



VALIDAÇÃO DOS MÉTODOS DE AGRUPAMENTO HIERÁRQUICO PARA AS INFORMAÇÕES MUNICIPAIS DE PASTAGEM NO ESPÍRITO SANTO

Allison Queiroz de Oliveira¹, Maristela de Oliveira Bauer², Juliana Di Giorgio Giannotti³

Universidade Federal do Espírito Santo/Departamento de Engenharia Rural, Alto Universitário, s/nº -
Guararema, Alegre - ES | CEP 29500-000

¹allison_queiroz@hotmail.com, ²marisufes@gmail.com, ³jggiannotti@gmail.com

Resumo - O objetivo deste trabalho foi avaliar, de acordo com o coeficiente de correlação cofenética (CCC) qual método de agrupamento hierárquico produziu um resultado mais consistente para reunir os municípios do estado do Espírito Santo de acordo com as informações de pastagem (áreas com pastagens naturais, degradadas e com forrageiras) presentes no censo agropecuário do IBGE. Os dados foram padronizados e submetidos à análise de agrupamento usando cinco métodos hierárquicos (Ward, ligação completa, ligação média, ligação simples, centróide). O CCC foi significativo para todos os métodos sendo que o valor mais próximo de um foi para o método da ligação simples. Recomenda-se o CCC para avaliar a consistência do método de agrupamento.

Palavras-chave: áreas com pasto, coeficiente de correlação cofenética, métodos hierárquicos de agrupamento.

Introdução

A análise multivariada pode ser definida como um conjunto de métodos estatísticos que analisam mais de uma medida, simultaneamente, sobre o objeto em estudo (HAIR et al., 2005). Dentro da análise multivariada, a análise de agrupamento é um método cuja finalidade é agregar objetos com base nas similaridades das características/variáveis que eles possuem (JOHNSON & WICHERN, 1998). Há diferentes maneiras de se agregar estes objetos, seja utilizando métodos hierárquicos ou não hierárquicos e um problema recorrente, nos dois métodos, é definir o número adequado de grupos. Uma medida de validação do método de agrupamento é o coeficiente de correlação cofenética (CCC), o qual compara as distâncias observadas e previstas entre os objetos (BARROSO & ARTES, 2003).

Neste trabalho o objetivo foi avaliar, mediante a aplicação do CCC, qual o método hierárquico de agrupamento produziu um resultado mais consistente com relação aos grupos de municípios agregados, considerando as variáveis de pastagem, em nível municipal, para o estado do Espírito Santo.



Metodologia

A base de dados foi formada pelas variáveis áreas (em hectares) dos municípios do Estado do Espírito Santo com: pastagens naturais; pastagens plantadas degradadas; pastagens plantadas em boas condições e; área plantada com forrageiras para corte. Essas informações foram obtidas do censo agropecuário 2006 (IBGE, 2017).

Iniciou-se a análise de agrupamento com a padronização das variáveis, subtração da média e divisão pelo desvio padrão para cada uma delas (HAIR et al., 2005). A matriz de distância euclidiana obtida foi usada como medida de similaridade para a análise de agrupamento dos municípios cujos métodos hierárquicos empregados foram: Ward; ligação completa; ligação média; ligação simples; centróide. Assim, para cada método foi gerado um dendrograma.

A fim de avaliar qual método apresentou melhor consistência no agrupamento calculou-se o CCC. Este coeficiente mede o quanto a matriz das distâncias euclidianas e a matriz resultante do método de agrupamento estão ajustadas. O melhor ajuste se dá para um maior valor do CCC, deste modo, valores próximos da unidade indicam a melhor adequação do método (BARROSO & ARTES, 2003). Comparou-se os resultados dos CCC dentro de cada método de agrupamento por meio do teste t, a 5% de probabilidade. Todas as análises estatísticas dos dados foram realizadas no programa computacional de acesso livre R (R DEVELOPMENT CORE TEAM, 2015).

Resultados e Discussão

O CCC foi significativo para todos os métodos (p-valor <0,05) e seu valor foi: 0,6949 para Ward; 0,9574 para ligação simples; 0,9314 para ligação completa; 0,9490 para ligação média; 0,9498 para centróide.

O valor mais alto do CCC foi para o método da ligação simples, assim pode-se recomendar este método para realizar a análise de agrupamento. Porém os métodos da ligação média, da ligação completa e do centróide apresentaram bons resultados quanto à magnitude do valor do CCC, todos acima de 0,70 e próximos de um.

Na Figura 1 estão os dendrogramas gerados para cada um dos métodos de agrupamento. Por este resultado pode-se observar que os dendrogramas gerados pelos métodos da ligação simples e da ligação média tem semelhanças, assim como os



28ª SEAGRO

dendrogramas dos métodos da ligação completa e do centróide. O método de Ward foi o que apresentou o pior resultado quanto à consistência do agrupamento, e seu dendrograma mostrou apresentou-se diferenças com relação aos demais.

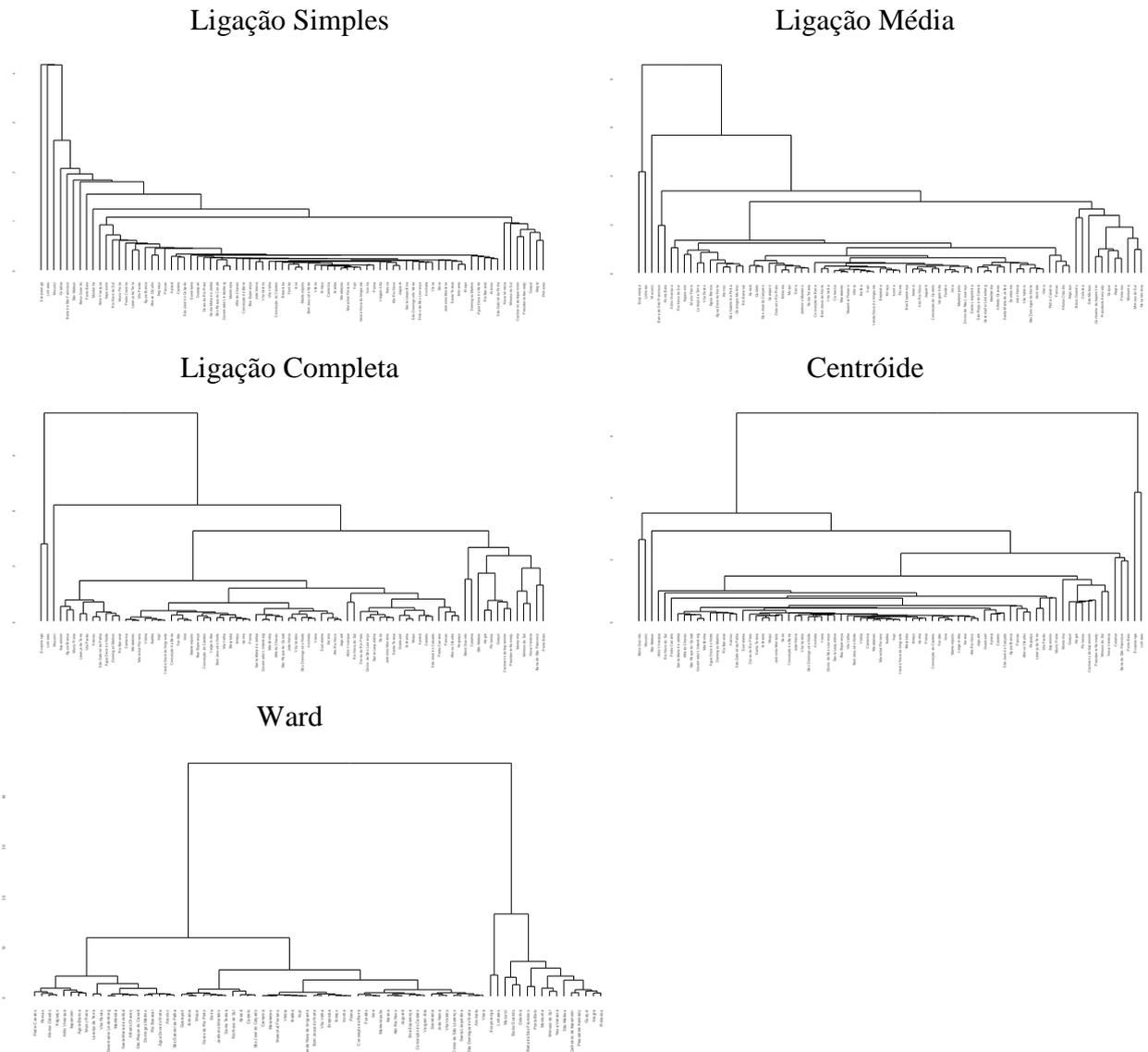


Figura 1- Dendrogramas do agrupamento dos municípios pelos métodos de agrupamento hierárquicos ligação simples, ligação completa, ligação média, centroide e Ward.

Fonte: Dados de pesquisa.

Cargnelutti Filho et al (2010) utilizaram a diferentes matrizes de similaridade e o CCC como método para avaliar a consistência de agrupamento em cultivares de feijão. Esses autores concluíram que o método da ligação média baseado na distância euclidiana foi o que apresentou maior coeficiente. Silva e Dias (2010) utilizaram o método de Tocher para



28ª SEAGRO

obtenção da correlação cofenética em estudos de cultivares de alho. Encontraram valores de correlação mais baixos para o método de Ward e mais alto para o método de Tocher, ambos baseados na distância euclidiana. E concluíram que comparações entre agrupamentos podem ser realizadas utilizando como critério valores do CCC. Tais estudos estão em consonância com os resultados encontrados neste trabalho.

Conclusão

Recomenda-se o uso do coeficiente de correlação cofenética como maneira de validar a consistência e escolher o método para realizar a análise de agrupamento.

Referências

BARROSO, L.P.; ARTES, R. **Análise Multivariada**. Lavras: UFLA, 2003.

CARGNELUTTI FILHO, A.; RIBEIRO, N. D.; BURIN, C. Consistência do padrão de agrupamento de cultivares de feijão conforme medidas de dissimilaridade e métodos de agrupamento. **Pesquisa agropecuária brasileira**, Brasília, v. 45, n. 3, p. 236-243, mar. 2010.

HAIR, J.F.JR.; ANDERSON, R.E.; TATHAM, R.L.; BLACK, W.C. **Análise multivariada de dados**. Porto Alegre: Bookman, 2005.

IBGE. Instituto Brasileiro de Geografia e Estatística. Censo Agropecuário, 2006. Disponível em: <<http://www.ibge.gov.br/home/estatistica/economia/agropecuaria/censoagro/default.shtm>>. Acesso em: 20 maio 2017.

JOHNSON, R.A.; WICHERN, D.W. **Applied Multivariate Statistical Analysis**. New Jersey: Editora Prentice Hall, 1998.

R CORE TEAM (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Disponível em: <<https://www.R-project.org/>>. Acesso em: 10 de julho 2017.

SILVA, A. R.; DIAS, C.T.S. A cophenetic correlation coefficient for Tocher's method. **Pesquisa agropecuária brasileira**, Brasília, v. 48, n. 6, p. 589-596, jun. 2013.