

AGRUPAMENTO DE GENÓTIPOS SEGUNDO A ADAPTABILIDADE AO AMBIENTE¹

JOÃO GILBERTO CORRÊA DA SILVA²

RESUMO - Em uma coleção de genótipos, em geral, podem ocorrer diversos padrões de adaptabilidade ao ambiente, caracterizados por diferentes combinações de distintos níveis de rendimento médio e de sensibilidade de resposta em ambientes desfavoráveis e em ambientes favoráveis. O método de regressão linear segmentada (Silva & Barreto, 1985; Silva, 1995a, 1995b) permite identificar e discriminar esses distintos comportamentos de genótipos, através de inferências referentes aos parâmetros de um modelo linear. Este artigo apresenta uma extensão deste método para o agrupamento de genótipos segundo a similaridade das estimativas dos parâmetros de adaptabilidade do modelo linear segmentado. Esse método de agrupamento pode ser conveniente em situações em que seja desejada a caracterização e classificação dos genótipos segundo os distintos padrões de adaptabilidade ao ambiente. O método baseia-se em um índice de dissimilaridade provido pela estatística F para teste de homogeneidade de linhas de regressão. O método é ilustrado através de sua aplicação a dados de um experimento de comparação de cultivares de arroz de sequeiro.

Termos para indexação: interação genótipo x ambiente, agrupamento de cultivares, regressão linear segmentada, análise de agrupamento.

GROUPING OF GENOTYPES ACCORDING TO ENVIRONMENTAL ADAPTATION

ABSTRACT - In a set of genotypes, in general, several patterns of environmental adaptation may occur, characterized by different combinations of distinct levels of average yield and sensitivity of response in unfavorable and favorable environments. The method of segmented linear regression (Silva & Barreto, 1985; Silva, 1995a, 1995b) allows the identification and discrimination of these distinct genotype behaviors, through inferences related to the parameters of a linear model. This paper presents an extension of this method for grouping genotypes according to the similarity of the estimates of the parameters of adaptability of the segmented linear model. This method of grouping may be convenient when the characterization and classification of the genotypes according to the distinct patterns of adaptation is desired. The method is based on a dissimilarity index provided by the F statistics for testing the hypothesis of homogeneity of regression lines. The method is illustrated through its application to data of an experiment of comparison of upland rice cultivars.

Index terms: genotype x environment interaction, grouping of cultivars, segmented linear regression, cluster analysis.

¹ Aceito para publicação em 3 de agosto de 1995.

² Eng. Agr., Ph.D., Prof. Titular, Dep. de Matemática, Estatística e Computação, Inst. de Física e Matemática, Univ. Fed. de Pelotas, Caixa Postal 354, CEP 96010-900 Pelotas, RS. Bolsista do CNPq.

INTRODUÇÃO

Genótipos de uma coleção suficientemente grande geralmente apresentam variados padrões de adaptabilidade ao ambiente que se distinguem por diferentes comportamentos médios e sensibilidades de resposta à variação ambiental. O genótipo teoricamente ideal é aquele a que correspondem resposta média elevada e sensibilidade de resposta baixa sob condições ambientais desfavoráveis, e elevada sob condições ambientais favoráveis. Outros padrões de adaptabilidade podem ser caracterizados pelas combinações de distintos níveis de comportamento médio e de sensibilidade de resposta à variação de condições ambientais em ambientes desfavoráveis e em ambientes favoráveis.

O método de regressão linear simples tem sido o mais usado para a análise da interação genótipo x ambiente com vistas ao estudo da adaptabilidade de genótipos. Este método relaciona as respostas individuais dos genótipos com o "índice de ambiente", provido pela resposta média dos próprios genótipos no experimento. Esta abordagem foi originalmente sugerida por Yates & Cochran (1938) e posteriormente desenvolvida, principalmente, por Finlay & Wilkinson (1963) e Eberhart & Russell (1966).

Entretanto, essa técnica de regressão linear simples não identifica os distintos padrões de adaptabilidade que resultam das combinações de diferentes níveis de sensibilidade de resposta em condições ambientais abaixo da média e acima da média. Em particular, ela não é adequada para determinar os genótipos com as características desejáveis, ou seja, genótipos que possuam a habilidade de tomar vantagem de ambientes favoráveis ou melhorados e que mantenham rendimentos razoáveis em ambientes adversos (Verma et al., 1978; Nor & Cady, 1979).

Com base nesse argumento, Silva & Barreto (1985) sugerem representar a resposta de cada genótipo à gama de ambientes em consideração por um gráfico composto de dois segmentos de reta, conectados no ponto correspondente ao índice de ambiente nulo. Esse modelo linear segmentado permite levar em conta a possibilidade de alteração da sensibilidade de resposta do genótipo entre ambientes desfavoráveis e favoráveis, demarcados pelo índice de ambiente zero.

Silva (1995a) apresenta os fundamentos do método de regressão linear segmentada e sua abordagem no contexto do modelo linear geral, que permite procedimentos de inferência que propiciam amplas possibilidades para a caracterização e comparação da adaptabilidade de cultivares. Silva (1995b) descreve a aplicação deste método, utilizando dados de um experimento de comparação de cultivares de arroz de sequeiro.

Segundo o método de regressão linear segmentada, a adaptabilidade de um genótipo ao ambiente é caracterizada por três parâmetros: o rendimento médio e os dois coeficientes de regressão parciais do modelo de regressão linear segmentada, estes dois últimos exprimindo a taxa de variação da resposta nos ambientes desfavoráveis e a alteração da taxa de resposta entre os ambientes desfavoráveis e favoráveis.

Através da aplicação deste método, Silva (1995b) classifica as oito cultivares de arroz de sequeiro de um experimento conduzido em vinte e um ambientes em dois grupos, caracterizados por distintos padrões de resposta nos ambientes desfavoráveis e nos ambientes favoráveis. Esse resultado pode sugerir que uma caracterização e classificação da adaptabilidade de genótipos pode ser obtida pela inspeção das linhas de regressão individuais dos genótipos. Entretanto, em geral, pode não ser tão simples e nítida a caracterização dos distintos padrões de adaptabilidade de um grupo de genótipos através da interpretação de conclusões de resultados de inferências derivadas separadamente para os distintos parâmetros de adaptabilidade.

A razão básica para essa dificuldade é que a resposta de um genótipo ao ambiente é multivariada. Uma saída para esse dilema é o agrupamento de genótipos de acordo com as estruturas de suas respostas, que logre sua classificação em subconjuntos de genótipos de adaptabilidade qualitativamente homogênea, em vez da ordenação da adaptabilidade através de medidas quantitativas. A vantagem da classificação é que, embora os genótipos sejam agrupados com base em um conjunto específico de dados, a relação relativa entre os genótipos pode ser independente do conjunto de dados analisado (Lin et al., 1986).

Francis & Kannenberg (1978) salientam que em estudos referentes ao fundamento fisiológico da estabilidade do rendimento tem sido mais prático caracterizar genótipos na base de grupos do que individualmente. Eles adicionam que o método de agrupamento também poderá ser útil no contexto do melhoramento de plantas.

Nessas circunstâncias, pode ser desejável um procedimento para classificação de genótipos em grupos, cada um dos quais constituído de genótipos com mesmo padrão de adaptabilidade ao ambiente. Westcott (1986) e Lin et al. (1986) efetuam uma revisão e discutem os diversos métodos de agrupamento de genótipos que têm aparecido na literatura.

O presente artigo apresenta um método para o agrupamento de genótipos que combina o método de regressão linear segmentada (Silva & Barreto, 1985; Silva 1995a, 1995b) com a técnica multivariada da análise de agrupamento ou análise de conglomerado (cluster analysis). Esse método efetua o agrupamento de genótipos segundo as similaridades das estimativas dos parâmetros que caracterizam a adaptabilidade pelo método de regressão linear segmentada. Trata-se de uma adaptação de um método sugerido por Lin & Thompson (1975). O método baseia-se em um índice de dissimilaridade provido pela estatística F para teste de homogeneidade de linhas de regressão. O método é ilustrado através de sua aplicação aos dados do experimento de comparação de cultivares de arroz de sequeiro considerado por Silva (1995b).

MÉTODO DE AGRUPAMENTO

O método de regressão linear segmentada (Silva, 1995a) exprime o rendimento médio observado do *i*-ésimo genótipo no *j*-ésimo ambiente pela seguinte equação:

$$y_{ij} = b_{0i} + b_{1i}x_{1j} + b_{2i}x_{2j} + e_{ij}, \quad i = 1, 2, \dots, C, \quad j = 1, 2, \dots, A,$$

onde: $x_{1j} = I_j$, $x_{2j} = \delta_j I_j$, sendo I_j o *índice de ambiente*, ou seja, para cada ambiente, o desvio de sua média em relação à média geral: $I_j = \bar{y}_j - \bar{y}$, e $\delta_j = 0$ se $I_j \leq 0$, $= 1$ se $I_j > 0$; b_{1i} é a declividade do segmento de reta nos ambientes desfavoráveis; b_{2i} é a alteração de declividade da reta segmentada entre os ambientes desfavoráveis e favoráveis; b_{0i} é a ordenada do ponto de interseção desses dois segmentos de reta; e $e_{ij} = \epsilon_{ij} + d_{ij}$, sendo ϵ_{ij} o desvio entre os rendimentos observado e esperado do i -ésimo genótipo no ambiente j , e d_{ij} o desvio entre o efeito da interação genótipo x ambiente e seu componente sistemático atribuível à relação linear com o efeito de ambiente.

Um método para o agrupamento de genótipos segundo a similaridade de seus parâmetros de adaptabilidade, ou seja, b_{0i} , b_{1i} e b_{2i} , $i = 1, 2, \dots, C$, é descrito a seguir.

A soma dos quadrados para o teste da homogeneidade simultânea desses parâmetros de adaptabilidade para um conjunto de genótipos será tão mais elevada quanto mais distintas forem as respectivas estimativas dos parâmetros para os genótipos. Assim, é razoável definir para índice de dissimilaridade de um conjunto de genótipos uma função da soma dos quadrados para o teste da homogeneidade simultânea desses parâmetros.

Denote-se por $SQD(G_i)$ a soma dos quadrados dos desvios da linha de regressão para o genótipo G_i , e, de modo mais geral, por $SQD(G_1, G_2, \dots, G_p)$, a soma dos quadrados dos desvios da linha de regressão comum para os genótipos G_1, G_2, \dots, G_p . Define-se como *índice de dissimilaridade* para o conjunto de genótipos G_1, G_2, \dots, G_p a seguinte estatística:

$$d(G_1, G_2, \dots, G_p) = \frac{SQD(G_1, G_2, \dots, G_p) - \sum_{i=1}^p SQD(G_i)}{3(p-1)s_{V|X, X_2}^2},$$

onde $s_{V|X, X_2}^2 = \sum_{i=1}^C SQD(G_i) / (C-1)(A-3)$, é o residuo combinado para os C genótipos.

A estatística $d(\cdot)$ é a razão de quadrados médios para testar a hipótese de homogeneidade das linhas de regressão dos p genótipos. Adota-se $(C-1)(A-3)$ graus de liberdade para a estimativa da variância do desvio da regressão combinado dos C genótipos, $s_{V|X, X_2}^2$, em vez de $C(A-3)$, para levar em conta o uso do índice de ambiente, estimado pelos próprios dados, como variável preditora (Silva, 1995a).

Observe-se que a estatística $d(\cdot)$ apresenta a seguinte propriedade: pequenos valores de $d(\cdot)$ correspondem a linhas de regressão mais semelhantes; valores crescentes de $d(\cdot)$ correspondem a linhas de regressão com dissimilaridade crescente. Demonstra-se, no Apêndice 2, que o índice de dissimilaridade para p genótipos é igual à média dos índices de dissimilaridade para todos os $p(p-1)/2$ pares dos p genótipos, ou seja:

$$d(G_1, G_2, \dots, G_p) = \frac{2}{p(p-1)} \sum_{i>j=1}^p d(G_i, G_j).$$

Também é demonstrado, no Apêndice 3, que a seqüência dos índices mínimos de dissimilaridade determinados para os $C-1$ ciclos do processo de agrupamento é monótona não decrescente, ou seja:

$$d'_1(\cdot) \leq d'_2(\cdot) \leq \dots \leq d'_{C-1}(\cdot).$$

onde $d'_n(.)$ é o índice de dissimilaridade mínimo no n-ésimo ciclo do processo de agrupamento.

Logo, o índice de dissimilaridade $d(.)$ satisfaz às condições para a aplicação do método de agrupamento por pares não ponderado (Sokal & Michener, 1958).

O método de agrupamento por pares não ponderado consiste em uma seqüência de ciclos, em cada um dos quais são determinados os coeficientes de dissimilaridade para todos os pares de grupos de genótipos construídos no ciclo anterior. O processo inicia com C grupos, constituídos pelos próprios genótipos, individualmente. Nesse primeiro ciclo, são determinados os índices de dissimilaridade para todos os $C(C-1)/2$ pares de genótipos. O par com o menor índice é reunido em um grupo. Com esse par constituindo um grupo e os demais genótipos constituindo grupos individuais, determinam-se, no segundo ciclo, os índices de dissimilaridade para os novos $(C-1)(C-2)/2$ pares de grupos. Nesse segundo ciclo, um novo grupo é formado, pela inclusão de um outro genótipo no grupo formado no primeiro ciclo, ou pela reunião de outros dois genótipos em um novo grupo. De qualquer forma, no terceiro ciclo são determinados os índices de dissimilaridade para $(C-2)(C-3)/2$ pares de grupos. O processo prossegue, com a diminuição de um grupo após cada ciclo, até o ciclo C-1, quando todos os C genótipos são incluídos em um único grupo.

A definição de um critério apropriado para a determinação do número adequado de grupos, ou seja, do ponto de parada do processo seqüencial de agrupamento, é, em geral, um problema dos métodos de agrupamento. Adota-se, aqui, o critério sugerido por Lin & Thompson (1975), baseado no teste F da hipótese de homogeneidade de linhas de regressão. O processo de agrupamento é parado no ciclo em que o índice de dissimilaridade mínimo exceder o ponto $100\alpha\%$ superior da distribuição F para $3(p-1)$ e $(C-1)(A-3)$ graus de liberdade, para uma probabilidade α de erro tipo I previamente estabelecida.

Os índices de dissimilaridade, em cada ciclo do processo, são determinados através do processo de teste de hipótese do modelo linear geral. Por exemplo, o numerador do índice de dissimilaridade para os genótipos G_1 e G_2 , ou seja, $d(G_1, G_2)$, é a diferença entre a soma de quadrados dos desvios da regressão conjunta para os genótipos G_1 e G_2 sob a hipótese de igualdade simultânea dos correspondentes parâmetros de adaptabilidade, e a soma de quadrados dos desvios da regressão conjunta sob a hipótese alternativa, ou seja, de diferenças entre os dois genótipos quanto a pelo menos um dos três parâmetros.

Os cálculos podem ser facilmente efetuados com o recurso de um "pacote" de análise estatística que disponha de um procedimento geral para análise de modelos lineares, como o procedimento GLM do SAS (SAS Institute, 1985).

ILUSTRAÇÃO DO MÉTODO

A aplicação do método é ilustrada com o uso dos dados de um experimento do programa de melhoramento genético de arroz de sequeiro do Instituto Agrônomo de Campinas, conduzido em 3 anos agrícolas (1974/75, 1975/76 e 1976/77), em 9 locais do Estado de São Paulo, compreendendo 21 ambientes. Silveira (1980) e Silveira & Vencovsky (1988) apresentam a caracterização das 8 cultivares do experimento, dos locais e de outros aspectos relevantes do experimento.

Silva (1995b) utiliza esses dados para ilustração da análise de adaptabilidade pelo método de regressão linear segmentada e apresenta os resultados dessa análise.

Como a razão entre a maior e a menor das estimativas de variância dos desvios das regressões para as 8 cultivares é inferior a sete, pode-se admitir a homogeneidade da variância dos desvios para todas as 8 cultivares (Pimentel-Gomes,

1990). A estimativa da variância do desvio da regressão comum para as 8 cultivares (quadrado médio do Resíduo Combinado) é $s_{V|X_1X_2}^2 = 299.371$, com 126 graus de liberdade (Silva, 1995b).

A matriz dos índices de dissimilaridade do primeiro ciclo do processo de agrupamento é apresentada na Tabela 1. O menor índice de dissimilaridade desse ciclo, $d_1(G_6, G_7) = 0,005$ determina o agrupamento das cultivares 6 e 7. Uma matriz de dissimilaridade semelhante é formada em cada um dos ciclos sucessivos, em cada um dos quais o menor índice de dissimilaridade determina a formação de um novo grupo de cultivares. Os resultados dos sete ciclos do processo de agrupamento são sumarizados na Tabela 2.

O índice de dissimilaridade mínimo excede o valor de $F[3(p-1); 126]$ para o nível de probabilidade $\alpha=0,05$ apenas no sétimo ciclo. Então, o critério adotado determina a formação de dois grupos de cultivares: um grupo constituído pelas cultivares 1, 2 e 3 e o outro, pelas cultivares 4, 5, 6, 7 e 8.

O resultado desse processo de agrupamento das cultivares quanto a padrões de adaptabilidade pode ser apresentado, de forma mais compreensível, em um diagrama de árvore, apresentado na Fig. 1.

Os dois grupos formados pelo método de agrupamento por pares não ponderado são nitidamente salientados pela distinção das estimativas dos parâmetros de adaptabilidade mostradas na Tabela 3, onde os grupos 1: (1,2,3) e 2: (4,5,6,7,8) são indicados na última coluna. Este resultado é consistente com o obtido por Silva (1995b).

Os resultados da análise da variação conjunta das médias das 8 cultivares nos 21 ambientes, com a decomposição da variação atribuível à Cultivar e à interação Cultivar x Ambiente que leva em conta o agrupamento formado, são apresentados na Tabela 4. Nesta Tabela, também são apresentadas as partições da interação Cultivar x

TABELA 1. Matriz de dissimilaridade para o primeiro ciclo do processo de agrupamento.

Grupo	Grupo						
	2	3	4	5	6	7	8
1	0,135	0,542	7,229	7,851	8,200	8,403	8,559
2	-	0,590	7,335	7,890	8,144	8,391	8,678
3	-	-	7,046	7,882	8,036	8,219	7,332
4	-	-	-	0,053	0,081	0,071	0,522
5	-	-	-	-	0,031	0,029	0,807
6	-	-	-	-	-	<u>0,005</u>	0,675
7	-	-	-	-	-	-	0,627

TABELA 2. Resumo dos resultados dos ciclos do processo de agrupamento de cultivares.

Ciclo	Grupos formados	$d'()$	Cultivares no Grupo (p)	$F[3(p-1); 126]$ $\alpha=0,05$
1	(6,7),1,2,3,4,5,8	0,005	2	2,68
2	(5,6,7),1,2,3,4,8	0,022	3	2,17
3	(4,5,6,7),1,2,3,8	0,045	4	1,95
4	(4,5,6,7),(1,2),3,8	0,135	2	2,68
5	(4,5,6,7,8),(1,2),3	0,290	5	1,83
6	(4,5,6,7,8),(1,2,3)	0,422	3	2,17
7	(1,2,3,4,5,6,7,8)	4,406	8	1,64

Ambiente e de seus componentes Grupo x Ambiente e Cultivar x Ambiente dentro de Grupo nos correspondentes componentes atribuíveis à heterogeneidade das retas segmentadas, $Heter.R(X_1, X_2)$, heterogeneidade das retas (não-segmentadas) para o conjunto total dos 21 ambientes, $Heter.R(X_1)$, e heterogeneidade das alterações de declividade dos segmentos de reta entre os ambientes desfavoráveis e favoráveis, $Heter.R(X_2|X_1)$.

Os resultados na Tabela 4 revelam que a maior parte da variação altamente significativa ($P < 0,0001$) atribuível à heterogeneidade das retas segmentadas para as

CULTIVAR

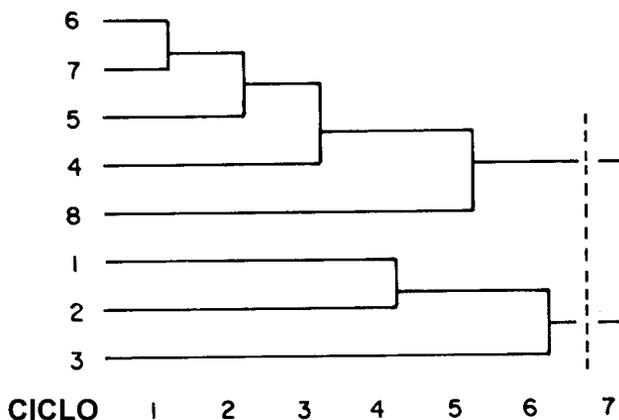


FIG. 1. Diagrama de árvore dos resultados do agrupamento das cultivares de arroz de sequeiro pelo método de agrupamento por pares não ponderado.

TABELA 3. Estimativas dos parâmetros de adaptabilidade das 8 cultivares determinadas pelo método de regressão linear segmentada, e de cada um dos dois grupos formados pelo método de agrupamento por pares não ponderado.

Cultivar	Média $\hat{b}_0 + \hat{b}_2 \bar{X}_2$	Taxa de resposta		Grupo
		\hat{b}_1	\hat{b}_2	
1	2.076	-0,368	1,957	1
2	2.087	-0,295	2,004	1
3	2.276	-0,161	1,669	1
4	1.932	1,645	-0,985	2
5	1.877	1,720	-1,047	2
6	1.908	1,801	-1,117	2
7	1.908	1,821	-1,173	2
8	2.125	1,837	-1,308	2
Grupo 1	2.146	-0,275	1,877	
Grupo 2	1.950	1,765	-1,126	

8 cultivares é levada em conta pela heterogeneidade significativa ($P < 0,01$) das retas segmentadas para os dois grupos. A heterogeneidade dentro dos grupos revela-se não-significativa ($P > 0,05$). A parte relevante da variação altamente significativa ($P < 0,0001$) devida à heterogeneidade das alterações de declividade dos segmentos de reta entre ambientes desfavoráveis e favoráveis também é levada em conta pela heterogeneidade significativa ($P < 0,01$) entre os dois grupos. A correspondente heterogeneidade dentro dos grupos mostra-se não-significativa. Esses resultados refletem o satisfatório agrupamento das cultivares.

TABELA 4. Análise da variação conjunta das médias das oito cultivares nos vinte e um ambientes, levando em conta o agrupamento das cultivares.

Fonte de variação	GL	SQ	QM	Prob.>F
Ambiente	20	174.156.609	8.707.831	< 0,0001
Cultivar	7	2.888.439	412.634	0,4921
Grupo	1	1.516.700	1.516.700	0,0687
Cultivar (Grupo)	6	1.371.739	228.623	0,7979
Cultivar x Amb.	140(90)	62.530.916	446.649	< 0,0001
Heter.R (X_1, X_2)	14	24.810.218	1.772.158	< 0,0001
Heter.R (X_1)	7	658.098	94.014	0,9845
Heter.R ($X_2 X_1$)	7	24.152.120	3.450.303	< 0,0001
Resíduo combin.	126	37.720.698	299.371	< 0,0001
Grupo x Ambiente	20	57.022.662	2.851.133	< 0,0001
Heter.R(X_1, X_2)	2	24.380.873	12.190.437	0,0066
Heter.R(X_1)	1	409.528	409.528	0,6404
Heter.R($X_2 X_1$)	1	23.971.345	23.971.345	0,0019
Resíduo combin.	18	32.641.789	1.813.433	< 0,0001
Cuit.xAmb.(Grupo)	120	5.508.254	45.902	< 0,0001
Heter.R(X_1, X_2)	12	429.345	35.779	0,6889
Heter.R(X_1)	6	248.570	41.428	0,5116
Heter.R($X_2 X_1$)	6	180.775	30.129	0,6974
Resíduo combin.	108	5.078.909	47.027	< 0,0001
Erro combinado	882 (557)		22.965,2	

DISCUSSÃO

Diferenças de genótipos quanto à adaptabilidade ao ambiente podem resultar de diferenças de resposta média à variação ambiental e de diferenças de sensibilidade de resposta em ambientes desfavoráveis e em ambientes favoráveis.

A análise da interação genótipo x ambiente através do método de regressão linear segmentada permite distinguir esses diversos comportamentos de resposta dos genótipos e separar os correspondentes componentes sistemáticos da interação da variação aleatória.

O presente método de agrupamento estende a abordagem de regressão linear segmentada a uma técnica descritiva que efetua o agrupamento de genótipos em subconjuntos homogêneos, cada um dos quais constituído de genótipos com linhas de regressão comuns, ou seja, com mesma média, mesma taxa de resposta em ambientes desfavoráveis e mesma taxa de resposta em ambientes favoráveis. Os diversos padrões de adaptabilidade dos genótipos são objetivamente caracterizados e clas-

sificados, levando em conta, simultaneamente, os três parâmetros de adaptabilidade do modelo linear segmentado.

Esse procedimento contorna o possivelmente complicado processo de síntese de resultados de testes de hipóteses referentes aos distintos parâmetros de adaptabilidade. Além disso, o método de agrupamento sugerido simplifica a caracterização da adaptabilidade relativa dos genótipos, com as seguintes vantagens (Lin & Thompson, 1975; Lin et al., 1986):

(i) O número de linhas de regressão para os genótipos é reduzido ao número de grupos. Dessa forma, a caracterização da adaptabilidade dos genótipos pode ser efetuada mais facilmente na base de grupos, o que é conveniente em algumas aplicações.

(ii) Os genótipos em um mesmo grupo têm o mesmo comportamento de resposta ao ambiente. Se um genótipo de desempenho conhecido é incluído em um grupo, ele pode servir de paradigma para os outros genótipos do grupo. Estes genótipos podem ser considerados como possuindo as mesmas características gerais do genótipo conhecido, o que pode permitir extrapolação referente ao desempenho destes genótipos para uma amplitude de ambientes muito maior do que aquela em que o experimento foi conduzido.

(iii) A adequabilidade e a eficácia do agrupamento podem ser avaliadas através da análise da variação, pela separação da variação atribuível às características dos grupos, tanto da variação entre genótipos como da variação devida à heterogeneidade das linhas de regressão para os genótipos.

O método de agrupamento usa como medida de dissimilaridade dos genótipos a estatística F para teste de diferenças entre linhas de regressão e o correspondente valor crítico de F como critério de parada para definição do agrupamento. Dessa forma, a variância do desvio da regressão deve ser homogênea para todos os genótipos. Essa condição, em geral, pode ser aceita se a razão entre o maior e o menor dos quadrados médios dos resíduos das regressões individuais para os genótipos for inferior a sete (Pimentel-Gomes, 1990). Se a variância do desvio da regressão revelar-se heterogênea, os métodos de agrupamento propostos por Lin (1982) e Ramey & Rosielle (1983) poderão ser alternativas mais apropriadas. Esses métodos, entretanto, não identificam e caracterizam objetivamente os padrões de adaptabilidade dos grupos formados através de parâmetros significativos.

O método de agrupamento exposto baseia-se conjuntamente nos três parâmetros de adaptabilidade do modelo linear segmentado, ou seja, na média e nas taxas de resposta dos genótipos nos ambientes desfavoráveis e nos ambientes favoráveis. Entretanto, ele pode ser adaptado para levar em conta apenas as distinções de taxa de resposta dos genótipos nos ambientes desfavoráveis e nos ambientes favoráveis. Ele também pode ser adaptado para o agrupamento de genótipos quando o modelo de regressão linear simples de Finlay & Wilkinson (1963) e Eberhart & Russell (1966) revelar-se apropriado.

APÊNDICE

1. Sejam G_1, G_2, \dots, G_p , $p > 1$, p grupos de genótipos, o grupo G_k constituído de v_k genótipos ($k=1, 2, \dots, p$; $v_1 + v_2 + \dots + v_p = v$, $v \leq C$). Então,

$$d(G_1, G_2, \dots, G_p) = \frac{1}{(p-1)v} \sum_{k > k'}^p (v_k + v_{k'}) d(G_k, G_{k'})$$

Proceder-se-á a prova em quatro passos:

(i) O estimador do rendimento médio dos genótipos do grupo G_k no j -ésimo ambiente é:

$$\hat{y}_{ijk} = \bar{y}_k + \hat{b}_{1k} x_{1j} + \hat{b}_{2k} (x_{2j} - \bar{x}_2), \quad i = 1, 2, \dots, v_k, \quad k = 1, 2, \dots, p, \quad j = 1, 2, \dots, A,$$

onde \hat{b}_{1k} e \hat{b}_{2k} são as taxas de resposta médias estimadas dos v_k genótipos do grupo G_k , respectivamente nos ambientes desfavoráveis e favoráveis. Então, a soma dos quadrados dos desvios da linha de regressão comum para os genótipos do grupo G_k é:

$$\begin{aligned} \text{SQD}(G_k) &= \sum_{i=1}^{v_k} \sum_{j=1}^A (y_{ijk} - \hat{y}_{ijk})^2 = \\ &= \sum_{i=1}^{v_k} \sum_{j=1}^A \left[y_{ijk} - \bar{y}_k - \hat{b}_{1k} x_{1j} - \hat{b}_{2k} (x_{2j} - \bar{x}_2) \right]^2 = \\ &= \sum_{i=1}^{v_k} \sum_{j=1}^A (y_{ijk} - \bar{y}_k)^2 - v_k \hat{b}_{1k}^2 \sum_{j=1}^A x_{1j}^2 - v_k \hat{b}_{2k}^2 \sum_{j=1}^A (x_{2j} - \bar{x}_2)^2 - \\ &\quad - 2v_k \hat{b}_{1k} \hat{b}_{2k} \sum_{j=1}^A x_{1j} (x_{2j} - \bar{x}_2). \end{aligned}$$

(ii) O estimador comum do rendimento médio de um genótipo dos grupos G_1, G_2, \dots, G_p é:

$$\hat{y}_{ijk} = \bar{y}_{(p)} + \hat{b}_{1(p)} x_{1j} + \hat{b}_{2(p)} (x_{2j} - \bar{x}_2), \quad i = 1, 2, \dots, v_k, \quad k = 1, 2, \dots, p, \quad j = 1, 2, \dots, A,$$

$$\text{onde: } \bar{y}_{(p)} = \frac{v_1 \bar{y}_1 + v_2 \bar{y}_2 + \dots + v_p \bar{y}_p}{v_1 + v_2 + \dots + v_p}, \quad \hat{b}_{1(p)} = \frac{v_1 \hat{b}_{11} + v_2 \hat{b}_{12} + \dots + v_p \hat{b}_{1p}}{v_1 + v_2 + \dots + v_p} \quad e$$

$$\hat{b}_{2(p)} = \frac{v_1 \hat{b}_{21} + v_2 \hat{b}_{22} + \dots + v_p \hat{b}_{2p}}{v_1 + v_2 + \dots + v_p}.$$

Então, a soma dos quadrados dos desvios da linha de regressão comum para os grupos de genótipos G_1, G_2, \dots, G_p é:

$$\begin{aligned} \text{SQD}(G_1, G_2, \dots, G_p) &= \sum_{k=1}^p \sum_{i=1}^{v_k} \sum_{j=1}^A (y_{ijk} - \hat{y}_{ijk})^2 = \\ &= \sum_{k=1}^p \sum_{i=1}^{v_k} \sum_{j=1}^A (y_{ijk} - \hat{y}_{(p)})^2 - \sum_{k=1}^p v_k \hat{b}_{1(p)}^2 \sum_{j=1}^A x_{1j}^2 - \\ &\quad - \sum_{k=1}^p v_k \hat{b}_{2(p)}^2 \sum_{j=1}^A (x_{2j} - \bar{x}_2)^2 - 2 \sum_{k=1}^p v_k \hat{b}_{1(p)} \hat{b}_{2(p)} \sum_{j=1}^A x_{1j} (x_{2j} - \bar{x}_2) = \\ &= \sum_{k=1}^p \sum_{i=1}^{v_k} \sum_{j=1}^A (y_{ijk} - \bar{y}_k)^2 + \frac{A}{v^2} \sum_{k \neq k'}^p v_k v_{k'} (\bar{y}_k - \bar{y}_{k'})^2 - \frac{(\sum v_k \hat{b}_{1k})^2}{v} \sum_{j=1}^A x_{1j}^2 - \\ &\quad - \frac{(\sum v_k \hat{b}_{2k})^2}{v} \sum_{j=1}^A (x_{2j} - \bar{x}_2)^2 - 2 \frac{(\sum v_k \hat{b}_{1k})(\sum v_k \hat{b}_{2k})}{v} \sum_{j=1}^A x_{1j} (x_{2j} - \bar{x}_2). \end{aligned}$$

(iii) Por definição,

$$\begin{aligned}
 3S_{Y|X_1X_2}^2 d(G_k, G_{k'}) &= \text{SQD}(G_k, G_{k'}) - \text{SQD}(G_k) - \text{SQD}(G_{k'}) = \\
 &= \frac{A}{(v_k + v_{k'})^2} \sum_{k \neq k'} v_k v_{k'} (\bar{y}_k - \bar{y}_{k'})^2 + \left[v_k \hat{b}_{1k}^2 + v_{k'} \hat{b}_{1k'}^2 - \frac{(v_k \hat{b}_{1k} + v_{k'} \hat{b}_{1k'})}{v_k + v_{k'}} \right] \sum_{j=1}^A x_{1j}^2 + \\
 &+ \left[v_k \hat{b}_{2k}^2 + v_{k'} \hat{b}_{2k'}^2 - \frac{(v_k \hat{b}_{2k} + v_{k'} \hat{b}_{2k'})}{v_k + v_{k'}} \right] \sum_{j=1}^A (x_{2j} - \bar{x}_2)^2 + \\
 &+ 2 \left[v_k \hat{b}_{1k} \hat{b}_{2k} + v_{k'} \hat{b}_{1k'} \hat{b}_{2k'} - \frac{(v_k \hat{b}_{1k} + v_{k'} \hat{b}_{1k'}) (v_k \hat{b}_{2k} + v_{k'} \hat{b}_{2k'})}{v_k + v_{k'}} \right] \sum_{j=1}^A x_{1j} (x_{2j} - \bar{x}_2) = \\
 &= \frac{v_k v_{k'}}{v_k + v_{k'}} \left[A(\bar{y}_k - \bar{y}_{k'})^2 + (\hat{b}_{1k} - \hat{b}_{1k'})^2 \sum_{j=1}^A x_{1j}^2 + (\hat{b}_{2k} - \hat{b}_{2k'})^2 \sum_{j=1}^A (x_{2j} - \bar{x}_2)^2 + \right. \\
 &\quad \left. + 2(\hat{b}_{1k} - \hat{b}_{1k'}) (\hat{b}_{2k} - \hat{b}_{2k'}) \sum_{j=1}^A x_{1j} (x_{2j} - \bar{x}_2) \right].
 \end{aligned}$$

(iv) Por definição,

$$\begin{aligned}
 3(p-1)S_{Y|X_1X_2}^2 d(G_1, G_2, \dots, G_p) &= \text{SQD}(G_1, G_2, \dots, G_p) - \sum_{k=1}^p \text{SQD}(G_k) = \\
 &= \frac{A}{v^2} \sum_{k \neq k'} v_k v_{k'} (\bar{y}_k - \bar{y}_{k'})^2 + \left[\sum_{k=1}^p v_k \hat{b}_{1k}^2 - \frac{(\sum v_k \hat{b}_{1k})^2}{v} \right] \sum_{j=1}^A x_{1j}^2 + \\
 &+ \left[\sum_{k=1}^p v_k \hat{b}_{2k}^2 - \frac{(\sum v_k \hat{b}_{2k})^2}{v} \right] \sum_{j=1}^A (x_{2j} - \bar{x}_2)^2 + \\
 &+ 2 \left[\sum_{k=1}^p v_k \hat{b}_{1k} \hat{b}_{2k} - \frac{(\sum v_k \hat{b}_{1k})(\sum v_k \hat{b}_{2k})}{v} \right] \sum_{j=1}^A x_{1j} (x_{2j} - \bar{x}_2) = \\
 &= \frac{1}{v} \sum_{k > k'} v_k v_{k'} \left[A(\bar{y}_k - \bar{y}_{k'})^2 + (\hat{b}_{1k} - \hat{b}_{1k'})^2 \sum_{j=1}^A x_{1j}^2 + (\hat{b}_{2k} - \hat{b}_{2k'})^2 \sum_{j=1}^A (x_{2j} - \bar{x}_2)^2 + \right. \\
 &\quad \left. + 2(\hat{b}_{1k} - \hat{b}_{1k'}) (\hat{b}_{2k} - \hat{b}_{2k'}) \sum_{j=1}^A x_{1j} (x_{2j} - \bar{x}_2) \right].
 \end{aligned}$$

Tendo em conta o resultado do passo (iii), obtém-se:

$$3(p-1)s_{Y|X_1X_2}^2 d(G_1, G_2, \dots, G_p) = \frac{1}{v} \sum_{k > k'} (v_k + v_{k'}) 3s_{Y|X_1X_2}^2 d(G_k, G_{k'}).$$

Portanto,

$$d(G_1, G_2, \dots, G_p) = \frac{1}{(p-1)v} \sum_{k>k'}^p (v_k + v_{k'}) d(G_k, G_{k'})$$

2. Se G_1, G_2, \dots, G_p são p genótipos individuais ($p>1$), então:

$$d(G_1, G_2, \dots, G_p) = \frac{2}{p(p-1)} \sum_{i>i'=1}^p d(G_i, G_{i'})$$

A prova é imediata do resultado do Apêndice 1, considerando cada um dos grupos G_1, G_2, \dots, G_p constituído de um único genótipo, ou seja, $v_k=1$, $k=1, 2, \dots, p$.

$$3. d'_1(\cdot) \leq d'_2(\cdot) \leq \dots \leq d'_{c-1}(\cdot).$$

Sejam $G_1, G_2, \dots, G_{c-n+1}$ os grupos de genótipos no ciclo $n-1$ ($1 \leq n \leq c-2$) do processo de agrupamento. (No ciclo zero, todos os grupos são constituídos de um único genótipo.) Seja (G_{k_1}, G_{k_2}) o grupo formado no n -ésimo ciclo, a que corresponde o índice mínimo de dissimilaridade $d'_n(\cdot) = d(G_{k_1}, G_{k_2})$. O novo grupo formado no ciclo $n+1$, com índice mínimo de dissimilaridade $d'_{n+1}(\cdot)$, provirá ou da inclusão de um outro grupo do ciclo $n-1$ no grupo G_{k_1}, G_{k_2} ou da fusão de outros dois grupos deste ciclo.

(i) Se o grupo formado no ciclo $n+1$ resultar da inclusão de um grupo G_{k_3} do ciclo $n-1$ no grupo (G_{k_1}, G_{k_2}) , então, segundo o resultado do Apêndice 1:

$$\begin{aligned} d'_{n+1}(\cdot) &= d(G_{k_1}, G_{k_2}, G_{k_3}) = \\ &= \frac{1}{2(v_{k_1} + v_{k_2} + v_{k_3})} \left[(v_{k_1} + v_{k_2}) d(G_{k_1}, G_{k_2}) + (v_{k_1} + v_{k_3}) d(G_{k_1}, G_{k_3}) + \right. \\ &\quad \left. + (v_{k_2} + v_{k_3}) d(G_{k_2}, G_{k_3}) \right]. \end{aligned}$$

Como $d'_n(\cdot) = d(G_{k_1}, G_{k_2})$ é o menor índice de dissimilaridade do ciclo n , segue-se que:

$$\begin{aligned} d'_{n+1}(\cdot) &\geq \frac{1}{2(v_{k_1} + v_{k_2} + v_{k_3})} \left[(v_{k_1} + v_{k_2}) d'_n(\cdot) + (v_{k_1} + v_{k_3}) d'_n(\cdot) + (v_{k_2} + v_{k_3}) d'_n(\cdot) \right] = \\ &= d'_n(\cdot). \end{aligned}$$

(ii) Se o grupo formado no ciclo $n+1$ provir da fusão de dois grupos do ciclo $n-1$ diferentes de G_{k_1} e G_{k_2} , seu índice de dissimilaridade deve ser, por definição, maior que $d'_n(\cdot)$. Assim,

$$d'_n(\cdot) \leq d'_{n+1}(\cdot), \quad 1 \leq n \leq c-2.$$

Portanto,

$$d'_1(\cdot) \leq d'_2(\cdot) \leq \dots \leq d'_{c-1}(\cdot)$$

REFERÊNCIAS

- EBERHART, S.A.; RUSSELL, W.A. Stability parameters for comparing varieties. **Crop Science**, Madison, v.6, p.36-40, 1966.
- FINLAY, K.W.; WILKINSON, G.N. The analysis of adaptation in a plant-breeding programme. **Australian Journal of Agricultural Research**, Victoria, v.14, p.742-754, 1963.
- FRANCIS, T.R.; KANNENBERG, L.W. Yield stability studies in short-season maize. I. A descriptive method for grouping genotypes. **Canadian Journal of Plant Science**, v.58, p.1029-1034, 1978.
- LIN, C.S. Grouping genotypes by a cluster method directly related to genotype-environment interaction mean square. **Theoretical and Applied Genetics**, v.62, p.277-280, 1982.
- LIN, C.S.; THOMPSON, B. An empirical method of grouping genotypes based on a linear function of the genotype-environment interaction. **Heredity**, Edinburgh, v.34, n.2, p.255-263, 1975.
- LIN, C.S.; BINNS, M.R.; LEFKOVITCH, L.P. Stability analysis: Where do we stand? **Crop Science**, Madison, v.26, p.894-900, 1986.
- NOR, K.M.; CADY, F.B. Methodology for identifying wide adaptability in crops. **Agronomy Journal**, Madison, v.71, p.556-559, 1979.
- PIMENTEL-GOMES, F. **Curso de Estatística Experimental**. 13. ed. São Paulo: Livraria Nobel, 1990. 467p.
- RAMEY, T.B.; ROSIELLE, A.A. HASS cluster analysis: A new method of grouping genotypes or environments in plant breeding. **Theoretical and Applied Genetics**, v.66, p.131-133, 1983.
- SAS INSTITUTE INC. **SAS User's Guide: Statistics**, Version 5. ed. Cary, NC: SAS Institute Inc., 1985. 956p.
- SILVA, J.G.C. da. Análise da adaptabilidade através de regressão linear segmentada: 1. Fundamentos. **Pesquisa Agropecuária Brasileira**, Brasília, v.30, n.4, p.435-448, 1995a.
- SILVA, J.G.C. da. Análise da adaptabilidade através de regressão linear segmentada: 2. Aplicação. **Pesquisa Agropecuária Brasileira**, Brasília, v.30, n.4, p.449-462, 1995b.
- SILVA, J.G.C. da; BARRETO, J.N. An application of segmented linear regression to the study of genotype x environment interaction. **Biometrics**, v.41, n.4, p.1093, 1985.
- SILVEIRA, E.P. **Interação genótipo por locais em arroz de sequeiro (*Oriza sativa* L.) no Estado de São Paulo**. Piracicaba: ESALQ-USP, 1980. 97p. Dissertação de Mestrado.
- SILVEIRA, E.P.; VENCOVSKY, R. Interaction of genotypes and environment for grain yield in upland rice. **Pesquisa Agropecuária Brasileira**, Brasília, v.23, n.12, p.1407-1412, 1988.

- SOKAL, R.R.; MICHENER, C.D. A statistical method for evaluating systematic relationships. **University of Kansas Science Bulletin**, n.38, p.1409-1438, 1958.
- VERMA, M.M.; CHAAL, G.S.; MURTY, B.R. Limitations of conventional regression analysis. A proposed modification. **Theoretical and Applied Genetics**, v.53, p.89-91, 1978.
- WESTCOTT, B. Some methods of analyzing genotype-environment interaction. **Heredity**, Edinburgh, v.56, p.243-253, 1986.
- YATES, F.; COCHRAN, W.G. The analysis of groups of experiments. **Journal of Agricultural Science**, Cambridge, v.28, p.556-580, 1938.