

ESTABILIDADE EM ANÁLISE DE AGRUPAMENTO: ESTUDO DE CASO EM CIÊNCIA FLORESTAL¹

Márcio Augusto de Albuquerque², Rinaldo Luiz Caraciolo Ferreira³, José Antônio Aleixo da Silva³, Eufrázio de Souza Santos⁴, Borko Stosic⁴ e Agostinho Lopes de Souza⁵

RESUMO – Objetivou-se, neste trabalho, propor uma sistemática para o estudo e interpretação da estabilidade dos métodos de análise de agrupamento, através de vários algoritmos de agrupamento em dados de vegetação. Utilizaram-se dados provenientes de levantamento na Mata da Silvicultura, da Universidade Federal de Viçosa, em Viçosa, MG. Para a análise de agrupamento, foram estimadas as matrizes de distância de Mahalanobis com base nos dados originais e via reamostragem “bootstrap”, bem como aplicados os métodos da ligação simples, ligação completa e médias das distâncias, do centróide, da mediana e do Ward. Para a detecção de associação entre os métodos, foi aplicado o teste Qui-Quadrado (χ^2) a 1 e 5% de probabilidade. Para os diversos métodos de agrupamento foi obtida a correlação cofenética. Os resultados de associação dos métodos foram semelhantes, indicando, em princípio, que qualquer algoritmo de agrupamento estudado está estabilizado e existem, de fato, grupos entre os indivíduos observados. No entanto, verificou-se que os métodos são coincidentes, exceto os métodos do centróide e Ward e os métodos do centróide e mediana, em comparação com o de Ward, respectivamente, com base nas matrizes de Mahalanobis a partir dos dados originais e “bootstrap”. A sistemática proposta é promissora para o estudo e interpretação da estabilidade dos métodos de análise de agrupamento em dados de vegetação.

Palavras-chave: Análise multivariada, “bootstrap” e métodos hierárquicos aglomerativos.

STABILITY IN CLUSTER ANALYSIS: STUDY OF CASE IN FOREST SCIENCE

ABSTRACT – The main objective of this research was to propose a system to the study and interpretation of stability in cluster analysis through several cluster algorithms in vegetation data. The data set used derived from a survey in the Silviculture Forest at Federal University of Viçosa – MG. To perform the cluster analysis, the Mahalanobis distance matrices were estimated on basis of original data and bootstrap resampling. Also, the single linkage, complete linkage, average distances, centroid, median and Ward methods were used. Chi-square test was applied to detect the association among the methods. A co-phenetic correlation was obtained for the cluster methods. The results for the method associations were very similar, indicating that any algorithm of the studied clusters is stabilized, and in fact, that groups exist among the analyzed individuals. However, it was verified that the methods are coincident, except for the centroid and Ward, and also the centroid and median methods, when compared to Ward, respectively, based on the Mahalanobis matrices derived from the original data set and bootstrap. The methodology proposed is promising to the study and interpretation of the stability of cluster analysis methods in vegetation data.

Keywords: Multivariate analysis, bootstrap and agglomerative hierarchical methods.

¹ Recebido em 29.03.2005 e aceito para publicação em 10.11.2005.

² Departamento de Matemática, Estatística e Informática da Universidade Estadual da Paraíba. Rua Juvêncio Arruda, s/n, Bodocongó, 58109-790 Campina Grande-PB. E-mail: <macioaa@uol.com.br>.

³ Departamento de Ciência Florestal da UFRPE, Rua Manoel de Medeiros, s/n, Dois Irmãos, 52171-900 Recife-PE. E-mail: <rinaldof@ufrpe.br>; <jaaleixo@uol.com.br>.

⁴ Departamento de Estatística e Informática da UFRPE. E-mail: <eufrazio@ufrpe.br>; <borko@ufrpe.br>.

⁵ Departamento de Engenharia Florestal da Universidade Federal de Viçosa. E-mail: <alsouza@ufv.br>.

1. INTRODUÇÃO

A análise de agrupamento é uma técnica multivariada amplamente utilizada para diversos fins em ciência florestal (FONSECA e RODRIGUES, 2000; SCUDELLER et al., 2001; MARTINS et al., 2003; SOUZA et al., 2003; FERRAZ et al., 2004; SANTOS et al., 2004; ARAÚJO et al., 2004).

A análise de agrupamento tem por finalidade reunir, por algum critério de classificação, as unidades amostrais em grupos, de tal forma que exista homogeneidade dentro do grupo e heterogeneidade entre grupos (JOHNSON e WICHERN, 1992; CRUZ e REGAZZI, 1994).

Vários são os tipos de técnicas de agrupamento encontradas na literatura (MARDIA et al., 1997), dos quais o pesquisador tem de decidir qual o mais adequado ao seu propósito, uma vez que as diferentes técnicas podem levar a diferentes soluções.

As técnicas de agrupamento podem ser classificadas em hierárquicas e não-hierárquicas (CORMACK, 1971). A técnica hierárquica consiste em uma série de sucessivos agrupamentos ou sucessivas divisões de elementos, em que os elementos são agregados ou desagregados. A técnica não-hierárquica foi desenvolvida para agrupar elementos em K grupos, em que K é a quantidade de grupos definida previamente.

As técnicas hierárquicas são as mais amplamente difundidas (SIEGMUND et al., 2004) e envolvem basicamente duas etapas. A primeira se refere à estimação de uma medida de similaridade ou dissimilaridade entre os indivíduos e a segunda, à adoção de uma técnica de formação de grupos (SANTANA e MALINOVSKI, 2002).

Um grande número de medidas de similaridade ou de dissimilaridade tem sido proposto e utilizado em análise de agrupamento, sendo a escolha entre elas baseada na preferência e, ou, na conveniência do pesquisador (BUSSAB et al., 1990).

Com a definição da medida de dissimilaridade a ser utilizada, a etapa seguinte é a adoção de uma técnica de agrupamento para formação dos grupos. Para realização dessa tarefa, existe um grande número de métodos disponíveis, dos quais o pesquisador tem de decidir qual o mais adequado ao seu propósito, uma vez que as diferentes técnicas podem levar a diferentes soluções (SOUZA et al., 1997).

As técnicas de análise de agrupamento exigem de seus usuários a tomada de uma série de decisões independentes, que requerem o conhecimento das propriedades dos diversos algoritmos à disposição e que podem representar diferentes agrupamentos. Além disso, o resultado dos agrupamentos pode ser influenciado pela escolha da medida de dissimilaridade, bem como pela definição do número de grupos (GOWER e LEGENDRE, 1986; JACKSON et al., 1989; DUARTE et al., 1999).

Recentes avanços da ciência da computação permitiram o desenvolvimento de sistemas interativos de processamentos de dados, com algoritmos rápidos e precisos. Com isso, muitos pesquisadores estão desenvolvendo metodologias estatísticas, com a finalidade de estudar e avaliar a estabilidade dos agrupamentos obtidos a partir de matrizes de dissimilaridade. Entre essas metodologias, destaca-se o procedimento de reamostragem “bootstrap”, que pode fornecer um ponto de equilíbrio que permite uma estimativa precisa dos grupos (LAVORANTI, 2003).

Assim, objetivou-se propor uma sistemática para o estudo e a interpretação da estabilidade dos métodos em análise de agrupamento, através de vários algoritmos de agrupamento em dados de vegetação.

2. MATERIAL E MÉTODOS

Foram utilizados dados de um levantamento da vegetação da Mata da Silvicultura (Quadro 1), da Universidade Federal de Viçosa, em Viçosa, MG, retirado de Souza et al. (1997).

Para realizar a análise de agrupamento foi utilizada como medida de dissimilaridade à distância de Mahalanobis (D^2), calculada conforme a seguinte expressão:

$$D^2 = (\underline{X}_i - \underline{X}_j)' \cdot \Sigma^{-1} (\underline{X}_i - \underline{X}_j)$$

em que Σ^{-1} é a inversa da matriz de co-variância residual de X, \underline{X}_i é o vetor referente à parcela i, \underline{X}_j é o vetor referente à parcela j, $(\underline{X}_i - \underline{X}_j)'$ é o vetor transposto da diferença entre \underline{X}_i e \underline{X}_j e D^2 tem a característica de ser invariante para qualquer transformação linear não-singular.

Para se estabilizarem os métodos em análise de agrupamentos via “bootstrap”, foram seguidos os seguintes passos:

1. Considerou-se a seguinte matriz X, denominada matriz de dados ou matriz original (primária).

$$X_{(17 \times 11)} = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{111} \\ X_{21} & X_{22} & \dots & X_{211} \\ \vdots & \vdots & \ddots & \vdots \\ X_{171} & X_{172} & \dots & X_{1711} \end{bmatrix}$$

2. Com a matriz primária, encontrou-se a matriz de distância de Mahalanobis, para aplicação dos algoritmos de agrupamento.

3. De posse da matriz de Mahalanobis, aplicou-se "bootstrap" e calculou-se uma nova matriz de distância Mahalanobis, para aplicação dos algoritmos de agrupamento e comparação com a aplicação do item 2.

4. Construção de quadros de contingência 2x2 para comparação entre algoritmos de agrupamentos (número de observações que se agrupam no mesmo grupo para o mesmo número de grupo).

$$\text{Porcentagem} = \frac{\text{Número de observações}}{\text{Número de grupos}} \cdot 100\%$$

5. Cálculo do indicador do grau de associação entre dois métodos, obtido pela expressão:

$$C = \sqrt{\frac{\chi^2_{\text{calculado}}}{\chi^2_{\text{calculado}} + N}}$$

Esse coeficiente pode variar entre [0,1], estando mais associados os métodos quanto maior foi o valor de C.

Os algoritmos de agrupamento utilizados foram Método da Ligação Simples, Método da Ligação Completa, Método da Centróide, Método da Mediana, Método das Médias das Distâncias e Método de Ward, conforme descrito por Johnson e Wichern (1992) e Mardia et al. (1997). Esses métodos foram utilizados por serem os mais usados na prática e pela facilidade de serem encontrados nos mais diversos programas computacionais.

A seqüência de fusão dos agrupamentos, conforme o método utilizado, foi representada graficamente por dendrogramas, os quais foram divididos com a estatística descritiva usando o percentil, com um corte de 50% da distância de Mahalanobis máxima de fusão, para determinar o número de grupos. Os diferentes dendrogramas obtidos foram, então, comparados para possibilitar a análise da associação entre métodos.

Quadro 1 – Densidade de 17 espécies da Mata da Silvicultura, em parcelas de 20 X 50 m, Universidade Federal de Viçosa, Viçosa, MG, 1993

Table 1 – Density of 17 species from the Silviculture Forest, in 20 X 50 m sample units at Federal University of Viçosa, Viçosa, MG, 1993

Espécies	Parcelas										
	1	2	3	4	5	6	7	8	9	10	11
<i>Casearia decandra</i> Jacq.	8	1	27	0	1	9	2	3	22	15	7
<i>Anadenanthera peregrina</i> Spreng.	0	0	0	0	0	0	12	1	17	1	9
<i>Apuleia leiocarpa</i> (Vog.) Macbr.	3	9	4	6	22	9	5	2	7	4	4
<i>Mabea fistulifera</i> Mart.	6	3	3	4	29	12	0	4	4	4	4
<i>Anadenanthera macrocarpa</i> (Benth.) Brenan.	0	12	0	1	0	0	1	0	2	0	0
<i>Platypodium elegans</i> Vog.	0	0	1	1	9	1	0	0	5	11	1
<i>Machaerium floridum</i> (Benth.) Ducke	0	0	10	1	9	2	1	0	0	11	5
<i>Copaifera langsdorffii</i> Desf.	1	1	0	2	1	13	0	0	0	3	1
<i>Ocotea pretiosa</i> Mez.	2	0	2	2	2	6	0	5	0	2	2
<i>Cabrlea cangerana</i> Saldanha	1	0	0	2	0	0	1	6	2	3	1
<i>Piptadenia gonoacantha</i> Macbr.	0	0	0	0	0	0	6	0	1	0	5
<i>Dalbergia nigra</i> Allem. ex Benth.	5	0	7	0	5	0	0	0	0	1	0
<i>Luehea divaricata</i> Mart.	0	0	1	0	0	0	2	0	0	5	2
<i>Cecropia hololeuca</i> Miq.	7	0	0	0	0	1	0	1	0	0	0
<i>Melanoxylon brauna</i> Schott.	0	0	0	0	0	0	0	0	0	2	1
<i>Cedrela fissilis</i> Vell.	0	0	0	0	0	0	1	0	0	0	0
<i>Croton flribundus</i> Spreng.	0	0	1	0	0	0	0	0	0	0	0

Fonte: Souza et al. (1997).



Com base nos dendrogramas foram construídas tabelas de contingência bi-dimensionais, nas quais uma amostra de N observações foi classificada com relação a dois métodos de agrupamento aplicados. Dessa forma, foi possível cruzar as diversas características relevantes aos métodos pesquisados com diversas variáveis, tomadas duas a duas.

Para a detecção de associação entre os métodos, ou seja, saber se as diferenças observadas entre métodos são significativas o suficiente para serem atribuídas a outros fatores que não aleatórios, foi aplicado o teste qui-quadrado (χ^2) a 1 e 5% de probabilidade.

Para os diversos métodos de agrupamento utilizados foram obtidas as respectivas matrizes cofenéticas resultantes da simplificação proporcionada pelo método. Com base nas matrizes de dissimilaridade original e cofenética, foi obtida a correlação cofenética, conforme a expressão (BUSSAB et al., 1990):

$$r_{\text{cof}} = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n (c_{ij} - \bar{c})(d_{ij} - \bar{d})}{\sqrt{\sum_{i=1}^{n-1} \sum_{j=i+1}^n (c_{ij} - \bar{c})^2} \sqrt{\sum_{i=1}^{n-1} \sum_{j=i+1}^n (d_{ij} - \bar{d})^2}}$$

em que: c_{ij} = valor de dissimilaridade entre os indivíduos i e j, obtidos a partir da matriz cofenética; e d_{ij} = valor de dissimilaridade entre os indivíduos i e j, obtidos a partir da matriz de dissimilaridade.

Todos os gráficos e as análises ao longo deste trabalho foram implementados através dos programas

computacionais EXCEL, STATISTICA, MINITAB e pela construção de um programa na linguagem C.

3. RESULTADOS E DISCUSSÃO

Com base na matriz de dissimilaridade de Mahalanobis obtida a partir dos dados originais e via “bootstrap” (Quadro 2), foram aplicados os métodos da ligação simples, da ligação completa, do centróide, da mediana, da média das distâncias e de Ward e obtidos os respectivos dendrogramas (Figuras de 1 a 6).

Embora a estrutura geral dos agrupamentos seja bastante similar, pode-se observar que há pequenas alterações nos níveis em que os indivíduos são agrupados, ou seja, os indivíduos que estão dentro de um mesmo grupo podem ser agrupados em outra ordem, quando se mudam os métodos.

De forma geral, os dendrogramas apresentaram estruturas de agrupamentos de objetos homogêneos, embora não exista critério objetivo para determinar um ponto de corte no dendrograma, ou seja, para determinar quais os grupos foram formados.

Observa-se que nos métodos de ligação simples (Figura 1A) e Ward (Figura 6A), ligação completa (Figura 2A), mediana (Figura 4A) e médias das distâncias (Figura 5A) e centróide (Figura 3A) foram obtidos 9, 10 e 11 grupos, respectivamente. As Figuras 3B e 4B apresentam 11 grupos e as Figuras 1B, 2B e 5B, 10 grupos, enquanto a Figura 6B, a formação de nove grupos.

Quadro 2 – Matrizes de distância de Mahalanobis dos dados originais e via “bootstrap” para as 11 parcelas da Mata da Silvicultura, Universidade Federal de Viçosa, Viçosa, MG

Table 2 – Matrices of Mahalanobis distance from the original data set and bootstrap according to the 11 sample units of the Silviculture Forest at the Federal University of Viçosa, Viçosa, MG

Parcela	2	3	4	5	6	7	8	9	10	11
1	16,81	20,01	15,93	20,35	14,40	19,50	13,37	20,11	18,53	17,66
2	20,01	14,40	20,11	16,81	17,66	20,35	17,66	19,50	16,81	20,11
3		18,50	7,721	18,51	13,81	18,80	18,59	15,59	20,89	19,03
4		18,51	18,50	18,80	15,59	18,59	15,59	19,03	20,89	18,51
5			17,68	20,80	18,21	22,27	20,08	18,48	20,55	19,10
6			20,80	18,21	20,80	19,10	19,10	20,08	22,27	19,10
7				12,57	10,32	16,73	7,092	14,54	15,04	13,84
8					14,54	10,32	16,73	16,73	13,84	7,09
9					17,65	22,86	20,99	19,20	20,37	17,55
10					17,55	20,99	19,20	19,20	20,99	17,65
11					19,90	16,21	16,89	18,27	18,27	13,12
					16,21	19,90	19,90	16,21	16,21	16,21
						20,02	17,36	22,69	22,69	16,47
						22,69	22,69	16,47	16,47	16,47
							15,11	18,29	18,29	14,05
							14,05	14,05	14,05	18,29
								14,87	14,87	13,20
								13,20	13,20	14,87
									11,90	11,90
										11,90

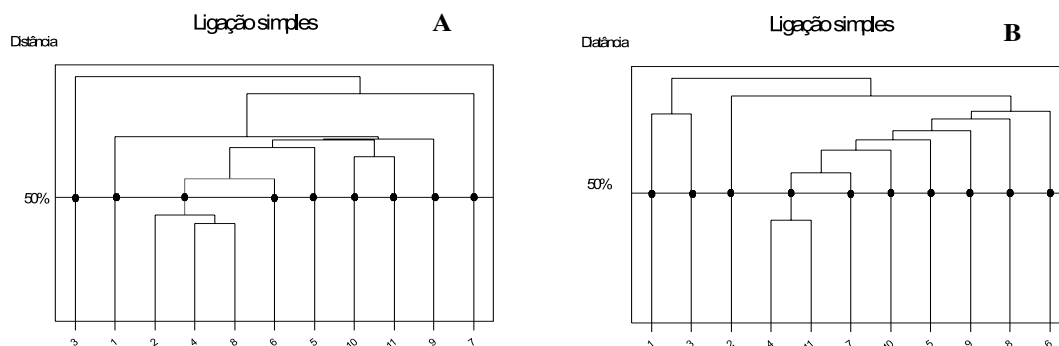


Figura 1 – Dendrogramas representando as seqüências das fusões das parcelas, obtidos pelo emprego do método da ligação simples, com base nas distâncias de Mahalanobis, a partir dos dados originais (A) e “bootstrap”(B).

Figure 1 – Dendrograms representing the sequences of sample unit fusions, obtained through the single linkage method based on the Mahalanobis distance of the original data set (A) and by bootstrap (B).

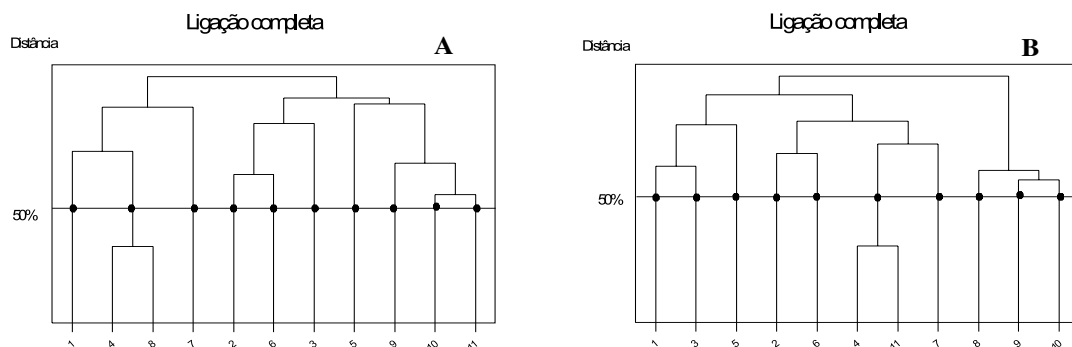


Figura 2 – Dendrogramas representando as seqüências das fusões das parcelas, obtidos pelo emprego do método da ligação completa, com base nas distâncias de Mahalanobis, a partir dos dados originais (A) e “bootstrap”(B).

Figure 2 – Dendrograms representing the sequences of sample unit fusions, obtained through the complete linkage method based on the Mahalanobis distance of the original data set (A) and bootstrap (B).

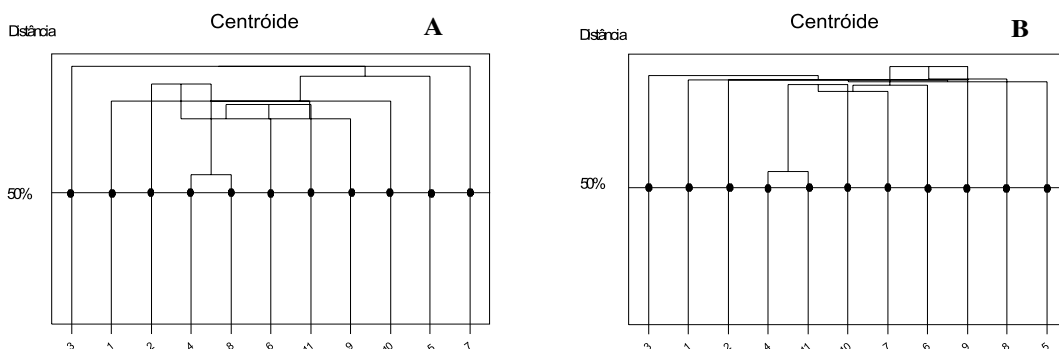


Figura 3 – Dendrogramas representando as seqüências das fusões das parcelas, obtidos pelo emprego do método da centróide, com base nas distâncias de Mahalanobis a partir dos dados originais (A) e “bootstrap”(B).

Figure 3 – Dendrograms representing the sequences of sample unit fusions, obtained through the centroid method based on the Mahalanobis distance of the original data set (A) and bootstrap (B).

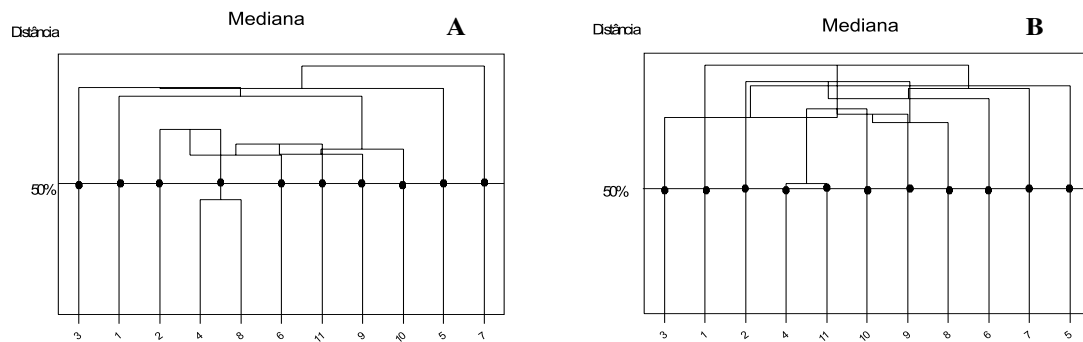


Figura 4 – Dendrogramas representando as seqüências das fusões das parcelas, obtidos pelo emprego do método da mediana, com base nas distâncias de Mahalanobis, a partir dos dados originais (A) e “bootstrap”(B).

Figure 4 – Dendrograms representing the sequences of sample unit fusions, obtained through the median method based on the Mahalanobis distance of the original data set (A) and bootstrap (B).

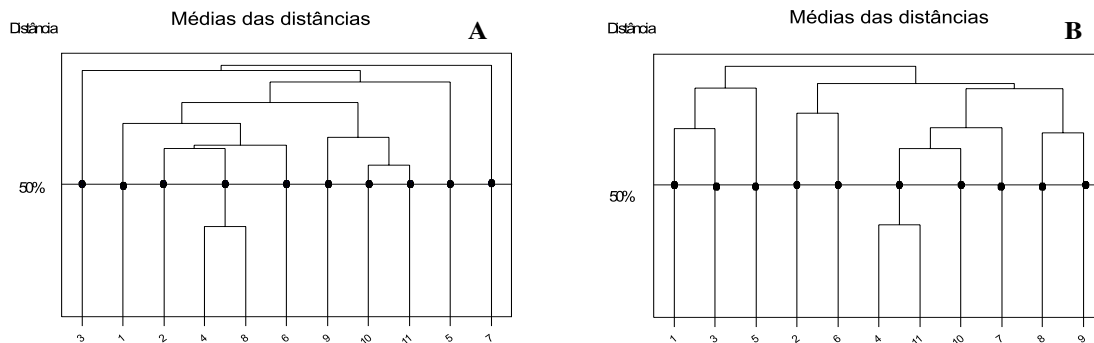


Figura 5 – Dendrogramas representando as seqüências das fusões das parcelas, obtidos pelo emprego do método das médias da distância, com base nas distâncias de Mahalanobis, a partir dos dados originais (A) e “bootstrap”(B).

Figure 5 – Dendrograms representing the sequences of sample unit fusions, obtained through the average distances method based on the Mahalanobis distance of the original data set (A) and bootstrap (B).

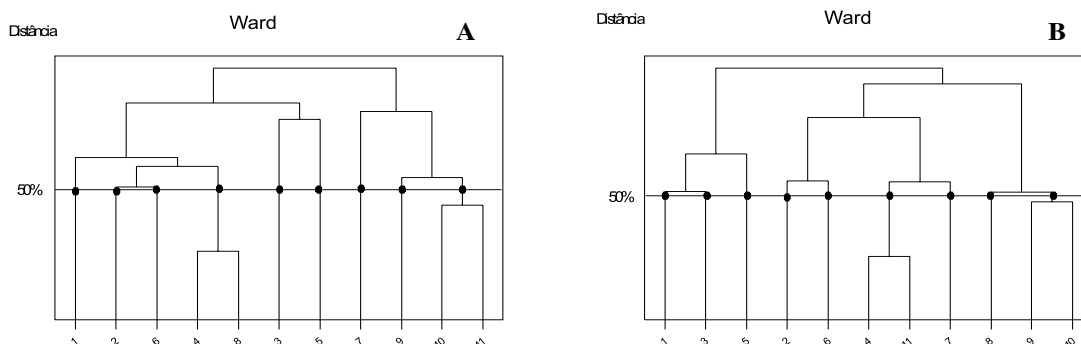


Figura 6 – Dendrogramas representando as seqüências das fusões das parcelas, obtidos pelo emprego do método do Ward, com base nas distâncias de Mahalanobis, a partir dos dados originais (A) e “bootstrap”(B).

Figure 6 – Dendrograms representing the sequences of sample unit fusions, obtained through the Ward method based on the Mahalanobis distance of the original data set (A) and by bootstrap (B).

Na análise com a distância baseada nos dados originais, verificou-se que os métodos médias das distâncias e mediana, médias das distâncias e ligação completa, mediana e ligação completa formaram 10 agrupamentos, apresentando semelhança de 100%. Os métodos médias das distâncias e centróide, centróide e mediana, centróide e ligação completa, médias das distâncias e ligação simples, médias das distâncias e Ward, mediana e ligação simples, mediana e Ward, ligação simples e ligação completa, ligação completa e Ward, apresentaram 86 e 84% de semelhança. E os métodos centróide e ligação simples, centróide e Ward, ligação simples e Ward apresentaram 80, 70 e 67% de semelhança, respectivamente (Quadro 3).

Para análise com a distância baseada nos dados de reamostragem “bootstrap”, verificou-se que os métodos médias das distâncias e ligação simples, médias das distâncias e ligação completa, com 10 grupos, centróide e mediana com 11 grupos, tiveram 100% de semelhança. Os métodos médias das distâncias e centróides, médias das distâncias e mediana, centróide e ligação simples, centróide e ligação completa, mediana

e ligação simples, mediana e ligação completa tiveram 86% de semelhança e os métodos médias das distâncias e Ward, ligação simples e Ward, ligação completa e Ward, 84% de semelhança. Já os métodos ligação simples e ligação completa e centróide e Ward, mediana e Ward atingiram 80 e 70% de semelhança, respectivamente (Quadro 3). Esses resultados apontaram que existe boa estabilidade entre os métodos.

Observou-se, ainda, que os resultados de associação dos métodos foram semelhantes e o nível de significância, relativamente alto, sendo possível concluir que, em princípio, qualquer algoritmo de agrupamento estudado está estabilizado e existem, de fato, grupos entre os indivíduos verificados. No entanto, com o teste qui-quadrado (Quadro 4) para níveis de significância de 1 e 5%, com graus de liberdade de 3,84 e 6,64, respectivamente, conclui-se que os métodos são coincidentes, exceto os métodos do centróide e Ward e os métodos do centróide e mediana, quando comparados com o de Ward, com base nas matrizes de Mahalanobis a partir dos dados originais e “bootstrap”.

Quadro 3 – Porcentagem de grupos coincidentes entre métodos de agrupamento com base nas matrizes de Mahalanobis, a partir dos dados originais e “bootstrap”

Table 3 – Percentage of similar groups among the cluster methods based on the Mahalanobis matrix of the original data set and bootstrap

Métodos	Médias das distâncias		Centróide		Mediana		Ligação simples		Ligação completa	
	Original	Bootstrap	Original	Bootstrap	Original	Bootstrap	Original	Bootstrap	Original	Bootstrap
Centróide	86%	86%								
	(0,58)	(0,58)								
Mediana	100%	86%	86%	100%						
	(0,71)	(0,58)	(0,58)	(0,71)						
Ligação Simples	84%	100%	80%	86%	84%	86%				
	(0,57)	(0,71)	(0,52)	(0,58)	(0,59)	(0,58)				
Ligação Completa	100%	100%	86%	86%	100%	86%	84%	80%		
	(0,71)	(0,71)	(0,58)	(0,58)	(0,71)	(0,58)	(0,57)	(0,51)		
Ward	84%	84%	70%	70%	84%	70%	67%	67%	84%	84%
	(0,57)	(0,57)	(0,38)	(0,38)	(0,57)	(0,38)	(0,61)	(0,57)	(0,57)	(0,57)

Quadro 4 – Valores estimados de χ^2 para estudo da associação dos métodos de agrupamentos com base nas matrizes de Mahalanobis a partir dos dados originais e “bootstrap”

Table 4 – χ^2 values estimated by the study of cluster method associations based on the Mahalanobis matrices derived from the original data and bootstrap

Métodos	Médias das distâncias		Centróide		Mediana		Ligação simples		Ligação completa	
	Original	Bootstrap	Original	Bootstrap	Original	Bootstrap	Original	Bootstrap	Original	Bootstrap
Centróide	10,84	10,82								
Mediana	20,00	10,82	10,82	22,00						
Ligação simples	8,99	20,00	7,60	10,82	10,42	10,82				
Ligação Completa	20,00	20,00	10,82	10,82	20,00	10,82	9,00	7,2		
Ward	9,00	9,00	3,43	3,48	9,00	3,46	11,00	9,00	9,00	9,00

Os valores das correlações cofenéticas (Quadro 5) foram todos de magnitude elevada, nos dados originais e “bootstrap”, indicando que há boa representação das matrizes de dissimilaridade na forma de dendrogramas e que isso independe do método usado.

Quadro 5 – Correlações cofenética entre as matrizes cofenéticas e a de dissimilaridade obtidas conforme o método de agrupamento utilizado

Table 5 – *Co-phenetic correlations between the co-phenetic matrices and dissimilarity matrices according to the cluster method used*

Métodos de Agrupamento	Matriz	
	Original	Bootstrap
Ligação simples	0,99	0,99
Ligação completa	0,98	0,99
Centróide	0,99	0,99
Mediana	0,99	0,99
Média das distâncias	0,99	0,99
Ward	0,99	0,99

4. CONCLUSÃO

A sistemática proposta é promissora para o estudo e interpretação da estabilidade dos métodos de análise de agrupamento em dados de vegetação.

5. AGRADECIMENTOS

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela concessão de bolsa de produtividade aos professores Agostinho Lopes de Souza, José Antônio Aleixo da Silva e Rinaldo Luiz Caraciolo Ferreira.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- ARAÚJO, M.M. et al. Análise de agrupamento da vegetação de um fragmento de floresta estacional decidual aluvial, Cachoeira do Sul, RS, Brasil. **Ciência Florestal**, v. 14, n. 1, p. 133-147, 2004.
- BUSSAB, W. O.; MIAZAKI, E. S.; ANDRADE, D. **Introdução à análise de agrupamentos**. São Paulo: Associação Brasileira de Estatística, 1990. 105p.
- CORMARCK, R. A review of classification. **Journal of the Royal Statistical Society (Series A)**, v.134, p.321-367, 1971.
- CRUZ, C. D.; REGAZZI, A. J. Divergência genética. In: CRUZ, C. D.; REGAZZI, A. J. **Métodos biométricos aplicados ao melhoramento genético**. Viçosa, MG: Universidade Federal de Viçosa. 1994. p. 287-323.
- DUARTE, M. C.; SANTOS, J. B.; MELO, L. C. Comparison of similarity coefficients based on RAPD markers in the common bean. **Genetics and Molecular Biology**, v.22, n.3, p.427-432, 1999.
- FERRAZ, E.M.N.; ARAÚJO, E.L.; SILVA, S.I. Floristic similarities between lowland and montane areas of Atlantic Coastal Forest in Northeastern Brazil. **Plant Ecology**, v.174, n.1, p.59-70, 2004.
- FONSECA, R.C.B.; RODRIGUES, R.R. Análise estrutural e aspectos do mosaico sucessional de uma floresta semidecídua em Botucatu, SP. **Scientia Forestalis**, n. 57, p. 27-43, 2000.
- GOWER, J. C.; LEGENDRE, P. Metric and euclidean properties of dissimilarity coefficients, **Journal of Classification**, v. 3, p. 5-48, 1986.
- JACKSON, A. A.; SOMERS, K. M.; HARVERY, H. H. Similarity coefficients: measures for co-occurrence and association or simply measures of occurrence? **American Naturalist**, v.133, p. 436-453, 1989.
- JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 3. ed. New Jersey: Prantice Hall, 1992. 642p.
- LAVORANTI, O. J. **Estabilidade e adaptabilidade fenotípica através da reamostragem “Bootstrap” no modelo AMMI**. 2003. 166f. Tese (Doutorado em Estatística e Experimentação Agronômica) – Escola Superior de Agricultura “Luiz de Queiroz”, Piracicaba, 2003.
- MARDIA, A.K.V.; KENT, J.T.; BIBBY, J.M. **Multivariate analysis**. London: Academic Press, 1997. 518p.
- MARTINS, S.V. et al. Distribuição de espécies arbóreas em um gradiente topográfico de Floresta Estacional Semidecidual em Viçosa-MG. **Scientia Forestalis**, n. 64, p. 172-181, 2003.
- SANTANA, C. M.; MALINOVSKI, J. R. Uso da análise multivariada no estudo de fatores humanos em operadores de motosserra, **Cerne**, v. 8, n. 2, p. 101-107, 2002.

SANTOS, J.H.S. et al. Distinção de grupos ecológicos de espécies florestais por meio de técnicas multivariadas. **Revista Árvore**, v.28, n.3, p.387-396, 2004.

SCUDELLER, V.V.; MARTINS, F.R.; SHEPHERD, G.J. Distribution and abundance of arboreal species in the atlantic ombrophilous dense forest in Southeastern Brazil. **Plant Ecology**, v. 152, n.2, p.185-199, 2001.

SIEGMUND, K.D.; LAIRD, P.W.; LAIRD-OFFRINGA, I.A. A comparison of cluster analysis methods using DNA methylation data. **Bioinformatics**, v. 20, n. 12, p.1896-1904, 2004.

SOUZA, A. L.; FERREIRA, R. L. C.; XAVIER, A. **Análise de agrupamento aplicada à ciência florestal**, Viçosa, MG: SIF, 1997. 109p., (Documento SIF, 16).