{B}$, $\mathbf{P}_{\mathbf{BB}'}$ e $\mathbf{P}_{\mathbf{B}'\mathbf{B}}$, matrizes ortogonais com os autovetores de \mathbf{BB}' e $\mathbf{B}'\mathbf{B}$, em suas colunas e linhas, respectivamente.

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

3.2.4. A seleção do modelo *AMMI*

a) Testes estatísticos para os componentes da $SQ_{G \times E}$:

Os critérios adotados para a definição do número de eixos a serem retidos, de maneira a explicar o *padrão* relacionado à interação, têm sido objeto de várias pesquisas (Gollob, 1968; Mandel, 1971; Gauch, 1988; Gauch & Zobel, 1988; Piepho, 1994; Piepho, 1995; entre outros). Quaisquer destes critérios, à semelhança da *ACP*, levam em consideração, pelo menos indiretamente, a proporção da $SQ_{G \times E}$ acumulada até o n -ésimo eixo ($\sum_{k=1}^n \lambda_k^2 / SQ_{G \times E}$).

Um dos procedimentos usuais consiste em determinar os graus de liberdade associados à parcela da $SQ_{G \times E}$ relacionada a cada membro da família de modelos *AMMI*. Obtém-se, então, o quadrado médio (QM) correspondente a cada parcela (ou modelo). Em seguida, é obtido um teste F avaliando-se a significância de cada componente em relação ao $QM_{\text{Erro médio}}$. Isso resulta num quadro de análise de variância semelhante ao tradicional, com desdobramento para a fonte de variação $G \times E$. Considerando tratar-se de um desdobramento ortogonal da soma de quadrados da interação, assume-se, por este procedimento, que o quociente entre o quadrado médio de um dos componentes da interação e o quadrado médio do erro segue uma distribuição F , com os correspondentes graus de liberdade (desde que garantida uma aproximação normal para a variável resposta original). Dessa forma, o ponto de parada que determina a seleção do modelo (*AMMI0*, *AMMI1*, ..., ou *AMMIF*) baseia-se na significância do teste F para os sucessivos termos da interação. O *resíduo AMMI*, reunindo os demais (últimos) termos da $SQ_{G \times E}$ (descartados), também pode ser testado de maneira a assegurar o seu caráter desprezível.

Gauch & Zobel (1996) mencionam a existência de vários sistemas para atribuir graus de liberdade a um modelo *AMMI*, entre os quais dois se popularizaram, o de Gollob (1968) e o de Mandel (1971). Os autores alertam que, infelizmente, não há boa concordância entre esses sistemas, de forma que a escolha por um deles acaba recaindo sobre considerações de natureza teórica e prática. O sistema aproximado de Gollob (1968) é muito fácil de ser aplicado,

requerendo uma aritmética trivial visto que: $GL_{IPCA_n} = g + a - 1 - 2n$. Provavelmente por isso, esse sistema tenha sido grandemente utilizado enquanto outros exigem extensas simulações.

Mandel (1971) parte de uma definição geral de *graus de liberdade* associados a cada parcela λ_k^2 que é: $GL_{IPCA_k} = E[\lambda_k^2]/\sigma^2$ (com $E[]$ denotando o operador esperança matemática e σ^2 a variância aleatória populacional). Desta expressão exata e da suposição de normalidade, são feitas simulações que resultam em três tabelas relacionadas ao número de graus de liberdade. Tais tabelas, entretanto, exigem regras especiais para consulta, não sendo também exaustivas. Razões como estas acabam reduzindo a utilização desse sistema em trabalhos de aplicação. Diferentemente de Gollob (1968), o sistema de Mandel (1971) geralmente resulta num decréscimo não linear dos graus de liberdade (para os sucessivos termos de interação), os quais podem ainda ser fracionários.

O trabalho de Mandel (1971) representa uma das melhores descrições da análise *AMMI*, embora ele a referisse apenas como *modelo para dados não aditivos*. O autor faz uma advertência importante àqueles impulsivamente tentados a aplicar o teste *F*. Para o autor, não há uma forma simples de encontrar os graus de liberdade associados a cada parcela λ_k^2 . Pois, embora os valores $\lambda_1^2, \lambda_2^2, \dots, \lambda_p^2$ sejam legítimos componentes aditivos de $\sum_{i,j} (g\hat{e})_{ij}^2$, a $SQ_{G \times E}$, eles não têm distribuição de qui-quadrado (χ^2). Isso pode, em princípio, ser explicado pelas covariâncias não nulas entre os termos da matriz **GE**, mantidas nas p matrizes parciais. Note-se que, apesar de estas matrizes serem ortogonais e, portanto, independentes entre si (o mesmo valendo em relação a $\lambda_1^2, \lambda_2^2, \dots$ e λ_p^2), os elementos dentro de cada uma não o são. Sendo λ_k^2 equivalente à soma de quadrados dos elementos de A_k (k -ésima matriz parcial), não sendo estes elementos independentes, mesmo sob normalidade, $\lambda_k^2 = \sum_{i,j} (a_k)_{ij}^2$ não terá uma distribuição χ^2 original, a menos que os graus de liberdade sofram o devido ajuste.

Por alguns anos, a atribuição de graus de liberdade seguindo a abordagem proposta por Mandel (1971) foi considerada exata e, portanto, a mais correta. Entretanto, atualmente, essa proposta tem recebido críticas consistentes que a torna menos indicada do que o próprio sistema aproximado de Gollob (1968). As razões para isto assentam-se nas pressuposições assumidas por Mandel em suas simulações (Gauch, 1992; p.114-129).

Como coloca Gauch (1992), o sistema de Gollob utiliza a definição estrutural clássica de graus de liberdade, isto é, o número de parâmetros estimados menos o número de restrições. Do modelo *AMMI* (equação 2), cada eixo de interação ($IPCA_k$) tem *um* valor singular λ_k , g elementos para o vetor γ_k e a elementos para o vetor α_k , totalizando $[g+a+1]$ parâmetros estimados. Após a remoção dos efeitos principais, os autovetores (ou vetores singulares γ_k e α_k) ficam sujeitos a 2 restrições: $\sum_i \gamma_{ik} = 0$ e $\sum_j \alpha_{jk} = 0$. Adicionalmente, escalonando-os como vetores de comprimento unitário, impõem-se outras 2 restrições: $\sum_i \gamma_{ik}^2 = 1$ e $\sum_j \alpha_{jk}^2 = 1$. Finalmente, requerendo que cada eixo $IPCA_n$, para $n > 1$, seja ortogonal a todos os $[n-1]$ eixos anteriores, impõe-se $2(n-1)$ restrições adicionais: $\sum_i \gamma_{in} \gamma_{ik} = 0$ e $\sum_j \alpha_{jn} \alpha_{jk} = 0$, para $k=1, 2, \dots, n-1$. Por isso, conforme já enunciado, o eixo $IPCA_n$ terá:

$$GL_{IPCA_n} = g + a + 1 - [2 + 2 + 2(n-1)] = g + a - 1 - 2n.$$

Para ilustrar esta definição Gauch (1992) considera uma matriz Y de 7 linhas (genótipos) e 10 colunas (ambientes), totalizando 70 graus de liberdade (GL). Estes 70 GL são distribuídos, 1 para a média geral, 6 para os desvios de genótipos, 9 para os desvios de ambientes, o que totaliza 16 GL na parte aditiva do modelo $AMMI$, restando, portanto, 54 GL para a interação de genótipos com ambientes original ou clássica. Seguindo a expressão anterior, os valores de GL para os eixos de interação (desdobramento dos 54 GL) ficam atribuídos da seguinte forma: 14 para $IPCA_1$, 12 para $IPCA_2$, 10 para $IPCA_3$, 8 para $IPCA_4$, 6 para $IPCA_5$ e 4 para $IPCA_6$ (dado que, neste caso: $p = \min\{n^o.linhas-1, n^o.colunas-1\} = 6$). Assim, se o desdobramento da interação parar, por exemplo, em 3 componentes, o padrão da interação $G \times E$ terá 36 GL (14+12+10) e o resíduo $AMMI$ terá 18 GL . Tomando-se esta situação, correspondente à escolha do modelo $AMMI3$, o autor apresenta um desenho esquemático da matriz Y para ilustrar a atribuição dos graus de liberdade segundo a definição estrutural (Figura 3). Nesta ilustração, o número de células delimitadas por um efeito (linhas contínuas) define os graus de liberdade associados àquele efeito.

média geral				médias	de	colunas			
					$IPCA_1$				
médias					$IPCA_2$				
de					$IPCA_3$				
linhas					<i>Resíduo</i>	<i>AMMI</i>			

Figura 3. Definição estrutural dos graus de liberdade da análise $AMMI$, no caso de uma matriz $Y_{(7 \times 10)}$ (Gauch, 1992).

Adotando-se, portanto, o sistema de Gollob, o quadro completo da análise conjunta de variância (obtida a partir de médias) fica estruturado da seguinte maneira (assumindo um delineamento experimental inteiramente ao acaso):

F. Variação	G.L.	S.Q.
Genótipos (<i>G</i>)	<i>g</i> -1	SQ_G
Ambientes (<i>E</i>)	<i>a</i> -1	SQ_E
Interação <i>GxE</i>	$(g-1)(a-1)$	$SQ_{G \times E}$
<i>IPCA</i> 1	$g+a-1-(2.1)$	λ_1^2
<i>IPCA</i> 2	$g+a-1-(2.2)$	λ_2^2
<i>IPCA</i> 3	$g+a-1-(2.3)$	λ_3^2
...
<i>IPCA</i> <i>p</i>	$g+a-1-(2.p)$	λ_p^2
Erro médio/ _{<i>r</i>}	$ga(r-1)$	SQ_{Erro}
Total	$gar-1$	SQ_{TOTAL}

Piepho (1995) investigou a robustez de alguns testes alternativos usados para selecionar um modelo *AMMI*, quanto às suposições de *homocedasticidade* e *normalidade* dos erros. O autor comenta que o teste *F* aplicado de acordo com o critério de Gollob (1968), é bastante liberal em selecionar mais termos multiplicativos do que o verdadeiro modelo contempla. Entre quatro alternativas estudadas, incluindo Gollob (1968), o teste F_R proposto por Cornelius *et al.* (1992) mostrou-se o mais robusto. Assim, se outro destes testes for utilizado o autor recomenda avaliações preliminares para verificar a validade das suposições.

A estatística F_R , sob a hipótese nula de que não haja mais do que *n* termos determinando a interação, tem uma distribuição *F* aproximada com f_2 graus de liberdade e $GL_{\text{Erro médio}}$. Sob esta hipótese, o numerador da expressão apresentada a seguir ($SQ_{G \times E}$ para o resíduo *AMMI*) é, aproximadamente, uma variável qui-quadrado (Piepho, 1995).

$$F_{R,n} = (SQ_{G \times E} - \sum_{k=1}^n \lambda_k^2) / (f_2 \text{QM}_{\text{Erro médio}}) ; \text{ com: } f_2 = (g - 1 - n) (a - 1 - n).$$

Assim, um resultado significativo pelo teste sugere que pelo menos um termo multiplicativo ainda deve ser adicionado aos *n* já ajustados. Logo, F_R pode ser visto como um teste para a significância dos *n+1* primeiros termos da interação (similar ao teste de falta de ajuste em regressão linear). Observa-se que para *n=0* (quando nenhum termo multiplicativo é ajustado) o teste é equivalente ao *F* para a interação *GxE* global, na ANOVA conjunta, que é um teste exato. Nota-se também que os graus de liberdade do numerador de F_R é igual aos graus de liberdade para toda a interação menos os graus de liberdade atribuídos por Gollob (1968) para os *n* primeiros termos. Conclui-se, portanto, que a aplicação de F_R é equivalente ao teste do resíduo *AMMI* para *GxE*, conforme sugerido anteriormente.

Gauch (1992) apresenta uma discussão ampla para a questão de atribuir graus de liberdade aos componentes multiplicativos de um modelo AMMI. O autor conclui que sempre que simulações rigorosas parecerem desnecessárias ou impraticáveis, a recomendação geral é que se use o sistema de Gollob, de maneira a permitir um teste F aproximado; mantendo-se em mente que o procedimento é um mero guia para a intuição. Nos casos em que parece haver uma clara divisão entre grandes componentes determinando o *padrão* e pequenos componentes ricos em *ruídos*, o autor sugere que uma atribuição igualitária de graus de liberdade ($GL_{IPCA\ k} = [(g-1)(a-1)]/p$) é especialmente interessante para os primeiros componentes; mesmo porque normalmente haverá pouco interesse em partições adicionais do resíduo (rico em *ruídos*). Acrescenta também que em determinadas questões de pesquisa, a atribuição exata dos graus de liberdade a cada termo multiplicativo pode ser uma necessidade antes que um simples preciosismo.

b) Avaliação preditiva por validação cruzada:

Em razão de dificuldades de natureza teórica ou prática, como as levantadas anteriormente, procedimentos estatísticos computacionalmente intensivos, livres de distribuições teóricas, vem ganhando aplicação crescente. Entre eles destacam-se os baseados em reamostragem como *jackknife*, *bootstrap* e *validação cruzada*. Na análise AMMI, Gauch (1988) introduziu a chamada *avaliação preditiva*, por meio de *validação cruzada* (Wold, 1978), para fins de seleção do modelo (AMMI0, AMMI1, AMMI2, ... ou AMMIF).

Gauch & Zobel (1988) comentam que a avaliação pelo teste F é “*posdictiva*”, na medida em que busca um modelo que explique uma grande parte da variação mostrada por todos os dados observados (com coeficiente de determinação elevado). Assim, argumentam que tal abordagem mostra-se menos 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). Para ilustrar os dois enfoques os autores exemplificam que, se uma SQ_{TRAT} for constituída de 70% de *padrão* e 30% de *ruídos*, a meta de uma análise estatística não deveria ser recuperar 100% desta soma de quadrados. Mas, de preferência apenas 70% e, em especial, aquela porção da variabilidade que representa o *padrão*. Recuperar padrão implica em aumentar a precisão e melhorar a qualidade das predições; por outro lado, recuperar ruídos implica em reduzir essa precisão cada vez mais. Assim, um modelo que seletivamente recupera o padrão e relega ruídos a um resíduo desconsiderado na predição de respostas, pode resultar em melhor precisão do que os próprios dados. Este é o princípio subjacente à proposta de Gauch (1988) para seleção do modelo AMMI.

Dessa forma, através da *validação cruzada*, os dados de repetições, para cada combinação de genótipos e ambientes, são divididos, por um critério

aleatório, em dois subconjuntos: (i) *dados para o ajuste do modelo AMMI*; e (ii) *dados de validação*. As respostas previstas por um determinado modelo da família *AMMI* são confrontadas com os respectivos dados de validação, calculando-se as diferenças entre esses valores. Obtém-se, então, a soma de quadrados destas diferenças, dividindo-se o resultado pelo número delas. À raiz quadrada deste resultado os autores chamaram de *diferença preditiva média (RMSPD)*. Esse procedimento deve ser repetido, segundo Crossa *et al.* (1991), por cerca de dez vezes, obtendo-se uma média dos resultados para cada membro da família de modelos. Um pequeno valor de *RMSPD* indica um bom sucesso preditivo do modelo. O modelo escolhido é, então, utilizado para analisar os dados de todas as *r* repetições, conjuntamente, numa análise definitiva. Assim, a seqüência completa da análise *AMMI* fica definida conforme o esquema ilustrado na Figura 4. Esta seqüência tem sido a mais utilizada em aplicações, nos dias atuais.

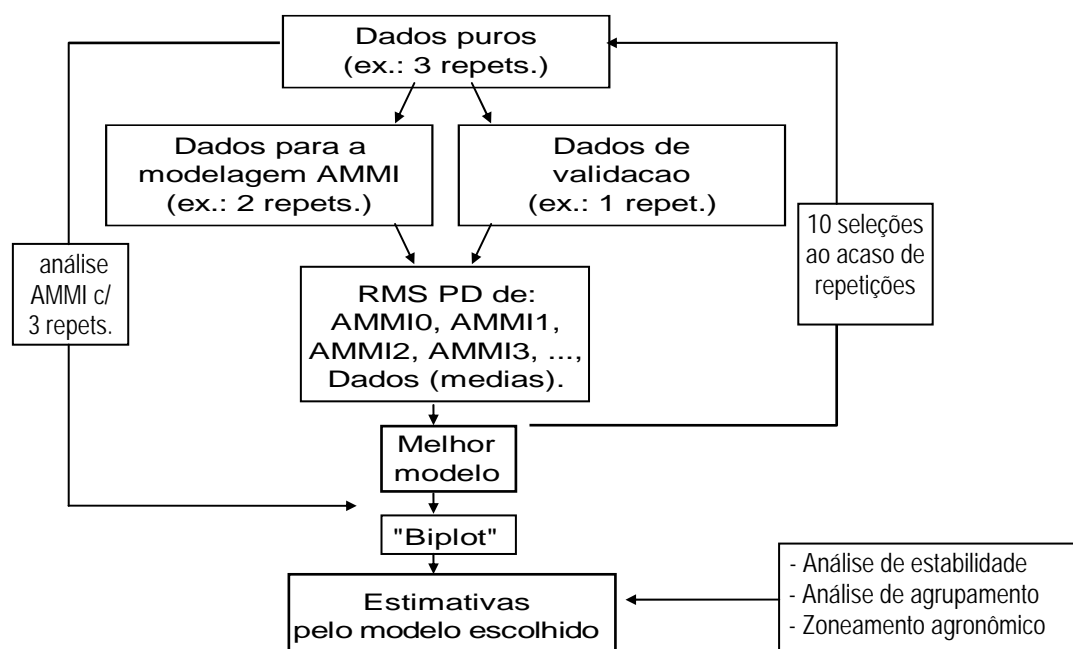


Figura 4. Seqüência da análise estatística *AMMI* com validação cruzada da família de modelos (Crossa *et al.*, 1991).

Embora o procedimento original de Gauch (1988) venha sendo amplamente utilizado, algumas modificações já têm sido propostas. Piepho (1994) sugere que a obtenção do valor médio de *RMSPD* seja obtido a partir de 1000 randomizações diferentes, ao invés de 10 como sugerem Crossa *et al.* (1991). O autor propõe ainda uma modificação para a partição aleatória dos dados (modelagem e validação) quando estes são provenientes de experimentos delineados em blocos. Neste caso, recomenda que sejam feitos sorteios de blocos inteiros dentro de experimentos e não de parcelas para cada combinação de genótipo e ambiente (tratamento). Assim, fica preservada a estrutura original de blocos. Contudo, apesar da coerência lógica desse tipo de proposta, estudos confirmando sua eficácia ainda não são suficientemente disponíveis. Em Gauch &

Zobel (1996) é sugerido ainda que se faça o conjunto de dados de validação sempre com uma só observação para cada tratamento. Acredita-se, com isso, que seja mais provável, a partir de $r-1$ dados, encontrar um modelo que mais se aproxime do ideal para analisar o conjunto completo dos r dados por tratamento.

Deve-se observar que, ao avaliar o modelo por validação cruzada, a análise *AMMI* deve partir das observações individuais propriamente ditas (dados de cada repetição dentro de experimentos). Por outro lado, se o modelo for avaliado por um teste F a análise pode ser feita a partir das médias dos genótipos nos ambientes (experimentos), desde que se disponha dos quadrados médios residuais, obtidos nas análises de variâncias de cada experimento.

3.2.5. Predição de respostas genotípicas por ambiente

Uma vez selecionado o membro da família de modelos (*AMMI0*, *AMMI1*, *AMMI2* ... ou *AMMI F*) que melhor descreve a estrutura dos dados, é possível prever a resposta de cada genótipo em cada ambiente (\hat{Y}_{ij}) com base no modelo escolhido:

$$\hat{Y}_{ij} = \bar{Y}_{i.} + \bar{Y}_{.j} - \bar{Y}_{..} + \sum_{k=1}^n \lambda_k \gamma_{ik} \alpha_{jk} \quad (k = 1, 2, \dots, p)$$

em que: n é o número de eixos ou componentes principais selecionados para descrever o *padrão* da interação $G \times E$ ($n \leq p$)

Vale lembrar que, se *AMMI0* for o modelo selecionado, nenhum eixo de interação (último termo da expressão) deve ser incluído na predição. Assim, de uma maneira mais específica a expressão anterior assume as seguintes formas, conforme o modelo escolhido:

Modelo <i>AMMI</i> selecionado	Expressão da resposta esperada do genótipo i no ambiente j (\hat{Y}_{ij})
<i>AMMI0</i>	$\hat{Y}_{ij} = \bar{Y}_{i.} + \bar{Y}_{.j} - \bar{Y}_{..}$
<i>AMMI1</i>	$\hat{Y}_{ij} = \bar{Y}_{i.} + \bar{Y}_{.j} - \bar{Y}_{..} + \lambda_1 \gamma_{i1} \alpha_{j1}$
<i>AMMI2</i>	$\hat{Y}_{ij} = \bar{Y}_{i.} + \bar{Y}_{.j} - \bar{Y}_{..} + \lambda_1 \gamma_{i1} \alpha_{j1} + \lambda_2 \gamma_{i2} \alpha_{j2}$
...	...
<i>AMMIF</i>	$\hat{Y}_{ij} = \bar{Y}_{i.} + \bar{Y}_{.j} - \bar{Y}_{..} + \lambda_1 \gamma_{i1} \alpha_{j1} + \lambda_2 \gamma_{i2} \alpha_{j2} + \dots + \lambda_p \gamma_{ip} \alpha_{jp}$

Se a seleção recair sobre *AMMI F* (modelo completo com todos os termos multiplicativos), a resposta esperada equivale à média da casela relativa ao genótipo i e no ambiente j . Ou seja, a média das r repetições do genótipo naquele ambiente, o que pode ser facilmente verificado:

$$\hat{Y}_{ij} = \bar{Y}_{i.} + \bar{Y}_{.j} - \bar{Y}_{..} + (g\hat{e})_{ij}; \quad \text{pois: } \sum_{k=1}^p \lambda_k \gamma_{ik} \alpha_{jk} = (g\hat{e})_{ij};$$

$$\text{mas: } (g\hat{e})_{ij} = Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..};$$

$$\text{logo: } \hat{Y}_{ij} = \bar{Y}_{i.} + \bar{Y}_{.j} - \bar{Y}_{..} + (Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..}) = Y_{ij}.$$

Deve ficar claro, portanto, que um modelo *AMMI* não se restringe à descrição da interação, pois sempre contempla também os efeitos principais. É possível sim não incluir termo algum de interação (*AMMI0*), quando nenhum deles se mostrar significativo. Entretanto, na maior parte das pesquisas em genética e melhoramento de plantas a interação *GxE* é significativa, de maneira que pelos menos um termo multiplicativo normalmente deve ser incluído na predição das respostas.

Estimativas dessa natureza são tidas como mais precisas do que as próprias médias obtidas das repetições de cada genótipo em cada ambiente. A razão disso, segundo Gauch (1990), está no simples fato de o método *AMMI* considerar e usar mais dados. Numa abordagem tradicional, a estimativa considera apenas o que Gauch (1992) denomina de *dados diretos*, enquanto uma estimativa *AMMI* inclui também os chamados *dados indiretos*. Por exemplo, considere-se um grupo de experimentos balanceados testando 25 genótipos em 20 ambientes, com 4 repetições por experimento. Para estimar a resposta de um certo genótipo num dado ambiente o método *AMMI* dispõe de 4 *observações diretas* e 1996 *observações indiretas*, perfazendo 2000 *observações totais*, tomadas de forma integrada. O procedimento convencional de médias de caselas o faz apenas com as 4 *observações diretas*.

Neste sentido, Gauch & Zobel (1988) mostram que uma análise *AMMI* com dados não repetidos, quando associada à *validação cruzada* para a seleção do modelo, é preditivamente mais precisa do que as médias de tratamentos obtidas de duas repetições. Esse benefício aumenta com o tamanho do ensaio (aumentando o número de genótipos e ambientes) e quanto maior o nível de perturbação (*ruído*) presente nos dados. Segundo estes autores, isto sugere que, em situações apropriadas, esse tipo de análise pode reduzir a quantidade de repetições requeridas em torno de duas a cinco vezes, permitindo elevar o número de genótipos a serem avaliados. Ademais, o uso das informações indiretas permite à análise *AMMI* que se estime também a resposta de um genótipo num ambiente em que ele não tenha sido avaliado (caso de casela vazia). Isto, entretanto, exige um tratamento estatístico mais sofisticado do que o apresentado neste texto, denominado algoritmo *EM-AMMI* (Gauch & Zobel, 1990; Gauch, 1992).

3.3. A representação gráfica em “*biplot*”

3.3.1. Conceituação e origem do *biplot* AMMI

Uma virtude do método *AMMI* está na possibilidade de representar, num único gráfico, os efeitos de interação para cada genótipo e cada ambiente. O gráfico, denominado *biplot* (Gabriel, 1971), baseia-se na aproximação *DVS* de uma matriz, por outra de posto inferior. Para permitir a construção do gráfico o posto da matriz aproximada deverá, na realidade, ser igual a *um*, *dois* ou *três*, resultando num *biplot* em uma, duas ou três dimensões, respectivamente. No Apêndice 7.2, ao final deste texto, apresentam-se os princípios deste tipo de representação, ilustrando o seu relacionamento com a técnica *DVS*. É incluído também um exemplo numérico simples com comentários gerais sobre suas propriedades, especialmente aquelas úteis à análise *AMMI*, evidenciando a riqueza de informações geradas por esse tipo de gráfico.

Na análise *AMMI* o uso do *biplot* é garantido em razão da *DVS* aplicada à matriz de interações, que será representada, na maioria das aplicações, por meio de sua aproximação de posto *um* ou *dois*. Seja, então, a seguinte aproximação *DVS* para ***GE*** (matriz de posto conhecido *p*): $\mathbf{G\hat{E}}_{(g \times a)} = \sum_{k=1}^n \lambda_k \gamma_k \alpha'_k = \tilde{\mathbf{U}} \tilde{\mathbf{S}} \tilde{\mathbf{V}}'$; onde, neste momento, $\tilde{\mathbf{U}}_{(g \times n)}$ tem em suas colunas apenas os *n* primeiros vetores $\gamma_{(g \times 1)}$; $\tilde{\mathbf{V}}'_{(n \times a)}$, tem em suas linhas os *n* primeiros vetores $\alpha'_{(1 \times a)}$; e $\tilde{\mathbf{S}}_{(n)}$ é a matriz diagonal com os primeiros valores singulares: $\lambda_1, \lambda_2, \dots$ e λ_n (com $n < p$). A expressão pode também ser escrita como: $\mathbf{G\hat{E}}_{(g \times a)} = (\tilde{\mathbf{U}} \tilde{\mathbf{S}}^{1/2})(\tilde{\mathbf{S}}^{1/2} \tilde{\mathbf{V}}')$ ou ainda por: $\mathbf{G\hat{E}}_{(g \times a)} = (\tilde{\mathbf{G}})(\tilde{\mathbf{H}}')$; bastando para isto fazer: $\tilde{\mathbf{G}}_{(g \times n)} = (\tilde{\mathbf{U}} \tilde{\mathbf{S}}^{1/2})$ e $\tilde{\mathbf{H}}'_{(n \times a)} = (\tilde{\mathbf{S}}^{1/2} \tilde{\mathbf{V}}')$. A matriz $\tilde{\mathbf{G}}$ terá, então, *g* vetores denominados marcadores de linhas, cada um com *n* componentes (as coordenadas dos genótipos para cada eixo singular selecionado). Da mesma forma, $\tilde{\mathbf{H}}'$ terá *a* marcadores de colunas, cada um com *n* componentes, correspondendo às coordenadas de ambientes para os mesmos *n* eixos. Dada a ortogonalidade dos eixos singulares, pode-se plotar os *g* genótipos e os *a* ambientes num sistema cartesiano de *n* eixos perpendiculares (na prática, $n \leq 3$).

Demonstra-se, dessa forma, que o *padrão* subjacente à interação *GxE* passa a ser expresso, inclusive graficamente, como o resultado da multiplicação de um termo relativo aos genótipos (efeitos de linhas) e outro relativo aos ambientes (efeitos de colunas). Ambos os termos expressos em unidades de raiz quadrada da unidade de Y_{ij} . Logo, multiplicando-se uma linha *i* de $\tilde{\mathbf{G}}$ por uma coluna *j* de $\tilde{\mathbf{H}}'$ (produto interno de vetores) obtém-se, nesse tipo de representação, a estimativa *AMMI* para a interação do genótipo *i* com o ambiente *j*, ou seja: $\sum_{k=1}^n \lambda_k \gamma_{ik} \alpha_{jk}$, conforme discutido na seção anterior (este passo é ilustrado em detalhe no Apêndice 7.2). Abre-se, com isso, a possibilidade de capitalizar, na seleção, interações positivas de genótipos e ambientes específicos, o que corresponderá, no gráfico, a combinações de genótipos e ambientes com coordenadas (ou escores) de mesmo sinal. Assim, de uma maneira esquemática,

o raciocínio subjacente à representação em *biplot*, no caso da análise *AMMI*, pode ser resumido nos enunciados a seguir:

$$\begin{aligned}
 \mathbf{GE}_{(g \times a)} &= \mathbf{U} \mathbf{S} \mathbf{V}' = \underbrace{\mathbf{U} \mathbf{S}^{1/2}}_{\mathbf{G}} \underbrace{\mathbf{S}^{1/2} \mathbf{V}'}_{\mathbf{H}'} = \mathbf{G} \mathbf{H}' \rightarrow \boxed{\text{decomposição exata}} \\
 \mathbf{G}\hat{\mathbf{E}}_{(g \times a)} &= \tilde{\mathbf{U}} \tilde{\mathbf{S}} \tilde{\mathbf{V}}' = \tilde{\mathbf{G}} \tilde{\mathbf{H}}' = \sum_{k=1}^n (\lambda_k^{1/2} \gamma_k) (\lambda_k^{1/2} \alpha'_k) \rightarrow \\
 &\quad \boxed{\text{decomposição aproximada por } \underline{n} \text{ componentes}} \\
 \tilde{\mathbf{G}}_{(g \times n)} &\rightarrow \text{marcadores de linhas (efeitos de genótipos).} \\
 \tilde{\mathbf{H}}'_{(n \times a)} &\rightarrow \text{marcadores de colunas (efeitos de ambientes).}
 \end{aligned}$$

Deve ficar claro que o termo *biplot*, em alguns artigos tomado como sinônimo de *AMMI* (embora não seja exclusivo deste tipo de análise), não se refere a qualquer modelo estatístico particular, mas apenas a um tipo de gráfico contendo duas categorias de pontos ou marcadores. Aqui, uma delas referindo-se aos genótipos e a outra aos ambientes. Ademais, embora a maioria destes gráficos seja construída em duas dimensões, não é isto que determina o nome *biplot*, mas os dois tipos de marcadores, os quais podem ser representados, na prática, em *uma*, *duas* ou em até *três* dimensões (Gauch, 1992; p. 206).

3.3.2. Os tipos mais usados de *biplots AMMI*

Os gráficos em *biplot* mostrados em artigos que utilizam a análise *AMMI* no estudo da interação $G \times E$, normalmente apresentam apenas o primeiro eixo de interação ($IPCA_1$). Mesmo assim os gráficos são construídos em duas dimensões, utilizando-se o eixo das abscissas para representar os efeitos principais (médias de genótipos e de ambientes) e as ordenadas para expressar os escores de genótipos e ambientes para o referido eixo de interação. A Figura 5, retirada de Zobel *et al.* (1988), ilustra este tipo de gráfico, denominado a partir deste momento como *biplot AMMI1*.

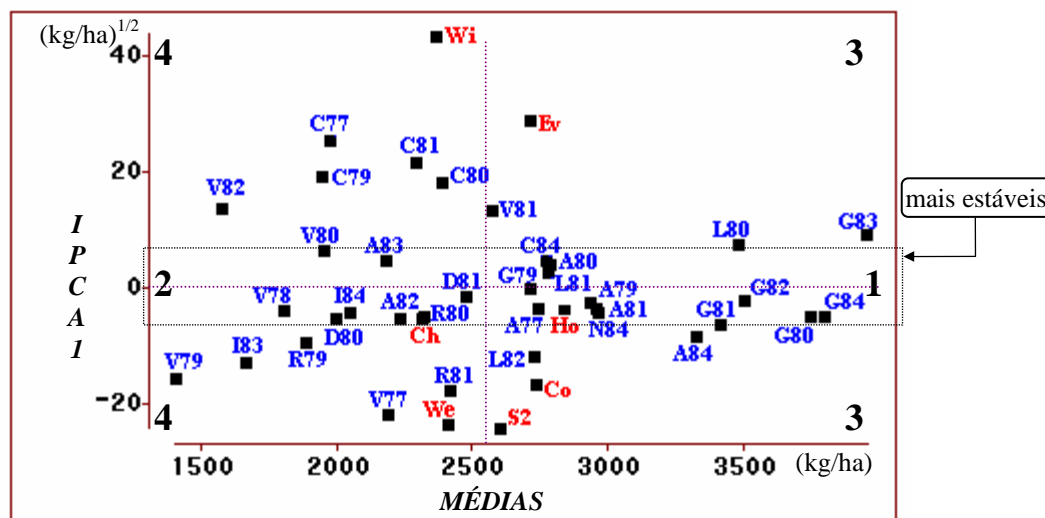


Figura 5. *Biplot* da análise AMMI para ensaios de rendimento de grãos (kg/ha), em soja, com 7 genótipos (vermelho) e 35 ambientes (azul) (obtido de Zobel *et al.*, 1988); os números de 1 a 4 nas bordas do gráfico determinam zonas de classificação dos genótipos e ambientes, segundo Alvarez & Eyhérabide (1996).

A princípio esta representação só se justificaria quando o primeiro eixo (parte do modelo *AMMI1*) explicasse uma alta proporção da $SQ_{G \times E}$, à semelhança do que se faz numa *ACP* tradicional. Entretanto, o *padrão* de $G \times E$ (interação propriamente dita) pode ser pouco expressivo, de forma que grande parte desta SQ esteja associada a *ruídos*. E mesmo nesta situação, uma alta proporção deste *padrão* estará presente no primeiro eixo. Crossa *et al.* (1991) comentam que uma característica marcante de modelos multivariados, incluindo a análise *AMMI*, é que eles captam grande parte do *padrão* nos primeiros eixos principais, relegando aos eixos subsequentes cada vez menos *padrão* e mais *ruídos*. Observa-se, à luz de Gauch & Zobel (1988), que a exigência de um alto coeficiente de determinação caracteriza um critério de avaliação “*posdictiva*”, enquanto o raciocínio posterior segue um enfoque *preditivo* de avaliação. Logo, conclui-se que uma inspeção ao *biplot AMMI1* será sempre conveniente.

No trabalho de Zobel *et al.* (1988), um dos precursores na aplicação desse tipo de análise no melhoramento de plantas, o primeiro componente captou 71% da $SQ_{G \times E}$, indicando que este eixo concentra a maior proporção do *padrão* de resposta diferencial dos genótipos diante dos ambientes. Por conseguinte, as médias ajustadas pelo modelo *AMMI1* são também as que melhor caracterizam esse *padrão*, o que justificou plenamente a adoção do *biplot AMMI1*. Entretanto, vários outros autores utilizaram o mesmo tipo de representação apesar de terem verificado uma menor proporção da $SQ_{G \times E}$ explicada pelo primeiro eixo singular, a saber: 60% (Eyhérabide *et al.*, 1997); 57,6% (Flores *et al.*, 1996); 54,6% (Crossa *et al.*, 1990); 44,6% (Pereira & Costa, 1998); 28,6% (Arias, 1996); 27,1% (Crossa *et al.*, 1991).

De Cauwer & Ortiz (1998) adotaram dois tipos de gráficos para apresentar os resultados da análise *AMMI* aplicada a dados de rendimento e peso de cacho num estudo com bananeiras. O primeiro foi o tradicional *biplot AMMI1* (efeitos principais vs. interação- $IPCA_1$), já descrito anteriormente. O outro, denominado *biplot AMMI2*, representou somente efeitos de interação: $IPCA_1$ vs. $IPCA_2$ (contendo 73% da $SQ_{G \times E}$, 27% a mais do que *AMMI1*). Uma alternativa para incluir informações sobre os efeitos principais, num *biplot AMMI2*, é indexar cada ponto (genótipo ou ambiente), por um número correspondente à sua ordem no respectivo ranqueamento de médias, para genótipos ou ambientes. A Figura 6, representando 74% da $SQ_{G \times E}$ (51% em $IPCA_1$ e 23% em $IPCA_2$), ilustra este tipo de gráfico. Neste caso, apesar de não ser possível avaliar a magnitude das diferenças entre médias, a representação no plano ainda garante vantagens sobre uma terceira alternativa que seria a representação tridimensional ($IPCA_1$ vs. $IPCA_2$ vs. médias).

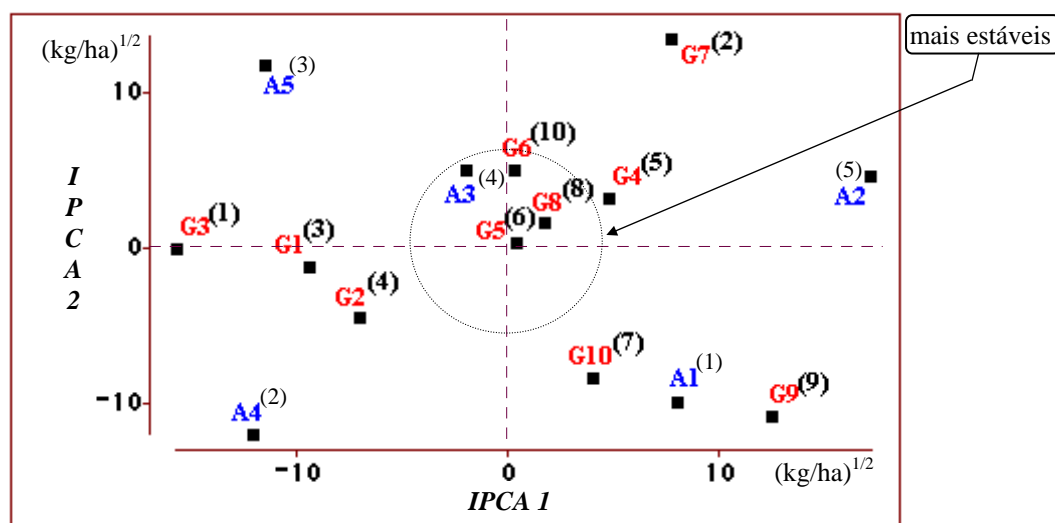


Figura 6. *Biplot AMMI2* para dados de produtividade de grãos (kg/ha), em feijoeiro; o expoente indica o posto do genótipo (G) ou do ambiente (A), nas listas de médias em ordem decrescente (análise ilustrada no item 4 - Exemplo Numérico).

3.3.3. Interpretando um *biplot AMMI*

A interpretação de um *biplot* quanto à interação $G \times E$ é feita observando-se a magnitude e o sinal dos escores de genótipos e ambientes para o(s) eixo(s) de interação. Assim, escores baixos (próximos de zero) são próprios de genótipos e ambientes que contribuíram pouco ou quase nada para a interação, caracterizando-os como *estáveis*. Num *biplot AMMI1*, a exemplo do mostrado na Figura 5, a estabilidade é avaliada inspecionando-se as ordenadas (primeiro eixo de

interação – $IPCA_1$). Logo, os pontos situados na faixa horizontal em torno de zero em relação ao eixo $IPCA_1$ (retângulo pontilhado) correspondem aos genótipos e ambientes mais estáveis. Entre estes estão os genótipos *Ho* e *Ch*, e os ambientes *G82*, *L81*, *G79*, *D81*, *V78* (a letra codifica o local e o número o ano de teste). Num *biplot* AMMI2, genótipos e ambientes estáveis (com pequena contribuição para a $SQ_{G \times E}$) são aqueles cujos pontos situam-se próximos à origem, ou seja, com escores praticamente nulos para os dois eixos de interação ($IPCA_1$ e $IPCA_2$). Este é o caso dos genótipos *G5*, *G8* e *G6*, bem como do ambiente *A3*, delimitados pelo círculo na Figura 6.

Pelo lado dos genótipos, a estabilidade aqui avaliada é um indicativo de suas respectivas amplitudes adaptativas, ou seja, genótipos estáveis mostraram-se amplamente adaptados aos ambientes de teste. Entretanto, para fins de recomendação de cultivares, genótipos estáveis devem também apresentar uma performance desejável (ex: alta produtividade), o que é avaliado através de suas médias (efeitos principais). Assim, entre os genótipos ilustrados na Figura 5, destaca-se como estável e de produtividade média elevada o genótipo *Ho* (mais à direita no retângulo pontilhado). Já entre aqueles mostrados na Figura 6, os mais estáveis tiveram uma má classificação em produtividade média (*G5* foi o sexto em produtividade, *G8* foi o oitavo e *G6* foi o menos produtivo de todo o grupo). Logo, para esse conjunto específico de dados, produtividades médias elevadas parecem estar associadas a adaptações específicas.

Por outro lado, o ordenamento (*ranking*) dos genótipos num ambiente estável será de maior confiança para o melhorista, haja vista que tal classificação será determinada basicamente pelos efeitos genotípicos (interação praticamente nula). O ambiente *A3*, na Figura 6, pode ser tomado como um exemplo deste tipo. Além disso, no caso de ambientes que representam um mesmo local em diferentes anos (ou épocas), valores baixos para o(s) eixo(s) de interação indicam que a classificação dos genótipos de um ano para outro, neste local, é pouco variável, resultando numa classificação mais consistente dos genótipos. Logicamente, a uma maior repetibilidade do ordenamento ao longo dos anos (ou épocas) está associada uma maior confiabilidade na classificação genotípica. Na ilustração dos resultados de Zobel *et al.* (1988), Figura 5, o local *Aurora* (*A77* a *A84*) exhibe esta propriedade, contrariamente ao local *Valatie* (*V77* a *V82*). O primeiro, embora apresente alta variabilidade de ano para ano, em termos de efeitos principais, mostra pequena variação em interações $G \times E$. Já o segundo mostra-se bastante variável também nas interações. Logo, torna-se mais difícil fazer uma recomendação para *Valatie* do que para *Aurora*. Portanto, os melhoristas podem também selecionar locais de teste conforme seus interesses.

Em decorrência de informações desta natureza a análise AMMI é apontada como de grande utilidade não só na identificação de genótipos de alta produtividade e ampla adaptação, mas também para a realização do chamado *zoneamento agrônomo*, incluindo a seleção de locais “chave” (locais de teste, como referem Gauch & Zobel, 1996).

Alvarez & Eyherabide (1996) resumem as informações de um *biplot* AMMI1, classificando os genótipos e ambientes em quatro grupos: **1)** com rendimentos

médios elevados (acima da média) e de baixa interação $G \times E$ (estáveis); **2**) com baixos rendimentos (inferiores à média) e estáveis; **3**) com rendimentos elevados, porém, mostrando maior interação $G \times E$ (instáveis); e **4**) com baixos rendimentos e instáveis. A interpretação desta classificação, também indicada na Figura 5, é perfeitamente coerente com o que foi anteriormente discutido.

Além disso, a análise *AMMI* permite, como aventado anteriormente, explorar também os efeitos positivos da interação $G \times E$. Assim, genótipos “instáveis” não devem ser prontamente descartados em função desta sua adaptabilidade restrita. Num *biplot* (em qualquer dimensão) é possível identificar tais especificidades, permitindo ao melhorista aproveitá-las vantajosamente na seleção e recomendação de genótipos. Como já comentado, uma propriedade desse tipo de representação é que o produto interno entre um i -ésimo marcador de linha e um j -ésimo marcador de coluna reproduz uma aproximação para o i,j -ésimo elemento da matriz de interações⁷. Dessa forma, o produto entre os escores de um dado genótipo e de um certo ambiente fornece uma estimativa do efeito da interação $G \times E$ para aquela combinação. Caso a estimativa seja positiva o melhorista estará diante de uma combinação favorável de genótipo e ambiente, podendo capitalizar este efeito benéfico da interação, na sua seleção. Isso leva, portanto, a uma recomendação regionalizada.

Essas relações adaptativas podem ser facilmente percebidas num *biplot AMMI* observando-se apenas os sinais dos escores para cada par de genótipos e ambientes. Assim, genótipos e ambientes com escores de mesmo sinal devem interagir positivamente evidenciando um sinergismo adaptativo a ser aproveitado na seleção. Já aqueles com sinais opostos devem interagir negativamente, sugerindo certo antagonismo, ou seja, uma combinação desfavorável de genótipo e ambiente. Desta maneira, informações fundamentais ao processo de seleção tornam-se prontamente disponíveis aumentando, por conseguinte, a probabilidade de sucesso nas recomendações de cultivares.

Uma questão adicional que se coloca é se os eixos principais de interação, selecionados por determinado critério, são interpretáveis em termos de propriedades genótípicas e ambientais conhecidas. As tentativas de nominar ou identificar aspectos biológicos descritos pelos eixos de interação, à semelhança do que se faz em *ACP*, também são feitas para as aplicações do modelo *AMMI*. A partir disso, pode-se avaliar a coerência da análise em captar padrões estatística e agronomicamente importantes nos eixos da interação $G \times E$. Alvarez & Eyhérabide (1996) avaliaram híbridos comerciais de milho, na Argentina, constatando que o

⁷/ - Esse componente de interação: $(g\tilde{e})_{ij} = \sum_{k=1}^n (\lambda_k^{1/2} \gamma_{ik})(\lambda_k^{1/2} \alpha_{jk})$, pode também ser observado direta-mente na matriz de interações estimada (aproximada) pelo modelo *AMMI* escolhido. O uso da matriz será especialmente útil quando os primeiros eixos não captarem uma alta proporção da SQ_{GE} , de forma a despertar uma menor confiança na representatividade de um *biplot AMMI1* ou *AMMI2*. Se a esta estimativa de interação forem adicionadas aquelas para os efeitos principais, obtém-se a estimativa da resposta para uma dada combinação de genótipo e ambiente, conforme apresentado no item 3.2.5 deste texto.

primeiro eixo foi capaz de captar efeitos de ciclo e da constituição genética dos híbridos (simples, duplos e triplos).

A análise de Zobel *et al.* (1988), aplicada a dados de rendimento de grãos em cultivares de soja, constatou que a relação entre o comprimento do dia (fotoperíodo : latitude), no local e época de teste, e o ciclo do cultivar (grupo de maturação), foi responsável pela porção significativa da interação $G \times E$ capturada pelo primeiro eixo da ACP. Observando a Figura 5, os cultivares Wilkin (Wi) e Evans (Ev) são ambos do grupo 0 de maturação (materiais mais precoces), enquanto SRF200 (S2), Wells (We) e Corsoy (Co), no extremo oposto do eixo de interação, pertencem ao grupo II (mais tardios). Ademais, os genótipos Chippewa 64 (Ch) e Hodgson (Ho) são de ciclo intermediário (grupo I). Concomitantemente, o local Chazy (C77, C79, etc.), ao norte dos EUA e na parte alta da figura, favorece o cultivo de materiais do grupo 0 (locais e cultivares com escores positivos para o eixo de interação). Enquanto os locais na parte baixa da figura são mais apropriados para genótipos do grupo II (ambos, locais e cultivares com escores negativos para a interação).

Uma estratégia para tentar estabelecer esse tipo de associação é medir, em todos os ambientes, fatores como a precipitação pluviométrica, temperaturas máximas e mínimas, etc, correlacionando-os aos escores dos ambientes para os eixos de interação (Gauch & Zobel, 1996). Estes autores relatam trabalhos, em trigo, em que foram encontrados resultados neste sentido. Por exemplo: forte correlação entre o primeiro eixo ($IPCA_1$) e o número de dias frios no período de formação da panícula, e entre o segundo eixo ($IPCA_2$) e o estresse hídrico (Annicchiarico & Perenzin, 1994); $IPCA_1$ associado ao conteúdo de argila e precipitação pluviométrica no verão (Annicchiarico, 1992); e $IPCA_1$ refletindo altitude e irrigação (Nachit *et al.*, 1992). Gauch & Zobel (1996) reportam também trabalhos que confirmam relações desse tipo, captadas pela análise AMMI, através de outras metodologias (Van Eeuwijk & Elgrma, 1993; Van Eeuwijk, 1992). A identificação segura dessas relações ajudará cada vez mais a entender os mecanismos que determinam o fenômeno da interação $G \times E$.

3.3.4. O *biplot* para diagnose do modelo de análise

A configuração dos pontos num *biplot* de médias nas abcissas e o primeiro eixo singular (ou principal) nas ordenadas, pode também ser utilizada como instrumento diagnóstico na seleção de modelos para análise de dados em tabelas de dupla entrada (Bradu & Gabriel, 1978). Gauch (1992; p.207-209) apresenta vários padrões que podem emergir de um *biplot* AMMI1. Por exemplo, uma disposição horizontal dos pontos (genótipos e ambientes) ao longo das abcissas evidencia a predominância de efeitos principais, ou seja, o conjunto de dados mostra apenas efeitos aditivos sem interação alguma. Logo, o ajuste dos efeitos principais por ANOVA (submodelo AMMI0) será perfeito. Por outro lado, uma disposição ao longo das ordenadas (vertical) evidencia uma variação determinada basicamente por efeitos de interação, sem qualquer ação aditiva (dos efeitos

principais). Neste caso, o submodelo *DVS* (ou *ACP*) do *AMMI* ajusta-se perfeitamente aos dados.

Estas duas situações seriam também facilmente diagnosticadas inspecionando-se o quadro da análise conjunta de variância: $SQ_{G \times E} \cong 0$ para o primeiro caso e $SQ_G \cong SQ_E \cong 0$ para o segundo. No entanto, para casos mais complexos o *biplot* pode oferecer uma ajuda substancial.

Outras configurações permitem ainda escolher entre um modelo de regressão linear (ex: Finlay & Wilkinson, 1963; Eberhart & Russell, 1966) e um modelo *AMMI* de análise. Assim, se os pontos dispuserem-se ao longo da linha bissetriz dos quadrantes ímpares (45° com o eixo das abcissas), recomenda-se o uso de um modelo de regressão linear conjunta: $Y_{ij} = \mu + g_i + e_j + \beta g_i e_j + \delta_{ij} + \varepsilon_{ij}$; onde β é o coeficiente da regressão e δ_{ij} é o seu respectivo desvio. Se apenas os genótipos estiverem assim alinhados o modelo deverá ajustar (além dos efeitos principais) regressões lineares das respostas genotípicas (β_i) sobre índices da qualidade ambiental (ex: $I_j = \bar{Y}_{.j} - \bar{Y}_{..} = e_j$), ou seja: $Y_{ij} = \mu + g_i + e_j + \beta_i e_j + \delta_{ij} + \varepsilon_{ij}$. Se, porém, esta disposição verificar-se somente para os ambientes o modelo deveria ser: $Y_{ij} = \mu + g_i + e_j + \beta_j g_i + \delta_{ij} + \varepsilon_{ij}$. Neste caso, ajustando respostas ambientais (β_j) sobre índices genotípicos ($g_i = \bar{Y}_{i.} - \bar{Y}_{..}$). Finalmente, quando os pontos se dispersam no *biplot*, nenhum modelo simplificado poderá ajustar convenientemente os dados, de forma que a análise *AMMI* mostra-se apropriada (Gauch, 1992; p. 62-65).

4. EXEMPLO NUMÉRICO

Nesta seção será considerado um exemplo numérico da análise *AMMI* aplicada ao estudo da interação $G \times E$. Os dados de partida, neste caso, correspondem já à matriz de médias ($Y_{(g \times a)}$), mas poderiam ser as observações ao nível de parcelas dos experimentos (repetições individuais). Seja, então, a matriz Y a seguir, construída com dados de produtividade média de grãos (kg/ha) de 10 genótipos de feijoeiro comum, avaliados em 5 ambientes (Ramalho *et al.*, 1993):

	A1	A2	A3	A4	A5	
G1	1587,9	560,9	1020,6	1449,7	1489,6	$= Y_{(10 \times 5)}$
G2	1386,7	691,7	1160,8	1591,2	1235,3	
G3	1572,9	770,4	1316,2	1835,9	1698,0	
G4	1139,7	734,5	862,7	1143,8	974,9	
G5	1348,9	458,2	663,8	1021,2	1116,3	
G6	1061,5	459,4	576,7	948,3	1013,7	
G7	1590,9	1104,0	1286,3	1250,9	1502,7	
G8	1223,9	546,9	674,3	1053,7	1053,4	
G9	1402,0	570,9	664,4	993,4	672,9	
G10	1313,6	580,1	596,3	1185,9	913,7	

Assumindo satisfeitas as pressuposições básicas da análise de variância, o resultado do ajuste dos efeitos principais por ANOVA (primeira etapa da análise *AMMI*, item 3.2.2), aos dados da matriz Y , é o seguinte:

F.V.	GL	SQ	QM	F
Genótipos (G)	9	2230894,83	247877,20	10,81**
Ambientes (E)	4	3398566,23	849641,56	37,07**
Resíduo = $G \times E$	36	825226,64	22922,96	--
TOTAL (TRATAMENTOS)	49	6454687,71	--	--

Note-se que esse quadro corresponde à parte da análise conjunta de variância, efetuada ao nível de médias. Logo, é possível testar a significância da interação $G \times E$, desde que se disponha de uma estimativa da variância do erro experimental σ^2 (neste caso, dada pelo $QM_{\text{Erro médio}}$). Um primeiro resultado de grande interesse neste quadro é a soma de quadrados: $SQ_{G \times E} = 825.226,64$ (objeto da decomposição *DVS*, na segunda etapa da análise), que, para este conjunto de dados, representou 13% da SQ_{TOTAL} (de tratamentos: $G + E + G \times E$).

As estimativas das médias de genótipos e ambientes, bem como da média geral, ajustadas pelo modelo (sem interação), são dadas por:

Genótipo	Média ($\bar{Y}_{i.}$)	Ambiente	Média ($\bar{Y}_{.j}$)
G1	1221,74	A1	1362,80
G2	1213,14	A2	647,70
G3	1438,68	A3	882,21
G4	971,12	A4	1247,40
G5	921,68	A5	1167,05
G6	811,92		
G7	1346,96		
G8	910,44		
G9	860,72		
G10	917,92		
		Média Geral ($\bar{Y}_{..}$)	
		1061,43	

Considerando ainda as relações apresentadas no item 3.2.2, o resíduo do ajuste dos efeitos principais à matriz de médias, corresponde exatamente ao termo geral de interação $(g\hat{e})_{ij}$. Ou seja:

$$\hat{e}_{ij} = Y_{ij} - \bar{Y}_{ij} \Leftrightarrow (g\hat{e})_{ij} = Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..}$$

A aplicação desta expressão à matriz **Y**, uma vez disponíveis as estimativas de médias, resulta na matriz de interações (**GE**_(gxa)) a seguir:

	A1	A2	A3	A4	A5	
G1	64,792	-247,108	-21,918	41,992	162,242	= GE _(10x5)
G2	-127,808	-107,708	126,882	192,092	-83,458	
G3	-167,148	-254,548	56,742	211,252	153,702	
G4	-132,788	177,112	70,802	-13,288	-101,838	
G5	125,852	-49,748	-78,658	-86,448	89,002	
G6	-51,788	61,212	-55,998	-49,588	96,162	
G7	-57,428	170,772	118,562	-282,028	50,122	
G8	12,092	50,192	-56,918	-42,708	37,342	
G9	239,912	123,912	-17,098	-53,288	-293,438	
G10	94,312	75,912	-142,398	82,012	-109,838	

Note-se que este resultado pode ser obtido operando-se diretamente com as quatro matrizes correspondentes aos termos Y_{ij} , $\bar{Y}_{i.}$, $\bar{Y}_{.j}$ e $\bar{Y}_{..}$. Ou seja, denotando-se **L**, **C** e **M** as respectivas matrizes (todas de ordem gxa, como **Y** e **GE**), com as médias dos genótipos em suas linhas (**L**), com as médias dos ambientes em suas colunas (**C**) e com a média geral em todas as posições (**M**), tem-se: **GE** = **Y** - **L** - **C** + **M**; em que as matrizes **L**, **C** e **M** são dadas por:

$$\mathbf{L} = \begin{bmatrix} \bar{Y}_{1.} & \bar{Y}_{1.} & \dots & \bar{Y}_{1.} \\ \bar{Y}_{2.} & \bar{Y}_{2.} & \dots & \bar{Y}_{2.} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{Y}_{g.} & \bar{Y}_{g.} & \dots & \bar{Y}_{g.} \end{bmatrix}; \mathbf{C} = \begin{bmatrix} \bar{Y}_{.1} & \bar{Y}_{.2} & \dots & \bar{Y}_{.a} \\ \bar{Y}_{.1} & \bar{Y}_{.2} & \dots & \bar{Y}_{.a} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{Y}_{.1} & \bar{Y}_{.2} & \dots & \bar{Y}_{.a} \end{bmatrix}; \text{ e } \mathbf{M} = \begin{bmatrix} \bar{Y}_{..} & \bar{Y}_{..} & \dots & \bar{Y}_{..} \\ \bar{Y}_{..} & \bar{Y}_{..} & \dots & \bar{Y}_{..} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{Y}_{..} & \bar{Y}_{..} & \dots & \bar{Y}_{..} \end{bmatrix}.$$

A etapa seguinte da análise AMMI corresponde ao ajuste da interação por DVS (item 3.2.3), aplicada à matriz **GE**. Desde já, pode-se calcular o posto de **GE**, aqui denotado por p e que determina o número de parcelas em que se pode decompor a $SQ_{G \times E}$. Tem-se, então: $p = \min\{g-1, a-1\} = \min\{9, 4\} = 4$. Logo, neste caso, a $SQ_{G \times E}$ pode ser decomposta em até 4 componentes ortogonais (somadas de quadrados parciais). A operação fundamental para obter estes componentes, segundo a abordagem adotada neste texto, é a determinação da DVS da matriz **GE**. Dada a dimensão desta matriz recomenda-se o uso de um *software* matemático de álgebra linear (ex: *MATLAB*, *SAS/IML*, etc) para a realização dessa tarefa. No *SAS/IML*, por exemplo, isto é feito através do comando “CALL SVD (U,S,V,GE)”. Assim, para esse conjunto de dados, as matrizes resultantes que determinam a DVS da matriz de interações (**GE** = **U.S.V'**), são dadas por:

$\mathbf{U}_{(10 \times 4)} =$	-0,368996	-0,056027	0,438540	0,286434	;
	-0,273018	-0,207555	-0,518337	0,268881	
	-0,617939	0,002892	-0,150374	-0,006240	
	0,193135	0,162270	-0,503047	-0,178907	
	0,019971	0,022439	0,485867	0,048160	
	0,014459	0,247645	0,091445	-0,415796	
	0,308265	0,657367	-0,029672	0,356981	
	0,070478	0,085858	0,124541	-0,254904	
	0,494042	-0,519068	0,012380	0,417016	
	0,159603	-0,395823	0,048656	-0,521625	
	↓	↓	↓	↓	
	γ_1	γ_2	γ_3	γ_4	

$$S_{(4)} = \begin{bmatrix} 648,1521 & 0,0000 & 0,0000 & 0,0000 \\ 0,0000 & 437,0842 & 0,0000 & 0,0000 \\ 0,0000 & 0,0000 & 409,1752 & 0,0000 \\ 0,0000 & 0,0000 & 0,0000 & 216,0060 \end{bmatrix} ; e$$

$$V'_{(4 \times 5)} = \begin{bmatrix} 0,319555 & 0,681160 & -0,075538 & -0,473155 & -0,452022 \\ -0,475216 & 0,229817 & 0,249848 & -0,576211 & 0,571761 \\ 0,620212 & -0,282319 & -0,441409 & -0,357774 & 0,461290 \\ 0,295621 & -0,451111 & 0,732822 & -0,340739 & -0,236593 \end{bmatrix} \rightarrow \begin{matrix} \alpha'_1 \\ \alpha'_2 \\ \alpha'_3 \\ \alpha'_4 \end{matrix}$$

Estas matrizes são, portanto, a matriz com os *vetores singulares-coluna* γ_k (U), a matriz com os *valores singulares* λ_k (S) e a matriz com os *vetores singulares-linha* α'_k (V'), com $k = 1, 2, 3, 4$.

Em seqüência, obtém-se os elementos do desdobramento da $SQ_{G \times E}$ por DVS, correspondentes aos quadrados dos quatro valores singulares (λ_k^2) ou, equivalentemente, aos quatro autovalores de $(GE)(GE')$ ou de $(GE')(GE)$:

Eixo singular	Autovalor (λ_k^2)	Proporção da $SQ_{G \times E}$ / Eixo	Proporção acumulada (%)
1	420101,0907	0,5091	50,9074
2	191042,6293	0,2315	74,0577
3	167424,3290	0,2029	94,3460
4	46658,5922	0,0565	100,0000
Total ($\sum_k \lambda_k^2$)	825226,6400	1,0000	--

Observa-se que o primeiro eixo singular (ou principal) de interação captura 51% da $SQ_{G \times E}$, o segundo 23%, o terceiro 20% e o quarto e último apenas 6%. Logo, um modelo *AMMI2*, por exemplo, explica 74% desta soma de quadrados, dividindo-a em: $SQ_{G \times E} [\text{padrão}] = 611.143,72$ (com os 74%) e $SQ_{G \times E} [\text{ruídos}] = 214.082,92$ (com os 26% restantes). Note-se que a soma dos autovalores ($\sum_k \lambda_k^2 = 825.226,64$) corresponde exatamente à $SQ_{G \times E}$, demonstrando o seu perfeito desdobramento nos *quatro* componentes. Assim, um modelo *AMMI4*, por exemplo, contemplaria os quatro eixos de interação, resgatando toda a $SQ_{G \times E}$.

Neste momento é possível montar, de forma completa, o quadro da análise conjunta de variância, incluindo o desdobramento da interação $G \times E$. Logo, atribuindo-se graus de liberdade aos componentes da interação ($IPCA_k$), conforme o sistema de Gollob (1968), ou seja: $GL_{IPCA_n} = g + a - 1 - 2n$, tem-se o seguinte quadro:

F.V.	GL	SQ	QM	F _{Gollob}	GL _{Res.AMMI}	QM _{Res.AMMI}	F _{Cornelius}
Genótipos (G)	9	2230894,83	247877,20	10,81**	--	--	--
Ambientes (E)	4	3398566,23	849641,56	37,06**	--	--	--
Interação GxE	36	825226,64	22922,96	2,37**	36	22922,96	2,37**
<i>IPCA 1</i>	12	420101,09	35008,42	3,62**	24	16880,23	1,74*
<i>IPCA 2</i>	10	191042,63	19104,26	1,97*	14	15291,64	1,58 ^{ns}
<i>IPCA 3</i>	8	167424,33	20928,04	2,16*	6	7776,43	0,80 ^{ns}
<i>IPCA 4</i>	6	46658,59	7776,43	0,80 ^{ns}	0	0	--
Erro médio _r	90	871163,73	9679,5970	--	--	--	--

Obs.: O Erro médio é originário das análises individuais de variância (média ponderada dos QM_{Erro}'s dos 5 experimentos, reduzida ao nível de médias).

O quadro expressa também os resultados do *resíduo AMMI* da interação, correspondente a cada membro da família de modelos. Assim, o resíduo de interação para *AMMI0* é toda a interação $G \times E$, com 36 GL; já para o modelo *AMMI1* é o restante da interação, após retirados os 12 GL e a soma de quadrados

atribuída ao primeiro eixo ($IPCA_1$), ou seja, é a soma dos efeitos atribuídos aos demais eixos ($IPCA_2$ a $IPCA_4$); e assim por diante. Estes resultados são de interesse para a avaliação dos modelos através do teste F_R (Cornelius *et al.*, 1992), comentado no item 3.2.4.

Os resultados permitem, portanto, avaliar a família de modelos (AMMI0, AMMI1, AMMI2, AMMI3 e AMMI4) com relação a dois critérios. Pelo teste F , aplicado segundo a regra de Gollob, três dos quatro eixos de interação são significativos (5% de probabilidade), o que levaria à seleção do modelo AMMI3. Considerando tratar-se de um critério liberal na seleção de eixos (Piepho, 1995), uma alternativa seria aumentar o rigor do teste. Assim, por exemplo, para o nível de significância de 1%, o modelo selecionado seria AMMI1. Entretanto, uma resposta neste sentido ainda não está disponível na literatura. Dessa forma, outra saída é a aplicação de testes alternativos como por exemplo o teste F_R , defendido por Piepho (1995) como um dos mais robustos. Por este critério o modelo selecionado é o AMMI2. Diante da maior simplicidade representativa deste modelo e das propriedades do teste F_R optar-se-á, neste exemplo, por AMMI2 como o melhor descritor do padrão de resposta diferencial dos genótipos aos ambientes. Ademais, cada um de seus 22 GL relativos à interação, explica, em média, 3,4% da $SQ_{G \times E}$ ($74\% \div 22$ GL). Com um modelo AMMI3, de interpretação mais complexa e provavelmente incorporando uma maior parcela de ruídos, esta relação é de 3,1% ($94\% \div 30$ GL).

Pode-se notar que, no caso da análise a partir da matriz de médias (não dispondo das observações de repetições), não é possível fazer uso do critério de *avaliação preditiva* através da *validação cruzada*. Assim, a seleção do modelo é feita somente a partir de testes estatísticos, os quais ainda carecem de alguns esclarecimentos teóricos (comentários no item 3.2.4). Embora as lacunas também existam para a chamada *avaliação preditiva*, este critério segue uma abordagem completamente distinta, livre de suposições teóricas acerca da distribuição de probabilidade para as variáveis envolvidas.

Diante desse quadro, a possibilidade de utilizar, simultaneamente os dois tipos de avaliação, baseados em enfoques independentes, aumenta as chances de escolher corretamente o melhor modelo de análise. O uso da *avaliação preditiva* altera muito pouco a seqüência operacional ilustrada neste exemplo. A menos que, em princípio, a análise é executada várias vezes tomando-se subamostras do conjunto original de dados, para fins de validação dos modelos⁸.

⁸/ - Num experimento balanceado, com r repetições por casela (cada combinação de genótipos e ambientes), a avaliação preditiva pode ser feita sorteando-se $r-1$ dados por casela para construir uma matriz de médias ($Y_{(g \times a)}$). A outra observação em cada casela formará a matriz de dados de validação ($Y_{v(g \times a)}$). À matriz Y aplica-se a análise AMMI aqui ilustrada, obtendo-se as predições de resposta por casela, para cada membro da família de modelos, ou seja, as matrizes ora denotadas: $\hat{Y}_0, \hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_p$ (todas de ordem $g \times a$). Em seguida, obtém-se as diferenças matriciais: $D_k = \hat{Y}_k - Y_v$; com $k = 0, 1, 2,$

\dots, p . Para cada matriz D_k ($g \times a$) calcula-se, então, a soma dos quadrados de seus elementos, dividindo-a por ga . Os resultados são os escalares $RMSPD_k$, que refletem as respectivas qualidades preditivas dos modelos AMMI0, AMMI1, AMMI2, ..., AMMI p (item 3.2.4, letra b). Entretanto, para aumentar a confiabilidade da avaliação realizam-se vários ciclos de sorteios, obtendo-se um valor médio para $RMSPD_k$. Ao menor $RMSPD_k$ médio corresponde o melhor modelo em capacidade preditiva. Uma vez selecionado o melhor modelo (AMMI n) a análise é aplicada novamente, agora à matriz Y construída com médias de todas as r repetições por casela. Só então, faz-se a avaliação dos genótipos e ambientes, com base nos resultados do modelo AMMI n desta análise final.

Um passo seguinte na análise, após a escolha do modelo *AMMI*, é a predição de respostas para cada combinação de genótipos e ambientes. Considerando a seleção do modelo *AMMI2* isto deverá ser feito a partir da seguinte expressão (item 3.2.5):

$$\hat{Y}_{ij} = \bar{Y}_{i.} + \bar{Y}_{.j} - \bar{Y}_{..} + \left(\sum_{k=1}^{n=2} \lambda_k \gamma_{ik} \alpha_{jk} \right), \text{ ou}$$

$$\hat{Y}_{ij} = \bar{Y}_{i.} + \bar{Y}_{.j} - \bar{Y}_{..} + \lambda_1 \gamma_{i1} \alpha_{j1} + \lambda_2 \gamma_{i2} \alpha_{j2} .$$

Logo, a estimativa *AMMI2* da resposta de um genótipo *i* num ambiente *j* depende das estimativas de média do genótipo, do ambiente, da média geral, bem como de uma aproximação *DVS* de posto *dois* para o termo específico de interação ($g\hat{e}$)_{ij}. Este último termo reflete o *padrão* da interação *GxE* presente no conjunto de dados, sendo função dos *dois* primeiros valores singulares (λ_1 e λ_2), bem como de elementos dos dois pares de vetores singulares a eles associados (γ_1, γ_2 e α_1, α_2). Assim, de posse das matrizes *U*, *S* e *V* é possível obter esta aproximação para qualquer combinação de genótipos e ambientes. A seguir, são dados três exemplos ilustrativos da determinação de estimativas de resposta:

i) *Genótipo 5 no Ambiente 1:*

$$\begin{aligned} \hat{Y}_{51} &= \bar{Y}_{5.} + \bar{Y}_{.1} - \bar{Y}_{..} + \lambda_1 \gamma_{51} \alpha_{11} + \lambda_2 \gamma_{52} \alpha_{12} = 921,68 + 1362,80 - 1061,43 + \\ &+ [648,1521.(0,019971).(0,319555) + 437,0842.(0,022439).(-0,475216)] = \\ &= 1223,05 + [-0,5244] = \mathbf{1222,52 \text{ kg/ha.}} \end{aligned}$$

Observe-se que esta combinação de genótipo e ambiente tem sua resposta (1222,52 kg/ha) determinada basicamente pelos efeitos principais (1223,05 kg/ha), pois o termo de interação específica (-0,52 kg/ha) é praticamente nulo. Mais adiante poderá ser constatado que isto se mantém para os demais ambientes, com relação ao genótipo *G5*, caracterizando-o como de alta estabilidade.

ii) *Genótipo 7 no Ambiente 2:*

$$\begin{aligned} \hat{Y}_{72} &= \bar{Y}_{7.} + \bar{Y}_{.2} - \bar{Y}_{..} + \lambda_1 \gamma_{71} \alpha_{21} + \lambda_2 \gamma_{72} \alpha_{22} = 1346,96 + 647,70 - 1061,43 + \\ &+ [648,1521.(0,308265).(0,681160) + 437,0842.(0,657367).(0,229817)] = \\ &= 933,23 + [202,1296] = \mathbf{1135,36 \text{ kg/ha.}} \end{aligned}$$

Neste caso a combinação de genótipo e ambiente tem sua resposta (1135,36 kg/ha) fortemente determinada pelo efeito da interação específica (202,13 kg/ha), o que caracteriza uma boa adaptação do genótipo *G7* ao ambiente *A2*.

iii) *Genótipo 9 no Ambiente 5:*

$$\begin{aligned} \hat{Y}_{95} &= \bar{Y}_{9.} + \bar{Y}_{.5} - \bar{Y}_{..} + \lambda_1 \gamma_{91} \alpha_{51} + \lambda_2 \gamma_{92} \alpha_{52} = 860,72 + 1167,05 - 1061,43 + \\ &+ [648,1521.(0,494042).(-0,452022) + 437,0842.(-0,519068).(0,571761)] = \\ &= 966,34 + [-274,4630] = \mathbf{691,88 \text{ kg/ha.}} \end{aligned}$$

Esta combinação de genótipo e ambiente mostra-se indesejável, haja vista os efeitos principais relativamente baixos (966,34 kg/ha) e o forte componente

negativo da interação específica (-274,46 kg/ha), o que caracteriza uma falta de adaptação do genótipo G9 ao ambiente A5.

O cálculo destas estimativas, matricialmente, torna-se conveniente e bastante simplificado. Isto porque com uma única operação é possível prever todas as respostas. Para isto, além das matrizes **L**, **C** e **M**, anteriormente definidas, a predição envolve ainda uma aproximação de posto *dois* para a matriz **GE**, aqui denotada por **GE₂**. Observe-se que o procedimento corresponde a obter uma estimativa da desconhecida matriz populacional de respostas verdadeiras (da qual **Y** representa uma amostra), com base no modelo escolhido. Dada a seleção do modelo *AMMI2* será associada à estimativa a notação **Ŷ₂**, cuja expressão matricial, equivalente àquela anteriormente manipulada, é a seguinte:

$$\hat{Y}_2 = L + C - M + GE_2 ;$$

$$\text{com: } GE_2 = \sum_{k=1}^{n=2} \lambda_k \gamma_k \alpha_k .$$

Note-se que o termo [**L** + **C** - **M**] corresponde à matriz de respostas ajustadas pelo modelo *AMMI0* (sem interação), aqui denotada **Ŷ₀**. Logo, pode-se escrever ainda: **Ŷ₂** = **Ŷ₀** + **GE₂**; o que implica nos seguintes resultados:

$$\begin{aligned} \hat{Y}_0 \text{ (10x5)} &= \begin{bmatrix} 1523,11 & 808,01 & 1042,52 & 1407,71 & 1327,36 \\ 1514,51 & 799,41 & 1033,92 & 1399,11 & 1318,76 \\ 1740,05 & 1024,95 & 1259,46 & 1624,65 & 1544,30 \\ 1272,49 & 557,39 & 791,90 & 1157,09 & 1076,74 \\ 1223,05 & 507,95 & 742,46 & 1107,65 & 1027,30 \\ 1113,29 & 398,19 & 632,70 & 997,89 & 917,54 \\ 1648,33 & 933,23 & 1167,74 & 1532,93 & 1452,58 \\ 1211,81 & 496,71 & 731,22 & 1096,41 & 1016,06 \\ 1162,09 & 446,99 & 681,50 & 1046,69 & 966,34 \\ 1219,29 & 504,19 & 738,70 & 1103,89 & 1023,54 \end{bmatrix} ; \\ \\ GE_2 \text{ (10x5)} &= \begin{bmatrix} -64,7892 & -168,5378 & 11,9477 & 127,2729 & 94,1065 \\ -13,4364 & -141,3848 & -9,2990 & 136,0014 & 28,1188 \\ -128,5884 & -272,5269 & 30,5702 & 188,7792 & 181,7659 \\ 6,2971 & 101,5679 & 8,2648 & -100,0980 & -16,0318 \\ -0,5244 & 11,0712 & 1,4727 & -11,7761 & -0,2434 \\ -48,4435 & 31,2593 & 26,3361 & -66,8043 & 57,6524 \\ -72,6934 & 202,1296 & 56,6950 & -260,0973 & 73,9662 \\ -3,2362 & 39,7401 & 5,9255 & -43,2376 & 0,8081 \\ 210,1413 & 165,9773 & -80,8730 & -20,7826 & -274,4630 \\ 115,2732 & 30,7040 & -51,0399 & 50,7424 & -145,6797 \end{bmatrix} ; e \\ \\ \hat{Y}_2 \text{ (10x5)} &= \begin{bmatrix} 1458,32 & 639,47 & 1054,47 & 1534,98 & 1421,46 \\ 1501,07 & 658,02 & 1024,62 & 1535,11 & 1346,88 \\ 1611,46 & 752,42 & 1290,03 & 1813,43 & 1726,06 \\ 1278,79 & 658,96 & 800,16 & 1056,99 & 1060,71 \\ 1222,52 & 519,02 & 743,93 & 1095,87 & 1027,05 \\ 1064,84 & 429,45 & 659,03 & 931,08 & 975,19 \\ 1575,63 & 1135,36 & 1224,43 & 1272,83 & 1526,54 \\ 1208,57 & 536,45 & 737,14 & 1053,17 & 1016,87 \\ 1372,23 & 612,97 & 600,63 & 1025,91 & 691,88 \\ 1334,56 & 534,89 & 687,66 & 1154,63 & 877,86 \end{bmatrix} . \end{aligned}$$

A matriz **Ŷ₀** expressa, portanto, as respostas esperadas de cada combinação de genótipos e ambientes devido somente aos efeitos principais. Já a matriz **GE₂** mostra a magnitude das interações específicas para cada uma destas

abordagem *AMMI*, é responsável por captar substancialmente o *padrão* da interação *GxE*.

Todas as operações algébricas e estatísticas, utilizadas no desenvolvimento do presente exemplo numérico, podem ser processadas através do programa SAS® listado ao final deste texto, no Apêndice 7.3.

A exemplo do que fazem De Cauwer & Ortiz (1998), serão aqui construídos o *biplot AMMI1* (médias *vs* *IPCA*₁), de importância retratada no item 3.3.2, e o *biplot AMMI2* (*IPCA*₁ *vs* *IPCA*₂), devido à sua correspondência com o modelo escolhido. Assim, as coordenadas dos eixos de abscissas (X) e de ordenadas (Y), para os dois gráficos, são dadas por:

Pontos	Genótipos e Ambientes	<i>biplot AMMI1</i>		<i>biplot AMMI2</i>		
		X (Médias)	Y <i>IPCA</i> 1	X <i>IPCA</i> 1	Y <i>IPCA</i> 2	Ordem p/ médias
1	G1	1221,74	-9,3942	-9,3942	-1,1713	(3)
2	G2	1213,14	-6,9507	-6,9507	-4,3393	(4)
3	G3	1438,68	-15,7320	-15,7320	0,0605	(1)
4	G4	971,12	4,9170	4,9170	3,3925	(5)
5	G5	921,68	0,5084	0,5084	0,4691	(6)
6	G6	811,92	0,3681	0,3681	5,1774	(10)
7	G7	1346,96	7,8481	7,8481	13,7433	(2)
8	G8	910,44	1,7943	1,7943	1,7950	(8)
9	G9	860,72	12,5777	12,5777	-10,8519	(9)
10	G10	917,92	4,0633	4,0633	-8,2753	(7)
11	A1	1362,80	8,1355	8,1355	-9,9351	(1)
12	A2	647,70	17,3415	17,3415	4,8047	(5)
13	A3	882,21	-1,9231	-1,9231	5,2235	(4)
14	A4	1247,40	-12,0460	-12,0460	-12,0466	(2)
15	A5	1167,05	-11,5079	-11,5079	11,9536	(3)

As Figuras 7 e 8 ilustram as duas representações (os *biplots AMMI1* e *AMMI2*, respectivamente) resultantes destes dois conjuntos de coordenadas. A partir delas fazem-se, então, as devidas interpretações, procurando identificar genótipos e ambientes que menos contribuíram para a interação *GxE*; combinações de genótipos e ambientes desejáveis em termos de adaptabilidade; relações entre os eixos de interação e características genotípicas e ambientais conhecidas, etc (conforme discussão apresentada no item 3.3.3). Vale mencionar que a Figura 8 é a mesma apresentada naquele item como exemplo de *biplot AMMI2*.

Pelo primeiro destes gráficos evidencia-se que os genótipos que menos contribuíram para a interação *GxE*, captada pelo primeiro eixo (*IPCA*₁), foram G5, G8 e G6; enquanto entre os ambientes destacou-se, neste sentido, o A3. Estes resultados são confirmados no *biplot AMMI2* (Figura 8), bem como pela matriz *GE*₂.

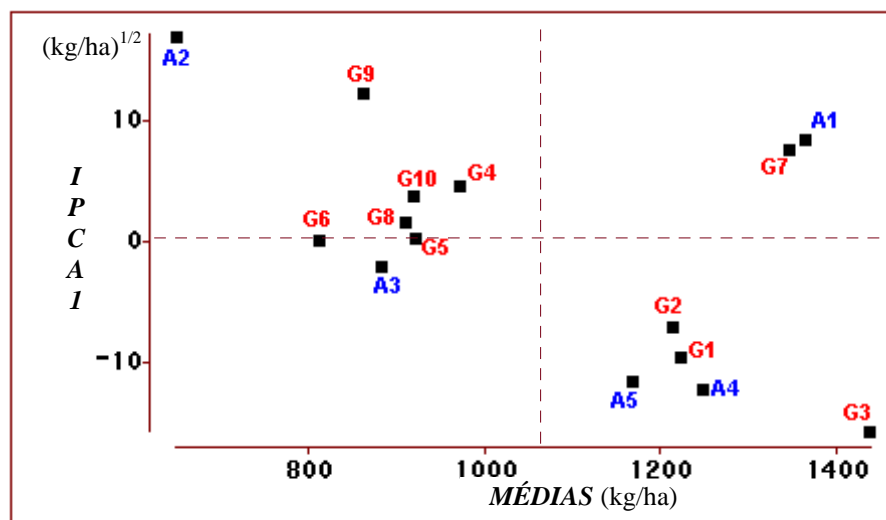


Figura 7. *Biplot AMMI1* para dados de produtividade de grãos (kg/ha), em feijoeiro, com dez genótipos (G) e cinco ambientes (A) (análise feita a partir de dados de Ramalho *et al.*, 1993).

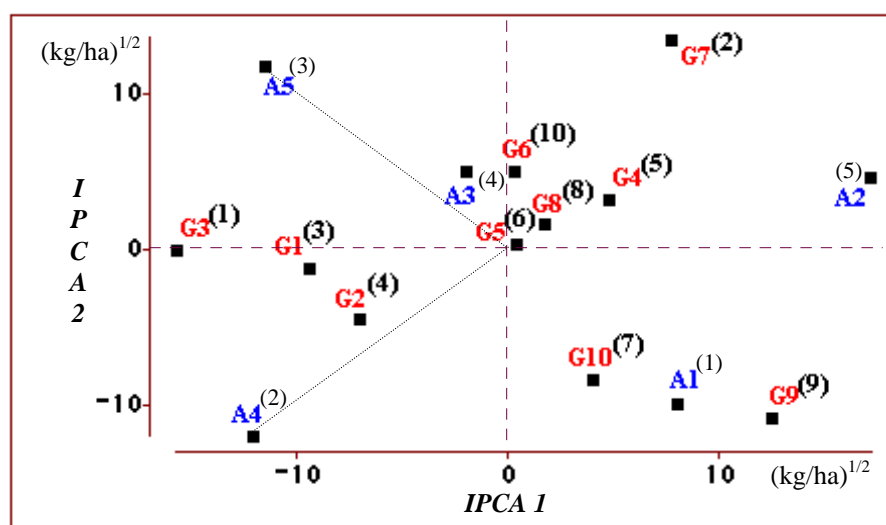


Figura 8. *Biplot AMMI2* para dados de produtividade de grãos (kg/ha), em feijoeiro. O expoente indica o posto do genótipo (G) ou do ambiente (A), nas listas de médias em ordem decrescente (análise feita a partir de dados de Ramalho *et al.*, 1993).

Por outro lado, a aparente adaptação específica de G7 ao ambiente A1, representada no *biplot AMMI1*, não é confirmada quando se incrementam as informações de interação capturadas pelo segundo eixo (*biplot AMMI2*). A estimativa de resposta do genótipo G7 no ambiente A1, observada na matriz \hat{Y}_2 ,

também não confirma uma boa adaptação. A elevada média observada para esta combinação (Y_{71}) parece advir principalmente dos efeitos principais, conforme é mostrado na matriz \hat{Y}_0 . Na realidade, como constatado anteriormente, a especificidade adaptativa do genótipo G7 parece estar associada ao ambiente A2 e não ao A1. Isto sim pode ser visto na Figura 8.

Este tipo de constatação reflete um problema inerente da simplificação representativa de um conjunto p -dimensional de dados num espaço de dimensão mais reduzida. Note-se que as projeções dos dois pontos sobre o eixo $ICPA_1$, na Figura 8, praticamente coincidem, provocando a ilusão de proximidade, embora os pontos estejam distantes numa outra dimensão ($ICPA_2$). O problema merece atenção especial nas situações em que resta uma razoável variabilidade no sentido dos eixos desconsiderados na representação. Logo, no presente caso, o fato ajuda a confirmar que o segundo eixo contém informação relevante para a interação $G \times E$, uma vez que o problema não teria surgido se a variabilidade nesta dimensão fosse diminuta.

Pode-se notar, entretanto, que as concordâncias entre os dois *biplots* prevalecem sobre esta discordância. Observe-se que a falta de adaptação do genótipo G9 ao ambiente A5, já comentada anteriormente, é nítida nos dois *biplots* (vetores apontando em direções opostas). Além disso, as adaptações especiais dos genótipos G1, G2 e G3 aos ambientes A4 e A5, sugeridas na Figura 7, estão também evidentes na Figura 8. Note-se que os vetores (pontos) relativos a estes três genótipos apontam para uma direção comum, que pode ser orientada pelos marcadores dos ambientes A4 e A5. Na Figura 8, estes marcadores estão representados pelas linhas pontilhadas. O mesmo pode-se dizer com relação ao genótipo G9 e os ambientes A1 e A2. Todos estes resultados podem ser confirmados inspecionando-se a matriz $G\hat{E}_2$.

Outro ponto de interesse aplicado é a tentativa de identificar aspectos característicos dos genótipos e dos ambientes, determinando os eixos de interação. Neste sentido, observa-se que o primeiro eixo singular deve ser determinado por características contrastantes entre o ambiente A2 e o par A4 - A5 (similares para tais aspectos). Já o segundo eixo parece ser resultante principalmente das diferenças entre A4 e A5. Em termos genotípicos, o primeiro eixo parece estar relacionado a aspectos determinantes da divergência entre os genótipos G3 e G9; enquanto o segundo à divergência entre G9 e G7. Caberá, portanto, ao melhorista identificar no seu material (genótipos e ambientes sob teste) tais características, de maneira a discernir melhor os mecanismos determinantes da interação $G \times E$.

5. CONSIDERAÇÕES FINAIS

É oportuno considerar que a análise *AMMI*, embora tenha sido aqui enfocada para o estudo da interação de genótipos com ambientes, pode, certamente, ser utilizada para quaisquer conjuntos de dados provenientes de experimentos com dois fatores de classificação cruzada. Ou de forma mais genérica, para qualquer conjunto de dados que possa ser disposto num tabela de dupla entrada. Basta para isso, que as pressuposições mínimas para a aplicação de uma *análise de variância* sejam atendidas.

Neste sentido, em trabalhos de melhoramento de plantas como a avaliação de linhagens cruzadas com testadores, a análise *AMMI* pode ser aplicada buscando-se identificar híbridos promissores (Charcosset *et al.*, 1993, citado por Gauch & Zobel, 1996). Da mesma forma, o método pode ser útil também para um aprimoramento do controle local em delineamentos, subsidiando métodos estatísticos de análise espacial de experimentos. Isto baseia-se na possibilidade de isolar algum componente sistemático na grade (matriz) que determina a disposição das parcelas no campo experimental (Eisenberg *et al.*, 1996).

Dado o enfoque dirigido para tabelas de dupla entrada, uma questão que se coloca é como a análise *AMMI* trataria experimentos com mais de dois fatores cruzados. Por exemplo, uma série de experimentos com genótipos (*G*), locais (*L*) e anos (*A*), o que leva a uma ANOVA com as seguintes partições: *G*, *L*, *A*, *GxL*, *GxA*, *LxA* e *GxLxA*. Caso a interação tripla (*GxLxA*) seja não significativa, há duas maneiras de tratar as duas interações *GxE* restantes. Uma delas é avaliar conjuntamente os efeitos de locais e anos, num *biplot*, nominando os ambientes do seguinte modo, por exemplo: *RV95*, *RV96*, *RV97*, *Go95*, *Go96* e *Go97* (locais: *Rio Verde* e *Goiânia*, anos: 1994 a 1997). Assim, a dispersão dos pontos em relação ao(s) eixo(s) *IPCA*(s), para um mesmo local, refletirá o efeito de anos sobre a interação *GxE* (item 3.3.3). Por outro lado, os pontos correspondentes a locais distintos, num mesmo ano, evidenciam o efeito dos locais sobre a interação. Outra maneira seria aplicar a análise *AMMI*, separadamente, às duas matrizes de médias correspondentes: *genótipos x locais* e *genótipos x anos*. Se ambas forem importantes, serão necessárias duas análises para elucidar os *padrões* de interação. Contudo, se apenas uma delas for relevante (ex: *GxL*) a análise torna-se simplificada.

Quando a interação tripla for significativa, contando com uma elevada *SQ*, infelizmente, como alertam Gauch & Zobel (1996), tanto a seleção de genótipos, como a recomendação de variedades, possuem pouca generalidade através dos locais e anos. Estes autores consideram também que uma elevada *SQ* para *GxLxA* pode originar-se de duas causas bem distintas, a saber: a devida a fatores reais de relevada importância causal ou a que decorre do grande número de graus de liberdade incorporando uma porção considerável de *ruídos*. Para estimar a quantidade de *ruídos* em SQ_{GxLxA} , os autores recomendam que se multiplique o valor dos graus de liberdade desta interação, $(g-1)(l-1)(a-1)$, pelo $QM_{\text{Erro médio}}$. Denotando-se o resultado por $SQ_{[\text{ruídos}, \text{GLA}]}$, se uma alta proporção da SQ_{GxLxA} for

devida a *ruídos*, por exemplo $(SQ_{[ruídos, GLA]})(SQ_{G \times L \times A}) = 0,95$, a interação tripla poderia ser ignorada. E, neste caso, o tratamento estatístico volta a considerar apenas para as interações simples ($G \times L$ e $G \times A$).

Finalmente, deve-se mencionar que a análise *AMMI*, embora já bastante elaborada, carece ainda de desenvolvimentos consideráveis. Pesquisas para melhor esclarecer a atribuição de graus de liberdade aos componentes da interação são requeridas. O procedimento de validação cruzada para seleção do modelo *AMMI* também tem merecido investigação. Os problemas de desbalançamento e da presença de caselas vazias ainda merecem pesquisas adicionais. Ademais, a possibilidade de se optar por um modelo contínuo que pondere cada vez menos os p eixos de interação, ao invés do modelo truncado em n eixos ($n < p$), tem sido postulada. Procedimentos analíticos que permitam assumir uma estrutura de erros mais genérica do que a clássica condição $\varepsilon_{ijk} \sim N(0, \sigma^2)$, também vem sendo implementados através da versão *AMMI* generalizada, denotada *GAMMI* (Gauch & Zobel, 1996). Contudo, apesar de lacunas ainda por preencher, o estágio atual de desenvolvimento deste tema já é suficiente para garantir um conjunto inovador de informações na interpretação das interações de fatores.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- ALVAREZ, M. P. & EYHÉRABIDE, G.H. Estabilidad del rendimiento de cultivares de híbridos comerciales de maíz en la área de la EEA Pergamino. **Rev. Tecnología Agropec. INTA Pergamino**. n. may/ago, p. 17-21, 1996.
- ARIAS, E.R.A. Adaptabilidade e estabilidade das cultivares de milho avaliadas no Estado do Mato Grosso do Sul e avanço genético obtido no período de 1986/87 a 1993/94. Lavras, 1996. 118 p. Tese (Doutorado) - UFLA.
- BRADU, D. & GABRIEL, K.R. The biplot as a diagnostic tool for models of two-way tables. **Technometrics**. v. 20, n. 1, p. 47-68, 1978.
- BRASIL, E. M. Comparação de métodos de controle da interação genótipo x ambiente em milho (*Zea mays* L.). Goiânia, 1990. 214 p. Dissertação (Mestrado) - UFG.
- CARVALHO FILHO, A.A. de A. A decomposição da matriz de planejamento por valores singulares em modelos de posto incompleto. Piracicaba, 1987. 116 p. Dissertação (Mestrado) - ESALQ/USP.
- CROSSA, J. Statistical analyses of multilocation trials. **Advances in Agronomy**. v. 44, p. 55-85, 1990.
- CROSSA, J.; GAUCH, H.G. & ZOBEL, R.W. Additive main effects and multiplicative interaction analysis of two international maize cultivar trials. **Crop Science**. v. 30, n. 3, p. 493-500, 1990.
- CROSSA, J.; FOX, P.N.; PFEIFER, W.H.; RAJARAM, S. GAUCH, H.G. AMMI adjustment for statistical analysis of an international wheat yield trial. **Theoretical Applied of Genetics**. v. 81, p. 27-37, 1991.
- CRUZ, C.D.; TORRES, R.A.A. & VENCOVSKY, R. An alternative to the stability analysis proposed by Silva and Barreto. **Rev. Brasil. Genet.** v. 12, p. 567-580, 1989.
- DE CAUWER, I. & ORTIZ, R. Analysis of the genotype x environment interaction in *Musa* trials. **Experimental Agric.** v. 34, p. 177-188, 1998.
- DUARTE, J.B. & ZIMMERMANN, M.J. de O. Selection of locations for common bean (*Phaseolus vulgaris* L.) germoplasm evaluation. **Rev. Brasil. Genética**. v. 14, n. 3, p. 765-770, 1991.
- EBERHART, S.A.; RUSSELL, W.A. Stability parameters for comparing varieties. **Crop Science**. v. 6, p. 36-40, 1966.

- ECKART, C. & YOUNG, G. The approximation of one matrix by another of lower rank. **Psychometrika**. v. 1, n. 3, p. 211-218, 1936.
- EISENBERG, B.E.; GAUCH, H.G.; ZOBEL, R.W. & KILIAN, W. Spatial analysis of field experiments: fertilizer experiments with wheat (*Triticum aestivum*) and tea (*Camellia sinensis*). In: KANG, M.S.; GAUCH, H.G., ed. **Genotype by environment interaction**. Boca Raton: CRC Press, 1996. Cap. 14, p. 373-404.
- EYHÉRABIDE, G.H.; ALVAREZ, M. P.; PRESELLO, D. COLAZO, J.C. DAMILANO, A. & FERNÁNDEZ, A. Estabilidad del rendimiento de cultivares de maíz en la área de la EEA Pergamino en el trienio 1994/95 - 1996/97. **Rev. Tecnología Agropec. INTA Pergamino**. n. may/ago, p. 51-54, 1997.
- FINLAY, K.W.; WILKINSON, G.N. The analysis of adaptation in a plant-breeding programme. **Austr. J. Agric. Res.** v. 14, p. 742-754, 1963.
- FLORES, F.; MORENO, M.T.; MARTINEZ, A. & CUBERO, J.I. Genotype-environment interaction in faba bean: comparison of AMMI and principal coordinate models. **Field Crops Research**. v. 47, p. 117-127, 1996.
- GABRIEL, K.R. The biplot graphic display of matrices with application to principal component analysis. **Biometrika**. v. 58, p. 453-467, 1971.
- GABRIEL, K.R. Least squares approximation of matrices by additive and multiplicative models. **J. R. Statist. Soc.- Ser. B**. v. 40, n. 2, p. 186-196, 1978.
- GAUCH, H.G. Model selection and validation for yield trials with interaction. **Biometrics**. v. 44, p. 705-715, 1988.
- GAUCH, H.G. Full and reduced models for yield trials. **Theoretical Applied of Genetics**. v. 80, p. 153-160, 1990.
- GAUCH, H.G. **Statistical analysis of regional yield trials**: AMMI analysis of factorial designs. New York: Elsevier Science, 1992. 278 p.
- GAUCH, H.G. & ZOBEL, R.W. Predictive and postdictive success of statistical analysis of yield trials. **Theoretical Applied of Genetics**. v. 76, p. 1 -10, 1988.
- GAUCH, H.G. & ZOBEL, R.W. Accuracy and selection success in yield trial analyses. **Theoretical Applied of Genetics**. v. 77, p. 473-481, 1989.
- GAUCH, H.G. & ZOBEL, R.W. Imputing missing yield trial data. **Theoretical Applied of Genetics**. v. 79, p. 753-761, 1990.
- GAUCH, H.G. & ZOBEL, R.W. AMMI analysis of yield trials. In: KANG, M.S.; GAUCH, H.G., ed. **Genotype by environment interaction**. Boca Raton: CRC Press, 1996. Cap. 4, p. 85-122.

- GOLLOB, H.F. A statistical model which combines features of factor analytic and analysis of variance techniques. **Psychometrika**. v. 33, n. 1, p. 73-115, 1968.
- GOOD, I. J. Some applications of the singular decomposition of a matrix. **Technometrics**. v. 11, n. 4, p. 823-831, 1969.
- GREENBERG, B.G. Biplot . In: KOTZ, S. & JOHNSON, N.L., ed. **Encyclopedia of Statistical Science**. v. 1. New York: J. Wiley & Sons, 1982. p. 263-271.
- IEMMA, A.F. **Matrizes para estatística: um texto para profissionais de ciências aplicadas**. Piracicaba: DME/ESALQ-USP, 1988. 339 p.
- KANG, M.S.; MAGARI, R. New developments in selecting for phenotypic stability in crop breeding. In: KANG, M.S.; GAUCH, H.G., ed. **Genotype by environment interaction**. Boca Raton: CRC Press, 1996. Cap. 1, p. 1-14.
- KEMPTON, R.A. The use of biplots in interpreting variety by environment interactions. **Journal Agric. Science**. v. 103, p. 123-135, 1984.
- MAGARI, R. & KANG, M.S. SAS-STABLE: Stability analyses of balanced and unbalanced data. **Agronomy Journal**. v. 89, p. 929-932, 1997.
- MANDEL, J. A new analysis of variance model for non-additive data. **Technometrics**. v. 13, n. 1, p. 1-18, 1971.
- MANDEL, J. Use of the singular value decomposition in regression analysis. **The American Statistician**. v. 36, n. 1, p. 15-24, 1982.
- PEREIRA, A. da S. & COSTA, D. M. da . Análise de estabilidade de produção de genótipos de batata no Rio Grande do Sul. **Pesq. Agrop. Bras**. v. 33, n.4, p. 405-409, 1998.
- PIEPHO, H.P. Best linear unbiased prediction (BLUP) for regional yield trials: a comparison to additive main effects and multiplicative interaction (AMMI) analysis. **Theoretical Applied of Genetics**. v. 89, p. 647-654, 1994.
- PIEPHO, H.P. Robustness of statistical test for multiplicative terms in the additive main effects and multiplicative interaction model for cultivar trial. **Theoretical Applied of Genetics**. v. 90, p. 438-443, 1995.
- SHUKLA, G.K. Some statistical aspects of partitioning genotype-environmental components of variability . **Heredity**. v. 29, p. 237-245, 1972.
- SILVA, J.G.C.; BARRETO, J.N. Aplicação de regressão linear segmentada em estudos da interação gnótipo x ambiente. In: SIMP. ESTATÍSTICA APLIC. À EXPERIMENTAÇÃO AGRONÔMICA, 1. Campinas, Cargill. 1985. **Res**. p. 49-50.

- SOUZA, J. de. **Análise em componentes principais e suas aplicações**: métodos estatísticos nas ciências psicossociais. v. II. Brasília: Thesaurus, 1988. 68 p.
- WEBER, W.E.; WRICKE, G.; WESTERMANN, T. Selection of genotypes and prediction of performance y analysing genotype-by-environment interactions. In: KANG, M.S. & GAUCH, H.G., ed. **Genotype by environment interaction**. Boca Raton: CRC Press, 1996. Cap. 13, p. 353-371.
- WOLD, S. Cross-validatory estimation of the number of components in factor and principal components models. **Technometrics**. v. 20, n. 4, p. 397-405, 1978.
- WRICKE, G. Zur berechnung der okovalens bei sommerweizen und hafer. **Pflanzenzuchtung**. v. 52, p.127-138, 1965.
- ZOBEL, R.W.; WRIGHT, M.J.; GAUCH, H.G. Statistical analysis of a yield trial. **Agronomy Journal**. v. 80, p. 388-393. 1988.

7. APÊNDICES

7.1. Decomposição por valores singulares (DVS) - Exemplos e Propriedades

Exemplo 1:

Considere inicialmente a DVS de uma matriz quadrada simples ($\mathbf{A}_{(2 \times 2)}$):

$$\mathbf{A}_{(2)} = \begin{bmatrix} 4 & 2 \\ 1 & 3 \end{bmatrix} = \mathbf{U} \mathbf{S} \mathbf{V}' = \sum_k \lambda_k \mathbf{u}_k \mathbf{v}_k' = \lambda_1 \mathbf{u}_1 \mathbf{v}_1' + \lambda_2 \mathbf{u}_2 \mathbf{v}_2' = \mathbf{A}_1 + \mathbf{A}_2.$$

$k=1,2, \dots, p$; onde: $p = \min\{n_{\text{linhas LI}}, n_{\text{colunas LI}}\}=2$, é o posto de \mathbf{A} .

Para obter essa decomposição faz-se necessário, em princípio, encontrar os p escalares λ_k , bem como os p vetores \mathbf{u}_k e \mathbf{v}_k' (neste caso $k=1,2$; pois $p=2$), denominados *valores singulares* e *vetores singulares coluna* e *linha*, respectivamente. Uma forma de obter os *valores singulares* de uma matriz (\mathbf{A}) é extraíndo-se a raiz quadrada dos autovalores (λ_k^2) de uma das correspondentes matrizes simétricas $\mathbf{A}\mathbf{A}'$ ou $\mathbf{A}'\mathbf{A}$, que possuem os mesmos autovalores (Iemba, 1988). Os autovetores de $\mathbf{A}\mathbf{A}'$ associados a cada λ_k^2 , dispostos em ordem decrescente destes autovalores, são exatamente os p vetores-coluna que formam a matriz \mathbf{U} , ou seja, os *vetores singulares-coluna* de \mathbf{A} . Analogamente, os p autovetores de $\mathbf{A}'\mathbf{A}$ são os *vetores singulares-linha* de \mathbf{A} , dispostos como linhas da matriz \mathbf{V}' . As matrizes $\mathbf{A}\mathbf{A}'$ e $\mathbf{A}'\mathbf{A}$, neste caso, são dadas por:

$$\mathbf{A}\mathbf{A}' = \begin{bmatrix} 20 & 10 \\ 10 & 10 \end{bmatrix}, \text{ com polinômio característico dado por: } (20 - \lambda^2)(10 - \lambda^2) - (10 \cdot 10) = 0$$

$$\mathbf{A}'\mathbf{A} = \begin{bmatrix} 17 & 11 \\ 11 & 13 \end{bmatrix}, \text{ com polinômio característico dado por: } (17 - \lambda^2)(13 - \lambda^2) - (11 \cdot 11) = 0$$

Tomando-se o primeiro polinômio obtém-se as duas *raízes características* (ou *autovalores*) da matriz $\mathbf{A}\mathbf{A}'$:

$$200 - 20\lambda^2 - 10\lambda^2 + (\lambda^2)^2 - 100 = 0 \quad \therefore$$

$$(\lambda^2)^2 - 30\lambda^2 + 100 = 0 \Rightarrow \lambda_1^2 = 26,18034 \text{ e } \lambda_2^2 = 3,81966.$$

$$\text{Note que: } \lambda_1^2 + \lambda_2^2 = 26,18 + 3,82 = 30 = \sum_{i,j} a_{ij}^2 = (4)^2 + (2)^2 + (1)^2 + (3)^2 = 30.$$

Da mesma forma, as duas *raízes* ou *autovalores* de $\mathbf{A}'\mathbf{A}$ são obtidas a partir de:

$$221 - 17\lambda^2 - 13\lambda^2 + (\lambda^2)^2 - 121 = 0 \quad \therefore$$

$$(\lambda^2)^2 - 30\lambda^2 + 100 = 0 \Rightarrow \text{mesmas raízes } (\lambda_1^2 = 26,18034 \text{ e } \lambda_2^2 = 3,81966).$$

Os p autovetores normalizados (de comprimento unitário) de $\mathbf{A}\mathbf{A}'$ associados a cada autovalor λ_k^2 (neste caso: $k=1,2$) são obtidos por:

$$k=1) \begin{bmatrix} 20-26,18 & 10 \\ 10 & 10-26,18 \end{bmatrix} \begin{bmatrix} u_{11} \\ u_{12} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} -6,18 & 10 \\ 10 & -16,18 \end{bmatrix} \begin{bmatrix} u_{11} \\ u_{12} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \mathbf{u}_1 = \begin{bmatrix} 0,8506508 \\ 0,5257311 \end{bmatrix}, \text{ com } u_{11}^2 + u_{12}^2 = 1.$$

$$k=2) \begin{bmatrix} 20-3,82 & 10 \\ 10 & 10-3,82 \end{bmatrix} \begin{bmatrix} u_{21} \\ u_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 16,18 & 10 \\ 10 & 6,18 \end{bmatrix} \begin{bmatrix} u_{21} \\ u_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \mathbf{u}_2 = \begin{bmatrix} -0,5257311 \\ 0,8506508 \end{bmatrix}, \text{ com } u_{21}^2 + u_{22}^2 = 1.$$

Da mesma forma, os p autovetores normalizados de $\mathbf{A}'\mathbf{A}$ associados a cada λ_k^2 são obtidos por:

$$\begin{aligned}
 k=1) \quad & \begin{bmatrix} 17-26,18 & 11 \\ 11 & 13-26,18 \end{bmatrix} \begin{bmatrix} v_{11} \\ v_{12} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\
 & \begin{bmatrix} -9,18 & 11 \\ 11 & -13,18 \end{bmatrix} \begin{bmatrix} v_{11} \\ v_{12} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \mathbf{v}_1 = \begin{bmatrix} 0,7677517 \\ 0,6407474 \end{bmatrix}, \text{ com } v_{11}^2 + v_{12}^2 = 1. \\
 k=2) \quad & \begin{bmatrix} 17-3,82 & 11 \\ 11 & 13-3,82 \end{bmatrix} \begin{bmatrix} v_{21} \\ v_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\
 & \begin{bmatrix} 13,18 & 11 \\ 11 & 9,18 \end{bmatrix} \begin{bmatrix} v_{21} \\ v_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \mathbf{v}_2 = \begin{bmatrix} -0,6407474 \\ 0,7677517 \end{bmatrix}, \text{ com } v_{21}^2 + v_{22}^2 = 1.
 \end{aligned}$$

A decomposição por valores singulares (DVS) para a matriz \mathbf{A} é, portanto, dada pelas seguintes parcelas (neste caso duas, em razão do posto 2 de \mathbf{A}):

$$\begin{aligned}
 \lambda_1 \mathbf{u}_1 \mathbf{v}_1' &= (26,18)^{1/2} \cdot \begin{bmatrix} 0,8506 \\ 0,5257 \end{bmatrix} \cdot \begin{bmatrix} 0,7677 & 0,6407 \end{bmatrix} = \begin{bmatrix} 3,341640 & 2,788854 \\ 2,065225 & 1,723607 \end{bmatrix} = \mathbf{A}_1 \\
 \text{e} \\
 \lambda_2 \mathbf{u}_2 \mathbf{v}_2' &= (3,82)^{1/2} \cdot \begin{bmatrix} -0,5257 \\ 0,8506 \end{bmatrix} \cdot \begin{bmatrix} -0,6407 & 0,7677 \end{bmatrix} = \begin{bmatrix} 0,658360 & -0,78885 \\ -1,065225 & 1,276393 \end{bmatrix} = \mathbf{A}_2
 \end{aligned}$$

Fazendo-se: $\mathbf{A}_1 + \mathbf{A}_2$, obtém-se exatamente a matriz \mathbf{A} . Ademais, tomando-se a soma de quadrados dos elementos de \mathbf{A}_1 obtém exatamente λ_1^2 ($\sum_{ij} a_{1ij}^2 = \lambda_1^2 = 26,18034$), o mesmo verificando-se para a soma de quadrados dos elementos de \mathbf{A}_2 ($\sum_{ij} a_{2ij}^2 = \lambda_2^2 = 3,81966$). Note que \mathbf{A}_1 representa ainda uma aproximação para \mathbf{A} , no caso, a aproximação DVS de posto unitário para \mathbf{A} . Observe que esta aproximação (\mathbf{A}_1) é especialmente boa em termos da soma de quadrados dos elementos de \mathbf{A} (26,18 versus 30). Good (1969) mostra que uma aproximação DVS de posto n , com $n < p$, resulta também na menor soma de quadrados de desvios entre os elementos das duas matrizes. Ou seja, $\sum_{ij} (a_{ij} - a_{n,ij})^2$ é mínima para \mathbf{A}_n obtida pela soma das n primeiras parcelas da DVS de \mathbf{A} . O que significa dizer que nenhuma outra aproximação de posto n para a matriz \mathbf{A} resultará num valor igual ou inferior a este. No presente caso, nenhuma outra matriz de posto unitário é capaz de aproximar-se tanto de \mathbf{A} quanto \mathbf{A}_1 , em termos do parâmetro de dispersão $\sum_{ij} (a_{ij} - a_{1ij})^2$.

Equivalentemente, como já mencionado, a DVS de \mathbf{A} pode também ser obtida multiplicando-se as matrizes \mathbf{U} , \mathbf{S} e \mathbf{V}' , nesta ordem ($\mathbf{A} = \mathbf{U} \mathbf{S} \mathbf{V}'$), fazendo-se:

$$\mathbf{U} = \begin{bmatrix} 0,8506 & -0,5257 \\ 0,5257 & 0,8506 \end{bmatrix}; \quad \mathbf{S} = \begin{bmatrix} (26,18)^{1/2} & 0 \\ 0 & (3,82)^{1/2} \end{bmatrix}; \quad \text{e} \quad \mathbf{V} = \begin{bmatrix} 0,7677 & 0,6407 \\ -0,6407 & 0,7677 \end{bmatrix} \begin{matrix} \rightarrow \mathbf{v}_1' \\ \rightarrow \mathbf{v}_2' \end{matrix}$$

$\downarrow \qquad \qquad \downarrow \qquad \qquad \downarrow \qquad \qquad \downarrow$
 $\mathbf{u}_1 \qquad \qquad \mathbf{u}_2 \qquad \qquad \lambda_1 \qquad \qquad \lambda_2$

Exemplo 2:

Considere agora outra matriz $\mathbf{A}_{(3 \times 2)}$ de maior ordem e sua decomposição singular (DVS):

$$\mathbf{A} = \begin{bmatrix} 2 & 3 \\ 1 & 7 \\ 1 & 5 \end{bmatrix} = \mathbf{U} \mathbf{S} \mathbf{V}' = \sum_k \lambda_k \mathbf{u}_k \mathbf{v}_k' = \lambda_1 \mathbf{u}_1 \mathbf{v}_1' + \lambda_2 \mathbf{u}_2 \mathbf{v}_2' = \mathbf{A}_1 + \mathbf{A}_2.$$

$k=1,2, \dots, p(\mathbf{A})$; $p(\mathbf{A}) = \min\{n_{\text{linhas LI}}, n_{\text{colunas LI}}\}=2$;
 $\sum_{i,j} a_{ij}^2 = 89.$

Para matrizes de ordem elevada, a tarefa de encontrar autovalores e conseqüentemente valores singulares, bem como os respectivos autovetores e vetores singulares, torna-se impraticável manualmente. Entretanto, através de um programa computacional para álgebra de matrizes pode-se obter com facilidade as matrizes \mathbf{U} , \mathbf{S} e \mathbf{V} que determinam a DVS. No **SAS/ Proc IML**, por exemplo, fazendo-se uso do comando “*CALL SVD (U,S,V,A)*”, obtém-se as seguintes matrizes como resultado da DVS de \mathbf{A} :

$$\mathbf{U} = \begin{bmatrix} 0,3605 & 0,9204 \\ 0,7559 & -0,3835 \\ 0,5465 & -0,0767 \end{bmatrix} \quad \mathbf{S} = \begin{bmatrix} 9,3274 & 0,0000 \\ 0,0000 & 1,4142 \end{bmatrix}; \text{ e } \mathbf{V} = \begin{bmatrix} 0,2169 & 0,9762 \\ 0,9762 & -0,2169 \end{bmatrix}$$

$\downarrow \quad \downarrow \quad \downarrow \quad \downarrow$
 $\mathbf{u}_1 \quad \mathbf{u}_2 \quad \lambda_1 \quad \lambda_2 \quad \mathbf{v}_1 \quad \mathbf{v}_2$
 (vetores singulares-coluna de \mathbf{A}) (valores singulares de \mathbf{A}) (vetores singulares-linha de \mathbf{A})
 (autovetores de $\mathbf{A}\mathbf{A}'$) (autovalores de $\mathbf{A}\mathbf{A}'$ e $\mathbf{A}'\mathbf{A}$)^{1/2} (autovetores de $\mathbf{A}'\mathbf{A}$)
 $\hookrightarrow \lambda_1^2 + \lambda_2^2 = 87 + 2 = 89.$

Assim, \mathbf{A} é decomposta em duas parcelas (matrizes de posto unitário), quais sejam:

$$\mathbf{A}_1 = \lambda_1 \mathbf{u}_1 \mathbf{v}_1' = \begin{bmatrix} 0,7294 & 3,2824 \\ 1,5294 & 6,8824 \\ 1,1059 & 4,9765 \end{bmatrix}; \sum_{i,j} a_{1ij}^2 = 87.$$

$$\Rightarrow \mathbf{A}_1 + \mathbf{A}_2 = \begin{bmatrix} 2 & 3 \\ 1 & 7 \\ 1 & 5 \end{bmatrix} = \mathbf{A}$$

$$\mathbf{A}_2 = \lambda_2 \mathbf{u}_2 \mathbf{v}_2' = \begin{bmatrix} 1,2706 & -0,2824 \\ -0,5294 & 0,1176 \\ -0,1059 & 0,0235 \end{bmatrix}; \sum_{i,j} a_{2ij}^2 = 2.$$

Algumas propriedades dos resultados obtidos:

1)- A decomposição por valores singulares (DVS) corresponde a uma partição da soma de quadrados dos elementos da matriz decomposta ($\sum_{ij} a_{ij}^2 = \sum_k \lambda_k^2$; Mandel, 1971). Ademais esta partição é ortogonal, pois: $\mathbf{U}'\mathbf{U}=\mathbf{V}'\mathbf{V}=\mathbf{I} \Rightarrow \mathbf{A}_1'\mathbf{A}_2=\mathbf{A}_2'\mathbf{A}_1=\mathbf{0}$, sem perda de generalidade ($\mathbf{A}_i'\mathbf{A}_j=\mathbf{0}$, $\forall i \neq j$).

2)- Foram necessários dois termos (duas matrizes parciais) para reproduzir a matriz \mathbf{A} exatamente. Isso deve-se ao posto de \mathbf{A} ser 2 (duas colunas independentes). Adicionando-se uma linha e/ou uma coluna linearmente dependentes, a decomposição continua sendo feita com base em apenas dois termos, pois apenas dois de seus valores singulares (e autovalores de $\mathbf{A}\mathbf{A}'$ e de $\mathbf{A}'\mathbf{A}$) serão não nulos.

3)- Tomando-se \mathbf{A}_1 tem-se uma aproximação de posto ‘um’ para \mathbf{A} . Assim, se existir alguma lei regendo a construção da matriz \mathbf{A} e se tal lei tiver algum efeito sobre a soma de quadrados de seus elementos, não há dúvidas de que \mathbf{A}_1 fornece uma boa aproximação para essa medida (Good, 1969; Gabriel, 1978). Assim, \mathbf{A}_1 representa a aproximação de mínimos quadrados, de posto ‘um’, para a matriz \mathbf{A} , com: $\lambda_1^2 / \sum_k \lambda_k^2 = 0,98$. Se $p(\mathbf{A}) > 2$, então, $\mathbf{A}_1 + \mathbf{A}_2$ seria a aproximação DVS de posto ‘dois’ para \mathbf{A} e assim por diante.

Exemplo 3:

Na análise *AMMI* a *DVS* é aplicada a uma matriz de desvios, a matriz de interações **GE**, com restrições do tipo $\sum_i a_{ij} = \sum_j a_{ij} = 0$. Considere, então, a matriz do exemplo anterior, agora, aumentada de uma linha e uma coluna, de forma que cada linha e coluna somem zero. Tem-se, então, **B**_(4x3) :

$$\mathbf{B} = \begin{bmatrix} 2 & 3 & -5 \\ 1 & 7 & -8 \\ 1 & 5 & -6 \\ -4 & -15 & 19 \end{bmatrix} ; \quad p(\mathbf{B}) = \min\{n_{\text{linhas LI}}, n_{\text{colunas LI}}\} = 2;$$

$$\sum_{i,j} b_{ij}^2 = 816 = \text{SQ}(\text{desvios}) = \text{SQ}(\mathbf{GxE})$$

Sua *DVS* pode ser escrita em função das matrizes **U**, **S** e **V** a seguir (obtidas pelo SAS/IML):

$$\mathbf{U} = \begin{bmatrix} -0,2112 & 0,8181 & -0,3238 \\ -0,3732 & -0,5344 & -0,6884 \\ -0,2759 & -0,1913 & 0,6242 \\ 0,8603 & -0,0923 & -0,1780 \end{bmatrix} ; \quad \mathbf{S} = \begin{bmatrix} 28,521 & 0 & 0 \\ 0 & 1,598 & 0 \\ 0 & 0 & 0 \end{bmatrix} ; \quad \mathbf{V} = \begin{bmatrix} 0,1582 & 0,8010 & 0,5774 \\ -0,6146 & -0,5375 & 0,5774 \\ 0,7728 & -0,2635 & 0,5774 \end{bmatrix}$$

$\downarrow \quad \downarrow \quad \downarrow$
 $\mathbf{u}_1 \quad \mathbf{u}_2 \quad \mathbf{u}_3$
 (autovetores de **BB'**)

$\downarrow \downarrow$
 (valores singulares de **B**)

$\mathbf{v}_1 \quad \mathbf{v}_2 \quad \mathbf{v}_3$
 (autovetores de **B'B**)

$$\lambda_1^2 + \lambda_2^2 + \lambda_3^2 = 813,45 + 2,55 + 0 = 816.$$

(autovalores de **BB'** e **B'B**)

Dado o posto “dois” de **B**, sua decomposição exata pode ser feita por duas matrizes de posto “um”:

$$\mathbf{B}_1 = \lambda_1 \mathbf{u}_1 \mathbf{v}_1' = \begin{bmatrix} 0,9531 & 3,7025 & -4,6556 \\ 1,6839 & 6,5411 & -8,2250 \\ 1,2449 & 4,8357 & -6,0806 \\ -3,8819 & -15,0793 & 18,9611 \end{bmatrix} ; \quad \sum_{i,j} b_{1ij}^2 = 813,4479.$$

$$\mathbf{B}_2 = \lambda_2 \mathbf{u}_2 \mathbf{v}_2' = \begin{bmatrix} 1,0469 & -0,7025 & -0,3444 \\ -0,6839 & 0,4589 & 0,2250 \\ -0,2449 & 0,1643 & 0,0805 \\ -0,1181 & 0,0793 & 0,0389 \end{bmatrix} ; \quad \sum_{i,j} b_{2ij}^2 = 2,5521.$$

$$\Rightarrow \mathbf{B}_1 + \mathbf{B}_2 = \begin{bmatrix} 2 & 3 & -5 \\ 1 & 7 & -8 \\ 1 & 5 & -6 \\ -4 & -15 & 19 \end{bmatrix} ;$$

Isto porque $\mathbf{B}_3 = \lambda_3 \mathbf{u}_3 \mathbf{v}_3' = \mathbf{0}_{(4 \times 3)}$; visto que $\lambda_3 = 0$. Logo: $\mathbf{B} = \lambda_1 \mathbf{u}_1 \mathbf{v}_1' + \lambda_2 \mathbf{u}_2 \mathbf{v}_2' = \mathbf{B}_1 + \mathbf{B}_2$.

Algumas propriedades dos resultados:

1)- Nota-se, neste caso, que as restrições impõem que o posto de **B** (ou a matriz de interações) passa a ser dado por: $p(\mathbf{B}) = \min\{n_{\text{linhas LI}} - 1, n_{\text{colunas LI}} - 1\}$.

2)- Os valores 813,45 e 2,55 representam partições ortogonais da $\text{SQ}(\text{desvios}) = 816$. Assim, se **B** for uma matriz de interações de genótipos com ambientes: $\text{SQ}(\mathbf{B}) = \text{SQ}(\mathbf{GxE})$, com $(n_{\text{linhas}} - 1)(n_{\text{colunas}} - 1)$ graus de liberdade conhecidos e com duas partições ortogonais dadas por: $\text{SQ}(\mathbf{B}_1)$ e $\text{SQ}(\mathbf{B}_2)$. Gollob (1968) apresenta uma aproximação para os *graus de liberdade* destas somas de quadrados parciais, dada por: $n_{\text{linhas}} + n_{\text{colunas}} - 1 - 2n$; onde n é o número de termos singulares incluídos na aproximação de **B** por *DVS* (ex: $GL_{\text{SQ}(\mathbf{B}_1)} = 4 + 3 - 1 - 2 \cdot 1 = 4$; $GL_{\text{SQ}(\mathbf{B}_2)} = 4 + 3 - 1 - 2 \cdot 2 = 2$). Neste caso, observa-se que o primeiro termo da partição responde por 99,7% da $\text{SQ}(\mathbf{B}$ ou **GE**), ou seja: $\lambda_1^2 / \sum_k \lambda_k^2 = 0,997$.

3)- Nota-se também que **B**₁ realmente representa uma aproximação para a matriz **B** (no caso, uma aproximação de posto ‘um’); veja que exceto o valor b_{111} que difere bastante de b_{11} , os demais são bastante próximos para as duas matrizes.

4)- A *inversa generalizada de Moore-Penrose* para **B** é dada diretamente por: $\mathbf{B}^+ = \mathbf{V} \mathbf{S}^{-1} \mathbf{U}'$.

7.2. Associação entre a DVS e a representação *biplot*

Dada uma matriz $\mathbf{A}_{(l \times c)}$, de posto p (mínimo entre l e c) e sua DVS ($\mathbf{A} = \mathbf{U} \mathbf{S} \mathbf{V}'$), pode-se ainda escrevê-la como:

$$\mathbf{A} = \mathbf{U} (\mathbf{S}^{1/2} \mathbf{S}^{1/2}) \mathbf{V}' = (\mathbf{U} \mathbf{S}^{1/2}) (\mathbf{S}^{1/2} \mathbf{V}');$$

denotando-se por: $\mathbf{G} = (\mathbf{U} \mathbf{S}^{1/2})$ e $\mathbf{H}' = (\mathbf{S}^{1/2} \mathbf{V}')$; tem-se: $\mathbf{A} = \mathbf{G} \mathbf{H}'$.

onde: \mathbf{G} é a matriz $l \times p$ com l vetores de p componentes, denominados *marcadores de linhas*, cujos elementos correspondem às coordenadas para as l linhas (ex: indivíduos) num sistema p -dimensional de eixos ortogonais (eixos singulares⁹); e

\mathbf{H}' é a matriz $p \times c$ com c vetores também de p componentes, os *marcadores de colunas*, contendo as coordenadas das c colunas (ex: variáveis) no mesmo sistema p -dimensional de eixos singulares (Gabriel, 1971).

Dessa forma, pode-se representar, simultaneamente, as *linhas* e as *colunas* da matriz \mathbf{A} (p.ex: indivíduos e variáveis, no contexto da análise multivariada) num único gráfico chamado *biplot de Gabriel* ou simplesmente *biplot*. Teoricamente a representação seria feita num sistema p -dimensional de eixos ortogonais. Entretanto, na prática, isso só é possível até 3 dimensões. Assim, para matrizes de posto superior a 3 só é possível representar graficamente suas aproximações DVS de postos 1, 2 ou 3. Note que a denominação *biplot* vem do fato de representar as *linhas* e as *colunas* num mesmo gráfico e não da dimensão do espaço de representação. Logo, na prática, pode-se construir *biplots* na reta ($p=1$), no plano ($p=2$) ou no espaço tridimensional ($p=3$).

Exemplo:

Considere a matriz $\mathbf{A}_{(3 \times 2)}$ utilizada no *Exemplo 2* do Apêndice 1, com sua DVS já determinada. Admita, para fins de ilustração, que suas linhas representem 3 *genótipos* e suas colunas correspondam a 2 *ambientes* de teste.

$$\mathbf{A} = \begin{bmatrix} 2 & 3 \\ 1 & 7 \\ 1 & 5 \end{bmatrix} = \mathbf{U} \mathbf{S} \mathbf{V}' = (\mathbf{U} \mathbf{S}^{1/2}) (\mathbf{S}^{1/2} \mathbf{V}') = \mathbf{G} \mathbf{H}'. \\ p(\mathbf{A}) = \min\{3, 2\} = 2.$$

O posto igual a 2 indica que a matriz pode ser representada, exatamente, num *plano* determinado por dois eixos perpendiculares (os eixos singulares da DVS, aqui denotados SING1 e SING2). As matrizes \mathbf{U} e \mathbf{V} já foram obtidas no *Exemplo 2* do Apêndice 1. Assim, as matrizes \mathbf{G} e \mathbf{H}' possuem, respectivamente, 3 marcadores de linhas (*genótipos*) e 2 marcadores de colunas (*ambientes*), sendo dadas por:

$$\mathbf{G} = \mathbf{U} \mathbf{S}^{1/2} = \begin{bmatrix} 0,3605 & 0,9204 \\ 0,7559 & -0,3835 \\ 0,5465 & -0,0767 \end{bmatrix} \cdot \begin{bmatrix} (9,3274)^{1/2} & 0,000 \\ 0,000 & (1,4142)^{1/2} \end{bmatrix} = \begin{bmatrix} 1,1010 & 1,0945 \\ 2,3084 & -0,4560 \\ 1,6692 & -0,0912 \end{bmatrix} \begin{matrix} \rightarrow \mathbf{g}_1 \\ \rightarrow \mathbf{g}_2 \\ \rightarrow \mathbf{g}_3 \end{matrix}$$

e

$$\mathbf{H}' = \mathbf{S}^{1/2} \mathbf{V}' = \begin{bmatrix} (9,3274)^{1/2} & 0,000 \\ 0,000 & (1,4142)^{1/2} \end{bmatrix} \cdot \begin{bmatrix} 0,2169 & 0,9762 \\ 0,9762 & -0,2169 \end{bmatrix} = \begin{bmatrix} 0,6625 & 2,9813 \\ 1,1609 & -0,2578 \end{bmatrix}$$

\downarrow
 \mathbf{h}_1

\downarrow
 \mathbf{h}_2

^{9/} - Os *eixos singulares* são equivalentes às *componentes principais*, se a abordagem adotada fosse a de *análise de componentes principais* ao invés da DVS.

As setas indicam os 3 marcadores de linhas na matriz **G** e os 2 marcadores de colunas na matriz **H'**, todos com duas coordenadas, o que permite a representação exata de **A** no plano. Tomando-se apenas a primeira coordenada de cada marcador pode-se representar a matriz **A**, por meio de sua aproximação *DVS* de posto unitário (matriz **A**₁; *Exemplo 2* - Apêndice 1), numa reta definida pelo primeiro eixo singular. Este eixo responde por 98% da soma de quadrados dos elementos da matriz **A** (comentário 3 daquele mesmo *Exemplo*). De forma que, por este critério, o *biplot* aproximado (em uma só dimensão) representaria a matriz **A** de maneira bastante razoável.

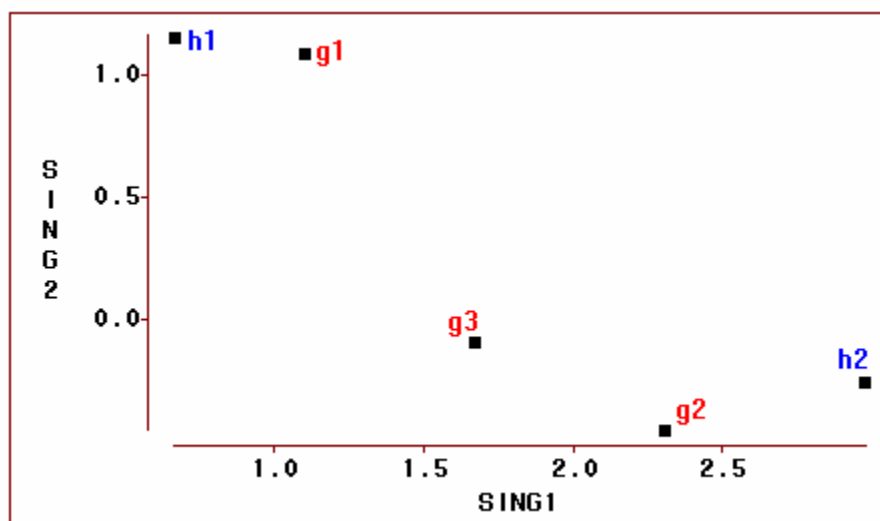
Considerando, então, as linhas de **A** como genótipos e suas colunas como ambientes tem-se:

$$\mathbf{G} = \begin{bmatrix} 1,1010 & 1,0945 \\ 2,3084 & -0,4560 \\ 1,6692 & -0,0912 \end{bmatrix} \begin{array}{l} \rightarrow \text{coordenadas para o genótipo 1.} \\ \rightarrow \text{coordenadas para o genótipo 2.} \\ \rightarrow \text{coordenadas para o genótipo 3.} \end{array}$$

$$\mathbf{H} = \begin{bmatrix} 0,6625 & 1,1609 \\ 2,9813 & -0,2578 \end{bmatrix} \begin{array}{l} \rightarrow \text{coordenadas para o ambiente 1.} \\ \rightarrow \text{coordenadas para o ambiente 2.} \end{array}$$

↓ ↓
SING1 SING2

De posse das coordenadas para os marcadores de genótipos e de ambientes para um sistema comum de eixos ortogonais, pode-se construir o *biplot*, objeto de representação da matriz **A**:



Uma propriedade desse tipo de representação é que, fazendo-se o produto interno (ou produto escalar) entre um marcador de linha, aqui denotado vetor g_i , e um marcador de coluna, denotado vetor h_j , reproduz-se o elemento a_{ij} da matriz **A**. O produto interno entre dois vetores u e v , com notações $\langle u, v \rangle$ ou $u'v$ (da álgebra linear), para $u = [u_1 \ u_2 \ \dots \ u_n]$ e $v = [v_1 \ v_2 \ \dots \ v_n]$, é dado por:

$$\langle u, v \rangle = u'v = u_1 v_1 + u_2 v_2 + \dots + u_n v_n = \sum_{i=1}^n u_i v_i ; \quad (i=1, 2, \dots, n \text{ componentes}).$$

Considere, então, alguns exemplos:

$$\text{genótipo 1 no ambiente 2: } g_1 = [1,1010 \quad 1,0945] \text{ e } h_2 = [2,9813 \quad -0,2578]$$

$$\langle g_1, h_2 \rangle = g_1' h_2 = (1,1010 \times 2,9813) + (1,0945 \times -0,2578) = 3,000 = a_{12} ;$$

genótipo 2 no ambiente 1: $g_2 = [2,3084 \quad -0,4560]$ e $h_1 = [0,6625 \quad 1,1609]$

$$\langle g_2, h_1 \rangle = g_2' h_1 = (2,3084 \times 0,6625) + (-0,4560 \times 1,1609) = 1,000 = a_{21} ; e$$

genótipo 3 no ambiente 2: $g_3 = [1,6692 \quad -0,0912]$ e $h_2 = [2,9813 \quad -0,2578]$

$$\langle g_3, h_2 \rangle = g_3' h_2 = (1,6692 \times 2,9813) + (-0,0912 \times -0,2578) = 5,000 = a_{32} .$$

Geometricamente, outras propriedades de interesse ainda podem ser exploradas nesse tipo de representação. Pode-se notar, por exemplo, que o eixo singular 1 é determinado, prioritariamente, por propriedades relacionadas ao ambiente 2 e aos genótipos 2 e 3. Por outro lado, o eixo 2 mostra-se fortemente relacionado a propriedades do ambiente 1 e do genótipo 2. Estas relações permitem, na prática, nominar estes eixos conforme as características ambientais e genéticas a eles associadas.

No contexto da análise multivariada, as colunas da matriz **A** normalmente representam variáveis para as quais um conjunto de indivíduos (as linhas de **A**) foram medidos. Assim, uma série de outras informações são ainda disponíveis. Por exemplo, o ângulo entre dois marcadores h_j e h_j' informa sobre a correlação linear entre esses dois vetores e, conseqüentemente, entre as variáveis a eles relacionadas (dado que: $\cos(\theta) = r_{(h_j, h_j')}$; onde θ é o ângulo entre h_j e h_j'). Logo, marcadores colineares sugerem alta correlação entre as variáveis correspondentes e marcadores perpendiculares sugerem independência entre as variáveis a eles associadas. Por este raciocínio, os ambientes aqui marcados por h_1 e h_2 , mostraram-se bastante independentes. Além disso, pode-se constatar que o genótipo g_1 mostrou-se mais adaptado ao ambiente h_1 , enquanto g_2 adaptou-se melhor a h_2 . Ademais, o comprimento de um vetor marcador relaciona-se à variância associada ao vetor ou variável correspondente na matriz **A**. Logo, o marcador g_1 , por exemplo, indica que o genótipo 1 apresentou respostas menos variáveis (menor variância) do que os genótipos 2 e 3, mostrando-se, portanto, mais estável.

A distância euclidiana entre dois marcadores g_i e g_i' (comprimento do vetor diferença: $g_i - g_i'$) informa ainda sobre a similaridade entre os indivíduos a que estes marcadores correspondem. Por exemplo, os genótipos g_2 e g_3 mostraram-se mais similares entre si do que os genótipos g_1 e g_2 . Multiplicando-se o comprimento do vetor resultante da projeção de g_i sobre h_j pelo comprimento do próprio h_j , obtém-se também o correspondente valor para a_{ij} (avaliação esta que pode ser feita diretamente no gráfico). Detalhes sobre estas e outras propriedades da representação em *biplot* são apresentados com o rigor formal nos trabalhos de Gabriel (1971), Bradu & Gabriel (1978), Greenberg (1982), Souza (1988), entre outros.

7.3. Um programa SAS® para análise AMMI

```

/*****
*   PROGRAMA PARA ANÁLISE DE INTERAÇÃO "GxE" PELO MÉTODO "AMMI"
*   (Dados relativos a médias de genótipos por ambientes)
*   -----
*   Os dados devem ser dispostos numa tabela completa com os genótipos
*   nas linhas e os ambientes nas colunas (matriz GxA).
*
*****
*   Para a execução do programa atualize os seguintes itens:
*   -----
*   a) o nome do arquivo a ser analisado (com sua localização),
*       na segunda linha de comandos (item '1' letra 'a');
*   b) os valores: A=_ e G=_ (números de ambientes e genótipos) na
*       terceira linha; e
*   c) os valores: GLERR=_ e QMERR=_ (GL e QM do Erro, ao nível de
*       médias, obtidos na análise conjunta), na segunda linha do item
*       '2' (letra 'a'); e também a quarta linha do mesmo item, para:
*       A1 A2 ... Aa (correspondente aos ambientes).
*****

```

1) ANOVA E DETERMINAÇÃO DA MATRIZ DE INTERAÇÕES GxE:

a) Leitura de dados:

```

----- */
data AMMIMED (drop=A G);
infile 'c:\dados\arquivo.PRN';
A=5; G=10;
do GENOT=1 to g;
  do AMB=1 to a;
    input PROD @@; output;
  end;
end;
options nodate;
/*

```

b) Ajuste dos efeitos principais e obtenção da matriz de interações GxE:

```

----- */
title "ESTUDO DE INTERAÇÃO GxE - MÉTODO AMMI";
title2 "Análise com dados de Médias (QM'originais/r)";
proc glm data=AMMIMED;
  class GENOT AMB;
  model PROD=GENOT AMB;
  run; title2 'Médias de Ambientes e Genótipos';
  lsmeans GENOT AMB;
  output out=GEAM r=GE;
run; title2;
proc sort data=AMMIMED;
by GENOT;
proc transpose data=AMMIMED out=MEDIAS (drop=_name_) prefix=A;
  var PROD;
  by GENOT;
run;

```

```

proc print data=MEDIAS (drop=GENOT);
  title2 "A matriz X de médias (gxa)";
run;
proc sort data=GEAM;
by GENOT;
proc transpose data=GEAM out=GXE (drop=_name_) prefix=A;
  var GE;
  by GENOT;
run;
proc print data=GXE (drop=GENOT);
  title2 "Matriz de interações GxE (gxa)";
run;
/*

```

2) A DECOMPOSIÇÃO DA INTERAÇÃO GxE :

a) DVS da matriz de interações GxE:

```

----- *;

proc iml;
GLERR=90; QMERR=9679.597;
use work.MEDIAS; setin work.MEDIAS;
read all var {A1 A2 A3 A4 A5} into X;
ng=nrow(X); na=ncol(X);
GM=X[, :]; AM=X[:, :]; GG=X[:, :];
CORR=(ng*na)*(GG##2); SQTR=ssq(X)-CORR;
SQGEN=na*ssq(GM)-CORR;
SQAMB=ng*ssq(AM)-CORR;
SQINT=SQTR-SQGEN-SQAMB;
GLMOD=(ng-1)+(na-1)+(ng-1)*(na-1);
XG=GM*J(1,na,1);XA=J(ng,1,1)*AM;
MG=J(ng,1,1)*GG*J(1,na,1);
GE=X-XG-XA+MG;
EIGV=EIGVAL(GE*GE`);
NNUL1=EIGV<-1E-9; NNUL2=EIGV>1E-9; NNUL=NNUL1+NNUL2;
r=(NNUL^=0)[+,];
title2 'Decomposição por Valores Singulares (A=USV`) da matriz GE';
SQGE=ssq(GE);
print 'SQ dos elementos da matriz GE = SQ (GxE):' SQGE;
print 'O posto da matriz GE:' r ;
CALL SVD(eG,VS,eA,GE);S=DIAG(VS[1:r,]);U=eG[,1:r];VT=eA[,1:r]`;
print 'Matriz de valores singulares (S):' S [format=12.4] /
  'Matriz de autovetores (U) associados a genótipos (AA`):' U [format=12.6] /
  'Matriz de autovetores (V`) associados a ambientes (A`A):' VT [format=12.6] /;
/*

```

b) Desdobramento da SQ(GxE) por AMMI (ACP ou DVS):

```

----- *;

Li=VS##2;AUTV=Li[1:r,];
SAUTV=SUM(Li);
PRP=INV(SAUTV)*Li;PROP=PRP[1:r,];
RESD1=SAUTV-Li[1:1,];
PACUM=CUSUM(PROP)*100;
LINH=(1:r)`;
RESUMO=LINH||AUTV||PROP||PACUM;
colu={'CP/Eixo' 'Autovalor' 'Prop./CP' '% Acumulada'};

```

```

title2 'Resultados do desdobramento da SQ(GxE) por AMMI (DVS ou ACP)';
print 'Notas:',
      '-----',
      'Autovalores da ACP <=> Desdobramento da SQ(GxE) por AMMI',
      'Proporção da SQ(GxE) atribuída a cada CP (ou eixo singular de interação)',
      'Percentagem da SQ(GxE) acumulada até o n-ésimo CP (ou eixo singular)'
      '-----';
print RESUMO [colname=colu format=12.4];
print, 'Soma dos Autovalores <=> SQ(GxE):' SAUTV /;
title2;
/*

```

c) Escores de genótipos e ambientes e aproximações sucessivas de GE e X:
 ----- *

```

n=0;
GXEI=j(ng,na,0); GLACU=0;
SQGEI=ssq(GXEI);
PROPI=INV(SAUTV)*SQGEI;
PSQTR=inv(SQTR)*(SQGEN+SQAMB);
XP0=XG+XA-MG;
print '=====',
      'Resultados da avaliação do modelo: AMMI' n;
print ' SQmod/SQtrat:' PSQTR ' SQge(mod)/SQge:' PROPI,
      '===== /;
print 'Matriz de respostas preditas (gxa): Modelo AMMI' n;
print XP0 [format=12.2] /;
do n=1 to r by 1; i=n;
VG=eG[,i]; ES=S[i,i]; VA=eA[,i];
IN=VG*ES*VA`;
G=VG*(sqrt(ES));ESGE=G`;
H=sqrt(ES)*VA;ESAM=H`;
if i=1 then ECP1=G//H;
if i=2 then ECP2=G//H;
if i=3 then ECP3=G//H;
GEI=G*H`; ZE=IN-GEI;
SQGEI=SQGEI+ssq(GEI);
PROPI=INV(SAUTV)*SQGEI;
GLAMI=ng+na-1-(2*i); QMAMI=inv(GLAMI)*(ssq(GEI));
GLACU=GLACU+GLAMI;
GLRAMI=(ng-1)*(na-1)-GLACU;
if i<r then QMRAMI=inv(GLRAMI)*(SAUTV-SQGEI); else QMRAMI=0;
FMODI=inv(QMERR)*QMAMI; FRESI=inv(QMERR)*QMRAMI;
GXEI=GXEI+GEI;
XPI=XG+XA-MG+GXEI;
PSQTR=inv(SQTR)*(SQGEN+SQAMB+SQGEI);
prob1=1-probf(FMODI,GLAMI,GLERR);
if i<r then prob2=1-probf(FRESI,GLRAMI,GLERR); else prob2={.};
FV1=GLAMI||QMAMI||FMODI||PROB1; FV2=GLRAMI||QMRAMI||FRESI||PROB2;
FV3=GLERR||QMERR||{.}||{.};
FV=FV1//FV2//FV3;
lin={'AMMI' n '(GxE)' 'Res.GxE/AMMI' n 'Erro médio'};
col={'GL' 'QM' 'teste F' 'Pr > F'};
print 'Resultados da avaliação do modelo: AMMI' n;
print FV [rowname=lin colname=col format=12.4];
print ' SQmod/SQtrat:' PSQTR ' SQge(mod)/SQge:' PROPI /;
print 'Escores para genótipos e ambientes / Modelo ajustado: AMMI' n,
      'Escores de genótipos' ESGE [format=12.4],
      'Escores de ambientes' ESAM [format=12.4] /;
print 'Aproximação DVS para a Matriz de interações (gxa): Modelo AMMI' n;
print GXEI [format=12.4] /;

```

```

print  'Matriz de respostas preditas (gxa): Modelo AMMI' n;
print  XPI [format=12.2] /;
end;
/*

```

d) Resultados para a representação gráfica (Biplot):
 ----- *;

```

GEN=1:ng;AMB=1:na;
OBJ=(GEN||AMB)`;
PROD=GM//AM`;
RES=OBJ||PROD||ECP1||ECP2||ECP3;
VAR={'GEN_AMB' 'MEDIAS' 'IPCA1' 'IPCA2' 'IPCA3'};
create RESULT from RES [colname=VAR];
append from RES;
close RESULT;
quit;
proc print data=RESULT;
title2 "Resultados úteis para a representação gráfica em Biplot";
title3 "Médias e escores (3 primeiros CP's) para genótipos e ambientes";
run;

```