

Os métodos Biplot em Escalonamento multidimensional

The Biplot methods in Multidimensional scaling

Édila Cristina de Souza ¹, Carlos Tadeu dos Santos Dias ², Francisco Bezerra Neto ³,
Glauber Henrique Sousa Nunes ⁴

¹*Depto. de Estatística -UFMT.*

²*Depto. de Ciências Exatas - USP/ESALQ.*

³*Depto. Fitotecnia, Núcleo de Pós-Graduação – UNIFERSA.*

⁴*Depto. de Ciências Vegetais - UNIFERSA.*

Introdução

A análise estatística multivariada envolve um conjunto de métodos estatísticos e matemáticos, destinada a descrever e interpretar os dados que provém da observação de várias variáveis estudadas conjuntamente e algumas estruturas de correlação (JOHNSON; WICHERN, 2006).

Biplot é uma técnica multivariada proposta por Gabriel (1971), com o objetivo de representar graficamente uma matriz de dados, de tal forma que esta representação permita visualizar em um plano as relações e inter-relações entre as linhas e colunas desta matriz. Fatorando a matriz de dados original pela DVS (Decomposição em Valores Singulares) como a soma de produtos de matrizes que contém os marcadores de linhas e colunas que constituem os elementos para a representação gráfica, consegue-se uma avaliação visual da estrutura da matriz de dados (GOWER; HAND, 1996).

Independente da forma de fatoração, as aplicações do método Biplot dividem-se em representação da matriz de dados, com finalidade descritiva e diagnóstico de modelos. A técnica é bastante útil, pois o gráfico utilizado para representar simultaneamente as linhas e colunas de uma matriz de dados pode indicar a existência de agrupamentos entre as observações, assim como mostrar as variâncias e correlações entre as variáveis (CÁRDENAS et al, 2007).

A avaliação de cultivares e a identificação de mega-ambientes são os objetivos mais importantes em ensaios de avaliação de cultivares. As análises AMMI (Additive Main Effects and Multiplicative Interaction) e SREG (Sites Regression), apresentam propriedades importantes e adequadas ao estudo da interação genótipo x ambiente, sendo recomendadas suas utilizações pelos melhoristas de plantas (YAN, 2002). Os gráficos Biplots captam a porção padrão da interação (G x A), eles mostram quais os genótipos e ambientes que menos contribuíram para a interação (estáveis), combinações de genótipos e ambientes desejáveis quanto à adaptabilidade. O presente trabalho tem o objetivo de propor a utilização da metodologia MDS (Multidimensional Scaling), determinando assim as associações entre os fatores observados nas análises em delineamentos experimentais.

Material e métodos

Os dados a serem utilizados são relativos ao experimento realizado no Pólo Agroindustrial Mossoró-Assu, situado no Estado do Rio Grande do Norte que é o principal produtor e exportador de melão (*Cucumis melo L.*) do Brasil. O melão do tipo Gália é caracterizado pela forma arredondada do fruto, coloração amarela, polpa esverdeada com rendilhamento e aroma suave. Foram coletados no período de agosto a outubro, nos anos de 2000, 2001 e 2002, totalizando doze ambientes:

- a) Agropólo Mossoró-Assu: MO-00 Mossoró, 2000 (A1); MO-01 Mossoró, 2001 (A2); MO-02 Mossoró, 2002 (A3);
- b) Baraúna: BA-00, 2000 (A4); BA-01, 2001 (A5); BA-02, 2002 (A6);
- c) Assu: AS-00, 2000 (A7); AS-01, 2001 (A8); AS-02, 2002 (A9);
- d) Alto do Rodrigues: AR-00, 2000 (A10); AR-01, 2001 (A11); AR-02, 2002 (A12).

Os híbridos simples avaliados do melão do tipo Gália foram: GPS 400 (G1), DRG 1531 (G2), DRG 1537 (G3), DRG 1531 (G4), Solar King (G5), Galileu (G6), Supra (G7), Num 1502 (G8) e Arava (G9). A variável medida foi PRODUÇÃO (t.ha⁻¹), adotou-se o delineamento experimental aleatorizado em blocos com três repetições.

No melhoramento genético do meloeiro, é relevante considerar os aspectos relacionados ao ciclo, à resistência a doenças e a tolerância de pragas, à capacidade produtiva e a qualidade de fruto. Genótipos precoces são desejáveis porque agregam valor econômico e agrônômico, ofertando o produto em menor prazo e com economia de defensivos no controle de pragas ou infecções de microorganismos patogênicos. Além disso, as plantas precisam ser altamente produtivas e prolíficas para permitirem maiores ganhos ao produtor. Essas características devem estar associadas àquelas que conferem qualidade ao fruto (VIANA, 2006).

Multidimensional Scaling (MDS), Escalonamento Multidimensional ou Análises de Proximidades, é um método que trata de representar as medidas de proximidade (similaridades ou dissimilaridades) entre pares de objetos como distâncias em um espaço multidimensional em dimensão reduzida, permitindo assim a inspeção visual da estrutura dos dados.

Os algoritmos MDS tornaram-se populares com o surgimento dos computadores, no qual apresenta uma solução analítica sem interações. O principal ganho da utilização da técnica MDS, é que ela pode ser utilizada tendo variáveis de qualquer escala (COX; COX, 2001).

Considere a matriz de dissimilaridade $\Delta = [\delta_{ij}]$, de ordem (n x n), em que δ_{ij} representa uma medida de proximidade entre o i-ésimo e o j-ésimo objeto. Um algoritmo de redução dimensional obtêm a configuração dos pontos (vetores de coordenadas) denominado por $x = (x_1, x_2, \dots, x_n)$ de ordem (n x q), em uma dimensão menor, ou seja (n > q), também deve-se verificar se a matriz de distância euclidiana $\mathbf{D} = [d_{ij}]$ de ordem (n x n), em que d_{ij} é definido por $d_{ij}^2 = (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T = \|\mathbf{x}_i - \mathbf{x}_j\|^2$, em que $i=(1, \dots, n)$ e $j=(1, \dots, p)$. A matriz \mathbf{D} é obtida a partir desse conjunto de pontos, se aproxime ao máximo da matriz de dissimilaridade original, ou seja, $\mathbf{D} \approx \Delta$, (TIMM, 2002).

A validação feita em MDS, é a partir da matriz de coordenadas \mathbf{X} , pode-se obter a matriz de distâncias \mathbf{D} . Usa-se então uma função de *Stress*, para medir a diferença entre a matriz de dissimilaridade original e a matriz de distância calculada. A escolha do algoritmo utilizado na MDS não depende da determinação do número de dimensões, tarefa importante para um bom desempenho da análise dos dados e interpretação dos resultados (BORG; GROENEN, 2005).

$$\mathbf{X}_{(nxq)} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1q} \\ x_{21} & x_{22} & \cdots & x_{2q} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nq} \end{pmatrix} \Rightarrow \mathbf{D}_{(nxn)} = \begin{pmatrix} 0 & d_{12} & \cdots & d_{1n} \\ d_{21} & 0 & \cdots & d_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ d_{n1} & d_{n2} & \cdots & 0 \end{pmatrix} \quad \mathbf{\Delta}_{(nxn)} = \begin{pmatrix} 0 & \delta_{12} & \cdots & \delta_{1n} \\ \delta_{21} & 0 & \cdots & \delta_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ \delta_{n1} & \delta_{n2} & \cdots & 0 \end{pmatrix}$$

Matriz de distâncias Matriz de dissimilaridade

Processo de validação $d_{ij} \approx \delta_{ij}$

Existem outras distâncias além da distância Euclidiana, por exemplo, a distância de Malahanobis, a distância Euclidiana ponderada, medida de City block, medida Minkowski etc, podem ser mais apropriadas em alguns casos, onde dados originais podem ser ordinais ou nominais. Em algumas situações os dados são descritos por funções (contínuas ou discretas), neste caso pode ser utilizada uma medida de correlação (MANLY, 2008).

Gower e Harding (1988) aplicaram a definição do Biplot para o escalonamento multidimensional, Gower e Krzanowski (1999) e Krzanowski (2002) descreveram-na com mais detalhes. Esta técnica pode ser usada para melhorar a análise de distância e utilizaram os resultados algébricos da ACP (Análise de componentes principais) chamando de Análise de Coordenadas Principais.

Resultados e discussão

Na Tabela (1) apresenta-se as Médias previstas da variável Produção. A maior média da variável Produção está no G9 no ambiente A12 e a menor média está no G3 no ambiente A5.

Tabela 1 - Médias previstas da variável Produção

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
G1	30,65	21,47	18,87	24,02	18,68	21,16	28,96	28,85	21,44	21,45	18,55	30,59
G2	29,34	23,19	17,86	22,96	20,45	16,63	25,65	30,45	27,77	23,04	23,06	25,28
G3	30,99	26,69	20,31	24,01	14,52	20,52	26,41	35,64	31,11	23,11	26,80	34,18
G4	32,26	26,12	16,03	28,31	17,91	22,71	32,59	31,33	33,35	26,94	16,10	32,38
G5	28,99	30,55	22,34	32,27	18,51	22,73	31,32	37,75	26,92	24,30	23,31	31,68
G6	26,75	26,17	20,14	28,24	23,67	18,83	33,42	35,96	26,96	21,83	24,56	40,14
G7	33,74	28,87	18,30	27,49	24,56	22,11	34,32	37,79	32,49	31,07	19,73	47,38
G8	31,97	27,29	20,83	29,26	23,33	24,02	34,56	39,71	30,43	28,69	24,12	47,28
G9	31,56	28,88	21,71	25,37	20,97	22,76	30,66	39,03	40,33	28,74	24,15	49,26

A Figura (1), refere-se ao Biplot com Escalonamento Multidimensional dos nove genótipos e dos doze ambientes. Na representação, as distâncias entre os pontos estão diretamente relacionadas com dissimilaridades entre os ambientes. Os ambientes A2 e A4; A1 e A7; A2 e A10 possuem os menores valores, portanto são mais semelhantes entre eles, do que em relação aos outros ambientes. Os ambientes A3 e A5 em relação ao ambiente A12 são os mais distantes, portanto apresentam maior diferença entre eles em relação à produção.

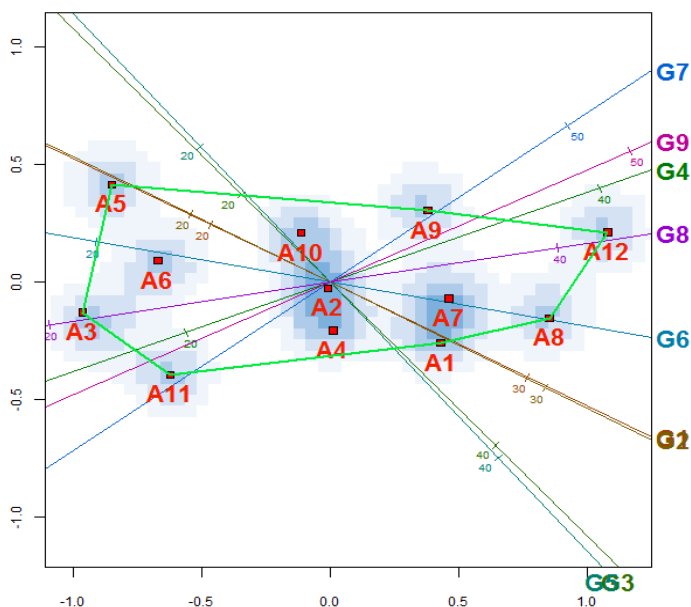


Figura 1 – Biplot com Escalonamento Multidimensional dos nove genótipos (G) e doze ambientes (A) para a variável produção.

Conclusões

A utilização da metodologia de MDS para os ambientes, mostrou-se eficiente em verificar os locais pela similaridade do desempenho dos cultivares, como técnica de redução dimensional se tentou preservar ao máximo as distâncias originais. O Biplot construído representou cada objeto (ambiente) por um ponto e cada variável (genótipo) por um vetor, permitindo assim, avaliar: as semelhanças dentro do conjunto de ambientes; as semelhanças dentro do conjunto de genótipos e as relações entre os ambientes e genótipos.

Referências bibliográficas

BORG, I.; GROENEN, P. J. F.; **Modern Multidimensional Scaling Theory and Applications**. New York, North Carolina: Springer Series in Statistics, 2005. 614 p.

COX, T. F.; COX, M. A. A.; **Multidimensional Scaling**. London: Chapman & Hall, 2001. 332 p.

CÁRDENAS, O.; GALINDO, P.; VICENTE-VILLARDÓN, J. L.; Los Métodos Biplot: evolución y aplicaciones. **Revista Venezolana de Análisis de Coyuntura**, Caracas, n.001, v. 13, p. 279–303, 2007.

GABRIEL, K. R.; The biplot graphic display of matrices with application to principal component analysis. **Biometrika**, Great Britain, vol.58, n.03, p.453–467, 1971.

GOWER, J. C.; HARDING S.; Nonlinear biplots **Biometrika**, Great Britain, n. 75, p. 445–455, 1988.

GOWER, J. C.; HAND, D. J.; **Biplots**. London: Chapman & Hall, 1996. 277 p.

GOWER, J. C.; KRZANOWSKI, W. J.; Analysis of distance for structured multivariate data and extensions to multivariate analysis of variance. **Applied Statistics**, London, n. 4, p. 505–519, 1999.

JOHNSON, R. A.; WICHERN, D. W.; **Applied Multivariate Statistical Analysis**. New Jersey: Prentice-Hall, Fifth Edition, 2002. 767 p.

KRZANOWSKI, W. J.; Multifactorial analysis of distance in studies of ecological community structure. **Journal of Agricultural, Biological and Environmental Statistics**, Alexandria, v.7, p. 222-232, 2002.

MANLY, B. J.; **Métodos Estatísticos Multivariados**. Porto Alegre: Bookman, 2008. 229 p.

TIMM, N. H.; **Applied multivariate analysis**. New York, Springer-Verlag, 2002. 718 p.

VIANA, S. S.; SILVA, L. M. R.; LIMA, P. V. P. S.; LEITE, L. A. S.; Competitividade do Ceará no mercado internacional de frutas: o caso do melão. **Revista Ciência Agronômica**, Fortaleza, v.37, n.1, p.25-31, 2006.

YAN, W.; KANG, M. S.; **GGE Biplot Analysis. A Graphical Tool for Breeders, Geneticists and Agronomists**. Florida: CRC Press, 2002. 271 p.