

APLICAÇÃO DE TÉCNICAS MULTIVARIADAS NA SELEÇÃO DE GENÓTIPOS DE PLANTAS PARA GRAMADOS RESISTENTES A INSETOS

Waldomiro Barioni Júnior¹, Marcos Rafael Gusmão², Francisco Humberto Dübbern de Souza², Frederico de Pina Matta², Sandra Elisa Fonrozo³

Introdução

O gênero *Paspalum* compreende espécies de alta variabilidade morfológica, com diferentes hábitos de crescimento e adaptadas às diferentes condições edafo-climáticas e bióticas. A vantagem das técnicas estatísticas multivariadas para análise exploratória é poder estudar as relações entre variáveis e semelhança entre indivíduos de forma gráfica, em dimensão reduzida, facilitando sua interpretação. Este trabalho teve o objetivo de avaliar, através de técnicas multivariadas, as variabilidades intra e interespecífica de *Paspalum* spp. quanto ao ataque por insetos (cigarrinhas-das-pastagens).

Material e Métodos

O experimento foi realizado em blocos casualizados, com 4 repetições e 35 tratamentos (genótipos de grama) no período de janeiro/2011 a julho/2013 na Embrapa Pecuária Sudeste, São Carlos, SP. Foram coletados os dados de sete variáveis de ataque de insetos: Dano0, Dano1-25, Dano26-50, Dano51-75, Dano76-100 (proporção da classe de dano), Adulto (nº de insetos adulto) e Espuma (nº de espuma do inseto). Considerando-se os dados médios, por tratamentos, realizou-se a análise dos dados através das técnicas de Análise de Agrupamentos (AC) e Análise de Componentes Principais (ACP). Na análise de agrupamento, visando encontrar k grupos de genótipos, usou-se o método Hierárquico de Cluster, que parte de uma matriz de similaridade $S_{n \times n}$, simétrica, que contém as medidas de proximidades entre os n indivíduos (genótipos), obtidas a partir da distância euclidiana [1] entre os pontos.

$$S_{n \times n} = \begin{bmatrix} 0 & & & & & \\ d(2,1) & 0 & & & & \\ d(3,1) & d(3,2) & 0 & & & \\ \vdots & \vdots & \vdots & \ddots & & \\ d(n,1) & d(n,2) & \dots & d(n,n-1) & 0 & \end{bmatrix}$$

¹Embrapa Pecuária Sudeste E-mail: waldomiro.barioni@embrapa.br

²Embrapa Pecuária Sudeste

³Estudante de Estatística – ICMC/USP

$$d(a, b) = \left[\sum_{j=1}^p (X_{aj} - X_{bj})^2 \right]^{1/2} \quad [1]$$

X_{aj} : valor da variável j para o indivíduo a;

X_{bj} : valor da variável j para o indivíduo b;

p: número de variáveis.

No método, calcula-se a primeira matriz de similaridade e então agrupam-se os indivíduos com menor distância observada. Em seguida, toma-se a média entre eles, e se repete o procedimento. A cada passo, forma-se um novo nível de agrupamento que podem ser representados através de um dendograma.

Após a formação dos k grupos, procedeu-se uma análise confirmatória, através de análise variância multivariada (Manova) seguida do teste F(n, k; $p \leq 0,01$), para testar a hipótese de nulidade (H_0 : homogeneidade dos grupos). Na rejeição de H_0 , aplicou-se o teste multivariado T^2 de Hotelling ($p \leq 0,01$) para verificar se existe diferença entre os vetores de médias (centróides), dois a dois (\bar{x} e \bar{y}), relativos aos k grupos formados. Sendo $n = n_x + n_y - 1$, o teste de Hotelling multivariado é realizado a partir da estatística F:

$$F = \frac{n - k}{k(n - 1)} T^2 \sim F(k, n - k)$$

$$S = \frac{(n_x - 1)S_x + (n_y - 1)S_y}{(n_x - 1) + (n_y - 1)}$$

$$T^2 = (\bar{x} - \bar{y})^T \left[S \left(\frac{1}{n_x} + \frac{1}{n_y} \right) \right]^{-1} (\bar{x} - \bar{y})$$

Realizou-se também a análise de componente principal (ACP) como complemento da análise de Agrupamento (AC) para conhecer e entender melhor a relação entre as variáveis e os grupos de genótipos criados. A ACP permitiu condensar a maior quantidade da informação original contida em p variáveis (p=7, neste estudo) em duas variáveis latentes ortogonais denominadas componentes principais, que são combinações lineares das variáveis originais criadas com os dois maiores autovalores da matriz de covariância dos dados (Hair, 2005). Na ACP, calcula-se a matriz **R** de covariâncias normalizada das observações e então se resolve a equação característica [2] dessa matriz, obtendo os autovalores λ_i .

$$\det[\mathbf{R} - \lambda \mathbf{I}] = 0 \quad [2]$$

Para cada autovalor λ_i calcula-se um autovetor \mathbf{a}_i .

$$\mathbf{a}_i = \begin{bmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{ip} \end{bmatrix}$$

O i -ésimo componente principal é dado por:

$$y_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p$$

Desta forma, o conjunto inicial de sete variáveis passou a ser caracterizado por duas novas variáveis latentes, o que possibilitou sua localização em um plano (ordenação dos genótipos por componentes principais). A adequação desta análise é verificada pela quantidade da informação total das variáveis originais retida pelos componentes principais que mostram autovalores superiores à unidade (Kaiser, 1958). Autovalores inferiores à unidade não dispõem de informação relevante. Todas as análises estatísticas foram processadas no software STATISTICA versão 7.0 (StatSoft, Tulsa, USA).

Resultado e Discussão

Análise de Cluster

A análise de agrupamento, pelo método hierárquico, possibilitou separar os 35 genótipos em três grupos. O ponto de corte na definição dos grupos foi na distância de ligação entre 10 e 15, na vertical. Este ponto separou os genótipos em três grupos como mostra o dendograma da Figura 1.

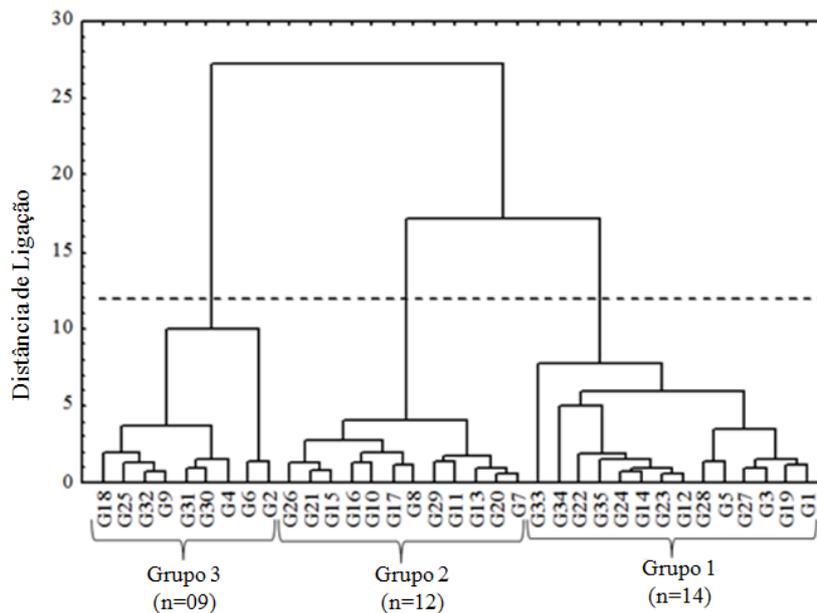


Figura 1. Dendrograma mostrando a estrutura de grupos para os 35 genótipos

Na tabela 1 são apresentados os três grupos, os quais diferiram entre si pelo teste T^2 de Hotelling ($p < 0,01$). O Grupo 1 é caracterizado por baixo dano e alta presença de inseto na fase adulta; Grupo 2

caracterizado por dano intermediário e baixa presença de insetos na fase adulta e de ninfa (*Espuma*) e o Grupo3 com alta incidência de dano e presença de insetos na fase de ninfa (*Espuma*).

Tabela1: Os 35 genótipos separados em três grupos distintos (I, II e III) pela análise de agrupamento

Grupos ¹	N	Genótipos selecionados	Características
I	14	G1, G19, G3, G27, G5, G28, G12, G23, G14, G24, G22, G34, G33 (<i>controle</i>) e G35	Baixo dano e presença de inseto Adulto
II	12	G26, G21, G15, G16, G10, G17, G8, G29, G11, G13, G7 e G20	Dano médio e baixa incidência de Adulto e Espuma
III	09	G2, G6, G4, G30, G31, G9, G25, G32 e G18	Alto dano e presença de Espuma

Análise de Componentes Principais

Os resultados da análise de componente principal (ACP), como complemento da análise de Agrupamento (AC) possibilitou conhecer e entender melhor a relação existente entre as variáveis e os três grupos de genótipos criados. Na figura2 encontram-se as projeções das 7 variáveis e dos 35 genótipos no plano bidimensional, formado pelo primeiro e segundo componentes principais (CP1 e CP2), com 69,33% de explicação da variância total.

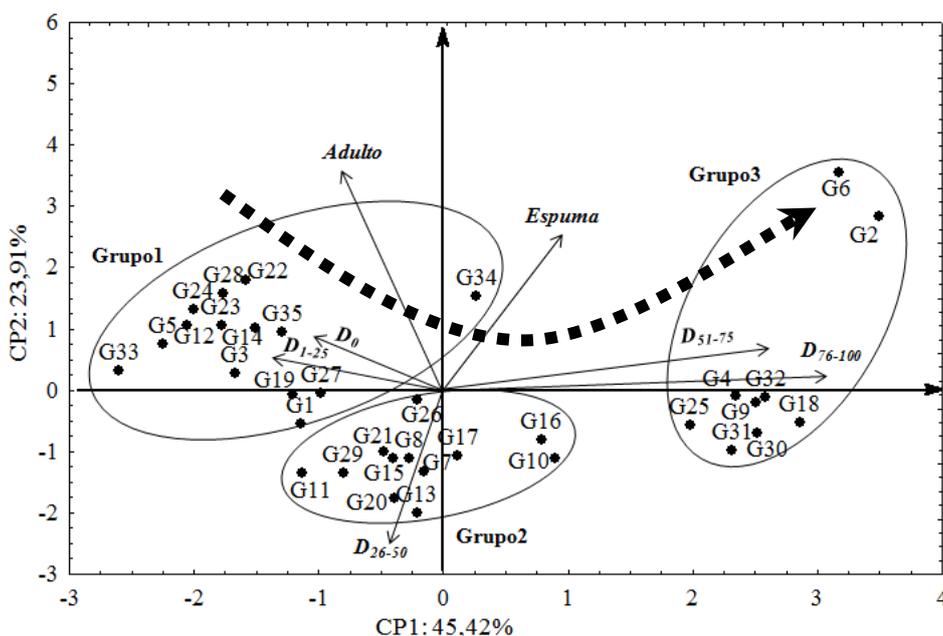


Figura 2. Projeção dos 35 genótipos, das sete variáveis de ataque de inseto no plano formado pelos dois primeiros componentes principais e a representação dos três grupos gerados na análise de agrupamento

Neste plano, também se encontram a localização e a região dos 3 grupos gerados na análise de agrupamento. Observa-se que o eixo horizontal (CP₁) representa a incidência de danos, cuja a intensidade aumenta da esquerda para a direita na direção da seta pontilhada.

Conclusão

Desta forma, conclui-se que as técnicas de análise multivariadas utilizadas permitiram caracterizar as variabilidades intra e interespecífica de genótipos de *Paspalum* spp., os quais foram separados em três grupos, em função do nível de ataque por cigarrinhas-das-pastagens.

Referências

1. HAIR, J. F.; ANDERSON, R. E.; TATHAM, R. L.; BLACK, W. Análise multivariada de dados. Porto Alegre, Bookman, 593 p., 2005.
2. KAISER H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, v.23, p.187-200, 1958.
3. VAL, J. E.; FERRAUDO, A. S.; BEZERRA, F.; CORRADO, M. P.; LÔBO, R. B.; FREITAS, M. A. R.; PANETO, J. C. C. Alternativas para seleção de touros da raça Nelore considerando características múltiplas de importância econômica. *Arq. Bras. Med. Vet. Zootec.*, v.60, p.705-712, 2008.
4. VARELLA, Carlos Alberto Alves. ANÁLISE MULTIVARIADA APLICADA AS CIÊNCIAS AGRÁRIAS. Disponível em: <<http://www.ufrj.br/institutos/it/deng/varella/Downloads-/multivariada%20aplicada%20as%20ciencias%20agrarias/Apresenta/analise%20de%20agrupamento.pptx>>. Acesso em: 30/03/2015.
5. VARELLA, Carlos Alberto Alves. ANÁLISE MULTIVARIADA APLICADA AS CIÊNCIAS AGRÁRIAS. Disponível em: <<http://www.ufrj.br/institutos/it/deng/varella/Downloads-/multivariada%20aplicada%20as%20ciencias%20agrarias/Aulas/analise%20de%20componentes%20principais.pdf>>. Acesso em: 30/03/2015.
6. ZAIONTZ, Charles. Hotelling's T-square Test for Two Independent Samples. Disponível em: <<http://www.real-statistics.com/multivariate-statistics/hotellings-t-square-statistic/hotellings-t-square-independent-samples/>>. Acesso em: 30/03/2015.