

Aplicação de técnica estatística multivariada em indicadores de sustentabilidade nos municípios do Marajó-PA

Diêgo Lima Crispim ^[1], Lindemberg Lima Fernandes ^[2], Roberta Luiza de Oliveira Albuquerque ^[3]

[1] dlimacrispim@gmail.com. Universidade Federal do Pará (UFPA) / Programa de Pós-graduação em Engenharia Civil (PPGEC) - Belém - PA. [2] lberge@ufpa.br. [3] robealbuquerque@gmail.com. Universidade Federal do Pará (UFPA) / Faculdade de Engenharia Sanitária e Ambiental (FAESA) - Belém - PA

RESUMO

Os indicadores são instrumentos importantes para orientar e auxiliar nas tomadas de decisões, bem como para conhecer o cenário de um determinado lugar e acompanhar seu desenvolvimento. Este estudo teve como finalidade analisar o comportamento dos municípios do Marajó-PA, através de indicadores que abrangem dimensões sociais, econômicas, habitacionais e de saneamento ambiental, utilizando a técnica estatística de análise multivariada, visando agrupar estes em um pequeno número de grupos homogêneos. Para escolha dos indicadores, foi realizado um *check-list* de trabalhos acadêmicos nacionais, regionais e locais que versam a respeito da sustentabilidade. Em seguida, foi feita uma padronização dos indicadores, em função das unidades e escalas de medidas distintas, não influenciando no resultado e apresentando pesos similares no cálculo do coeficiente de similaridade. A medida de dissimilaridade utilizada foi a distância euclidiana, e para composição dos agrupamentos foram aplicados os métodos de Ward e k-means. O resultado obtido na utilização do método de agrupamento hierárquico de Ward proporcionou a redução dos números de municípios para uma quantidade de 4 prováveis grupos, apresentando atributos semelhantes dentro do grupo e distintos entre os demais. Também apresentou um coeficiente de correlação cofenética (CCC) de $r=0,81$, indicando um bom grau de ajuste entre o dendrograma e a matriz de dissimilaridade. Além disso, indicou que a formação dos agrupamentos e os municípios integrados nestes apresentaram similaridade tanto no método hierárquico como no não hierárquico. No método k-means, constatou-se que quase todos os municípios que fazem divisão territorial ficaram dentro do mesmo grupo.

Palavras-chave: Coeficiente de correlação cofenética. Distância euclidiana. Método de Ward. Dendrograma. Método k-means.

ABSTRACT

Indicators are important tools to guide and assist decision-makers. They are also important to get to know the scenario of a given place and monitor its development. This study aimed to analyze the behavior of the municipalities of Marajó-PA through indicators that cover social, economic, housing and sanitation, using a statistical technique of multivariate analysis to group these into a small number of homogeneous groups. In order to choose the indicators, we carried out a checklist of national, regional and local academic papers dealing with sustainability. Then, the indicators were standardized according to the different units and scales of measurement, not influencing the result and presenting similar weights in the calculation of the similarity coefficient. The measure of dissimilarity used was the euclidean distance and for the composition of the groupings the Ward and k-Means methods were applied. The result obtained using Ward's hierarchical grouping method enabled the reduction of the numbers of municipalities to a number of 4 probable groups with similar attributes within the group and distinct among the others. It also presented a cophenetic correlation coefficient (CCC) of $(r = 0.81)$, indicating a good degree of fit between the dendrogram and the dissimilarity matrix. The results indicated that the formation of the clusters and the municipalities integrated in them presented similarity both in the hierarchical and non-hierarchical methods. In the k-means method it was found that almost all municipalities that make territorial division remained within the same group.

Keywords: Cophenetic correlation coefficient. Euclidean distance. Ward method. Dendrogram. k-means method.

1 Introdução

Segundo Macedo *et al.* (2016), a questão da sustentabilidade encontra-se em destaque por causa da urgência de procurar novas maneiras de ordenação do processo produtivo que privilegiem a continuidade da capacidade de estruturas ambientais e, deste modo, o bem-estar e qualidade de vida das atuais e futuras gerações.

Essa temática tem possibilitado a apresentação de técnicas que visam determinar, analisar e auxiliar as tomadas de decisão de entidades e/ou órgãos públicos e particulares, de forma a gerenciar eficazmente os recursos socialmente existentes (SILVA *et al.*, 2016).

De acordo com Carvalho *et al.* (2011), uma das grandes dificuldades na mensuração ou avaliação da sustentabilidade baseia-se na construção de técnicas apropriadas que possibilitem analisar realidades de cenários com alcances locais, regionais ou nacionais, haja vista que existem distintos atributos e especificidades intrínsecas aos componentes econômicos, sociais, ambientais e habitacionais.

Para Frainer *et al.* (2017), os indicadores são essenciais para nortear os tomadores de decisão diante da variedade de mecanismos para coordenação de políticas públicas, posto que as informações alcançadas por estes facilitam a tomada de decisão e podem colaborar para determinar a performance de ações voltadas para a sustentabilidade.

No âmbito municipal, a aplicação dos indicadores pelo poder público é importante para expor as ações indispensáveis, com a finalidade de reparar desconformidades descobertas, proporcionando o desenvolvimento sustentável em nível local, por conseguinte regional, alcançando um grau de sustentabilidade eficaz (REZENDE *et al.*, 2017).

Neste estudo, o grupo de municípios foi analisado com base em um conjunto de indicadores, com a finalidade de delinear uma análise de estrutura de agrupamento e verificar as semelhanças entre os municípios, bem como as suas distinções com os demais a partir da medida de dissimilaridade e do método empregado para formação dos grupos.

Para tal procedimento, foi empregada a técnica de análise multivariada, a saber, a análise de agrupamento, a qual consiste em um método que tem como finalidade ordenar objetos, lugares ou amostras (HAIR *et al.*, 2009), baseado nas características que estes apresentam, agrupando os municípios similares entre si em um grupo e os distintos em outros grupos (RODRIGUES; FACHEL; PASSUELLO, 2012).

Diante do exposto, este estudo teve como objetivo analisar o comportamento dos municípios do Marajó-PA, através de indicadores que abrangem componentes sociais, econômicos, habitacionais e de saneamento ambiental, aplicando um método estatístico de análise multivariada, com o propósito de agrupar estes em um pequeno número de grupos homogêneos.

2 Material e métodos

2.1. Área de estudo

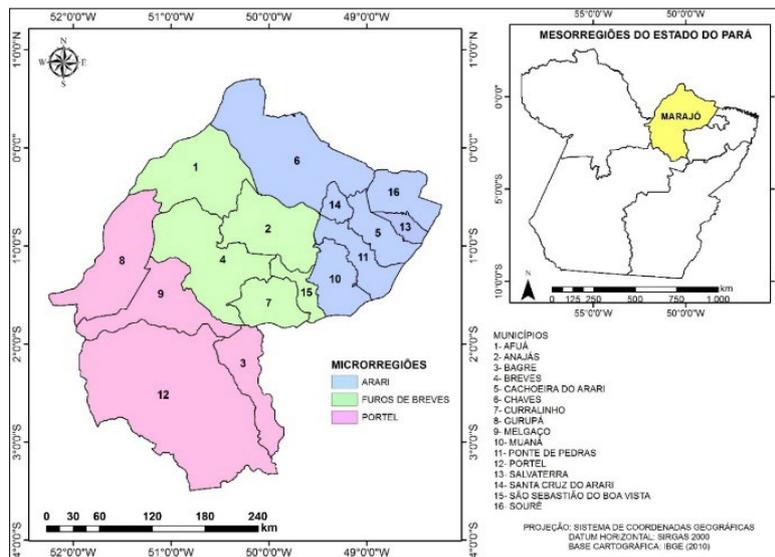
A área que compreende este estudo abrange os 16 municípios que constituem a mesorregião geográfica do Marajó, situada no estado do Pará (Figura 1). Esta mesorregião é formada pelas seguintes microrregiões: Arari, Furos de Breves e Portel, sendo as duas primeiras formadas por municípios situados integralmente na Ilha do Marajó e a última por municípios com sede em áreas continentais, tanto no trecho sul como no sudoeste da mesorregião (CRISPIM *et al.*, 2016; NASCIMENTO *et al.*, 2016).

Segundo informações do Instituto Brasileiro de Geografia e Estatística (IBGE, 2010), o número de habitantes da mesorregião do Marajó é cerca de 487.010, sendo 211.452 destes moradores da área urbana e 275.558 da área rural, para uma área territorial de 104.140 km².

Com base na classificação de Köppen-Geiger, a mesorregião do Marajó é delimitada em duas subdivisões climáticas: no trecho ocidental, predomina o clima equatorial sempre úmido (Af), não tendo estação seca, mas bastante chuvas estacionais, com uma média anual da precipitação registrada de 2.900 mm e mensal, em geral, acima de 60 mm (BEZERRA, 2014); no trecho oriental, onde o clima predominante é o equatorial monçônico (Am), existe um período moderadamente seco, todavia o acumulado anual da chuva é suficiente para sustentar este período, com uma precipitação média anual de aproximadamente 2.100 mm e uma precipitação mensal menor que 60 mm em alguns meses do ano (BEZERRA, 2014). Com relação às temperaturas registradas para a mesorregião, a média fica entre 26 a 27 °C, enquanto a umidade relativa do ar é superior a 80% (MDA, 2012).

A hidrografia da região é constituída por uma emaranhada rede de drenagem de novos canais, a saber: paleocanais, furos, baías, paranás, meandros abandonados, lagos e igarapés, ordenando um complexo em constante desenvolvimento, em que se sobressaem os rios Amazonas, Pará, Anapu, Jacundá e Anajás, com seus diversos afluentes (BEZERRA, 2014; CRISPIM *et al.*, 2016).

Figura 1 – Localização da área de estudo



Fonte: Autores (2019).

2.2. Indicadores de sustentabilidade

A preparação de uma estrutura para executar a escolha e a formação de indicadores de sustentabilidade é necessária para representar a condição de um determinado local no tocante ao desenvolvimento sustentável e a possíveis medidas ou deliberações a serem tomadas (FRAINER *et al.*, 2017).

Para seleção das dimensões e indicadores empregados na análise de agrupamento, foi realizado um *check-list* de trabalhos nacionais, regionais e municipais que abordam a temática da sustentabilidade (Quadro 1).

Quadro 1 – Dimensão e indicadores utilizados para análise de agrupamento com base em informações do PNUD (2010)

Dimensão	Indicador	Referência
Social	Esperança de vida ao nascer	Macedo et al. (2016)
	Expectativa de anos de estudo	Rezende et al. (2017)
	IDHM	Souza Júnior et al. (2017)
	Mortalidade Infantil	Cardoso, Toledo e Vieira (2016)
	População total	Macedo et al. (2016)
Econômico	Índice de GINI	Cardoso, Toledo e Vieira (2016)
	Porcentagem de extremamente pobres	Cardoso, Toledo e Vieira (2016)
	Renda per capita	Macedo et al. (2016)
Habitação	Porcentagem da população em domicílios com banheiro e água encanada	Sousa, Santos e Sousa (2016)
	Porcentagem de pessoas em domicílios com energia elétrica	Cardoso, Toledo e Vieira (2016)
Saneamento Ambiental	Porcentagem da população em domicílios com água encanada	Sousa, Santos e Sousa (2016)
	Porcentagem da população em domicílios com coleta de lixo	Cardoso, Toledo e Vieira (2016)
	Porcentagem de pessoas em domicílios com abastecimento de água e esgotamento sanitário inadequados	Rezende et al. (2017)

Fonte: Dados da pesquisa (2019)

Os critérios utilizados para seleção dos indicadores foram suas particularidades para escala municipal, disponibilidade em base de dados, simplicidade e clareza (JUWANA *et al.*, 2012). Foram utilizados dados secundários disponibilizados no Programa das Nações Unidas para o Desenvolvimento (PNUD, 2010) como referência. Desse modo, foram selecionados os indicadores a serem utilizados, por exemplo, Índice de Desenvolvimento Sustentável para Municípios (IDSM), Barômetro da sustentabilidade e Índice de Desenvolvimento Humano Municipal (IDHM).

Os indicadores selecionados foram utilizados para uma análise multivariada, especificamente a análise de agrupamento, com a finalidade de agrupar os municípios em um pequeno número de grupos semelhantes criados a partir de similaridades entre si e distinções com os demais. Assim, para os procedimentos de cálculos, foi utilizado o programa (R) versão 3.6.1, que consiste em um ambiente de software livre para computação estatística e gráficos.

2.3. Análise de agrupamento

A análise de agrupamento, também denominada de *cluster analysis*, consiste em um método que tem como finalidade agrupar elementos, componentes e objetos (HAIR *et al.*, 2009), assim como locais (GONÇALVES *et al.*, 2016), baseado em características ou atributos que estes têm (HAIR *et al.*, 2009), gerando elementos similares e homogêneos entre si em um grupo, todavia distintos dos elementos de outros grupos (RODRIGUES; FACHEL; PASSUELLO, 2012).

Neste estudo, foram empregadas duas técnicas de análise de agrupamento – um método hierárquico e outro não hierárquico – para determinar municípios homogêneos. A aplicação, a princípio, de uma técnica hierárquica visa realizar um processo exploratório; posteriormente, torna possível a utilização de um número definido de agrupamentos na técnica não hierárquica.

Inicialmente, foi utilizado o método hierárquico de Ward para definição dos grupos. Esta técnica tem como particularidade a formação de um diagrama com duas dimensões, que também pode ser denominado de Dendrograma ou Diagrama de Árvore (SANTOS; BLANCO; GONÇALVES, 2011).

Em seguida, foi aplicado o método não hierárquico k-means, bastante utilizado em vários estudos (ALENCAR, BARROSO, ABREU, 2013; MINGOTI, 2013; RIBEIRO *et al.*, 2014; DHANACHANDRA,

MANGLEM, CHANU, 2015), em especial quando existem diversas observações a serem agrupadas (SEIDEL *et al.*, 2008).

2.4. Medidas de similaridade ou distância (dissimilaridade)

Para Fávero *et al.* (2009), a determinação do método de similaridade para realizar a análise de agrupamento é essencial, pois a classificação dos grupos de elementos, indivíduos ou populações só é possível com a aplicação de alguma técnica de similaridade ou distância capaz de viabilizar a comparação direta entre os elementos.

Assim, as medidas possibilitam a comparação entre os elementos, tornando possível observar se um elemento A é mais similar a B ou a C. Desse modo, a partir da medida utilizada, os elementos similares são agrupados e os demais são arranjados em grupos distintos.

Segundo Hair *et al.* (2009), as medidas de dissimilaridade são frequentemente empregadas na literatura. Assim, neste estudo foi empregada a medida de dissimilaridade distância euclidiana.

2.5. Distância euclidiana

A distância euclidiana tem particularidades métricas, bem como é a mais utilizada para variáveis de divisões ou separações e medidas em uma escala de intervalo (EVERITT; DUNN, 2010). Essa medida é usualmente empregada quando as variáveis ou indicadores são integralmente quantitativos (SEIDEL *et al.*, 2008; SILVA *et al.*, 2016).

Para mensurar a distância entre os municípios alvos da pesquisa, foi utilizada a distância euclidiana. Logo, na Equação 1 é demonstrada a solução geral da medida da distância euclidiana (D_x, y) entre o elemento x e o elemento y, em um espaço n-dimensional, definida por:

$$D_{x,y} = \sqrt{\sum_{k=1}^n (P_{kx} - P_{ky})^2} \quad (1)$$

Em que: $D_{x,y}$ é a distância entre os elementos x e y; P_{kx} é o valor do indicador P_k para o elemento (x); P_{ky} é o valor do elemento para a observação (y); realiza-se a soma para todos os indicadores (p) considerados.

2.6. Padronização dos indicadores

De acordo com Fávero *et al.* (2009), a realização da padronização dos dados sempre é necessária quando os indicadores apresentam unidades de medidas e/ou escalas diferentes, haja vista que podem mudar ou alterar a estrutura do agrupamento. Logo, a padronização dos indicadores pode amenizar este problema (HAIR *et al.*, 2009).

Segundo Fávero *et al.* (2009), a padronização das unidades de medidas dos indicadores possibilitará que estes apresentem pesos similares no cálculo do coeficiente de semelhança. Desse modo, para tal procedimento foi utilizado o método range, o qual faz com que o indicador apresente uma variação de 0 a 1 (Equação 2):

$$Y = \frac{X - \text{mínimo}}{\text{amplitude}} \quad (2)$$

Em que: **Y** é o valor obtido através do método range; **X** é o valor da variável observada.

2.7. Determinar o melhor número de agrupamentos

Para obter a melhor quantidade de grupos de observações a serem constituídos, foi utilizado o método Elbow, também denominado de método do cotovelo. A concepção básica deste método consiste em determinar o melhor número de grupos que pode ser obtido, mesmo sem conhecer a resposta previamente (KODINARIYA; MAKWANA, 2013; SYAKUR *et al.*, 2018).

Este método é um dos mais clássicos, sendo empregado em diversos estudos (BHOLOWALIA; KUMAR, 2014; SILVA; NASSER, 2018; SYAKUR *et al.*, 2018). Os agrupamentos iniciais irão inserir bastante informação, todavia, em um determinado ponto, o ganho marginal vai descer de forma drástica e formar um ângulo no gráfico. Assim, o “k” ideal para estabelecer a quantidade de grupos é selecionado nesta condição, por isso se chama “critério de cotovelo”. Logo, para o cálculo deste método é utilizada a Soma do Erro Quadrático (Equação 3), conforme Syakur *et al.* (2018):

$$SSE = \sum_{k=1}^k \sum_{X_i \in S_k} |X_j - C_k|^2 \quad (3)$$

Em que: **K** é a quantidade de grupos originados; **C_k** é o i-ésimo cluster; **X** refere-se aos dados existentes em cada grupo.

2.8. Coeficiente de correlação cofenética

O coeficiente de correlação cofenética (CCC) tem a finalidade de mensurar o grau de ajuste entre o dendrograma elaborado e a matriz de dissimilaridade, isto é, a matriz de dissimilaridade e a matriz originada da simplificação correspondente pela técnica de agrupamento (CARVALHO *et al.*, 2009; GOBO, GALVANI, WOLLMANN, 2018). Assim, optou-se pela utilização do CCC neste estudo (Equação 4).

$$CCC = \frac{C\hat{o}v(F, C)}{\sqrt{V(F)V(C)}} \quad (3)$$

Em que: **F** representa a matriz fenética; **C** retrata a matriz cofenética.

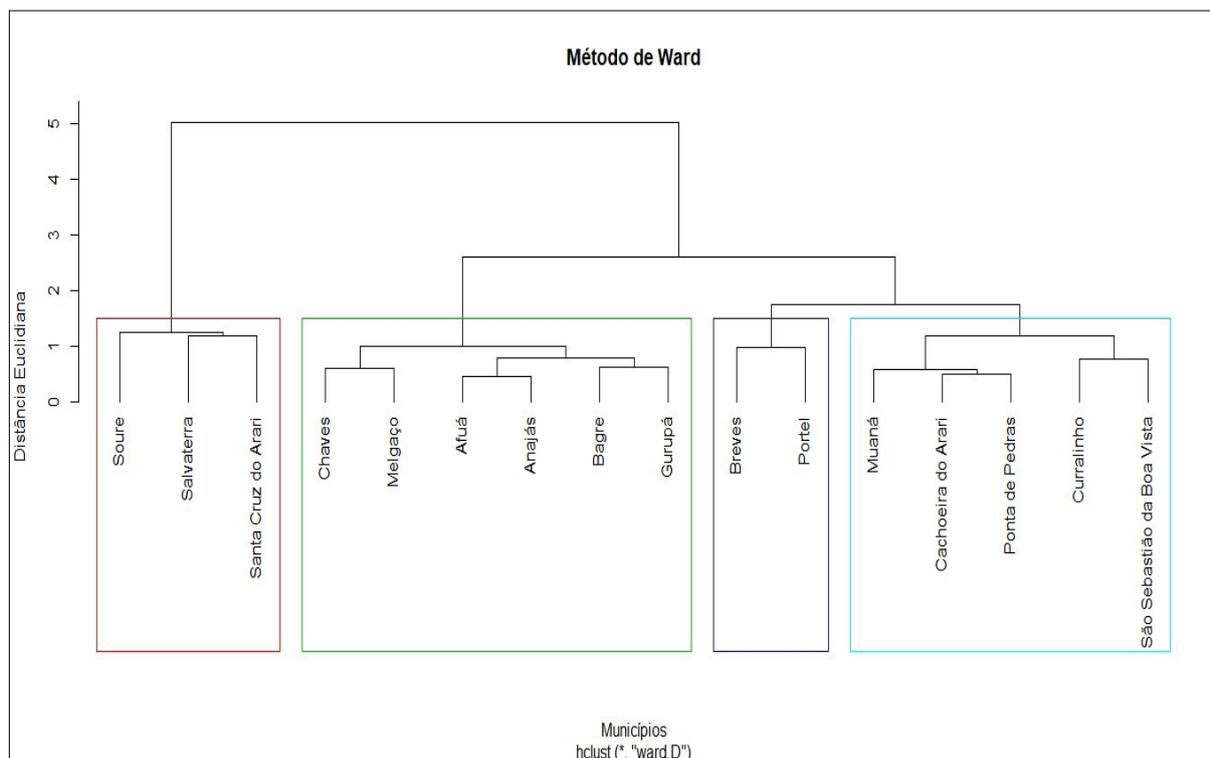
3 Resultados e discussão

O conjunto de dados submetidos à análise de agrupamento foi constituído por 16 observações nos municípios e por 13 indicadores, compreendidos pelos seus valores observados. Em seguida, os dados referentes aos indicadores foram organizados em uma planilha do Excel, na qual foi realizada uma padronização, tendo em vista que estes indicadores têm distintas unidades ou escalas de medida, e podem mudar ou alterar a estrutura do agrupamento.

A medida de dissimilaridade utilizada foi a distância euclidiana, e para composição dos agrupamentos foi empregado o método de Ward. O corte feito no eixo da dissimilaridade do dendrograma foi na altura 1,5, o que demonstrou a composição de 4 prováveis grupos, como pode ser constatado na Figura 2.

Com base no dendrograma, observa-se que no primeiro grupo foram considerados 3 municípios, com uma representação de 18,75% do total destes; para o segundo, foram indicados 6 municípios, com uma representatividade de 37,5% do total; já o terceiro grupo foi constituído por 2 municípios, retratando 12,5% do total; e, por fim, o quarto grupo, composto por 5 municípios, corresponde a 31,25% do total de municípios.

Figura 2 – Dendrograma resultante da análise de 16 municípios (baseado em 13 indicadores), obtido pelo método de Ward utilizando a distância euclidiana



Fonte: Dados da pesquisa (2019)

Os municípios mais semelhantes, com base na análise, são Cachoeira do Arari e Ponta de Pedras, seguidos de Afuá e Anajás. Assim, pressupõe-se que estes possuem maior similaridade entre si em termos do comportamento dos indicadores de sustentabilidade para o período observado e particularidades distintas dos demais.

O CCC do dendrograma gerado neste estudo foi de $r=0,81$, indicando um ajuste satisfatório entre a representação gráfica (dendrograma) e a matriz de dissimilaridade, tornando possível a execução de inferências através da análise visual do dendrograma. Segundo Rezende *et al.* (2016) e Streck *et al.* (2017), o coeficiente de correlação cofenética superior a 0,7 indica um bom ajuste do método de agrupamento.

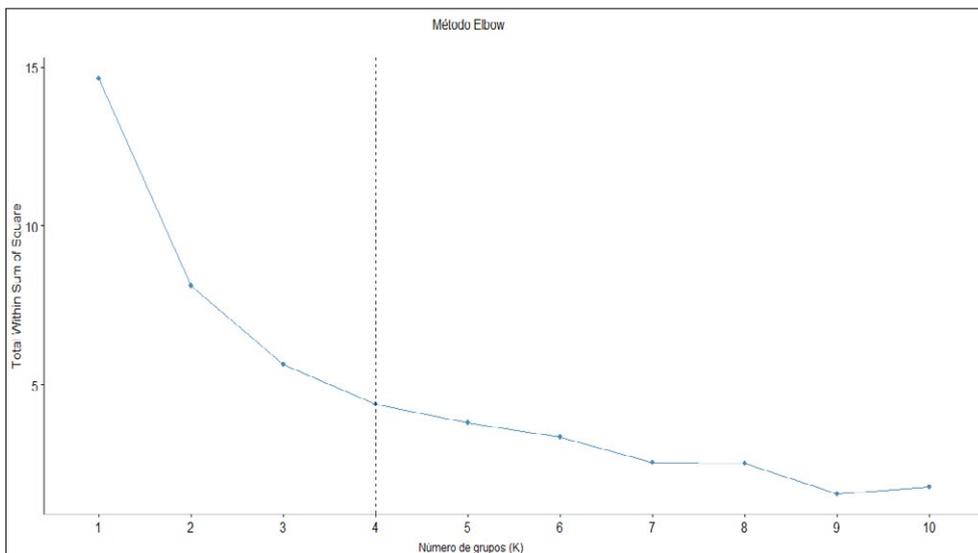
Assim, o resultado do CCC obtido nesta pesquisa mostra um ajuste adequado do método de agrupamento aplicado. Além disso, o desempenho do coeficiente de correlação cofenética neste estudo foi superior ao obtido na pesquisa de Streck *et al.* (2017), na qual obtiveram um $r=0,74$, um pouco abaixo

do resultado alcançado por Rezende *et al.* (2016), com $r=0,87$.

O resultado obtido através da aplicação do método do Elbow (Figura 3, na página seguinte) ilustra, de forma nítida, que existe uma variância entre o número de grupos 1 para 2, bem como existe uma distinção visual de 2 para 3 grupos, porém, considerando a metodologia Elbow, a quantidade de grupos adequado para este conjunto de informações é 4.

A aplicação do método Elbow contribuiu para definir o valor ideal de “k” a ser empregado na técnica não hierárquica k-means para formação de agrupamentos, haja vista que colabora na otimização dos agrupamentos (SYAKUR *et al.*, 2018). Assim, para utilização da técnica k-means, optou-se por selecionar 4 grupos, conforme a quantidade de agrupamentos determinado pelo método de Elbow. Os agrupamentos constituídos pela utilização do método k-means foram espacializados com o propósito de colaborar na compreensão e análise dos resultados (Figura 4).

Figura 3 – Imagem do gráfico originado pela aplicação do método Elbow

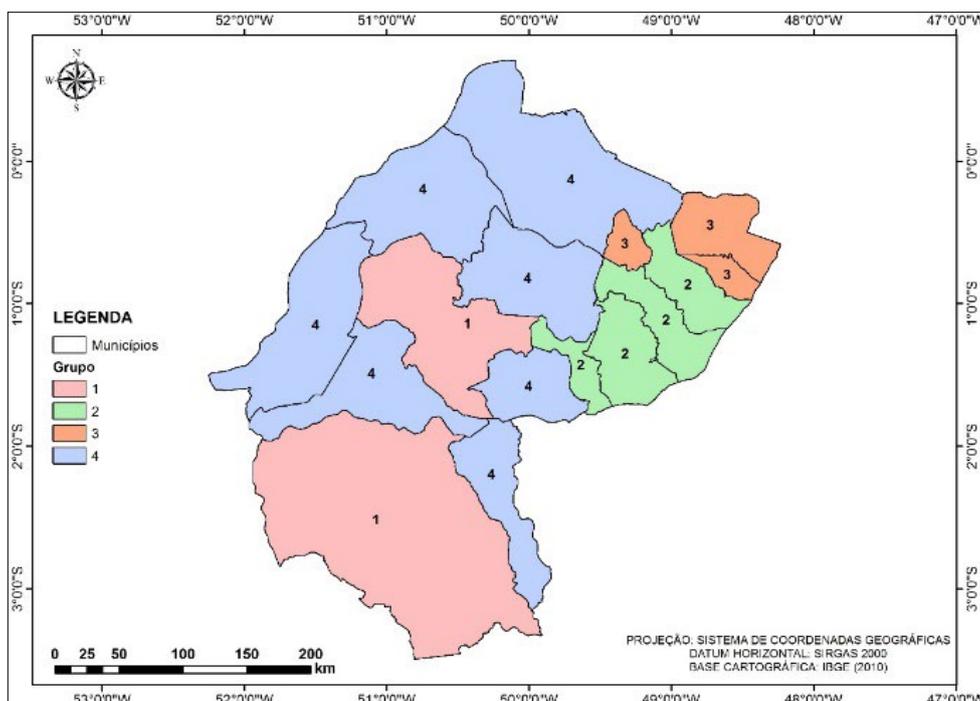


Fonte: Dados da pesquisa (2019)

A distribuição espacial dos municípios foi realizada com base no grupo ao qual pertencem, de acordo com a técnica que foi aplicada. Conforme o mapa, esta técnica possibilitou o agrupamento dos municípios entre si, apresentando alguns pontos dispersos,

possivelmente por causa de algumas particularidades locais distribuídas dentro do grupo. Todavia, constata-se a predominância de agrupamentos de municípios fronteiriços, ficando, estes, dentro de um mesmo grupo, conforme o método empregado.

Figura 4 – Distribuição espacial dos grupos determinados pelo método k-means



Fonte: Autores (2019)

Observando os resultados alcançados pela aplicação da técnica k-means, constata-se que estes apresentaram semelhança com os obtidos pelo método de Ward, pois os municípios constituídos por meio daquela técnica ficaram dentro dos mesmos grupos resultantes deste método, exceto um: Muaná.

4 Conclusão

O resultado obtido com o coeficiente de correlação cofenética (CCC) indicou um bom grau de ajuste entre o dendrograma e a matriz de dissimilaridade, possibilitando a realização de inferências a partir da representação gráfica.

A utilização da técnica de agrupamento possibilitou a diminuição da quantidade de municípios para um número de 4 grupos, compostos por municípios que provavelmente possuem características similares dentro de cada grupo e diferentes dos demais.

Neste estudo constatou-se que as composições dos municípios da mesorregião do Marajó nos agrupamentos foram semelhantes, tanto no método de agrupamento hierárquico como no não hierárquico. Além disso, percebeu-se, na aplicação do método k-means, que os municípios que fazem divisão territorial, em geral, ficaram dentro do mesmo grupo, exceto Breves e Santa Cruz do Arari.

Um dos estudos futuros será comparar técnicas de agrupamentos hierárquicos, baseado na medida de dissimilaridade (distância euclidiana), com a finalidade de obter dados a respeito de possíveis semelhanças e dissemelhanças entre os municípios inseridos na mesorregião do Marajó-PA, utilizando como dados de entrada indicadores de sustentabilidade. Além disso, para analisar os dendrogramas que serão formados e as matrizes de dissimilaridade, será empregado o CCC, visando observar se estes apresentam um ajuste adequado, selecionando o que apresentar o maior valor de CCC, o que traz simplicidade para a execução e é de fácil entendimento.

AGRADECIMENTOS

Agradecemos a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão de bolsa acadêmica do Processo de n.º 1848167. Ao Programa de Pós-Graduação em Engenharia Civil da Universidade Federal do Pará pelo suporte para realização da pesquisa.

REFERÊNCIAS

- ALENCAR, B. J.; BARROSO, L. C.; ABREU, J. F. Análise multivariada de dados no tratamento da informação espacial: uma abordagem com análise de agrupamentos. **Sistemas, Cibernética e Informática**, v. 10, n. 2, p. 1-7, 2013. Disponível em: [http://www.iiiisci.org/journal/CV\\$/ris-ci/pdfs/CA215JK13.pdf](http://www.iiiisci.org/journal/CV$/ris-ci/pdfs/CA215JK13.pdf). Acesso em: 8 maio 2019.
- BEZERRA, M. M. **Desenvolvimento Institucional da Educação Superior no Marajó**: um estudo sobre a implementação dos programas REUNI, PARFOR UAB e PROUNI no município de Breves (PA), no período de 2009 a 2013. 2014. Dissertação (Mestrado Profissional em Gestão Pública para o Desenvolvimento) –Núcleo de Altos Estudos Amazônicos, Universidade Federal do Pará, Belém, 2014. Disponível em: <http://www.naea.ufpa.br/naea/novosite/tcc/407>. Acesso em: 10 maio 2019.
- BHLOWALIA, P.; KUMAR, A. EBK-means: A clustering technique based on elbow method and k-means in WSN. **International Journal of Computer Applications**, v. 105, n. 9, p. 17-24, 2014.
- CARDOSO, A. S.; TOLEDO, P. M.; VIEIRA, I. C. G. Barômetro da sustentabilidade aplicado ao município de Moju, estado do Pará. **Revista Brasileira de Gestão e Desenvolvimento Regional**, v. 12, n. 1, p. 234-263, 2016.
- CARVALHO, J. R. M. *et al.* Proposta e validação de indicadores hidroambientais para bacias hidrográficas: estudo de caso na sub-bacia do alto curso do rio Paraíba, PB. **Sociedade & Natureza**, v. 23, n. 2, p. 295-310, 2011.
- CARVALHO, M. F. *et al.* Aplicação da análise multivariada em avaliações de divergência genética através de marcadores moleculares dominantes em plantas medicinais. **Revista Brasileira de Plantas Mediciniais**, v. 11, n. 3, p. 339-346, 2009.
- CRISPIM, D. L. *et al.* Especialização da cobertura do serviço de saneamento básico e do índice de desenvolvimento humano dos municípios do Marajó, Pará. **Revista Verde de Agroecologia e Desenvolvimento Sustentável**, v. 11, n. 4, p. 112-122, 2016.
- DHANACHANDRA, N.; MANGLEM, K.; CHANU, Y. J. Segmentação de imagens usando algoritmo de agrupamento K-means e Algoritmo de Clustering Subtrativo. **Procedia Computer Science**, v. 54, p. 764-771, 2015.

EVERITT, B. S.; DUNN, G. **Applied multivariate analysis**. 2. ed. London: Arnold, 2010. 354p.

FÁVERO, L. P. L. *et al.* **Análise de Dados: Modelagem multivariada para tomada de decisões**. Rio de Janeiro: Elsevier, 2009. 646p.

FRAINER, D. M. *et al.* Uma aplicação do Índice de Desenvolvimento Sustentável aos municípios do estado de Mato Grosso do Sul. **INTERAÇÕES**, v. 18, n. 2, p. 145-156, 2017.

GOBO, J. P. A.; GALVANI, E.; WOLLMANN, C. A. Influência do clima regional sobre o clima local a partir do diagnóstico de abrangência espacial e extrapolação escalar. **Revista Brasileira de Climatologia**, v. 22, p. 210-228, 2018.

GONÇALVES, M. F. *et al.* Identification of Rainfall Homogenous Regions taking into account El Niño and La Niña and Rainfall Decrease in the state of Pará, Brazilian Amazon. **Acta Scientiarum Technology**, v. 38, n. 2, p. 209-216, 2016.

IBGE. Instituto Brasileiro de Geografia e Estatística. **Censo Demográfico**. Rio de Janeiro: IBGE, 2010. Disponível em: <http://www.cidades.ibge.gov.br/xtras/temas.php?lang=&codmun=150450&idtema=16&search=para|melgaco|sintese dasinformacoes>. Acesso em: 19 maio 2019.

HAIR JR., J. F. *et al.* **Análise Multivariada de Dados**. 6. ed. Porto Alegre: Bookman, 2009. 688p.

JUWANA, I.; MUTIL, N.; PERERA, B. Indicator-based water sustainability assessment a review. **Science of the total environment**, v. 438, p. 357-371, 2012.

KODINARIYA, T. M.; MAKWANA, P. R. Review on determining number of Cluster in K-Means Clustering. **International Journal of Advance Research in Computer Science and Management Studies**, v. 1, n. 6, p. 90-95, 2013.

MACEDO, L. O. B. *et al.* Avaliação da sustentabilidade dos municípios do estado de Mato Grosso mediante o emprego do IDSM – Índice de Desenvolvimento Sustentável para Municípios. **Revista Brasileira de Gestão e Desenvolvimento Regional**, v. 12, n. 3, p. 323-345, 2016.

MDA. Ministério do Desenvolvimento Agrário. **Relatório Analítico do Território do Marajó**. Belém, 2012. 79p. Disponível em: <http://sit.mda.gov.br/download/ra/ra129.pdf>. Acesso em: 23 maio 2019.

MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. Belo Horizonte: Editora UFMG, 2013. 297p.

NASCIMENTO, E. C. *et al.* A influência do Programa Bolsa Família nas práticas alimentares das famílias do Território do Marajó, Pará, Brasil. **Revista Scientia Plena**, v. 12, n. 6, p. 1-11, 2016. Disponível em: <https://www.scientiaplenu.org.br/sp/article/view/3110/1492>. Acesso em: 20 maio 2019.

REZENDE, G. B. de M. *et al.* Sustentabilidade de Barra do Garças sob a ótica do índice de desenvolvimento sustentável para municípios. **Desenvolvimento em Questão**, n. 39, p. 203-235, 2017.

REZENDE, M. P. G. *et al.* Conformação corporal de equinos de diferentes grupos genéticos. **Ciência Animal Brasileira**, v. 17, n. 3, p. 316-326, 2016.

RIBEIRO, L. C. S. *et al.* A indústria do lazer no Brasil e sua relação com o desenvolvimento municipal. **Revista Brasileira de Estudos Regionais e Urbanos**, v. 8, n. 1, p. 77-91, 2014.

RODRIGUES, A.; FACHEL, J. M. G.; PASSUELLO, A. C. Estatística espacial e análise de cluster em dados de desastres naturais: mapeamento das inundações no Rio Grande do Sul entre 2003 e 2009. **Revista Iniciação Científica**, v. 10, n. 1, p. 48-67, 2012.

SANTOS, L. L.; BLANCO, C. J. C.; GONÇALVES, M. F. Identificação de Regiões Pluviométricas Homogêneas na Região Hidrográfica da Calha Norte - Pará. In: SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS, 19., 2011, Maceió. **Anais[...]**. Maceió: SBRH, 2011. Disponível em: https://www.abrh.org.br/SGCv3/index.php?PUB=3&ID=81&SUMARIO=245&ST=identificacao_de_regioes_pluviometricas_homogeneas_na_regiao_hidrografica_da_calha_norte_para. Acesso em: 10 maio 2019.

SEIDEL, E. J. *et al.* Comparação entre o método Ward e o método K-médias no agrupamento de produtores de leite. **Ciência e Natura**, v. 30, n. 1, p. 7- 15, 2008.

SILVA, F. P. *et al.* Grau de sustentabilidade do município de Rondonópolis de estado de Mato Grosso, a partir do Barômetro da Sustentabilidade. **Revista ESPACIOS**, v. 37, n. 23, p. 1-20, 2016.

SILVA, A. M.; NASSER, I. Uma metodologia para agrupamento de amostras de água da região Amazônica. **South American Journal of Basic Education, Technical and Technological**, v. 5, n. 2, p. 225-231, 2018.

SOUSA, L. C. R.; SANTOS; R. B. N.;
SOUSA, D. S. P. Desenvolvimento e pobreza
multidimensional na Amazônia Legal. **Revista
Espacios**, v. 37, n. 21, p. 26, 2016.

SOUZA JÚNIOR, F. C. *et al.* *Índice de Desenvolvimento
Humano Municipal e seus indicadores: O
desenvolvimento da Região do Médio Curso Superior
da Bacia Hidrográfica do Rio Apodi-Mossoró/RN.*
Revista Espacios, v. 38, n. 43, p. 1-16, 2017.

STRECK, E. A. *et al.* Variabilidade fenotípica de genótipos
de arroz irrigado via análise multivariada. **Revista
Ciência Agronômica**, v. 48, n. 1, p. 101-109, 2017.

SYAKUR, M. A. *et al.* Integration K-Means Clustering
Method and Elbow Method For Identification of The Best
Customer Profile Cluster. **IOP Conf. Series: Materials
Science and Engineering**, v. 336, p. 012017, 2018.