

**MÉTODOS PARA ESCOLHA DE COMPONENTES EM MODELO DE EFEITO
PRINCIPAL ADITIVO E INTERAÇÃO MULTIPLICATIVA (AMMI)**

CARLOS TADEU DOS SANTOS DIAS

**Tese apresentada à Escola Superior de
Agricultura “Luiz de Queiroz”, Universidade
de São Paulo, como parte dos requisitos do
Concurso para Professor Livre-docente no
Departamento de Ciências Exatas.**

PIRACICABA

**Estado de São Paulo – Brasil
Dezembro – 2005**

Dedico:

A minha família: Elvina Maria de Lima
Dias, esposa, amiga e companheira;
Victor Lima Dias e Laura Lima Dias,
filhos que me dão forças e alegrias.

AGRADECIMENTOS

Ao Departamento de Ciências Exatas da Escola Superior de Agricultura “Luiz de Queiroz” pela oportunidade que me foi conferida de desenvolver esse trabalho no exterior;

Ao Professor Wojtek Krzanowski, da Universidade de Exeter- Reino Unido, pelo seu espírito humano, pelo seu zelo à pesquisa, ética e respeito ao profissional, pela orientação segura e paciente, pela compreensão dispensada e sobretudo pelos ensinamentos;

Aos Professores Clarice Garcia Borges Demétrios e Roland Vencovsky pelo apoio e incentivo para desenvolver essa pesquisa no exterior.

Aos Professores Décio Barbin pelas sugestões e correções desse material e à Professora Roseli Aparecida Leandro pelo apoio quando da minha permanência no exterior.

As amigas: Helena Cardoso, Luciane Brajão, Milena Ramirez de Souza, Solange Sabadin e Expedita Azevedo pela ajuda na digitação e preparação do memorial.

Aos alunos Osmir José Lavoranti e Maria Clideana Cabral Maia, pela ajuda na preparação do material.

A todos os Professores e funcionários do Departamento de Ciências Exatas, que ficaram com o peso da carga de trabalho quando da minha saída ao exterior.

A minha família pelo apoio irrestrito, pelo convívio prazeroso, por suportar os momentos ausentes, pelo amor;

À FAPESP e CNPq, instituições financiadoras dos meus estudos.

SUMÁRIO

Página

RESUMO.....	vi
SUMMARY.....	vii
1 INTRODUÇÃO.....	1
2 REVISÃO DE LITERATURA.....	5
2.1 Interação genótipo ambiente.....	5
2.2 Conceitos de adaptabilidade e estabilidade fenotípica.....	8
2.3 Métodos de avaliação da estabilidade fenotípica.....	11
2.4 O modelo AMMI.....	14
2.5 Determinação do número ótimo de termos multiplicativos no modelo AMMI	16
2.6 Testes de significância de termos multiplicativos.....	17
2.7 Avaliação Preditiva Usando Validação Cruzada.....	21
2.8 Métodos “leave-one-out”.....	23
2.9 Histórico dos modelos AMMI.....	29
2.10 Métodos de reamostragem.....	32
3 MATERIAL E MÉTODOS.....	49
3.1 Aspectos técnicos.....	49
3.1.2 Dados e Modelo.....	49
3.1.3 Estudo de simulação.....	50
4 RESULTADOS E DISCUSSÃO.....	56
5 CONCLUSÕES.....	59
REFERÊNCIAS BIBLIOGRÁFICAS.....	60

MÉTODOS PARA ESCOLHA DE COMPONENTES EM MODELO DE EFEITO PRINCIPAL ADITIVO E INTERAÇÃO MULTIPLICATIVA

Autor: Carlos Tadeu dos Santos Dias

RESUMO

Esta pesquisa visou estudar alguns métodos mais utilizados para definir o número de componentes nos modelos de efeitos principais aditivos e interação multiplicativa (AMMI). Esses modelos permitem ao analista detectar interações entre linhas e colunas em uma tabela de dupla entrada. Entretanto, existem muitos métodos propostos na literatura para determinar o número de componentes multiplicativos a incluir no modelo AMMI. Esses métodos tipicamente fornecem diferentes resultados para um particular conjunto de dados, de tal forma que os usuários desses modelos necessitam de alguma orientação sobre qual método usar. Nessa pesquisa foram comparados quatro métodos comumente usados, cruzando-se uma amplitude de condições usando-se dados simulados baseados em experimentos reais e fornecendo-se algumas recomendações gerais. Mostrou-se que o método desenvolvido por Eastment-Krzanowski apresenta boa performance. Cada um dos outros métodos exibe baixo desempenho em pelo menos algumas das situações estudadas.

METHODS FOR CHOOSING COMPONENTS IN THE ADDITIVE MAIN EFFECT AND MULTIPLICATIVE INTERACTION (AMMI) MODEL

Author: Carlos Tadeu dos Santos Dias

SUMMARY

The additive main effect and multiplicative interaction (AMMI) model allows the analyst to detect interactions between rows and columns in a two-way table. However, there are many methods proposed in the literature to determine the number of multiplicative components to include in the AMMI model. These methods typically give different results for any particular data set, so the user needs some guidance as to which methods to use. In this paper we compare four commonly used methods using simulated data based on real experiments, and provide general recommendations. Our results show that the Eastment-Krzanowski method performs best; each of the other methods exhibits poor performance in at least some of the situations studied.

MÉTODOS PARA ESCOLHA DE COMPONENTES EM MODELO DE EFEITO PRINCIPAL ADITIVO E INTERAÇÃO MULTIPLICATIVA (AMMI)

1 INTRODUÇÃO

Muitos estudos observacionais ou experimentais produzem uma tabela de dupla entrada de dados a ser analisada. A causa mais comum de tais dados é um experimento de dois fatores; se um fator tem g níveis, o segundo tem e níveis, e há n observações repetidas em cada combinação de níveis de fator, então os resultados experimentais serão dispostos em uma tabela tendo $(g \times n)$ linhas e e colunas. Teoria estatística padrão, como colocada em qualquer livro elementar sobre modelagem ou análise (Steel e Torrie, 1980; Krzanowski, 1998), irá fornecer um mecanismo para investigar os efeitos principais dos fatores e a interação entre eles. Se os dados são quantitativos e normalmente distribuídos, então a análise de variância (ANOVA) é a estrutura apropriada. Mesmo se normalidade não pode ser assumida, a mesma estrutura pode ser adotada mas testes de hipóteses permutacionais devem substituir a suposição distribucional (Manly, 1997).

Para um teste de interação ser válido sob a estrutura ANOVA, devem existir no mínimo duas repetições em cada combinação do fator, isto é, deve-se ter $n > 1$. Muitos estudos práticos importantes são conduzidos, entretanto, em que essas condições não são satisfeitas mas onde a investigação da interação é importante. Um exemplo ocorre no melhoramento de plantas, onde ensaios multi-ambientes (MET-*multi-environment trials*) são importantes para testar adaptação geral e específica de cultivar. Uma cultivar desenvolvendo-se em diferentes ambientes frequentemente mostrará significativa flutuação na performance de produção relativa a outras cultivares. Essas mudanças são influenciadas por diferentes condições ambientais e são referidas como interação genótipo \times ambiente (GEI).

Um exemplo típico de uma tabela de dupla entrada de dados surge na análise do MET, onde as linhas da tabela são genótipos e as colunas são ambientes nos quais os genótipos são testados. A presença de GEI exclui modelos de interpretação simples que

tem somente efeitos principais aditivos de genótipos e ambientes (Mandel, 1971; Crossa, 1990; Kang e Magari, 1996). Por outro lado, adaptação específica de genótipo a subconjunto de ambientes é assunto fundamental a ser estudado em melhoramento de plantas porque um genótipo pode desempenhar bem sob condições ambientais específicas e pode ter um pobre desempenho sob outras condições.

Se $n = 1$, então o modelo de ANOVA padrão para os dados na tabela é

$$y_{ij} = \mu + g_i + e_j + \delta_{ij} \quad i = 1, \dots, g, \quad j = 1, \dots, e$$

onde:

y_{ij} é a observação na i -ésima linha e j -ésima coluna da tabela;

μ é o efeito da média geral

g_i representa os efeitos da i -ésima linha, e

e_j o efeito da j -ésima coluna.

O termo δ_{ij} pode ser visto tanto como resíduo como termos da interação entre linhas e colunas. O modelo acima é dito modelo aditivo. Entretanto, é possível que os termos da interação δ_{ij} ainda contenham alguma estrutura; Gollob (1968) e Mandel (1969, 1971) em conseqüência propuseram modelá-los por um conjunto de componentes multiplicativos mais erros residuais.

$$\delta_{ij} = \sum_{k=1}^m \alpha_{ik} \lambda_k \gamma_{jk} + \varepsilon_{ij}$$

o número de componentes $m \leq \min\{g - 1, e - 1\}$ é suposto ser escolhido de tal forma que o resíduo ε_{ij} represente ruído branco.

Combinando as últimas duas expressões obtém-se um modelo de efeito principal aditivo e interação multiplicativa (AMMI) para tabela de dupla entrada:

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

Efeitos não aditivos são freqüentemente observados em tabelas de dupla entrada, e, como Daniel (1976) chama a atenção, a não aditividade está associada com apenas poucas linhas ou colunas da tabela. Como observado por Snee (1982), a interpretação da não aditividade é um problema menor se observações repetidas estão presentes para cada uma das células da tabela. Quando existe somente uma observação por célula, entretanto, não é possível distinguir heterogeneidade de variância entre linhas ou colunas e interação apenas dos dados observados. Nesta situação, o uso de um modelo que detectará uma variedade de diferentes tipos de não aditividade em tabelas de dupla entradas pode ser muito útil.

Por essa razão, um objetivo primário aqui é determinar o número ótimo de termos multiplicativos no modelo AMMI. Também, boa predição da verdadeira resposta do ensaio em cada célula da tabela de dupla entrada pode ser alcançada truncando-se o modelo AMMI, tal que critérios para determinar o número de componentes necessários para explicar o padrão de interação tem sido objeto de algumas pesquisas.

Existem muitos testes na literatura para decidir quantos componentes devem ser retidos no modelo aditivo e interação multiplicativa. Eles incluem aqueles de Gollob (1968), Mandel (1969, 1971), Cornelius (1993), Corsten e van Eijnsberger (1972), Johnson e Graybill (1976), Hegemann e Johnson (1976), e Yochmowitz e Cornell (1978) entre outros. Muitos destes testes são baseados na ANOVA como decomposição da soma de quadrados, e requer associação de graus de liberdade para uso no teste F resultante. Este assunto é ainda objeto de debate (exemplo, witness the correspondece in *Biometrics* 1993, vol. 49, p. 952-954), mas ao longo dos anos parece que o teste devido a Gollob e Cornelius tem sido mais favorável em aplicações. Piepho (1995) investigou a robustez (para as suposições de homogeneidade e normalidade dos resíduos) de algum destes testes. Ele comenta que o teste F aplicado de acordo com o critério de Gollob (1968) é liberal, pois ele seleciona muitos termos multiplicativos. Dos quatro métodos que ele estudou o teste proposto por Cornelius foi o mais robusto. O autor então

recomenda que avaliações preliminares devam ser conduzidas para verificar a validade das suposições se algum dos outros testes for usado.

Mais recentemente, Dias e Krzanowski (2003) descreveram dois métodos 'leave-one-out' para determinar o número m de componentes de interação a ser retidos no modelo AMMI, e estes métodos parecem oferecer alguma vantagem sobre aqueles baseados nos testes F os testes de Gollob e Cornelius. Contudo, todos estes métodos tipicamente dão diferentes respostas quando aplicados a um conjunto de dados. Dessa forma, orientação de ordem prática é necessária no que diz respeito as suas propriedades. O propósito do presente trabalho é dessa forma conduzir uma comparação dos métodos para estabelecer condições sob as quais cada método pode apresentar desempenho satisfatório. Isso é feito por meio de um estudo de simulação baseado em dados experimentais reais.

2 REVISÃO DA LITERATURA

2.1 Interação entre genótipos e ambientes (G×E)

A manifestação fenotípica dos caracteres métricos resulta do efeito conjunto do genótipo, do ambiente específico e da interação destes fatores. Obviamente, se a amostra do material genético pesquisado fosse representativa da população original e sua avaliação fosse feita em um grande número de ambientes contrastantes, a média dos caracteres em estudo não representaria teoricamente, um desvio fenotípico significativo; neste caso, era de se esperar aleatoriedade dos efeitos e a média dos caracteres considerados se ajustaria à da população, admitindo-se não ter havido seleção, migração ou mutação. Na prática, contudo, tal procedimento torna-se inexequível, sendo que, basicamente, todos os programas de melhoramento genético envolvem pelo menos três etapas: escolha dos parentais que darão origem à população-base, seleção das progênes superiores dessa população e sua avaliação em um grande número de ambientes. Quando se avaliam materiais geneticamente distintos em uma série de ambientes, comumente, o componente interação entre genótipos e ambientes (G×E) aparece afetando o ganho com a seleção.

A interação (G×E) é definida como o comportamento diferencial dos genótipos em função da diversidade ambiental. Neste sentido, na presença da interação, os resultados das avaliações podem mudar de um ambiente para outro, ocasionando mudanças na posição relativa dos genótipos ou mesmo na magnitude das suas diferenças (Falconer & Mackay, 1996). Para Chaves (2001), a interação (G×E) deve ser encarada como um fenômeno biológico em suas implicações no melhoramento de plantas e não como um simples efeito estatístico, cumprindo buscar a explicação evolutiva do evento se se quiser tirar proveito de seus efeitos benéficos, bem como para contornar seus efeitos indesejáveis sobre a avaliação de genótipos e recomendação de cultivares. Diferenças em adaptação de genótipos em populações resultam, evidentemente, de diferenças de constituição gênica para os caracteres importantes nesta adaptação. A reação diferencial às mudanças ambientais pode-se dar desde os mecanismos de regulação gênica até caracteres morfológicos finais.

Cockerham (1963) atribuiu o aparecimento de interações (G×E) como sendo devido a respostas diferenciais do mesmo conjunto gênico em ambientes distintos ou pela expressão de diferentes conjuntos gênicos em diferentes ambientes. Quando um mesmo conjunto de genes se expressa em diferentes ambientes, as diferenças nas respostas podem ser explicadas pela heterogeneidade das variâncias genéticas e experimentais ou por ambas, e, quando diferentes conjuntos de genes se expressam em ambientes distintos, as diferenças nas respostas explicam-se pela inconsistência das correlações genéticas entre os valores de um mesmo caráter em dois ambientes (Falconer, 1989). Segundo Cruz & Regazzi (1994), a interação (G×E) também pode surgir em função de fatores fisiológicos e bioquímicos próprios de cada genótipo cultivado. Chaves et al. (1989) relatam ainda que a falta de ajuste do modelo estatístico adotado ao conjunto de dados pode ser uma das causas da interação (G×E) significativa.

No que concerne à contribuição dos diferentes ambientes para a interação, sua variância é atribuída a uma porção previsível (fixa) que faz parte da variação e que pode ser medida em magnitude e em direção, seu caráter sistemático facilita razoável controle; já a sua porção imprevisível (aleatória), faz parte da variação experimental não controlada, por ter natureza indeterminada, isto é, de sentido e direção incertos. Cruz (2001) ressalta que o ambiente atua isoladamente ou em interação com os genótipos, de tal forma que, quando se considera uma série de ambientes, sua magnitude pode ser estimada e seus efeitos devem ser avaliados. Estudos sobre a interação genótipos x ambientes devem pormenorizar o comportamento tanto dos genótipos disponíveis, em estudos de adaptabilidade e estabilidade, quanto dos ambientes estudados. Para Fehr (1987), as variáveis imprevisíveis são as que mais contribuem para as interações genótipos x anos. Portanto, Chaves (2001) entende ser o efeito do ambiente, em realidade, um confundimento de efeitos fixos e aleatórios. Então, quão mais previsível for o ambiente da lavoura, maior será a possibilidade de aproveitamento dos efeitos da interação, sendo ideal que os efeitos de ambientes pudessem ser tomados como fixos ou, pelo menos, que a porção fixa da interação pudesse ser medida e os resultados utilizados na recomendação de cultivares.

Segundo Vencovsky & Barriga (1992), o fato de a interação (G×E) mostrar-se significativa não diz muito, havendo a necessidade de subdividi-la para saber o que ela contém. Quando um grupo de genótipos é avaliado no mínimo em dois ambientes é possível decompor a componente interação em duas partes (Vencovsky, 1987). A interação (G×E) neste caso, contém uma parte que é função da diferença entre a variância genética do caráter nos dois ambientes e outra parte que decorre da falta de correlação genética entre os tratamentos genéticos, de um ambiente para outro. A primeira parcela é denominada interação simples e não acarreta maiores problemas para a seleção; a segunda parcela é denominada complexa, e pela falta de correlação, representa dificuldade na seleção de genótipos de adaptação mais ampla.

Entre as aspirações dos melhoristas está a busca de materiais genéticos que, ao serem avaliados nos vários ambientes, não mostrem interações significativas com eles, conferindo, assim, uma maior segurança na recomendação de cultivares. No entanto, somente uma pequena porção da interação pode ser atribuída a determinantes ambientais conhecidos, sendo que a maior parte é inexplicável na análise estatística (Rosse, 1999). Existem na literatura várias alternativas para se atenuar o efeito significativo da interação (G×E).

Uma opção seria a identificação de cultivares específicos para cada ambiente; porém, seu alto custo de execução a inviabiliza (Ramalho et al. 1993; Ribeiro, 1999).

A segunda maneira é realizar zoneamento ecológico de uma área extensa, ou seja, promover estratificação de uma área heterogênea em sub-regiões mais uniformes, de forma que, dentro das regiões os genótipos não interajam significativamente com os ambientes (Rosse, 1999). A principal crítica feita a essa técnica está no fato de o agrupamento ser feito com base nas diferenças macro ambientais, fato que o torna vulnerável às oscilações imprevisíveis que possam ocorrer no ambiente. Além disso, o zoneamento ecológico permite a redução do número de locais para avaliação de cultivares, pois identifica aqueles com características mais semelhantes. Como agravante Rosse (1999) acrescenta que essa técnica não permite o controle da interação entre genótipos e anos. O agrupamento de ambientes pode ser feito por meio de características edafo-climáticas ou do comportamento dos genótipos (Millgan, 1994). No entanto, a

maioria dos métodos tem utilizado as informações referentes à interação G×E (May & Kozub, 1995; Atlin et al. 2000).

A última alternativa, que é a que tem sido mais empregada, consiste em identificar genótipos com maior estabilidade fenotípica (Vencovsky & Barriga, 1992; Ramalho et al. 1993). Um grande número de métodos estatístico-genéticos foi proposto para mensurar a estabilidade, auxiliando sobremaneira a seleção e descarte de cultivares. As diferenças entre estes vários métodos dependem dos conceitos de estabilidade adotados pelos pesquisadores e dos próprios procedimentos estatísticos empregados na sua estimação (Rosse, 1999).

2.2 Conceitos de adaptabilidade e estabilidade fenotípica

Adaptabilidade e estabilidade, embora sejam fenômenos relacionados, não devem ser considerados iguais (Vencovsky & Barriga, 1992).

Lin et al. (1986) classificaram o conceito de estabilidade em três tipos. No tipo 1, o genótipo é considerado estável se sua variância entre os ambientes for pequena. Este conceito acomoda de certa forma a definição de estabilidade feita por Lewis (1954) e previsibilidade de comportamento de Mariotti et al. (1976); é chamado de “estabilidade no sentido biológico” (Becker, 1981), que caracteriza um genótipo com comportamento constante com a variação do ambiente e está em concordância com o termo homeostase proposto por Lerner (1954) ou “canalização” de C. H. Waddington em 1953 (citado por Mettler & Gregg, 1973); está no mesmo sentido de estabilidade de produção designado por Heinrich et al. (1983) e na consistência no ranqueamento denotado por Yue et al. (1997); correspondente a genótipo “bem tamponado” ou “bem ajustado” de Allard & Bradshaw (1964), sendo também equivalente às definições de estabilidade absoluta propostas por Finlay & Wilkinson (1963) e coincidente à denominação de estabilidade de Eberhart & Russel (1966); análogo ainda ao termo estático de J. Léon de 1985 (citado por Becker & Léon, 1988). Vencovsky & Torres (1988) e Fox et al. (1997) usaram estabilidade temporal fazendo referência à variação não significativa do genótipo frente às flutuações climáticas ao longo de anos agrícolas,

dentro de uma localidade. Assim sendo, esse comportamento pode não ser desejado, pois o genótipo não responde à melhoria do ambiente com o aumento da produção, além de estar comumente relacionado a uma menor produtividade (Becker & León 1988 e Ramalho et al. 1993). A estabilidade do tipo 1 pode ser medida pela variância de cada genótipo nos diferentes ambientes e é útil para características que devem ser mantidas, tal como resistência a patógenos e pragas (Nunes, 2000).

Na estabilidade do tipo 2, o genótipo é considerado estável se sua resposta ao ambiente for paralela à resposta média de todos os materiais avaliados no experimento, o que ocorre quando o genótipo possui interações mínimas com o ambiente. É denominada por Becker (1981) como “estabilidade no sentido agrônômico”; contempla o conceito de estabilidade dinâmica de J. León de 1985 (citado por Becker & Leon, 1988); sendo chamada pelo termo adaptabilidade (Mariotti et al. 1976), representando um nível paralelo de resposta do genótipo de forma a aproveitar vantajosamente os estímulos ambientais; ajustando-se à denominação de adaptabilidade de Eberhart & Russel (1966). Ratificada posteriormente por (Cruz & Regazzi 1994); nesta mesma linha, Vencovsky & Torres (1988) apresentaram adaptabilidade como sinônimo de estabilidade espacial ou geográfica. Nestes termos, é esperado que não haja interações significativas entre o desempenho produtivo e as oscilações climáticas entre anos. Fox et al. (1997) adotaram este mesmo conceito ao se referirem à adaptabilidade como reflexa às dimensões espaciais. Ela também tem sido preferida por identificar genótipos com o potencial de se manterem estáveis entre os melhores em todos os ambientes. Entretanto, alguns autores comentam que as inferências neste tipo de estabilidade são exclusivas ao grupo de genótipos avaliados. Dessa forma, um material estável em determinado grupo de cultivares não o será necessariamente em um segundo grupo avaliado (Lin et al. 1986). Por conseguinte, se o interesse do melhorista é obter cultivares com adaptação específica ou ampla a uma gama de ambientes, é preciso que o faça testando a adaptabilidade dos materiais (Duarte, 1988).

A estabilidade do tipo 3 é aquela na qual o genótipo seria considerado estável se o quadrado médio dos desvios de regressão for pequeno, classificando-o como de alta confiabilidade de resposta. Segundo Lin et al. (1986), todos esses conceitos apresentam

restrição, em especial, aquele baseado nos desvios de regressão. Posteriormente, Lin & Binns (1988) propuseram um novo tipo de estabilidade, a qual denominaram de tipo 4. Para identificação da estabilidade tipo 4, é preciso que as cultivares sejam avaliadas em um determinado número de anos e alguns locais. A análise de variância é realizada estimando o quadrado médio do efeito de anos dentro de locais para cada cultivar; aquela que apresentar o menor quadrado médio será a mais estável às variações imprevisíveis dos anos.

Alliprandini (1992) definiu estabilidade com relação à estatística do coeficiente de determinação (R^2) e dos desvios de regressão (s^2d_i), e adaptabilidade com base no desempenho do genótipo em relação à média dos genótipos avaliados e da responsividade medida pelo coeficiente de regressão (b_i). Chaves (2001) menciona que valores elevados de R^2 estão, geralmente, associados a genótipos responsivos. Duarte & Zimmerman (1995) demonstraram que existe uma correlação intrínseca entre a medida do (R^2) com os valores de (b_i), inferindo que a utilização do R^2 como parâmetro de estabilidade deve ser evitada.

Além de existirem conceitos diversificados para adaptabilidade e estabilidade, acrescenta-se, de acordo com Vencovsky & Torres (1986), um problema concernente à falta ou insignificância da correlação entre a estabilidade temporal (anos) e a estabilidade geográfica (locais). Esses autores encontraram em milho, correlação não significativa, indicando que ambas apresentam controle genético independente. Vencovsky (1987) comenta que a estabilidade apresenta uma herdabilidade menor do que a obtida para produtividade de grãos. Assim sendo, mesmo a herdabilidade da produtividade sendo baixa, a seleção para esse caráter é mais efetiva do que para os parâmetros de estabilidade, pois a repetibilidade da produtividade tem se mostrado superior, tanto àquela do coeficiente de regressão, quanto da variância dos desvios de regressão. Verma et al. (1978) consideraram que um genótipo ideal seria aquele com alta resposta à melhoria das condições ambientais e, ao mesmo tempo, tolerante às condições de ambientes desfavoráveis. Vencovsky & Torres (1988) enfatizaram que, para o produtor rural, é mais importante que uma cultivar seja estável ao longo dos anos.

2.3 Métodos de avaliação da estabilidade fenotípica

Segundo Duarte & Vencovsky (1999), dentre os métodos que procuram identificar os genótipos que menos contribuem para a interação (genótipos estáveis) e que poderiam ser recomendados para toda população de ambientes, desde que mostrem também uma performance média desejável, incluem-se aqueles baseados na regressão linear simples (Yates & Cochran, 1938; Finlay & Wilkinson, 1963; Eberhart & Russel, 1966) ou múltipla (Verma et al. 1978; Silva & Barreto, 1985; Cruz et al. 1989; Storck, 1989; Brasil, 1990; entre outros).

Os primeiros relatos da exploração da técnica de regressão linear para estudo da resposta de cultivares à variação dos ambientes reporta ao trabalho de Yates & Cochran (1938).

O uso de técnicas que empregam métodos de regressão tem como objetivo traçar curvas para representar o comportamento de cada genótipo nos diferentes ambientes, adotando como variável independente um índice ambiental formado pelo contraste entre a média de cada ambiente e a média geral de todos eles e, como variável dependente, a produtividade média do genótipo em cada ambiente (Rosse, 1999). A simplicidade do uso juntamente com a facilidade na interpretação biológica dos resultados tem estimulado e generalizado seu emprego. Por estas mesmas razões, sua eficiência em fornecer informações seguras são questionáveis; inicialmente, por não obedecer à independência entre o índice ambiental e o conjunto de dados analisados, tornando o coeficiente de regressão um estimador inconsistente no que tange à teoria estatística. Uma restrição ao seu emprego pressupõe linearidade das variáveis; é fato, porém, que entidades biológicas são bastante complexas, sendo um erro primário admitir ajustamento a um modelo que considera componentes aditivos também para o efeito da interação. Segundo Vencovsky & Barriga (1992), este relacionamento pode mostrar-se até curvilíneo.

Para Duarte & Vencovsky (1999), a violação da pressuposição de independência da variável explanatória (o índice ambiental é obtido a partir dos próprios

dados), não parece relevante quando o número de genótipos é elevado. Ressaltam o fato de que, para se ajustarem regressões individuais dos genótipos às variações ambientais, o desdobramento tradicionalmente feito não toma apenas a variação devida à interação G×E, mas aquela decorrente das fontes de variação: Ambientes + Interação G×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×E em si.

Crossa (1990) argumenta que a análise de regressão linear não é informativa se a linearidade falhar, sendo fortemente dependente do grupo de genótipos e ambientes incluídos e tende a simplificar modelos de resposta por explicar a variação devida à interação em uma única dimensão, quando na realidade ela pode ser bastante complexa, alertando para o risco em sacrificar informações relevantes para facilitar interpretações estatísticas e biológicas.

A essa crítica, Duarte & Vencovsky (1999) acrescentam ainda 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.

Neste sentido, pode-se esperar, inicialmente, que o ajustamento de dois ou mais segmentos de reta poderia explicar melhor a resposta daqueles genótipos cujos desvios de regressão linear simples se mostrassem elevados (Rosse, 1999). Fundamentando-se nesta nova tendência foram desenvolvidos métodos segmentados (Verma et al., 1978; Silva & Barreto, 1985; Cruz et al., 1989; Storck, 1989; Brasil, 1990). Algumas implicações de ordem estatística e biológica levantadas relativas a esses procedimentos são, respectivamente: ineficiência se o número de ambientes é reduzido e dependência entre as linhas de regressão; inconsistência dos testes de hipótese para os parâmetros; tendenciosidade nas conclusões pelo inconveniente relacionado ao emprego do coeficiente de determinação (R^2); comprometimento das interpretações biológicas

quando os dados são representados graficamente, além de admitirem um ponto de fixação arbitrário para o índice de ambientes, o que faz com que as recomendações extraídas deles sejam vistas com alguma ressalva, prejudicando conseqüentemente sua aplicabilidade.

Duarte & Vencovsky (1999) informam que a implicação estatística básica para uma interação de fatores significativa está diretamente relacionada à impossibilidade de uma interpretação apenas aditiva dos efeitos principais (Mandel, 1971; Kang & Magari, 1996). Isso pode resultar simplesmente da desigualdade das variâncias nas respostas genóticas de um ambiente para outro, sem modificação no posicionamento relativo dos genótipos (interação simples), até a completa falta de correlação nos comportamentos genóticos entre os ambientes (interação complexa ou cruzada). Nesse último caso, a presença da interação determina uma inconsistência na ordenação (*ranking*) dos genótipos de um ambiente para outro. Diante desse quadro, vários métodos estatísticos têm sido empregados buscando um melhor entendimento dos efeitos da interação G×E.

2.4 O modelo AMMI

Suponha que um conjunto de g genótipos tenha sido testado experimentalmente em e ambientes. A média de cada combinação de genótipo e ambiente, obtida de n repetições de um experimento (um conjunto balanceado de dados), pode ser representado pela seguinte matriz:

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

em que os vetores marginais são os vetores de médias de linhas e colunas dos elementos da matriz \mathbf{Y} e $[\bar{Y}_{**}]$ é a média geral daqueles dados.

O modelo AMMI postula 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 do genótipo i em um ambiente j é modelada por:

$$\bar{Y}_{ij} = \mu + g_i + e_j + \sum_{k=1}^m \lambda_k \alpha_{ik} \gamma_{jk} + \rho_{ij} + \varepsilon_{ij},$$

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

$$\sum_{k=1}^m \lambda_k \alpha_{ik} \gamma_{jk} + \rho_{ij},$$

sob as restrições:

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

Estimativas da média geral (μ) e dos efeitos principais (g_i e e_j) são obtidas de uma simples ANOVA com dois fatores a partir da matriz de médias $Y_{(g \times e)} = [Y_{ij}]$. Os resíduos obtidos à partir dessa matriz constituem a matriz de interações

$$GE_{(g \times e)} = [(\hat{g}e)_{ij}]$$

onde $GE = \bar{Y}_{ij} - \bar{Y}_i - \bar{Y}_j + \bar{Y}_{**}$, e os termos da interação multiplicativa são estimados da decomposição em valor singular (DVS) dessa matriz. Assim, λ_k é estimado pelo k -ésimo valor singular de GE, α_{ik} é estimado pelo i -ésimo elemento do vetor singular esquerdo $\alpha_{k(g \times 1)}$ e γ_{jk} é estimado pelo j -ésimo elemento do vetor singular direito $\gamma'_{k(1 \times e)}$ associado com λ_k (Good, 1969; Mandel, 1971; Piepho, 1995). Correspondências entre DVS e análise de componentes principais (ACP) são como segue: λ_k é o k -ésimo valor singular ou a raiz quadrada do k -ésimo maior autovalor da matriz $(GE)(GE)^T$ e $(GE)^T(GE)$, as quais têm igual autovalores não nulos; α_{ik} é o i -ésimo elemento do autovetor de $(GE)(GE)^T$ associado com λ_k^2 ; γ_{jk} é o j -ésimo elemento do autovetor de $(GE)^T(GE)$ associado com λ_k^2 .

A GEI neste modelo é então expressa como uma soma de componentes, cada um multiplicado por λ_k , para um efeito genotípico (α_{ik}) e um efeito ambiental (γ_{jk}). O termo λ_k dá a proporção da variância devida à interação GEI no k -ésimo componente. Os efeitos α_{ik} e γ_{jk} representam pesos para o genótipo i e o ambiente j , naquele componente da interação (λ_k^2). O posto de GE é $s = \min\{g - 1, e - 1\}$, de tal forma que o índice k na soma de componentes multiplicativos pode variar de 1 até s . O uso de todos s componentes recupera toda a variação: $S.Q.(GEI) = \sum_{k=1}^s \lambda_k^2$ e o modelo é saturado de tal forma que ele produz um ajuste exato dos dados, sem termo residual para testar os efeitos (exceto na situação quando um resíduo independente é estimado). Quando $m < s$, o modelo é dito ser truncado. Contudo, para o AMMI, não se tenta recuperar toda a $SQ(GEI)$ mas somente aqueles componentes mais fortemente determinado por genótipos e ambientes. Conseqüentemente, o índice é geralmente colocado a variar até $m < s$, de tal forma que as estimativas são obtidas dos primeiros m termos da DVS da matriz GE (Good, 1969; Gabriel, 1978). Esta é uma análise de mínimos quadrados que conduz a

um resíduo denotado por ρ_{ij} . Assim, a interação do genótipo i com o ambiente j é descrito por $\sum_{k=1}^m \lambda_k \alpha_{ik} \gamma_{jk}$, descartando o ruído dado por $\sum_{k=m+1}^s \lambda_k \alpha_{ik} \gamma_{jk}$. Aqui como no ACP, o componente conta, sucessivamente, para proporções decrescentes da variação presente na matriz $GE(\lambda_1^2 \geq \lambda_2^2 \geq \dots \geq \lambda_k^2)$. Dessa forma, o método AMMI é visto como um procedimento capaz de separar sinal e ruído na análise do GEI (Weber et al. 1996).

2.5 Determinação do número ótimo de termos multiplicativos no modelo AMMI

O principal objetivo aqui é a predição da verdadeira característica da resposta na célula em tabelas de dupla entrada de genótipos e ambientes. Para alcançar isto, um modelo AMMI truncado deve ser usado e então critérios para determinar o número de componentes necessários para explicar o padrão nos termos da GEI têm sido objetivo de algumas pesquisas (Gollob, 1968; Mandel, 1971; Gauch e Zobel, 1988; Piepho, 1994 e 1995; Cornelius, 1993; Cornelius et al., 1996).

Dois procedimentos básicos têm sido envolvidos para determinar o número ótimo de termos multiplicativos a serem retidos nos componentes GEI.

Um desses procedimentos usa o método de validação cruzada no qual os dados são aleatoriamente divididos para modelagem dos dados e para validação dos dados. O AMMI é ajustado para modelar os dados e o quadrado médio do erro de predição (expresso como a raiz quadrada da diferença preditiva média, RMSPD) são determinadas dos dados de validação. O principal criticismo deste procedimento é que o melhor modelo preditivo computado de um subconjunto de dados pode não ser o melhor modelo quando todos os dados são considerados (Cornelius e Crossa, 1999); se validação cruzada é usada em dados MET, os dados devem ser ajustados para repetições diferentes dentro de ambientes (Cornelius e Crossa, 1999). O outro procedimento para determinar o melhor modelo truncado preditivo é usar testes de hipóteses sobre o k -ésimo componente, $H_{0k} : \lambda_k = 0$, usando o conjunto de dados completos (e não um

subconjunto como em procedimento de validação cruzada). Esses testes são baseados em somas de quadrados seqüenciais explicadas por termos multiplicativos.

Far-se-á agora uma revisão breve desses dois procedimentos. Pode ser notado que encolhimento de estimadores (shrinkage) de modelos multiplicativos foram mostrados recentemente como bons preditores da performance de cultivares em ambientes (Cornelius e Crossa, 1999), mas esses estimadores requerem estimativas de graus de liberdade que os autores acham problemática. Entretanto, outras classes de estimadores podem ser melhor que encolhimento de estimadores (shrinkage) (ver Venter e Steel, 1993). Portanto, esse assunto não será considerado aqui.

2.6 Testes de significância de termos multiplicativos

A soma de quadrados seqüencial dos modelos AMMI para o k -ésimo componente, S_k , é dado por $n\lambda_k^2$ para $k = 1, 2, \dots$, posto (GE) (onde $GE = \bar{Y}_{ij} - \bar{Y}_{i*} - \bar{Y}_{*j} + \bar{Y}_{**}$). Como em ACP, todos critérios dos testes envolvem pelo menos indiretamente, a razão da soma de quadrados acumulada para os primeiros m componentes da SQ(GEI), isto é, $\sum_{k=1}^m \lambda_k^2 / \text{SQ(GEI)}$.

Um dos procedimentos usual consiste em determinar os graus de liberdade associados com um particular componente da SQ(GEI) para cada membro da família de modelos AMMI. Isto permite que o quadrado médio seja computado para cada componente, juntamente com um quadrado médio do resíduo. Desde que se tem uma partição ortogonal da soma de quadrados da interação, a razão do quadrado médio de qualquer componente da interação e o quadrado médio do resíduo é então assumido seguir uma distribuição F com os correspondentes graus de liberdade. Isto implicitamente assume uma distribuição normal para a variável resposta original, e permite que componentes de interação individual estejam sujeitos a testes de significância. Entretanto, a validade da distribuição F nessas circunstâncias está sujeita a considerável dúvida. O outro valor λ_k^2 da matriz $(GE)(GE)^T$ (ou $(GE)^T(GE)$) são

distribuídos como autovalores de uma matriz de Wishart mas não tem distribuição qui-quadrado. Desde que λ_k^2 , não são variáveis aleatórias independentes seguindo uma distribuição qui-quadrado, um teste F não é válido. Não obstante, seleção do modelo ótimo é freqüentemente baseado no teste F para os termos sucessivos da interação, o número de termos incluídos corresponde ao número de componentes significativos. O teste F aproximado de Gollob (1968) assume que $n\hat{\lambda}_k^2/\sigma^2$ é distribuído como qui-quadrado e então obviamente não é válido. Simulações computacionais feitas por Cornelius (1993) mostraram que o teste de Gollob ao nível de 0,05 é muito liberal com a taxa de erro tipo I de 66% para testar $H_{01} : \lambda_1 = 0$. O teste F aproximado F_{GH1} , F_{GH2} (Cornelius et al., 1992; Cornelius et al., 1993), efetivamente controla a taxa de erro Tipo I, e são geralmente mais parcimoniosos que o teste de Gollob. Entretanto, esses testes são conservativos para testar termos multiplicativos para o qual os termos anteriores são pequenos. Testes de interação e simulação têm maior poder que os testes F_{GH1} , F_{GH2} com bom controle das taxas de erro Tipo I. O resíduo AMMI, obtido nos últimos termos da SQ(GEI), pode também ser testado para confirmar sua não significância.

De volta à questão dos graus de liberdade, Gauch e Zobel (1996) mencionam alguns métodos para atribuir graus de liberdade para componentes do modelo AMMI; aqueles devido a Gollob (1968) e Mandel (1971) são particularmente populares. Entretanto, os autores chamam a atenção para o fato de, infelizmente, existirem discordâncias entre esses métodos. A escolha de um deles requer considerações tanto de ordem teórica quanto prática. O procedimento de Gollob (1968) é muito fácil de aplicar, desde que o número de graus de liberdade para o m-ésimo componentes da interação é simplesmente definido como $GL(IPCA_m) = g + e - 1 - 2m$, enquanto muitos outros procedimentos requerem simulações extensivas antes de serem usadas.

Por exemplo, Mandel (1971) define o número de graus de liberdade do k-ésimo componente como $GL(IPCA_k) = E[\lambda_k^2]/\sigma^2$, onde σ^2 é a variância populacional. Entretanto, simulações são conduzidas para avaliar o número de graus de liberdade em casos particulares. Mandel fornece algumas tabelas derivadas de tais simulações para um

limitado conjunto de condições, enquanto Krzanowski (1979) fornece algumas versões extras. Essas tabelas, entretanto, não são exaustivas e isso reduz a utilidade prática do método.

Ao contrário de Gollob (1968), o sistema de Mandel (1971) geralmente resulta em um decréscimo não-linear nos graus de liberdade para os termos de interações sucessivas, o qual pode ainda ser fracionário.

Por alguns anos, os graus de liberdade foram obtidos pela proposta de Mandel (1971), a qual foi considerada exata e dessa forma correta. Entretanto, esta proposta tem recebido muitas críticas recentemente (exemplo, Gauch, 1992), e ela é agora considerada menos apropriada que o procedimento de Gollob (1968). A razão para esse criticismo baseia-se na suposição feita por Mandel em suas simulações, de que a matriz contém somente ruído e não sinal uma vez que a presença de sinal afeta o padrão dos componentes substancialmente.

Gauch (1992) discute a questão de obter os graus de liberdade para componentes multiplicativos de um modelo AMMI. Ele conclui que simulações rigorosas parecem ser desnecessárias ou impraticáveis, e geralmente recomenda o uso do sistema de Gollob que usa um teste F aproximado, lembrando que o procedimento é um guia intuitivo. Em casos onde parece existir uma clara divisão entre grandes componentes determinando a parte sistemática e pequenos componentes ruídos, o autor sugere que associar igual número de graus de liberdade $\{GL(IPCA_k) = [(g-1)(e-1)]/e\}$ é especialmente útil para os primeiros componentes porque normalmente existirá pouco interesse na partição dos componentes ruído. Em adição, definitivas questões de pesquisa requerem a exata associação dos graus de liberdade para cada termo multiplicativo. Assim, pelo sistema de Gollob, a análise da variância conjunta completa (computadas a partir das médias) tem a estrutura como mostrada na Tabela 1.

Piepho (1995) investigou a robustez (para as suposições de homogeneidade e normalidade dos erros) de alguns testes alternativos para selecionar componentes de um modelo AMMI. Ele comenta que o teste F aplicado de acordo com o critério de Gollob

Tabela 1 – Análise da variância conjunta completa computada à partir das médias usando o sistema de Gollob e Cornelius.

Fonte de variação	Grande liberdade	Soma de quadrado Gollob	G.L. Cornelius	SQ Cornelius
Genótipo (G)	$g - 1$	SQ(G)	-	-
Ambiente (E)	$e - 1$	SQ(E)	-	-
Interação (GEI)	$(g - 1)(e - 1)$	SQ(GEI)	-	-
$IPCA_1$	$g + e - 1 - (2 \times 1)$	λ_1^2	$(g - 1 - 1)(e - 1 - 1)$	$\sum_{k=2}^s \lambda_k^2$
$IPCA_2$	$g + e - 1 - (2 \times 2)$	λ_2^2	$(g - 1 - 2)(e - 1 - 2)$	$\sum_{k=3}^s \lambda_k^2$
$IPCA_3$	$g + e - 1 - (2 \times 3)$	λ_3^2	$(g - 1 - 3)(e - 1 - 3)$	$\sum_{k=1}^s \lambda_k^2$
...
$IPCA_s$	$g + e - 1 - (2 \times s)$	λ_s^2	-	-
Erro médio/n	$ge(n - 1)$	SQ(Erro médio)	-	-
Total	$gen - 1$	SQ (Total)	-	-

(1968) é liberal, no que ele seleciona muitos termos multiplicativos. Dos quatro métodos estudados pelo autor, incluindo aquele de Gollob (1968), o teste proposto por Cornelius et al. (1992) foi o mais robusto. O autor então recomenda que avaliações preliminares devem ser conduzidas para verificar a validade das suposições se um dos outros testes for usado.

A estatística teste de Cornelius com m termos multiplicativos no modelo é como segue:

$$F_{R,m} = \left(\text{SQ(GEI)} - \sum_{k=1}^m \lambda_k^2 \right) / (f_2 \text{QM(ErroMédio)});$$

com $f_2 = (g - 1 - m)(e - 1 - m)$.

Este é o teste F_R de Cornelius et al. (1992) que pode tornar-se liberal quando comparado com F_{GH1} , F_{GH} , ou testes iterativos de simulação. Sob a hipótese nula de

que não mais que m termos determinam a interação, o numerador (isto é, a SQ (GEI) residual para o modelo AMMI ajustado) é, aproximadamente, uma variável qui-quadrado (Piepho, 1995) de tal forma que o teste estatístico tem uma distribuição F com $f_2 gl$ e graus de liberdade do quadrado médio do resíduo.

Portanto, um resultado significativo para o teste sugere que no mínimo um ou mais termos multiplicativos devem ser adicionados aos m já incluídos. Isso pode ser visto como um teste de significância para os primeiros $m+1$ termos de interação (similar ao teste de falta de ajuste em regressão linear). Quando $m=0$, isto é, quando nenhum termo multiplicativo é incluído, o teste é equivalente ao teste F para a GEI global na ANOVA. Ele é um teste exato. Nota-se também que o número de 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 m primeiros termos. Dessa forma, conclui-se que a aplicação de F_R é equivalente ao teste do resíduo AMMI para GEI, como sugerido.

2.7 Avaliação Preditiva Usando Validação Cruzada

Gauch e Zobel (1988) comentam que avaliações tais como aquelas discutidas por meio das suposições distribucional via teste F pode ser dita “posdictiva” no que se procura por um modelo que explique uma grande parte da variação nos dados observados (com alto coeficiente de determinação). Então, eles afirmam que, tais modelos não são eficientes para selecionar modelos parcimoniosos e são susceptíveis de incluir ruído. Por outro lado, critério “preditivo” de avaliação capitaliza a habilidade de um modelo de fazer predições com dados não incluídos na análise, simulando futuras respostas ainda não medida, de forma que seria preferível basear a escolha do modelo em tal critério.

Para fazer predições, em geral, é necessário usar procedimentos estatísticos computacionalmente intensivos. Quanto menos a escolha do modelo ou avaliação da performance de um preditor é baseada em suposições distribucionais, mais geral é o

resultado. Assim, métodos que são essencialmente baseados em dados e livres de distribuições teóricas terão maiores generalidades. Tais métodos envolvem reamostragem de um dado conjunto de dados, usando técnicas tais como ‘jackknife’, ‘bootstrap’ e validação cruzada. Gauch (1988) introduziu o nome “avaliação preditiva” quando ela é baseada na validação cruzada (Stone, 1974; Wold, 1978), e esse é o princípio sob a proposta dele para seleção do número de componentes nos modelos AMMI.

Em seu método, as repetições para cada combinação de genótipos e ambientes são aleatoriamente divididas em dois subgrupos: (i) dados para o ajuste do modelo AMMI e (ii) dados para validação. As respostas são preditas para uma família de modelos AMMI (isto é, para diferentes valores de m) e esses são comparados com os respectivos dados de validação, calculando-se as diferenças entre esses valores. Então, a soma de quadrados dessas diferenças é obtida e o resultado é dividido pelo número de respostas preditas. Este método foi desenvolvido, posteriormente, por Crossa et al. (1991). Os autores denominaram a raiz quadrada, deste resultado, de diferença preditiva média (RMSPD) e sugeriram que o procedimento fosse repetido 10 vezes, tomando-se uma média dos resultados para cada membro da família de modelos. Um pequeno valor de RMSPD indica sucesso preditivo do modelo, tal que o melhor modelo é aquele com o menor RMSPD. A escolha do modelo é então usada para analisar os dados de todas n repetições, conjuntamente, em uma análise definitiva.

Modificações adicionais foram propostas em recentes anos. Piepho (1994) sugeriu obter o valor médio do RMSPD para 1000 diferentes aleatorizações, ao invés de 10 sugeridas por Crossa et al. (1991). O autor considera uma modificação da partição completamente aleatória dos dados (modelagem e validação) quando o ensaio é em blocos. Neste caso, ele recomenda sortear o bloco inteiro de um ensaio e não fazer componentes para cada combinação de genótipo e ambiente. Assim, a estrutura original de blocos é preservada. Contudo, apesar da coerência lógica desse tipo de proposta, estudos confirmando sua efetividade ainda não estão disponíveis. Gauch e Zobel (1996) sugerem que o conjunto de dados de validação deve sempre ser apenas uma observação

para cada tratamento. Isto é devido ao fato de que é mais provável, para $n-1$ dados encontrar um modelo que está mais próximo para a análise do conjunto completo de n dados. Dias e Krzanowski (2003) tomaram essa idéia e descreveram dois métodos que otimizam o processo de validação cruzada por validar o ajuste do modelo em cada um dos dados por vez e então combinar essa validação em uma medida simples e geral de ajuste.

Cornelius e Crossa (1999) usaram validação cruzada para comparar a performance de encolhimento de estimadores (shrinkage), modelos multiplicativos truncados e melhor estimador linear não viesado (BLUP) comparando o RMSPD como a raiz quadrada da diferença quadrada média entre o valor preditivo e seus correspondentes dados de validação em dados ajustados de cinco MET. Os autores usaram um critério de parada para o número de validações cruzadas que consistiu em calcular a diferença preditiva quadrada média conjunta na m -ésima execução do laço da estrutura de repetição computacional (PMSPD _{m}). A validação cruzada é considerada terminada quando o valor absoluto máximo do $\text{PMSPD}_m - \text{PMSPD}_{m-1}$ for menor que 0,01. O número máximo de validações cruzadas requeridas foi de 64 e o mínimo foi 39.

2.8 Método “leave-one-out”

Dias e Krzanowski (2003) propuseram dois métodos baseados em um procedimento “leave-one-out” completo, que otimiza o processo de validação cruzada. No que segue, assume-se que se deseja prever os elementos x_{ij} da matriz \mathbf{X} por meio do modelo:

$$x_{ij} = \sum_{k=1}^m d_k u_{ik} v_{jk} + \varepsilon_{ij}$$

Os métodos são aqueles esboçados por Krzanowski (1987) e Gabriel (2002) respectivamente, no qual prediz-se o valor \hat{x}_{ij}^m de x_{ij} ($i = 1, \dots, g$; $j = 1, \dots, e$) para cada possível escolha de m (o número de componentes), e a medida de discrepância entre o valor atual e predito como

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

Contudo, para evitar viés, os dados x_{ij} não devem ser usados nos cálculos de \hat{x}_{ij}^m para cada i e j . Como consequência, apelo a alguma forma de validação cruzada é indicado, e os dois procedimentos diferem na forma com que eles lidam com isso. Ambos, entretanto, assumem que a DVS de \mathbf{X} pode ser escrita como

$$\mathbf{X} = \mathbf{UDV}^T .$$

O procedimento de validação cruzada padrão subdivide \mathbf{X} em um certo número de grupos, deleta cada grupo por vez a partir dos dados, avalia os parâmetros do preditor a partir dos dados remanescentes, e prediz o valor deletado, Wold (1976, 1978). Krzanowski (1987) argumenta que a predição mais precisa resulta quando cada grupo deletado é tão pequeno quanto possível, que no presente caso seria um simples elemento de \mathbf{X} . Denota-se por $\mathbf{X}^{(-i)}$ a resultante após deletar a i -ésima linha de \mathbf{X} e centralizar em torno das médias das colunas. Denota-se por $\mathbf{X}_{(-j)}$ a resultante após deletar a j -ésima coluna de \mathbf{X} e centraliza-se em torno das médias das colunas, seguindo o esquema dado por Eastment and Krzanowski (1982). Então, pode-se escrever

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

e

$$\mathbf{X}_{(-j)} = \tilde{\mathbf{U}} \tilde{\mathbf{D}} \tilde{\mathbf{V}}^T \quad \text{com} \quad \tilde{\mathbf{U}} = (\tilde{u}_{st}), \quad \tilde{\mathbf{V}} = (\tilde{v}_{st}) \quad \text{e} \quad \tilde{\mathbf{D}} = \text{diag}(\tilde{d}_1, \dots, \tilde{d}_{p-1})$$

Agora, considere-se o preditor

$$\hat{x}_{ij}^m = \sum_{t=1}^m \left(\tilde{u}_{it} \sqrt{\tilde{d}_t} \right) \left(\bar{v}_{tj} \sqrt{\bar{d}_t} \right)$$

Cada elemento no lado direito desta equação é obtido da DVS de \mathbf{X} centrada na média após omitir a i -ésima linha e a j -ésima coluna. Assim, o valor x_{ij} não é usado no cálculo da predição, e o máximo uso é feito com os outros elementos de \mathbf{X} . Os

cálculos aqui são exatos, assim não há problema com a convergência como nos procedimentos de maximização que têm sido aplicados ao AMMI, mas que não garantem a convergência.

Gabriel (2002), por outro lado, tomou uma mistura de regressão e aproximação de uma matriz de posto inferior como a base para sua predição. O algoritmo para validação cruzada de aproximações de posto inferior proposto pelo autor é como segue:

Para dada matriz \mathbf{X} de GEI, usa-se a partição

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_{11} & \mathbf{X}_{1*}^T \\ \mathbf{x}_{*1} & \mathbf{X}_{111} \end{bmatrix}$$

e o ajuste aproximado da sub-matriz \mathbf{X}_{111} de posto m usando a DVS é:

$$\mathbf{X}_{111} = \sum_{k=1}^m \mathbf{u}_{(k)} \mathbf{d}_k \mathbf{v}_{(k)}^T = \mathbf{U} \mathbf{D} \mathbf{V}^T,$$

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

Então prediz-se x_{11} por

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

e obtém-se o resíduo de validação cruzada $e_{11} = x_{11} - \hat{x}_{11}$.

Similarmente, obtém-se o valor ajustado da validação cruzada x_{ij} e os resíduos $e_{ij} = x_{ij} - \hat{x}_{ij}$ para todos os outros elementos x_{ij} , $i = 1, \dots, g$; $j = 1, \dots, m$; $(i, j) \neq (1, 1)$. Cada um irá requerer uma partição diferente de \mathbf{X} .

Esses resíduos e valores ajustados podem ser sumarizados por

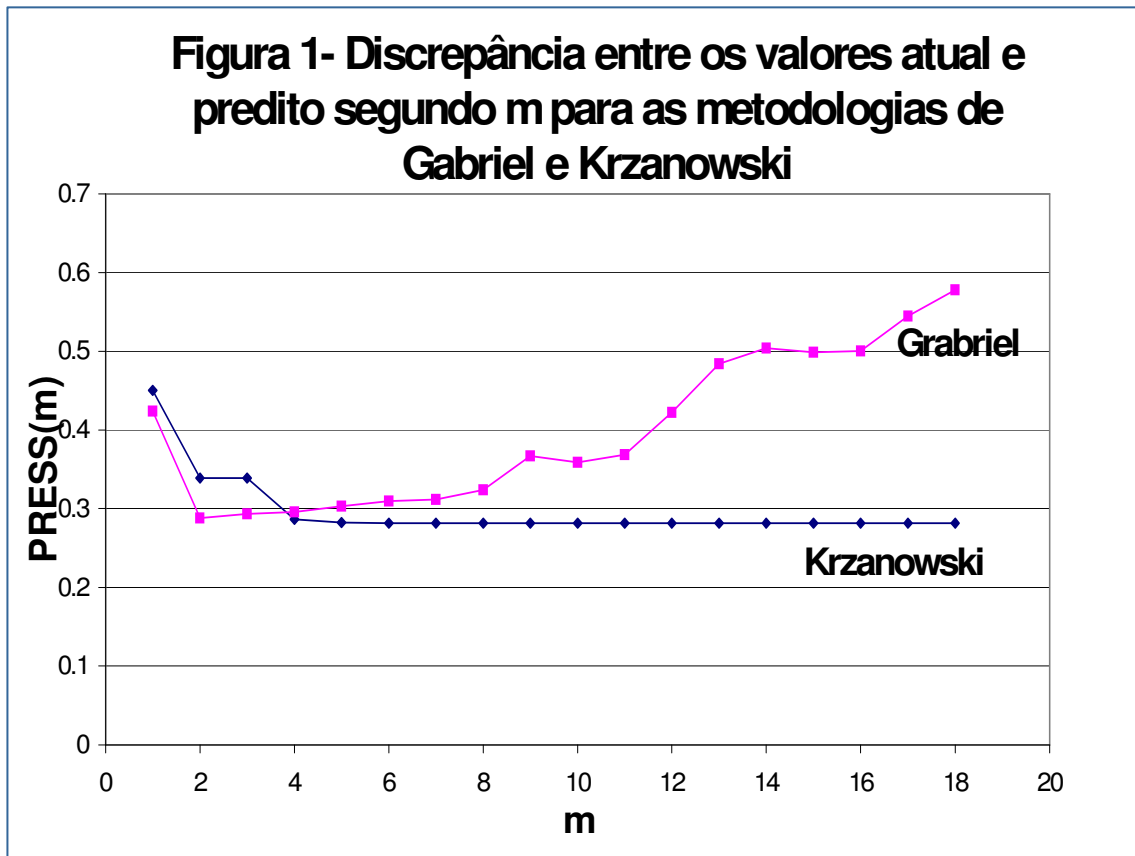
$$PRESS(m) = \frac{1}{ge} \sum_{i=1}^g \sum_{j=1}^e e_{ij}^2 \quad \text{e} \quad PRECORR(m) = Corr(x_{ij}, \hat{x}_{ij} | \forall i, j),$$

respectivamente.

Com cada método, a escolha de m pode ser baseada em alguma função apropriada de

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

Contudo, as características dessa estatística diferem para os dois métodos. O procedimento de Gabriel produz valores que primeiro decresce e então (usualmente) cresce com m . Por essa razão ele sugere que o valor ótimo de m seja aquele que produz o mínimo da função. O procedimento de Eastment-Krzanowski produz, geralmente, um conjunto de valores que é monotonicamente não-crescente com m , Figura 1.



Por isso, sugerem o uso de

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

onde D_m é o número de graus de liberdade requeridos para ajustar o m -ésimo componente e D_r é o número de graus de liberdade remanescentes após ajustar o m -ésimo componente. Considerações sobre o número de parâmetros a serem estimados juntos com todas as restrições nos autovetores em cada estágio, mostra que $D_m = g + e - 2m$. D_r pode ser obtido por sucessivas subtrações, dando $(g-1)e$ graus de liberdade na matriz centrada na média \mathbf{X} , isto é, $D_1 = (g-1)e$ e $D_r = D_{r-1} - [g + e - (m-1)2]$, $r = 2, 3, \dots, (g-1)$, (Wold, 1978). W_m representa o aumento na informação preditiva suprida pelo m -ésimo componente, dividido pela informação preditiva média em cada um dos componentes remanescentes. Assim,

importantes componentes devem produzir valores de W_m maiores que a unidade. Baseando-se a escolha de m em W_m pode ser vista como uma natural seleção de um melhor conjunto de variáveis regressoras ortogonais em análise de regressão múltipla.

Em nível computacional, a melhor precisão parece ser obtida quando as entradas (x_{ij}) em diferentes colunas de \mathbf{X} são comparáveis em tamanho e existe relativamente pouca variação entre os d_i . O procedimento mais estável é, portanto, aquele no qual a média \bar{x}_j e o desvio padrão s_j da coluna j ($j = 1, \dots, e$) são primeiro calculados dos valores presentes naquela coluna. As entradas existentes x_{ij} de \mathbf{X} são então padronizadas para $x'_{ij} = (x_{ij} - \bar{x}_j) / s_j$, e estimativas são obtidas pela aplicação de

$$\hat{x}_{ij} = \mathbf{x}_{i*}^T \mathbf{V} \mathbf{D}^{-1} \mathbf{U}^T \mathbf{x}_{*j}$$

aos dados padronizados, e então os valores finais são obtidos de

$$\hat{x}_{ij} = \bar{x}_j + s_j x'_{ij}.$$

Voltando ao caso dos dados de genótipo-ambiente, fica claro aqui que \mathbf{X} deve ser a matriz de interações previamente denotada por GE. Contudo, desde que se está simplesmente procurando pelo número apropriado de termos multiplicativos no modelo, e qualquer constante aditiva pode ser absorvida no componente ε_{ij} do modelo, pode-se aplicar o procedimento “leave-one-out” diretamente à matriz \mathbf{Y} de dados. De fato, isso pode freqüentemente ser preferível dado os valores pequenos tomados por muitos elementos de GE.

Cornelius et al. (1993) compararam resultados de validação cruzada com aqueles obtidos após calcular a estatística PRESS nos modelos multiplicativos em dados MET completos. A partição dos dados envolveu três repetições para modelagem e uma repetição para validação. Calcularam o RMSPD da estatística PRESS ajustando os valores de PRESS como $[PRESS / ge + 3s^2 / 4]^{1/2}$, onde g e e denotam o número de genótipos e ambientes no MET e s^2 é a variância residual conjunta dentro de ambientes.

O termo em s^2 é um ajuste para a diferença em variância da validação dos dados nas médias de caselas, para tornar os resultados comparáveis ao RMSPD da divisão 3 – 1 dos dados. Resultados em um MET com nove genótipos e vinte ambientes mostrou que PRESS é mais sensível a super ajuste do que os dados divididos. Tabela 2 mostra que PRESS diferencia mais claramente a forma do modelo do que a divisão dos dados. Para algumas formas de modelo (SHMM e SREG), o modelo com o menor PRESS para os dados preditos em uma casela deletada é melhor do que quando eles foram preditos por três repetições dos dados com todas caselas presentes. Por outro lado, muitos modelos superajustados dão predições muito pouco seguras das células deletadas.

2.9 Histórico dos modelos AMMI

A maioria dos dados coletados em agricultura experimental é de natureza multivariada porque vários atributos são medidos em cada um dos indivíduos incluídos nos experimentos, isto é, genótipos, tratamentos agrônômicos, etc. Tais dados podem ser arranjados em uma matriz \mathbf{X} , onde o (i, j) -ésimo elemento representa o valor observado para o j -ésimo atributo medido no i -ésimo indivíduo na amostra. Técnicas multivariadas comumente usadas para analisar tais dados incluem análise de componentes principais (ACP) se não existem a “priori” grupos de indivíduos ou variáveis; variáveis canônicas ou análise discriminante se os indivíduos na amostra formam um grupo a “priori”; análise de correlação canônica se as variáveis formam um grupo a “priori”; e análise de agrupamento se alguma partição da amostra é procurada.

Em melhoramento de plantas, ensaios multi-ambientes (MET-multienvironment trials) são importantes para testar adaptação geral e específica de cultivar. Um cultivar crescendo em diferentes ambientes irá frequentemente mostrar flutuações significantes na performance de produção relativa a outras cultivares. Essas mudanças são influenciadas por diferentes condições experimentais e são referidas como GEI. Um exemplo típico de uma matriz \mathbf{X} surge na análise MET, na qual as linhas de \mathbf{X} são os genótipos e as colunas são os ambientes onde os genótipos são testados. Presença de GEI exclui modelos interpretativos simples que têm somente efeitos principais

aditivos de genótipos e ambientes (Mandel 1971; Crossa, 1990; Kang e Magari, 1996). Por outro lado, adaptação específica de genótipos a subconjuntos de ambientes é um assunto fundamental a ser estudado em melhoramento de plantas porque um genótipo pode ter bom desempenho sob condições ambiental específica e pode ter um desempenho ruim ou pobre sob outras condições.

Crossa et al. (2002) fazem uma ampla revisão dos procedimentos para análise de GEI que inclui o modelo de análise da variância convencional de dois fatores fixos, o procedimento de regressão linear, e os modelos multiplicativos. A resposta média empírica, \bar{y}_{ij} , do i -ésimo genótipo no j -ésimo ambiente com n repetições em cada uma das $i \times j$ caselas é expressa como

$$\bar{y}_{ij} = \mu + g_i + e_j + (ge)_{ij} + \bar{\epsilon}_{ij}$$

onde μ é a média geral de todos genótipos e ambientes, g_i é o efeito aditivo do i -ésimo genótipo, e_j é o efeito aditivo do j -ésimo ambiente, $(ge)_{ij}$ é o componente de GEI para o i -ésimo genótipo no j -ésimo ambiente, e $\bar{\epsilon}_{ij}$ é o erro assumido NID $(0, \sigma^2 / n)$ (onde σ^2 é a variância residual dentro de ambiente, considerada constante). Este modelo não é parcimonioso, uma vez que cada casela da GEI tem seu próprio parâmetro de interação, e não informativo, porque os parâmetros de interação independentes são complicados e de difícil interpretação.

Yates e Cochran (1938) sugeriram tratar o termo GEI como sendo linearmente relacionados ao efeito de ambiente, isto é colocar $(ge)_{ij} = \xi_i e_j + d_{ij}$, onde ξ_i é o coeficiente de regressão linear do i -ésimo genótipo na média ambiental e d_{ij} é um desvio. Este procedimento foi posteriormente usado por Finlay e Wilkinson (1963) e ligeiramente modificado por Eberhart e Russell (1966). Tukey (1949) propôs um teste para GEI usando $(ge)_{ij} = K g_i e_j$ (onde K é uma constante). Mandel (1961) generalizou o modelo de Tukey colocando $(ge)_{ij} = \lambda \alpha_i e_j$ para genótipos $(ge)_{ij} = \lambda g_i \gamma_j$ para

ambientes e assim obtendo um feixe de linhas retas que podem ser testadas para concorrência (isto é, se os α_i ou os γ_j são todos os mesmos) ou não concorrência.

Gollob (1968) e Mandel (1969, 1971) propuseram um termo bilinear GEI

$(ge)_{ij} = \sum_{k=1}^s \lambda_k \alpha_{ik} \gamma_{jk}$ no qual $\lambda_1 \geq \dots \geq \lambda_s$ e α_{ik}, γ_{jk} satisfazem a restrição de ortogonalidade $\sum_i \alpha_{ik} \alpha_{ik'} = \sum_j \gamma_{jk} \gamma_{jk'} = 0$ para $k \neq k'$ e $\sum_i \alpha_{ik}^2 = \sum_j \gamma_{jk}^2 = 1$.

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

Uma família de modelos multiplicativos pode então ser gerada retirando-se o efeito principal de genótipo (modelo de regressão para ambientes, (SREG – Site Regression Model), o efeito principal de ambientes (GREG – Genotype Regression Model), ou ambos efeitos (COMM – Complete Multiplicative Model). Um outro modelo multiplicativo (SHMM – Shifted Multiplicative Model), (Seyedsadr e Cornelius, 1992) é útil para estudar crossover GEI (Crossa et al., 2002).

Contudo, um aspecto que ainda não foi completamente resolvido refere-se à determinação do número de componentes multiplicativos a serem retidos no modelo para, adequadamente, explicar o padrão na interação. Algumas propostas foram apresentadas por, entre outras, Gollob (1968), Mandel (1971), Gauch e Zobel (1988), Cornelius (1993) e Piepho (1994, 1995). Todas tomam em consideração a proporção da variância acumulada pelos componentes (Duarte e Vencovsky, 1999), e o mais recente foca a validação cruzada como metodologia preditiva baseada em dados (Dias e Krzanowski, 2003). Entretanto, alguns problemas ainda permanecem, notadamente em saber qual desses métodos é o melhor.

2.10 Métodos de reamostragem

Técnicas de reamostragem, conhecidas como métodos estatísticos computacionalmente intensivos, foram propostos à décadas, e se apresentam como alternativas interessantes para analisar dados de origem ecológica e evolutiva. Porém, sua difusão era limitada por exigirem grandes quantidades de cálculos. Devido a grande capacidade de processamento das máquinas e pelo desenvolvimento de programas de computadores, estes procedimentos vêm tendo aplicabilidade cada vez maior. Dentre estas metodologias, pode-se destacar o “jackknife” e o “bootstrap”, apresentados por Efron (1979).

O “jackknife” foi introduzido em 1949 por M. H. Quenouille, para reduzir o viés de um estimador de correlação, com base na divisão da amostra original em duas semi-amostras. Anos depois, este estudo foi completado com a generalização do método (Quenouille, 1956). Assim, a amostra original de tamanho n passa a ser dividida em g sub-amostras de tamanho h , de modo que se tem $n = gh$. O tamanho h fixo em um, faz com que a amostra original passe a ser dividida em n grupos com unidade observacional (Lavoranti, 2003). O termo “jackknife” foi criado por Tukey (1958), fazendo analogia aos canivetes de bolso com múltiplas funções, que embora versáteis, dificilmente realizam alguma das funções de modo ideal. Essa comparação refere-se à abordagem geral para testar hipóteses e construir intervalos de confiança quando outros métodos mais indicados não podem ser aplicados facilmente. Contudo, esse método produz melhores estimativas amostrais de alguns parâmetros de posição, como por exemplo, a média e a curtose, também fornecem freqüentemente variância e intervalos de confiança aproximados para estas estimativas.

É um processo de reamostragem, em que o parâmetro de interesse θ , em uma amostra com n observações é estimado apropriadamente como $\hat{\theta}$. Pseudovalores $\hat{\theta}_j$, associados a cada observação j , são obtidos como $\hat{\theta}_j = \hat{\theta} - (n - 1)(\hat{\theta}_{-j} - \hat{\theta})$, em que $\hat{\theta}_{-j}$ é o cálculo do parâmetro com a j -ésima observação excluída da amostra. O lado direito da expressão é a estimativa do parâmetro amostral menos um termo de tendenciosidade,

refletindo o desvio da j -ésima estimativa $\hat{\theta}_j$ desconsiderada da estimativa $\hat{\theta}$ da amostra total. A média dos pseudovalores $\hat{\theta}(\cdot)$ é a estimativa “jackknife” de $\hat{\theta}$. A diferença $\hat{\theta}_j - \hat{\theta}(\cdot)$, mede o viés amostral geral da estimativa original $\hat{\theta}$ (Crowley, 1992).

Para um melhor entendimento da técnica, Lavoranti (2003) propõe a seguinte situação clássica de análise de dados: seja $[x_1, x_2, \dots, x_n]$ uma amostra de variáveis aleatórias independentes e identicamente distribuídas (i.i.d.). Seja $p(X, \theta)$ a função de verossimilhança dessa amostra aleatória (a.a.), com dependência do parâmetro θ . Supondo-se que o estimador de θ seja a estatística $\hat{\theta} = T_n(x)$ e que $\hat{\theta}_i = T_{n-1}(x)$, seja o estimador baseado na amostra de tamanho $n - 1$ obtida retirando-se o i -ésimo ponto amostral. Com esses valores constroem-se os pseudovalores (Tukey, 1958):

$$\tilde{\theta} = n\hat{\theta} - (n-1)\hat{\theta}_i \quad i = 1, 2, \dots, n.$$

A média aritmética dos n pseudovalores $\tilde{\theta}_i$ dada por:

$$\tilde{\theta} = \frac{1}{n} \sum_{i=1}^n \tilde{\theta}_i,$$

é chamada de estimador “jackknife” do parâmetro θ .

De uma forma geral, o erro-padrão “jackknife” para a estatística $\hat{\theta}$ é dado por Tukey (1958):

$$\hat{SD} = \left[\frac{1}{n(n-1)} \sum_{i=1}^n (\tilde{\theta}_i - \tilde{\theta})^2 \right]^{\frac{1}{2}}.$$

É importante observar que dada a (a.a.) $[(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)]$ da distribuição bivariada de probabilidade F , é possível calcular uma estimativa da esperança do vetor aleatório bivariado e, além disso, pode-se avaliar sua precisão. Para tanto, basta calcular a expressão do vetor média amostral $[\bar{x}, \bar{y}]$ e a matriz de covariância amostral:

$$\begin{vmatrix} \hat{\sigma}_x^2 & \hat{\sigma}_{xy}^2 \\ \hat{\sigma}_{xy}^2 & \hat{\sigma}_y^2 \end{vmatrix}.$$

Contudo, quando se deseja avaliar a precisão de outros estimadores, como, por exemplo, a correlação amostral $\hat{\rho}_{XY} = \frac{\hat{\sigma}_{xy}}{\hat{\sigma}_x \hat{\sigma}_y}$, não é possível fazê-lo sem o conhecimento de F. A solução, neste caso, seria usar resultados assintóticos.

No caso da variância do estimador de $E[X]$, \bar{X} , não existe problema para a sua estimação. Conhecendo-se ou não a distribuição da variável aleatória (v.a.), tem-se acesso à precisão da estatística \bar{X} . Dada a (a.a.) $[x_1, x_2, \dots, x_n]$, da (v.a.) X, com distribuição desconhecida F, o estimador do parâmetro $\mu = E[X]$ é dado por:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Para o cálculo dos pseudovalores de $\tilde{\theta}_i$ precisa-se de $\hat{\theta}_{(i)}$ dados por:

$$\hat{\theta}_{(i)} = \bar{X}_{(i)} = \frac{n}{n-1} (\bar{X} - X_i) \quad (i=1, 2, \dots, n), \quad (1)$$

e, assim, têm-se os pseudovalores:

$$\tilde{\theta}_{(i)} = n \bar{X} - (n-1) \bar{X}_{(i)} \quad (i=1, 2, \dots, n), \quad (2)$$

que produzem o estimador “jackknife” de $E[X]$,

$$\tilde{\theta} = \frac{1}{n} \sum_{i=1}^n \hat{\theta}_i = n \bar{X} - (n-1) \frac{1}{n} \sum_{i=1}^n \bar{X}_{(i)}. \quad (3)$$

Pode-se mostrar através das equações (1) e (3) que, nesse caso, $\tilde{\theta} = \bar{X}$:

$$\begin{aligned}\tilde{\theta} &= n\bar{X} - (n-1)\frac{1}{n}\left[\frac{n\bar{X} - X_1}{n-1} + \frac{n\bar{X} - X_{2i}}{n-1} + \dots + \frac{n\bar{X} - X_n}{n-1}\right] \\ &= n\bar{X} - (n-1)\frac{1}{n(n-1)}\left[nn\bar{X} - \sum_{i=1}^n X_i\right] \\ &= \frac{1}{n}\sum_{i=1}^n X_i = \bar{X}\end{aligned}\quad (4)$$

e, ainda, por (2) e (4) tem-se: $\tilde{\theta}_{(i)} - \theta = X_i - \bar{X}$ pois,

$$\begin{aligned}\tilde{\theta}_{(i)} - \theta &= n\bar{X} - (n-1)\bar{X}_{(i)} - \bar{X} = (n-1)\bar{X} - (n-1)\bar{X}_{(i)} \\ &= (n-1)(\bar{X} - \bar{X}_{(i)}) = (n-1)\left[\bar{X} - \frac{n\bar{X} - X_i}{n-1}\right] \\ &= X_i - \bar{X}\end{aligned}\quad (5)$$

Logo, $V[\bar{X}]$ será dada por:

$$V[\bar{X}] = \left[\frac{1}{n(n-1)}\sum_{i=1}^n (\tilde{\theta}_{(i)} - \tilde{\theta})^2\right] = \left[\frac{1}{n(n-1)}\sum_{i=1}^n (X_i - \bar{X})^2\right] = \frac{s^2}{n}.\quad (6)$$

Portanto, seja $\tilde{\theta} = \frac{1}{n}\sum_{i=1}^n \tilde{\theta}_i$ o estimador “jackknife” correspondente a um estimador viciado do parâmetro θ , $\hat{\theta} = T_n(x)$. Então $\tilde{\theta}$ elimina o termo de ordem $\frac{1}{n}$ do vício de $\hat{\theta}$.

Segundo Tukey (1958), os n pseudovalores poderiam ser tratados como v.a.’s aproximadamente i.i.d. e, dessa forma, poder-se-ia construir a quantidade pivotal:

$$\frac{\tilde{\theta} - \theta}{\left[\frac{1}{n(n-1)} \left[\frac{\sum_{i=1}^n (\tilde{\theta}_i - \tilde{\theta})}{n-1} \right]^2 \right]^{\frac{1}{2}}}, \quad (7)$$

que tem aproximadamente distribuição t_{n-1} , e pode ser usada como pivô na estimação por intervalo do parâmetro θ .

Bickel & Freedman (1981) citam o vício como sendo um dos fatores importantes para a escolha de um estimador. Esse exprime a quantidade de erro sistemático que ocorre na estimação do parâmetro e, por isso, é preferível que tenha valor nulo e baixa variância. No entanto, isso nem sempre é possível, sendo assim, o conhecimento do vício é importante para avaliar a precisão das estimativas.

Dada (a.a.) $X = [x_1, x_2, \dots, x_n]$ da distribuição de probabilidade desconhecida F , é comum estimar-se o parâmetro de interesse θ , usando-se um estimador natural $\hat{\theta} = T_n(x)$. Caso o estimador seja viciado, uma aplicação de destaque do “jackknife” é o estimador do vício, $b(\theta, \hat{\theta})$, de um estimador $\hat{\theta}$. Dessa forma, tem-se:

$$\begin{aligned} E(\hat{\theta}) &= \theta + b(\theta, \hat{\theta}) \\ &= \theta + \frac{a_1(F)}{n} + \frac{a_2(F)}{n^2} + \frac{a_3(F)}{n^3} + \dots, \end{aligned}$$

como, $\hat{\theta} = \frac{1}{n} \sum_{i=1}^n \hat{\theta}_i$ e $\hat{\theta} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$, pela equação (1), tem-se:

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n \hat{\theta}_i = \frac{1}{n} \sum_{i=1}^n \left[\frac{n \bar{X} - X_i}{n-1} \right] = \frac{1}{n} \left[\frac{n^2 \bar{X} - n \sum_{i=1}^n X_i}{n-1} \right] = \left[\frac{n (n-1) \bar{X}}{n (n-1)} \right] = \bar{X},$$

portanto, $\hat{b}(\theta, \bar{X}) = (n-1)(\bar{X} - \bar{X}) = 0$.

Para a estimação “jackknife” do vício do estimador de máxima verossimilhança

$$T_n(x) = \hat{\theta}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2, \text{ da variância } \sigma^2, \text{ da distribuição F, conhecida ou não, tem-se:}$$

se:

$$E(\hat{\theta}^2) = E\left[\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2\right] = \frac{1}{n}(n-1)\sigma^2 = \sigma^2 - \frac{\sigma^2}{n}. \quad (8)$$

Logo, $b(\theta^2, \hat{\theta}^2) = \frac{-\sigma^2}{n}$, e da equação (8) vê-se que a aproximação vai somente até o termo $\frac{1}{n}$, sendo os demais nulos. E, como o estimador “jackknife” elimina esse termo da aproximação de $E[\hat{\sigma}^2]$, conclui-se que $\tilde{\theta}$ é não-viciado para σ^2 . A estimativa do vício é dada por:

$$\begin{aligned} b(\theta^2, \hat{\theta}^2) &= \hat{\sigma} - \tilde{\sigma} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 - \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \\ &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \end{aligned}$$

que, em alguns casos, são exatamente iguais aos da teoria clássica.

É importante destacar que, no caso da estimação do vício de σ^2 , obtém-se uma estatística cuja esperança é $\frac{-\sigma^2}{n}$:

$$E\left[b(\theta^2, \hat{\theta}^2)\right] = \frac{-1}{n(n-1)} E\left[\sum_{i=1}^n (X_i - \bar{X})^2\right] = \frac{-1}{n(n-1)} \sigma^2(n-1) = \frac{-\sigma^2}{n}.$$

Portanto, o “jackknife” fornece um estimador não-viciado para o vício de $\hat{\sigma}^2$.

Para Efron & Stein (1981), esse resultado ocorre sempre que a estatística for uma função quadrática.

Crowley (1992) lembra que para o parâmetro de interesse θ , é possível obter com “jackknife” a estimativa do parâmetro, de sua variância e seu intervalo de confiança. A variância “jackknife” é calculada considerando-se nulas as correlações entre pseudovalores, contudo, não se sabe sob que condições esse pressuposto é validado. O intervalo de confiança obtido é paramétrico, pois se parte do pressuposto de que as estimativas “jackknife” estão baseadas em erros com distribuição normal. Esta pressuposição é apoiada no Teorema Central do Limite se as amostras forem grandes, porém, nos casos de tamanho pequeno da amostra é difícil avaliar a validade desse pressuposto.

Manly (1997) considera uma das vantagens desse método como sendo a de transformar muitos problemas de estimação no simples problema de estimar a média. Basta que o estimador convencional seja função de n valores amostrais. Menciona que correlações entre pseudovalores são indesejáveis, causando viés na estimativa da variância, podendo esta ser sub ou superestimada.

Uma vantagem adicional deste método é que ao substituir o estimador convencional pelo estimador “jackknife” há uma redução de ordem $1/p$ no viés (Efron & Tibshirani, 1993); um “jackknife” de p -ésima ordem pode ser usado para remover viés de ordem $1/np$ (Manly, 1997).

Além disso, ao estimar o parâmetro de interesse sem uma observação de cada vez, o método “jackknife” permite identificar as observações mais divergentes na amostra, pois, é possível verificar quais observações promovem maiores alterações na estimativa quando são retiradas (Carlini-Garcia, 2001).

Para Efron & Tibshirani (1993), uma deficiência do método “jackknife” refere-se a estatísticas dadas por funções não-diferenciáveis, ou seja, aquelas que são pouco influenciadas por pequenas alterações no conjunto de dados, como por exemplo, a mediana. Neste caso, o estimador “jackknife” do erro-padrão dessas estatísticas torna-se inconsistente. Propõe remover d observações por reamostragem “jackknife” ao invés de apenas uma, tomando como aproximação ($\sqrt{n} \leq d \leq n$), sendo n o número de observações da amostra. Outra desvantagem importante consiste no fato de o método não fornecer a distribuição do parâmetro.

Segundo Efron & Tibshirani (1993), no caso do cálculo do viés, o “jackknife” é uma aproximação do “bootstrap”, sendo que se pode generalizar esse comportamento apenas para estatísticas do tipo não-linear.

Outro método computacionalmente intensivo que vem sendo cada vez mais utilizado é o “bootstrap”, este termo surgiu da frase “to pull oneself up one’s bootstrap” retirado do texto: “The Baron had fallen to the of a deep lake. Just when it looked like all was lost, he thought to pick himself up by his own bootstrap” de “Adventures of Baron Munchausem” de R. E. Raspe, século XVII. No qual relata uma situação em que o Barão estava afundando em um lago e vendo que tudo estava perdido, pensa que conseguirá emergir puxando os cadarços dos próprios sapatos (Efron & Tibshirani, 1993).

A metáfora “bootstrap” refere-se ao fato de os dados serem usados em sua própria análise estatística. Dessa forma, todo resultado “bootstrap” depende diretamente da amostra original observada, isto é, os resultados “bootstrap” são consistentes para a amostra original. Este método é baseado na reamostragem de dados reais com reposição, para revelar algum padrão-estrutural neles presentes. A noção básica é de que os dados em si, vistos como distribuição de frequências, representam a melhor imagem disponível da distribuição de frequências da qual eles são amostrados (Crowley, 1992).

Na estatística, as situações difíceis podem ser vistas como os problemas de soluções analíticas complexas, e, as mais variadas soluções possíveis devem ser tentadas; essas seriam a utilização de uma metodologia com grande quantidade de cálculos para analisar um pequeno conjunto de dados. A solução para esses casos, com o uso de métodos computacionalmente intensivos, é obtida substituindo-se o poder analítico das expressões teóricas pelo poder de processamento das máquinas. Operacionalmente, o procedimento “bootstrap” consiste na reamostragem de mesmo tamanho e com reposição dos dados da amostra original, e cálculo da estatística de interesse para cada reamostra “bootstrap” (pseudodados) (Lavoranti, 2003). Foi definido por Efron (1979) como técnica desenvolvida para fazer certos tipos de inferências estatísticas.

O método “bootstrap” é baseado no princípio “plug-in”, um método simples de estimar parâmetros a partir de amostras. Considerando o parâmetro θ e uma função t de probabilidade de uma distribuição F qualquer de uma variável aleatória, a estimativa “plug-in” do parâmetro $\theta = t(F)$ é dada por $\hat{\theta} = t(F)$. Assim, estima-se a função $\theta = t(F)$ da distribuição de probabilidades F pela mesma função da distribuição empírica F , $\hat{\theta} = t(F)$. A vantagem deste método consiste em produzir tendências e erros-padrões de forma automática, não importando quão complicada $\theta = t(F)$ seja. Este princípio é conveniente se a informação disponível sobre F é proveniente de uma amostra x . Nesta situação, $\hat{\theta} = t(F)$ não pode ser melhorada como um estimador de $\theta = t(F)$. Contudo é menos apropriado em situações onde há informações sobre F , além das fornecidas pela amostra x . Sendo assim, pode-se assumir que F é um membro da família paramétrica (Efron & Tibshirani, 1993).

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

Uma outra aplicação da metodologia “bootstrap” segundo Crowley (1992), é nos testes de hipóteses, por exemplo, com dados de cada nível de tratamento, amostrados separadamente, com reposição. Os testes podem ser formulados de acordo com a extensão de sobreposição entre intervalos de confiança ou por combinar amostras “bootstrap” para calcular um teste estatístico.

Lavoranti (2003) apresenta um esquema para melhor compreensão do método “bootstrap”, conforme é mostrado a seguir.

Sejam o parâmetro θ e seu estimador a estatística $T_n(X, F)$, em que $X = [x_1, x_2, \dots, x_n]$ é a (a.a.) disponível da v.a. com função de distribuição desconhecida $F(X \sim i.i.d. F)$, e

$$\hat{F}_{(n)}(x) = \frac{\left[\sum_{i=1}^n I(X_i \leq x) \right]}{n}, \quad (9)$$

em que:

$\hat{F}_{(n)}(x)$: estimador não-paramétrico de máxima verossimilhança de F ;

$I(X_i \leq x)$: função indicadora.

Assim, essa distribuição empírica é construída colocando-se massa probabilística $\frac{1}{n}$ em cada ponto amostral.

De $\hat{F}_{(n)}(x)$ (9) toma-se B amostras “bootstrap” de mesmo tamanho n ,

$$\begin{aligned} X_1^* &= [x_{11}^*, x_{12}^*, \dots, x_{1n}^*] & X_{1i}^* &\sim i.i.d. \hat{F}_n \\ X_2^* &= [x_{21}^*, x_{22}^*, \dots, x_{2n}^*] & X_{2i}^* &\sim i.i.d. \hat{F}_n \\ &\vdots & &\vdots \\ X_B^* &= [x_{B1}^*, x_{B2}^*, \dots, x_{Bn}^*] & X_{Bi}^* &\sim i.i.d. \hat{F}_n \end{aligned}$$

calculam-se as B estatísticas “bootstrap” $T_n^*(x_l^*)$ ($l = 1, 2, \dots, B$), correspondentes às B amostras “bootstrap”, e forma-se o conjunto:

$$\{ T_n^*(x_l^*) \}; \quad l = 1, 2, \dots, B \}, \quad (10)$$

que é uma simulação da verdadeira distribuição amostral da estatística $T_n(x, F)$.

Com o conjunto (10) pode-se obter a medida da variabilidade de $T_n(x, F)$, como erro-padrão “bootstrap” ($\hat{\theta}^* = s(x^*)$). A estimativa do vício da estatística $T_n(x, F)$ que é dada por:

$$b[\theta, T_n(x, F)] = T_n(x, F) - T_n^*,$$

em que,

$$T_n^* = \frac{\sum_{l=1}^B T(x_l^*)}{B}, \quad (11)$$

que corresponde à estimativa “bootstrap” do parâmetro de interesse.

Na Figura 1, está apresentado um diagrama esquemático ilustrando a construção da distribuição “bootstrap”

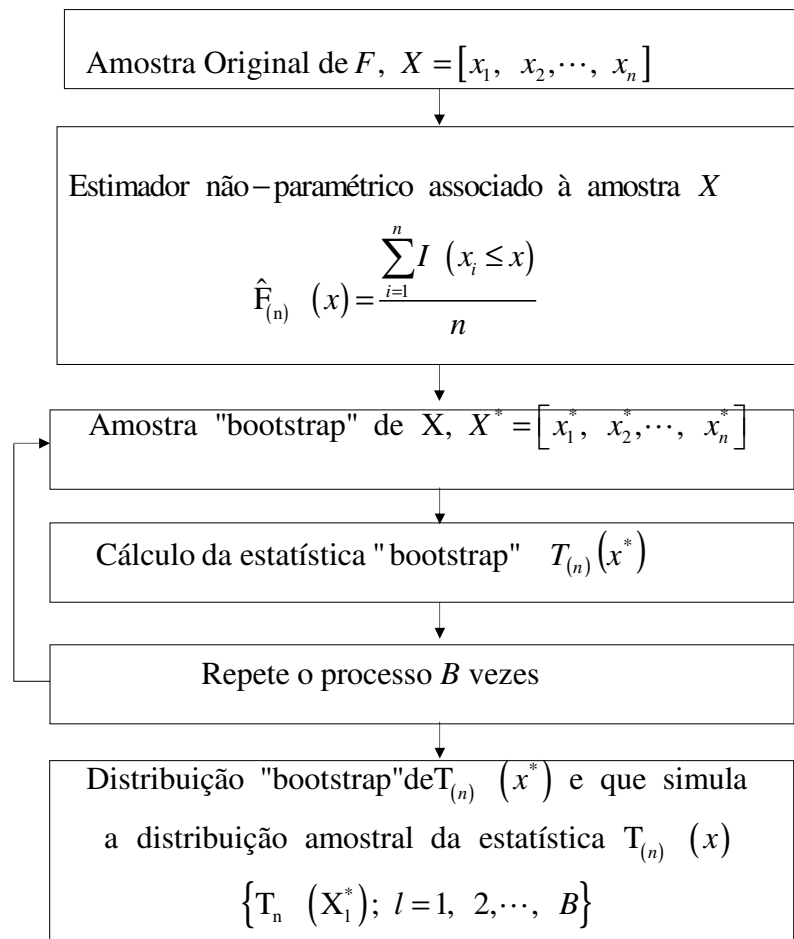


Figura 1. Diagrama da distribuição “bootstrap” da estatística $T_n(x, F)$ (Lavoranti, 2003)

Na prática, constrói-se a distribuição “bootstrap”, (9), de $T_n(x, F)$ por simulação Monte-Carlo com um número de repetições, B , suficientemente grande. Um indicador do tamanho adequado de B , independentemente do custo computacional, é a qualidade da convergência da estimativa “bootstrap” do parâmetro para a estimativa natural do parâmetro $T_n^*(B \rightarrow \infty) \mapsto T(x, F)$. A construção do algoritmo Monte-Carlo para obtenção da distribuição “bootstrap” das estatísticas usuais é em geral simples. Sua convergência está garantida pela lei dos Grandes Números, pois $T_n^*(X_1^*), T_n^*(X_2^*), \dots, T_n^*(X_B^*)$ nada mais são do que uma amostra de v.a.’s i.i.d. com distribuição condicional de $T_n^*(X, \hat{F})/X = x$.

Assim, quando B tende a infinito, a média amostral T_n^* (11) aproxima-se de $E[T_n(X, \hat{F})/X = x]$. Efron (1979) sugeriu que a distribuição condicional “bootstrap” de $T_n(X, \hat{F})/X = x$, pode ser usada como a distribuição de $T_n(X, F)/X = x$. Como se vê, no procedimento “bootstrap”, os pontos da amostra original $[x_1, x_2, \dots, x_n]$ são considerados como uma população com função de distribuição \hat{F} e média \bar{x} . A estatística “bootstrap” $T_n^* = T_n^*(x, \hat{F})$ é considerada como um estimador de $T_n(x, F)$.

A distribuição de $T_n(x^*, \hat{F})$ pode ser usada para aproximar a distribuição amostral desconhecida de $T_n(x, F)$, assim como, a distribuição de $\sqrt{n}(t_n^* - T_n)$ pode ser usada para aproximar a distribuição amostral de $\sqrt{n}(T_n - \theta)$ (Bickel & Freedman, 1981 e Efron & Tibshirani, 1993). A normalidade assintótica de $\sqrt{n}(T_n^* - T_n)$, e a convergência em probabilidade da variância “bootstrap” (12) foram apresentadas por Bickel & Freedman, (1981).

$$s^{*2} = \frac{1}{B-1} \sum_{l=1}^B [T_n(x_l^*) - T_n^*]^2 \quad (12)$$

O número de vezes em que o ponto amostral x_i é selecionado no procedimento “bootstrap” ($X^* = x^*$) é representado por:

$$N_i^* = \#\{X_i^* = x_i\},$$

conseqüentemente, $\sum_i^n N_i^* = n$, e o vetor $N^* = [N_1^*, N_2^*, \dots, N_n^*]$ tem distribuição

multinomial. Vê-se que $N_i^* \sim b\left(n, \frac{1}{n}\right)$ e, em correspondência a esses números existe

uma distribuição de freqüências relativas $f_i = \frac{N_i^*}{n}$, sendo que f_i assume valores no

conjunto $f \left\{0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n}{n}\right\}$. A v.a. f_i possui média e variância dadas por:

$$N_i^* \sim b\left(n, \frac{1}{n}\right) \rightarrow \begin{cases} E[f_i] = E\left[\frac{N_i^*}{n}\right] = \frac{1}{n} E[N_i^*] = \frac{1}{n} \\ V[f_i] = V\left[\frac{N_i^*}{n}\right] = \frac{1}{n^2} V[N_i^*] = \frac{n-1}{n^3} \end{cases}$$

Intervalos de confiança para parâmetros podem ser construídos com base em estatística “bootstrap” (Efron & Tibshirani, 1986).

Conforme Efron & Tibshirani (1993), a construção de intervalos de confiança baseados nas distribuições normal padronizada e t podem ser obtidos desde que se obedeça às pressuposições usuais necessárias para tanto. Com os intervalos baseados nas distribuições normal e t , os pontos percentuais são simétricos em relação a zero e, conseqüentemente, os intervalos de confiança resultantes são também simétricos em relação à estimativa pontual $\hat{\theta}$. Contudo, os percentis obtidos com esse tipo de intervalo

podem apresentar assimetria em relação ao zero, fazendo com que os intervalos se desloquem à direita ou à esquerda. Esses autores lembram que este procedimento pode ser empregado para estimar intervalos de confiança de estatística de posição. Porém, não é recomendável quando aplicado em casos mais gerais como por exemplo, para construção de intervalos de confiança para coeficiente de correlação. Salientam que podem fornecer resultados errôneos e podem sofrer influência de valores anômalos.

Dentre os vários tipos de intervalos de confiança que podem ser construídos a partir da reamostragem “bootstrap”, o mais simples dos intervalos “bootstrap” e que é também o mais difundido é o método percentil. Para Manly (1997), o uso desse tipo de intervalos baseia-se na tentativa de aproximar os percentis da distribuição de um estimador usando percentis gerados por “bootstrap”. Existem na literatura vários tipos de intervalos percentis, sendo o mais simples deles aquele que substitui a estimativa do erro-padrão, pela estimativa “bootstrap”. O autor ressalta que para intervalos percentis, é exigido número maior de amostras “bootstrap” em relação aos intervalos baseados na distribuição normal padronizada ou t , de forma que se obtenham estimativas acuradas dos pontos percentis da distribuição “bootstrap”.

Efron & Tibshirani (1986) propõem um procedimento para estabelecer um intervalo de confiança, com base em uma função monótona crescente f , tal que para $\varphi = f(\theta)$, $\hat{\varphi} = f(\hat{\theta})$ e $\hat{\varphi}^* = f(\hat{\theta}^*)$ tem-se as quantidades pivotais Gaussianas $\hat{\varphi} - \varphi$ e $\hat{\varphi}^* - \hat{\varphi}$, ou seja:

$$\hat{\varphi}^* - \hat{\varphi} \sim N(-z_0\sigma, \sigma^2) \quad e \quad \hat{\varphi} - \varphi \sim N(-z_0\sigma, \sigma^2) \quad (13)$$

para as constantes z_0 e σ . Portanto, $\hat{\varphi} - \varphi$ é uma quantidade pivotal Gaussiana sob F e \hat{F} . Assim sendo, se $\hat{\theta} = \theta(\hat{F})$ então, $\hat{\varphi} = f(\hat{\theta}) = f(\theta(\hat{F}))$ e, o intervalo de confiança de nível $(1 - \alpha)$ para φ , com base no pivô (13) é:

$$P\left[\frac{|\hat{\varphi} - \varphi - (-z_0\sigma)|}{\sigma} \leq z_{\frac{\alpha}{2}}\right] = 1 - \alpha \quad (14)$$

Como antes mencionado, o método mais simples para construir intervalos de confiança via “bootstrap” é o método percentil. Lavoranti (2003) apresenta ainda o esquema, a seguir, para melhor compreensão do mesmo: este método consiste em encontrar a distribuição \hat{F} “bootstrap” (9) e calcular os percentis da distribuição $\hat{F}^{-1}\left(\frac{\alpha}{2}\right)$ e $\hat{F}^{-1}\left(1-\frac{\alpha}{2}\right)$, que correspondem aos limites inferiores (L_I) e superiores (L_S), respectivamente, do intervalo de confiança $(1-\alpha)$ 100% (Efron & Tibshirani, 1993). Na prática, esses intervalos (L_I) e (L_S), podem ser obtidos pelos valores relativos à parte inteira de $B\frac{\alpha}{2}$ e $B\left(1-\frac{\alpha}{2}\right)+1$, respectivamente, do ranqueamento das estatísticas “bootstrap” $T_n(X^*)$, obtidas em (10).

Um refinamento do método do percentil, denominado BC_a “bias-corrected and accelerated” (Efron & Tibshirani, 1993), apresenta alta acurácia na estimação dos intervalos de confiança para dados obtidos da distribuição “bootstrap” (Diciccio & Efron, 1996). Esse método fixa os limites de confiança para o parâmetro θ em função dos percentis obtidos dos histogramas gerados pelas amostras “bootstrap”. Os percentis são dados por:

$$\hat{\theta}_{BC_a}[\alpha] = \hat{G}^{-1}\Phi\left(z_0 + \frac{z_0 + z^\alpha}{1 - \alpha \left(\frac{z_0 + z^\alpha}{z_0 + z^\alpha}\right)}\right);$$

em que, \hat{G}^{-1} é a inversa da função distribuição acumulada; Φ é a distribuição normal padronizada e $z^\alpha = \Phi^{-1}(\alpha)$.

Se a e z_0 são nulos, então $\hat{\theta}_{BC_a}[\alpha] = \hat{G}^{-1}(\alpha)$ é o centésimo percentil ‘bootstrap’. Assim, o intervalo de confiança fica determinado pelo 5° e 95° percentis “bootstrap”. Se \hat{G} for uma distribuição perfeitamente normal, então $\hat{\theta}_{BC_a} = \hat{\theta} + z^\alpha \hat{\alpha}$, ou seja, um intervalo de confiança padrão com estimativas “bootstrap” (Diciccio & Efron, 1996).

No caso de a e z_0 serem diferentes de zero, excelentes estimativas para a e z_0 são dadas por:

$$\hat{\alpha} = \text{Viés}_{\theta=\hat{\theta}} \frac{[l_{\theta}(\hat{\theta})]}{6},$$

em que, $l_{\theta}(\hat{\theta}) = \frac{\partial}{\partial \theta} \log \{g_{\theta}(\hat{\theta})\}$ e $g_{\theta}(\hat{\theta})$ é a densidade de $\frac{\partial G_{\theta}(\hat{\theta})}{\partial \hat{\theta}}$;

$$\hat{z}_0 = \Phi^{-1} \left\{ \frac{\# \{ \hat{\theta}^*(b) < \hat{\theta} \}}{B} \right\}.$$

Lavoranti (2003) menciona o método “bootstrap- t ” baseado na distribuição t -Student, como sendo bastante utilizado para determinar intervalos de confiança “bootstrap”. Os limites de confiança são dados por:

$$[T_n^* - t_{n-1}^{1-\alpha} s^*, T_n^* - t_{n-1}^{\alpha} s^*],$$

em que, T_n^* é definido na equação (11); s^* é definido na equação (12) em que t_{n-1}^{α} é o valor de t , da tabela t -Student, ao nível de $(1-\alpha)$ 100%, com $n-1$ graus de liberdade.

Incorporações ao método percentil foram propostas para construção de intervalos de confiança percentis corrigidos, como o intervalo de confiança percentil corrigido para viés “*bias-corrected percentile confidence limits*” e o intervalo de confiança percentil acelerado corrigido para o viés “*accelerated bias-corrected percentile confidence limits*”. Estes foram desenvolvidos em função das distribuições de probabilidade; é importante mencionar que “*bias-corrected percentile confidence limits*”, “*accelerated bias-corrected percentile confidence limits*” e bootstrap- t podem ser usados para a construção de intervalos de confiança não-paramétricos.

Métodos de reamostragem começam a se tornar ferramentas bastante úteis e importantes na construção de procedimentos inferenciais, evitando a obtenção de

fórmulas via argumentos analíticos, tornando-se alternativas eficientes para os métodos estatísticos tradicionais (Lavoranti, 2003).

3 MATERIAL E MÉTODOS

3.1 ASPECTOS TÉCNICOS

3.1.2 Dados e Modelo

Serão descritos o modelo e exemplos em termos de um conjunto de g genótipos que foi testado, experimentalmente, em e ambientes. A produção em todas as combinações de genótipos e ambientes é representada pela matriz $\mathbf{Y}_{(g \times e)} = [y_{ij}]$.

O modelo AMMI postula componentes aditivos para os efeitos principais de genótipos (g_i) e ambientes (e_j) e componentes multiplicativos para interação genótipo \times ambiente (ge_{ij}). Assim, a resposta média do genótipo i em um ambiente j é modelada por:

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

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

$$\sum_{k=1}^m \alpha_{ik} \lambda_k \gamma_{jk}$$

sob as restrições

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

Estimativas da média geral (μ) e efeitos principais (g_i e e_j) são obtidas de uma simples ANOVA da matriz de médias $Y_{(g \times e)}$. Os resíduos dessa matriz constituem então a matriz de interação $GE_{(g \times e)} = [(ge)_{ij}]$, e os termos de interação multiplicativa são estimados da decomposição em valor singular (DVS) desta matriz. Assim, λ_k é estimado pelo k -ésimo valor singular de GE, α_{ik} é estimado pelo i -ésimo elemento do vetor singular esquerdo $\mathbf{\alpha}_{k(g \times 1)}$, e γ_{jk} é estimado pelo j -ésimo elemento do vetor singular direito $\mathbf{\gamma}'_{k(1 \times e)}$ associado com λ_k (Good, 1969; Mandel, 1971; Piepho, 1995). Os dois vetores singulares α_{ik} e γ_{jk} representam os efeitos de genótipos e ambientes presentes

no k -ésimo componente de interação, enquanto o valor singular λ_k é um indicador da força desse componente.

Se a interação contém um ou mais termos ($m \geq 1$), então assume-se que os erros têm uma ou mais variâncias homogêneas após esses efeitos terem sido tomados em consideração.

Snee (1982) diz que não aditividade em tabela de dupla entrada com uma observação por casela pode ser devido a não homogeneidade de variância relacionada a linha ou coluna da matriz \mathbf{Y} , ou a interação entre os fatores da linha e coluna. Ele discute um exemplo envolvendo o efeito de quatro variedades de trigo em 13 localidades, e mostra que a não homogeneidade de variância relacionada à variedade é detectada por vários autores, é mais utilmente interpretada como uma interação nível dependente para as diferenças entre as produções de diferentes variedades de trigo sendo maiores naquelas localidades que têm altas produções. Também, comenta que a experiência tem indicado que, freqüentemente, somente um termo multiplicativo é necessário para descrever a não aditividade, vindo em favor da parcimoniosidade na escolha do número de componentes nos modelos AMMI.

3.1.3 Estudo de simulação

Estudou-se por simulação o comportamento dos principais métodos de seleção do número de componentes de interação nos modelos AMMI. Entretanto, para tornar as simulações realísticas, elas foram baseadas em conjuntos de dados reais para os quais o modelo é apropriado, de acordo com o seguinte esquema.

Supõe-se que $\mathbf{Y} = [y_{ij}]$ é a tabela de dupla entrada para um conjunto de dados escolhido, e que se ajuste o modelo AMMI

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

para um dado valor de m . Considere uma barra sobre a letra para denotar média, com $\bar{\cdot}$ indicando o sobescrito sobre o qual o somatório ocorre. Dessa forma,

$$\hat{\mu} = \bar{y}_{**} , \quad \hat{g}_i = \bar{y}_{i*} - \bar{y}_{**} , \quad \hat{e}_j = \bar{y}_{*j} - \bar{y}_{**}$$

e $\alpha_{ik} \lambda_k \gamma_{jk}$ é estimado do k -ésimo componente da DVS de $GE = \bar{y}_{ij} - \bar{y}_{i*} - \bar{y}_{*j} + \bar{y}_{**}$. O modelo assume $\varepsilon_{ij} \sim N(0, \sigma^2)$, e σ^2 é estimada da soma de quadrados do resíduo da ANOVA do conjunto de dados.

Para gerar uma repetição no estudo de simulação para um dado valor escolhido de m , simplesmente formam-se os valores.

$$\hat{y}_{ij} = \hat{\mu} + \hat{g}_i + \hat{e}_j + \sum_{k=1}^m \hat{\alpha}_{ik} \lambda_k \hat{\gamma}_{jk} + \hat{\varepsilon}_{ij}$$

onde o $\hat{\varepsilon}_{ij}$ são independentemente retirados de uma distribuição $N(0, \hat{\sigma}^2)$.

Por esse meio, é possível gerar repetições que imitem o padrão dos dados originais mas os quais têm embutido um número específico de componentes de interações (governada pela escolha inicial de m). Também por variarem alguns valores dos parâmetros, é possível forçar estruturas particulares nos dados. Por exemplo, trocando os valores de $\hat{\lambda}_i$, pode-se forçar alguns componentes a terem uma maior ou menor contribuição para a interação.

O primeiro estudo utiliza os dados de produção de trigo, Ensaio 3, dado por Cornelius e Crossa (1999) que descrevem cinco ensaios internacionais multiambiente de cultivares, todos em um delineamento aleatorizado em blocos. O ensaio 3 refere-se a milho (*Zea mays L.*) com 9 cultivares, 20 ambientes e $n = 4$ repetições. Calculou-se a média sobre as repetições para produzir uma única matriz de dados 20×9 .

A primeira série de simulações de experimentos foi conduzida para investigar a sensibilidade dos métodos ao incrementar-se o número de componentes de interação nos dados. Colocou-se o número de componentes de interação m sucessivamente igual a 0, 1, 2, ..., 9 e, então, usaram-se todos os parâmetros estimados dos dados originais, e o esquema delineado acima, para produzir os dados simulados. Geraram-se 100 repetições para cada um desses 10 experimentos, e usaram-se os quatro métodos descritos anteriormente para determinar, em cada caso, quantos componentes de interação devem

ser incluídos no modelo. A Tabela 2 mostra o valor mais freqüente escolhido para cada método e cada experimento.

Tabela 2. Número modal de termos multiplicativos escolhidos pelos métodos Eastment-Krzanowski, Gabriel, Gollob e Cornelius; 100 repetições das simulações baseadas nos dados do Ensaio 3 de Cornelius e Crossa (1999).

m	Eastment-Krzanowski	Gabriel	Gollob	Cornelius	Quadrado do valor singular original
0	1	1	1	1	140,32
1	2	2	2	1	37,71
2	2	3	2	2	26,05
3	2	4	3	2	21,53
4	2	3	4	3	12,37
5	1	3	5	4	6,90
6	1	3	5	4	2,51
7	1	3	5	4	2,90
8	1	2	5	-	0
Dados originais	2	7	2	2	

Então, conduziram-se três outras séries de experimentos, restringindo a ênfase dos componentes em cada caso de várias formas.

Na primeira série de experimentos, cuidou-se para que cada um dos nove componentes de interação tivesse aproximadamente igual peso. Para fazer isso cuidou-se para que os valores $\hat{\lambda}$ tivessem o seguinte padrão: $\hat{\lambda}_1^2 = c$, $\hat{\lambda}_2^2 = 0,99c$, $\hat{\lambda}_3^2 = 0,98c$, ... , $\hat{\lambda}_9^2 = 0,92c$, com c escolhido tal que $\sum_{i=1}^9 \hat{\lambda}_i^2$ fosse igual a soma de quadrados GEI. Uma

vez mais 100 repetições foram conduzidas para cada experimento, e a moda do número de componentes de interação é apresentada para cada método e cada experimento na Tabela 3.

Tabela 3. Número modal de termos multiplicativos escolhidos pelos métodos Eastment-Krzanowski, Gabriel, Gollob e Cornelius; 100 repetições das simulações baseadas nos dados do Ensaio 3 de Cornelius e Crossa (1999), termos de interação igualmente ponderados.

M	Eastment-Krzanowski	Gabriel	Gollob	Cornelius
0	1	1	1	1
1	2	2	2	1
2	3	3	2	2
3	3	4	3	3
4	1	5	4	4
5	1	6	5	5
6	1	7	6	6
7	1	1	7	7
8	1	1	8	-

Na segunda série de experimentos restringiram-se os primeiros três $\hat{\lambda}_i^2$ a terem altos valores e os $\hat{\lambda}_i^2$ remanescentes a serem proporcionais a seus atuais valores (sujeitos a $\sum_{i=1}^9 \hat{\lambda}_i^2$ igual a soma de quadrados da GEI). Na terceira série de experimentos, outra vez tiveram-se os três primeiros valores de $\hat{\lambda}_i^2$ dominantes, mas os remanescentes $\hat{\lambda}_i^2$ foram colocados, aproximadamente, constante. Resultados dessas duas séries são apresentados nas Tabelas 4 e 5.

Tabela 4. Número modal de termos multiplicativos escolhidos pelos métodos Eastment-Krzanowski, Gabriel, Gollob e Cornelius; 100 repetições das simulações baseadas nos dados do Ensaio 3 de Cornelius e Crossa (1999), três termos de interação altamente ponderados e os restantes proporcionais aos seus valores atuais.

M	Eastment-Krzanowski	Gabriel	Gollob	Cornelius
0	1	1	1	1
1	2	2	2	1
2	3	3	2	2
3	3	4	3	3
4	3	4	3	3
5	3	4	3	3
6	3	4	3	3
7	3	4	3	3
8	3	4	3	-

Tabela 5. Número modal de termos multiplicativos escolhidos pelos métodos Eastment-Krzanowski, Gabriel, Gollob e Cornelius; 100 repetições das simulações baseadas nos dados do Ensaio 3 de Cornelius e Crossa (1999), três termos de interação altamente ponderados e os restantes com seus valores constante.

M	Eastment-Krzanowski	Gabriel	Gollob	Cornelius
0	1	1	1	1
1	2	2	2	1
2	3	3	2	2
3	3	3	3	2
4	1	3	4	2
5	1	3	4	3
6	1	3	4	3
7	1	3	4	3
8	1	3	4	-

Finalmente, usaram-se os dados do Ensaio 5 de Cornelius e Crossa (1999) como a base para simulações, porque esses dados fornecem uma útil situação nula para a situação de GEI. Aqui, têm-se oito cultivares e cinquenta e nove ambientes, mas a ANOVA comum não detectou interação significativa para esse conjunto de dados. Por isso, conduziu-se somente a forma não restrita do experimento (isto é, como na Tabela 2), e os resultados são mostrados na Tabela 6.

Tabela 6. Número modal de termos multiplicativos escolhidos pelos métodos Eastment-Krzanowski, Gabriel, Gollob e Cornelius; 100 repetições das simulações baseadas nos dados do Ensaio 5 de Cornelius e Crossa (1999).

m	Eastment-Krzanowski	Gabriel	Gollob	Cornelius	Quadrado do valor singular original
0	1	1	1	1	65,53
1	1	2	2	1	48,04
2	1	3	2	2	43,56
3	1	4	3	3	37,79
4	1	5	4	3	23,85
5	1	1	5	4	22,78
6	1	1	5	5	21,61
7	1	1	6	6	0
Dados originais	1	1	2	3	

4 RESULTADOS E DISCUSSÃO

Admitiu-se no início que as investigações aqui são exaustivas (uma comparação de apenas quatro métodos sobre um relativamente pequeno conjunto de condições), e não foram delineados de tal forma que métodos ótimos podem ser determinados com relação a alguma função matemática objetiva bem definida. Em vez disso, partiu-se do ponto de vista que frequentemente há conflito na prática entre os métodos, e o que é requerido por um prático é alguma confirmação que um determinado método escolhido provavelmente desempenhe razoavelmente. Assim, nossas conclusões nesta discussão são essencialmente de natureza qualitativa, medidas contra o que acreditamos ser razoáveis interpretações de cada situação estudada.

Os quatro métodos para selecionar o número de componentes multiplicativos da interação produziram deferentes resultados nos conjuntos de dados usados neste trabalho. Além disso, os resultados para esses testes dão recomendações sobre um particular valor de m para um conjunto de dados. Esta característica pode ser vista claramente, nas Tabelas 2 - 6.

Em termos gerais, os dois métodos baseados no teste F, são conservativos e são propensos a selecionar um valor de m razoavelmente alto, com o método de Gollob geralmente dando um valor mais alto dentre os dois testes; o que está de acordo com os resultados de Piepho (1995). O método de Eastment-Krzanowski, por outro lado, é claramente o mais parcimonioso: ele não selecionou um valor de m maior do que 3, e usualmente, recomenda ou $m = 1$ ou $m = 2$. Este método portanto, seleciona modelos que estão mais de acordo com o que foi sugerido por Snee (1982) relativo a descrições de não aditividade.

Voltando aos resultados presentes, considere-se o primeiro experimento. O número de componentes de interação escolhido por cada método para os dados originais é dado no rodapé da Tabela 2, como reportado por Dias e Krzanowski, 2003.

Todos os métodos sugerem 2 componentes, exceto o método de Gabriel que sugere 7 componentes. A Tabela 3 então, mostra o número modal de componentes escolhidos por cada método, enquanto, progressivamente, mais termos de interação são

incluídos nos dados simulados. Agora, se um método seleciona r componentes para os dados originais, então, claramente, aqueles componentes além do r -ésimo são considerados sem importância. Assim, poder-se-ia esperar que o método escolhe o mesmo número de componentes enquanto mais do que r componentes são incluídos nos dados simulados, e então escolher r cada vez que mais do que r componentes são incluídos. Em resumo, procura-se por alguma forma de estabilidade no número selecionado na Tabela 3.

Desde que deve aparecer somente 2 componentes de interação nos dados originais, e desde que os componentes são adicionados de acordo com seus postos em ordem decrescente dos valores $\hat{\lambda}_i^2$ (última coluna da Tabela 2), espera-se que a adição do terceiro e sucessivos componentes tenha pouco efeito.

De qualquer maneira, os quatro métodos mostraram diferentes padrões de comportamento. O método de Eastment-Krzanowski é muito estável, um valor m igual a 2 para os dados originais comparado com 1 ou 2 para todas simulações. Os métodos de Gabriel e Gollob são muito inconstantes, o método de Gabriel escolheu $m=7$ para os dados originais mas 1, 2, 3 ou 4 nas simulações, e de Gollob teve todos valores entre 1 e 5 nas simulações enquanto escolheu $m=2$ para os dados originais. O método de Cornelius é menos variável, mas ainda exibe uma amplitude de 1 a 4.

Considere-se a seguir as simulações com restrição nos componentes da interação. Nas Tabelas 4 e 5 onde os componentes de interação foram restringidos por ponderar pesadamente os 3 primeiros componentes, de tal forma que seja provavelmente justo dizer que nenhum método deva escolher um valor maior do que 3. Os métodos de Eastment-Krzanowski e Cornelius estão de acordo com isso, enquanto aqueles de Gabriel e Gollob recomendam um valor 4 para uma proporção substancial de casos. Entretanto, em ambos os casos espera-se que uma vez o maior valor tenha sido alcançado, ele permaneça constante enquanto mais componentes são adicionados. Aqui o método de Eastment-Krzanowski falha (ver Tabela 4), onde m diminui de 3 para 1 à partir de $M=5$.

Na Tabela 3 os componentes têm claramente igual ponderação, de tal forma que se pode esperar uma escolha com incremento monotônico enquanto o número de componentes (M) incluídos no modelo aumenta. Todos os métodos exceto aquele de Eastment-krzanowski exibem tal incremento de conjunto de valores onde este método outra vez diminui de 3 para 1 à partir de $M=4$.

Finalmente, a Tabela 6 registra um caso onde os termos são sucessivamente adicionados na simulação enquanto m aumenta, mas onde os dados originais exibem pouca evidência de interação. A distribuição dos valores baixos dos quadrados dos autovalores (última coluna na Tabela 6) mostra esse ponto de vista. Conseqüentemente, somente termos “nulos” são sucessivamente adicionados nas simulações enquanto m cresce, e não deve existir nenhuma tendência crescente no número de componentes escolhidos enquanto m cresce. O método de Eastment-Krzanowski sugere apenas um componente, o menor valor possível para todos valores de M , enquanto todos outros métodos exibem uma tendência crescente no número de componentes escolhidos enquanto m cresce, e seleciona valores muito grandes para alguns desses casos. As Tabelas 2 – 5 não apresentam um número modal de termos multiplicativos AMMI para o método de Cornelius à partir de m igual a oito devido os graus de liberdade negativos.

5 CONCLUSÕES

Em resumo, mostra-se um quadro misto. O método de Eastment-Krzanowski é estável, lida corretamente com conjunto de dados com pouca interação, e se comporta apropriadamente em presença de relativamente um pequeno número de componentes “importantes”, mas geralmente subestima para, corretamente, identificar um grande número de componentes que contenha aproximadamente igual peso. O método de Gabriel é muito inconstante. Identifica importantes componentes no caso de pouca interação, e tende a escolher muitos componentes em cada uma das outras situações. Dos métodos que utilizam o teste F, aquele devido a Cornelius exhibe apropriado comportamento em presença de “importantes” componentes, mas é menos estável do que o método de Eastment-Krzanowski e detecta importantes componentes nos casos de pouca interação. Finalmente, a versão de Gollob é, de maneira geral, similar ao método de Cornelius mas com ligeiramente pior estabilidade, e tende a escolher mais componentes em algumas situações. Por fim, o método de Eastment-krzanowski parece ser o método de validação cruzada preferível enquanto o método de Cornelius seria o recomendado método baseado no teste F. O primeiro é preferível se parcimoniosidade for levada em consideração, mas o último é preferível se a presença de um grande número de componentes de interação é considerada existir. Dadas as diferentes performances dos dois métodos, ambos talvez devam ser usados antes de tomar uma decisão em aplicações práticas. Essas conclusões suportam aquelas obtidas por Dias e Krzanowski (2003) a partir da análise dos dados reais.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALLIPRANDINI, L.F. Estudos dos efeitos ambientais, estabilidade, adaptabilidade e ganho genético em linhagens de soja (*Glicine Max (L.) Merrill*) no Estado do Paraná. Londrina, 1992. 122p. Dissertação (Mestrado)-Universidade Estadual de Londrina.
- ALLARD, R.W.; BRADSHAW, A.D. Implications of genotype-environmental interactions in applied plant breeding. *Crop Science*, v.4, n.5, p.503-508, 1964.
- ATLIN, G.N.; MCRAE, K.B.; LU, X. Genotype \times region interaction for two-barley yield in Canada. *Crop Science*, v.40, n.6, p.1-6, 2000.
- BECKER, H.C. Correlations among some statistical measures of phenotypic stability. *Euphytica*, v.30, n.3, p.835-840, 1981.
- BECKER, H.C.; LÉON, J. Stability analysis in plant breeding. *Plant Breeding*, v.101, n.1, p.1-23, 1988.
- BICKEL, P.; FREEDMAN, D. Some asymptotic theory for the bootstrap. *Annals of Statistics*, v.1, n.9, p.1196-1197, 1981.
- BRASIL, E.M. Comparação de métodos no estudo da interação de genótipos com ambientes em milho (*Zea Mays L.*). Goiânia, 1990. 181p. Tese (Mestrado) – Universidade Federal de Goiânia – UFG.
- CARLINE-GARCIA, L.A. Uso do método de reamostragem bootstrap na estimação de parâmetros genéticos populacionais. Piracicaba, 2001. 162p. Tese Doutorado-Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo.
- CHAVES, L.J. Interação de genótipos com ambientes. In: Nass, L.L.; Valois,

- A.C.C.; Melo, I.S.; Valadares-Inglis, M.C. (Ed.). Recursos genéticos & melhoramento de plantas. Rondonópolis: Fundação MT, p. 816-858, 2001.
- CHAVES, L.J.; VENCOVSKY, R.; GERALDI, I.O. Modelo não linear aplicado ao estudo da interação de genótipos \times ambientes em milho. Pesquisa Agropecuária Brasileira, v.24, n.2, p. 259-269, 1989.
- COCKERHAM, C.C. Estimation of genetics variance. In: Hanson, W.D.; Robinson, H.F. Eds. Statistical genetics and plant breeding. Madison: National Academy of Sciences, 1963. p.53-94.
- CORNELIUS, P.L. Statistical tests and retention of terms in the additive main effects and multiplicative interaction model for cultivar trials. Crop Sci. 33:1186-1193, 1993.
- CORNELIUS, P.L., AND J. CROSSA. Prediction assessment of shrinkage estimators of multiplicative model for multi-environment cultivar trials. Crop Sci. 39:998-1009, 1999.
- CORNELIUS, P.L., J. CROSSA, and M.S. SEYEDSADR. Tests and estimators of multiplicative models for variety trials. p. 156-166. In Proceedings of Annual Kansas State University Conference on Applied Statistics in Agriculture, 5th. Manhattan, KS. 25-27 April 1993. Dep. of Statistics, Kansas State Univ. Manhattan, KS.
- CORNELIUS, P.L., J. CROSSA, and M.S. SEYEDSADR. Statistical tests and estimators of multiplicative models for genotype-by-environment interaction. p. 199-

234, 1996.. In M.S. Kang and H.G. Gauch (ed.) Genotype-by-environment interaction. CRC Press. Boca Raton, FL.

CORNELIUS, P.L., M. SEYEDSADR, and J. CROSSA. Using the shifted multiplicative model to search for “separability” in crop cultivar trials. *Theor. Appl. Genet.* 84:161- 172, 1992.

CORSTEN, L.C.A. and C.A. van EIJSBERGEN. Multiplicative effects in two-way analysis of variance. *Statistica Neerlandica*, 26, 61-68, 1972.

CROSSA, J. Statistical analyses of multilocation trials. *Adv. Agronomy* 44:55-85, 1990.

CROSSA, J., P.L. CORNELIUS, and W. YAN. Biplot of linear-bilinear models for studying crossover genotype x environment interaction. *Crop Sci.* 42:619-633, 2002.

CROSSA, J., P.N. FOX, W.H. PFEIFER, S. RAJARRAM, and H.G. GAUCH, AMMI adjustment for satatistical analysis of na international wheat yield trial. *Theor. Appl. Genet.* 81:27-37, 1991.

CROWLEY, P.H. Resampling methods for computation-intensive data analysis in ecology and evolution. *Ann. Rev. Ecol. Syst.* 23: 405-447, 1992.

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

- CRUZ, C.D.; TORRES, R.A.A.; VENCOVSKY, R. An alternative approach to the stability analysis proposed by Silva and Barreto. *Revista Brasileira de Genética*, v.12, n.2, p.567-580, 1989.
- CRUZ, C.D. A informática no melhoramento genético In: Nass, L.L.; Valois, A.C.C.; Melo, I.S.; Valadares-Ingliš, M.C. (Eds.). *Recursos genéticos & melhoramento de plantas*. Rondonópolis: Fundação MT, p. 1085-1118, 2001.
- DANIEL, C. *Application of statistics to industrial experimentation*, New York: John Wiley, 1976.
- DIAS, C.T.S., and W.J. KRZANOWSKI. Model selection and cross-validation in additive main effect and multiplicative interaction (AMMI) models. *Crop Sci.* 43:865-873, 2003.
- DICICCIO, T.J.; EFRON, B. Bootstrap confidence interval. *Statistical Science*, v.11, n.11, p.189-228, 1996.
- DUARTE, J.B. Estudo da adaptabilidade e estabilidade fenotípica em cultivares de feijão mulatinho (*Phaseolus vulgaris* L.). Goiânia, 1988. 155p. Dissertação (Mestrado)-Escola de Agronomia, Universidade Federal de Goiás.
- DUARTE, J.B., and R.VENCOVSKY. Interação genótipos x ambientes-uma introdução à análise "AMMI". Ribeirão Preto, S.P, 1999.

- DUARTE, J.B.; ZIMMERMAN, M.J.O. correlation among yield stability parameters in common bean. *Crop Science*, v.35, n.3, p.905-912, 1995.
- EASTMENT, H.T., and W.J. KRZANOWSKI. Cross-validatory choice of the number of components from a principal component analysis. *Technometrics* 24:73-77, 1982.
- EBERHART, S.A., and W.A. RUSSELL. Stability parameters for comparing varieties. *Crop Sci.* 6:36-40, 1966.
- EFRON, B. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7, 1-26, 1979.
- EFRON, B.; STEIN, C. The jackknife estimate of variance. *Annals of Statistics*, v.9, n.3, p.586-596, 1981.
- EFRON, B.; TIBSHIRANI, R.J. Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy. *Statistical Science*, v.1, n.1, p.54-76, 1986.
- EFRON, B.; TIBSHIRANI, R.J. *An introduction to the bootstrap*. London: Chapman and Hall, 1993. 573p.
- FALCONER, D.S.; MACKAY, T.F.C. *Introduction to quantitative genetics*. 4.ed. Harlow: Longman, 446p, 1996.

FALCONER, D.S. Introduction to quantitative genetics. 3.ed. Harlow: Longman, 438p, 1989.

FEHR, W.R. Principles of cultivar development. New York: Macmillan, 1987. cap.18, p.247-258.

FINLAY, K.W., and G.N. WILKINSON. The analysis of adaptation in a plant-breeding programme. Austr. J. Agric. Res. 14:742-754, 1963.

FOX, P.N.; CROSSA, J.; ROMAGOSA, I. Multi-environmental testing and genotype \times environment interaction. In: KENPTON, R.A.; FOX, P.N. Statistical methods for plant variety evaluation. London: Chapman & Hall, cap.8, p.117-138, 1997.

GABRIEL, K.R. Least squares approximation of matrices by additive and multiplicative models. Journal of Royal Statistical Society, Series B 40:186-196, 1978.

GABRIEL, K.R. Le biplot – outil d'exploration de données multidimensionnelles. Journal de la Societe Francaise de Statistique, 143, 2002. (to appear).

GAUCH, H.G. Model selection and validation for yield trials with interaction. Biometrics, 44:705-715, 1988.

GAUCH, H.G. Statistical analysis of regional yield trials; AMMI analysis of factorial designs. Elsevier Science. New York, 1992.

- GAUCH, H.G., and R.W. ZOBEL. Predictive and postdictive success of statistical analysis of yield trials. *Theor. Appl. Genet.* 76:1-10, 1988.
- GAUCH, H.G., and R.W. ZOBEL. AMMI analysis of yield trials. p. 85-122, 1996. In: M.S. Kang and H.G. Gauch, (ed). *Genotype by environment interaction*. CRC Press. Boca Raton, FL.
- GOLLOB, H.F. A statistical model which combines features of factor analytic and analysis of variance techniques. *Psychometrika* 33(1):73-115, 1968.
- GOOD, I.J. Some applications of the singular decomposition of a matrix. *Technometrics* 11(4):823-831, 1969.
- HEGEMANN, V., and D.E. JOHNSON. On analyzing two-way AOV with interaction. *Technometrics*, 18, 273-281, 1976.
- HEINRICH, G.M.; FRANCIS, C.A.; EASTIN, J.D. Stability of grain sorghum yield components across diverse environments. *Crop Science*, v.32, n.2, p.209-212, 1983.
- JEFFERS, J.N.R. Two case studies in the application of principal component analysis. *Applied Statistics* 16:225-236, 1967.
- JOHNSON, D.E., and F.A. GRAYBILL. An analysis of a two-way model with interaction and no replication. *Journal of the American Statistical Association*, 67, 862-868, 1972.

- KANG, M.S., and R. MAGARI. 1996. New developments in selecting for phenotypic stability in crop breeding. p. 1-14. IN M.S Kang and H.G. Gauch (ed). Genotype by environment interaction. CRC Press. Boca Raton, FL.
- KENDALL, M.G. Multivariate Analysis (2nd Ed.): Charles Griffin & Co., London, 1980.
- KRZANOWSKI, W.J. Some exact percentage points of a statistic useful in analysis of variance and principal component analysis. *Technometrics* 21:261-263, 1979.
- KRZANOWSKI, W.J. An Introduction to Statistical Modelling. Arnold, London, 1998.
- KRZANOWSKI, W.J. Cross-validation in principal component analysis. *Biometrics* 43: 575-584, 1987.
- KRZANOWSKI, W.J. Missing value imputation in multivariate data using the singular value decomposition of a matrix. *Listy Biometryczne-Biometrical Letters XXV* (1,2):31- 39, 1988.
- LAVORANTI, O.J. Estabilidade e adaptabilidade fenotípica através da reamostragem “Bootstrap” no modelo AMMI. Tese Doutorado Escola Superior de Agricultura Luiz de Queros, 166p. Piracicaba 2003.

LIN, C.S.; BINNS, M.R. A method of analyzing cultivars \times locations \times year experiment: a new stability parameter. *Theoretical and Applied genetics*, Berlin, v.76, n.1, p.425-430, 1988.

LIN, C.S.; BINNS, M.R.; LEFKOVICTH, L.P. Stability analysis: where do we stand? *Crop Science*, v.26, n.5, p.894-900, 1986.

LERNER, I.M. *Genetic homeostasis*. London, Oliver and Boyd, 134p, 1954.

LEWIS, D. Gene-environment interaction: A relationship between dominance heterosis phenotype stability and variability. *Heredity*, v.8, n.3, p.333-356, 1954.

MANDEL, J. Non-additivity in two-way analysis of variance. *J. Am. Statist. Assoc.* 56:878-888, 1961.

MANDEL, J. The partitioning of interaction in analysis of variance. *J. Res. Int. Bur. Stand. Sect. B* 73:309-328, 1969.

MANDEL, J. A new analysis of variance model for non-additive data. *Technometrics* 13(1):1-18, 1971.

MANLY, B.F.J. *Randomization and Monte Carlo Methods in Biology* (2nd Edition). London: Chapman and Hall, 1997.

MARIOTTI, J.A.; OYARZABAL, E.S.; OSA, J.M.; BULACIO, A.N.R.; ALMADA, G.H. Analisis de estabilidad y adaptabilidad de genótipos de caña de azucar. I.

- Interacciones dentro de una localidad experimental. *Revista Agronomica del Noroeste Argentino*, v.13, n.1-4, p. 105-127, 1976.
- MAY, K.W.; KOZUB, G.C. Genotype \times environment interactions for two-row barley grain yield and implications for selection of test locations. *Canadian Journal of Plant Science*, Ottawa, v.75, n.2, p.571-575, 1995.
- METTHER, L.E.; GREGG, T.G. Genética de populações e evolução. Tradução de R. Vencovsky, J.L. de Azevedo e G. Bandel. São Paulo: USP/Polígono, 1973.
- MILLGAN, S.B. Test sit within and among stages of a sugarcane breeding program. *Crop Science*, v.34, n.2, p.1184-1190, 1994.
- NUNES, G.H.S. Interação genótipo \times ambientes em eucalipto: implicações sobre a seleção e formas de atenuar seu efeito. Lavras, 2000. 160p. Tese(Doutorado)-Universidade Federal de Lavras.
- PIEPHO, H.P. Best linear unbiased prediction (BLUP) for regional yield trials: a comparison to additive main effects and multiplicative interaction (AMMI) analysis. *Theor. Appl. Genet.* 89:647-654, 1994.
- PIEPHO, H.P. Robustness of statistical test for multiplicative terms in additive main effects and multiplicative interaction model for cultivar trial. *Theor. Appl. Genet.* 90:438-443, 1995.
- QUENOUILLE, M.H. Notes on bias in estimation. *Biometrika*, **61**, 353-360, 1956.

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

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

ROSSE, L.N. Modelos de regressão não linear aplicado na avaliação da estabilidade fenotípica em plantas. Piracicaba, 179p, 1999. Tese (doutorado) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo.

SEYEDSADR M., AND P.L. CORNELIUS. Shifted multiplicative models for nonadditive two-way tables. Commun. Stat. B Simul. Comp. 21:807-832, 1992.

SILVA, J.G.C.; BARRETO, J.N. Aplicação da regressão linear segmentada em estudos da interação genótipo por ambiente. In: SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA, Piracicaba, 1985. Anais. Campinas: Fundação Cargill, 1985. p.49-50.

SNEE, R.D., Nonadditivity in a two-way classification: is it interaction or nonhomogeneous variance?, Journal of American Statistical Association, 77:515-519, 1982..

STEEL, R.G.D. and J.H. TORRIE. Principles and Procedures of Statistics, a Biometrical Approach. New York: McGraw-Hill, 1980.

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

STORK, L. Modelos de regressão bi-segmentados descontínuos com erros de medidas aplicadas na análise de estabilidade de cultivares. Piracicaba, 1989. 217p. Tese (Doutorado) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo.

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

TUKEY, J.W. Bias and confidence in note quite large sample, abstract. *Annals of Mathematics and Statistics*, v.1, n.29, p.614-623, 1958.

VENKCOVSKY, R. Herança quantitativa. In: Paterniani, E.; Viégas, G.P. (Coords.) *Melhoramento e produção de milho*. Campinas: Fundação Cargill, v.1, p137-214, 1987.

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

VENCOVSKY, R.; TORRES, R.A.A. Estabilidade geográfica e temporal de algumas cultivares de milho. In: *Congresso Nacional de Milho e Sorgo*, 16., Belo Horizonte, 1988. *Anais*. Belo Horizonte: EMBRAPA, CNPMS, p.294-300, 1988.

- VENTER, J.H., and S.J. STEEL. Simultaneous selection and estimation for the some zeros family of normal models. *Journal of Statistical Computation and simulation*. 45:129-146, 1993.
- VERMA, M.M.; CHAHAL, G.S.; MURTY, B.R. Limitations of conventional regression analysis, a proposed modification. *Theoretical and Applied Genetics*, v.53, n.2, p.89-91, 1978.
- VARGAS, M.V., and J. CROSSA. The AMMI analysis and graphing the biplot. *CIMMYT, INT. Mexico*, 39p, 2000.
- WEBER, W.E., G. WRICKE, and T. WESTERMANN. Selection of genotypes and predictions performance and analysing genotype-by-environment interactions. p. 353-371, 1996. In M.S. Kang and H.G. Gauch (ed.) *Genotype by environment interaction*. CRC Press. Boca Raton, FL.
- WOLD, S. Pattern recognition by means of disjoint principal component models. *Pattern Recognition* 8:127-139, 1976.
- WOLD, S. Cross-validatory estimation of the number of components in factor and principal component models. *Technometrics* 20:397-405, 1978.
- YATES, F., AND W.G. COCHRAN. The analysis of groups of experiments. *J. Agric. Sci. (Cambridge)* 28:556-580, 1938.
- YOCHMOWITZ, M.G., and R.G. CORNELL. Stepwise tests for multiplicative Components of interaction. *Technometrics*, 20, 79-84, 1978.

YUE, G.L.; ROOZEBOOM, K.L.; SCHAPAUGH Jr., W.T.; LIANG, G.H. Evaluation of soybean cultivars using parametric and nonparametric stability estimates. *Plant Breeding*, v.116, n.3, p.271-275, 1997.

ZOBEL, R.W., M.J. WRIGHT, and H.G. GAUCH, JR. Statistical analysis of a yield trial. *Agron. J.* 80:388-393, 1988.