

# ANÁLISE MULTIVARIADA DE DADOS NO TRATAMENTO DA INFORMAÇÃO ESPACIAL UMA ABORDAGEM COM A ANÁLISE DE AGRUPAMENTOS

Bernardo Jeunon de Alencar

Instituto de Informática – Programa de Pós Graduação em Geografia – Tratamento da Informação Espacial – PUC Minas  
Belo Horizonte – MG – 30535-901 - Brasil

Leônidas Conceição Barroso

Programa de Pós Graduação em Geografia – Tratamento da Informação Espacial – PUC Minas  
Belo Horizonte – MG – 30535-901 - Brasil

João Francisco de Abreu

Programa de Pós Graduação em Geografia – Tratamento da Informação Espacial – PUC Minas  
Belo Horizonte – MG – 30535-901 - Brasil

## Resumo

A organização e a análise de dados é um tema de grande importância na Geografia porque pode possibilitar uma maior facilidade no exame conjunto de informações que possam oferecer subsídios para a explicação de fenômenos geográficos de maneira a auxiliar o homem no planejamento de suas atividades, nos seus processos de tomada de decisões e em suas ações estratégicas.

A Análise Multivariada tem ganhado um significado cada vez mais amplo na Geografia por possibilitar uma maior facilidade no exame de grandes conjuntos de informações, tão necessários para a explicação de fenômenos geográficos, o estudo de tendências e padrões espaciais, a formulação de modelos e a elaboração de previsões.

A Análise de Agrupamentos é uma dessas técnicas multivariadas muito utilizadas na Geografia e esse artigo reúne diversos fundamentos matemáticos, estatísticos e computacionais que amparam a sua utilização no tratamento de dados espaciais. Ele faz um estudo de suas aplicações na Geografia e revela os algoritmos que tornam viáveis a sua computação. Ele também apresenta um exemplo de uso das técnicas na Geografia utilizando dados sócio-econômicos de 23 municípios pertencentes à Mesorregião do Vale do Mucuri, em Minas Gerais.

**Palavras-Chave** – Análise Multivariada – Análise de Agrupamentos – Classificação – Tratamento da Informação Espacial – Geografia

## Introdução

A Análise Multivariada de dados tem um significado cada vez mais amplo na Geografia porque possibilita uma maior facilidade no exame conjunto de informações necessárias ao fornecimento de subsídios que permitam a explicação de fenômenos geográficos, o estudo de tendências e padrões espaciais tais como a formulação de modelos e a elaboração de previsões. É necessário, disponibilizar, de forma rápida, precisa e organizada, informações que venham a auxiliar o homem no planejamento de suas atividades.

A organização, a classificação e a análise de dados na Geografia podem ser feitas por meio de várias técnicas

multivariadas. Esse trabalho discute uma delas, a Análise de Agrupamentos, que corresponde ao conjunto de diversos procedimentos e algoritmos que objetivam identificar e agrupar objetos segundo alguma similaridade existente entre eles.

## A Análise Multivariada

Segundo JOHNSON & WICHERN, 1988 [6], a Análise Multivariada se utiliza de métodos estatísticos com a finalidade de descrever e analisar dados de muitas variáveis simultaneamente. Entender o relacionamento entre essas diversas variáveis faz desse conjunto de técnicas uma metodologia de grande potencial de aplicação, principalmente com o acesso à computação, veloz e acessível como se observa atualmente.

Para HAIR *et al*, 2005 [5], a Análise Multivariada é um conjunto de técnicas voltadas para a análise de dados e é uma área em constante expansão. Dentre tantas já bastante estabelecidas, pode-se citar, por exemplo, a Análise de Componentes Principais e a Análise dos Fatores Comuns, a Análise de Agrupamentos, a Regressão Múltipla e a Correlação Múltipla, a Análise de Discriminante Múltipla, a Análise Multivariada de Variância e Covariância e a Correlação Canônica. Outras técnicas ainda emergentes podem ser citadas, por exemplo, a Análise de Correspondência, Modelos Lineares de Probabilidade e a Modelagem de Equações Simultâneas / Estruturais. A Área de Sistemas Multivariados, que envolve trabalhos em Mineração de Dados e Redes Neurais é outra em grande desenvolvimento.

A Análise de Agrupamentos é uma técnica analítica para criar grupos significativos de indivíduos ou objetos. Especificamente, o que se faz com essa técnica é classificar uma amostra de objetos em um número de grupos mutuamente excluídos com base nas similaridades entre seus atributos.

Muitas das técnicas multivariadas envolvem conceitos sobre classificação de dados. Classificar, em termos muito simplistas, é impor algum tipo de organização aos dados através da criação de grupos que tenham uma determinada característica semelhante. Esse conceito é antigo. Desde tempos imemoriais o homem faz isso. Ele observa seu lugar e as coisas desse lugar. Descreve e define seu espaço agrupando semelhanças que ele simplesmente observa.

A partir de 1960, a introdução das metodologias que utilizavam algoritmos e computadores para a construção de sistemas de classificação deu origem ao termo Taxonomia Numérica (SNEATH & SOKAL, 1973 [8]).

A Taxonomia Numérica pode ser definida como a formação de grupos baseados no estudo das características dos dados em análise e em suas semelhanças. Ela é feita através de processos numéricos, geralmente baseados em uma Matriz de Semelhanças que é reduzida à medida que grupos similares de objetos são construídos (SNEATH & SOKAL, 1973 [8]). Existem muitas vantagens da Taxonomia Numérica. SNEATH & SOKAL, 1973 [8], assim como JOHNSON & WICHERN, 1988 [6], apontam diversas delas, algumas citadas a seguir:

- A Taxonomia Numérica tem o poder de integrar dados de uma variedade de fontes, coisa difícil de ser feita por meio da Taxonomia Convencional;
- A possibilidade de automatização de processos taxonômicos propicia um aumento da eficiência na análise dos dados, exigindo menos trabalho e um número menor de pessoas envolvidas;
- Os dados trabalhados, como estão na forma numérica, podem ser integrados com sistemas de processamento de dados e utilizados na elaboração de descrições, gráficos, mapas e outros documentos;
- Sendo quantitativos, os métodos fornecem maior discriminação ao longo do espectro de diferenças taxonômicas e são mais sensíveis na delimitação de grupos, fornecendo consequentemente melhores agrupamentos;
- A criação de tabelas de dados, feitas de maneira explícita na Taxonomia Numérica, força seus usuários a utilizarem informações mais bem descritas, melhorando a qualidade dos resultados como um todo;

Como metodologia a Taxonomia Numérica não produz novos dados. Trata-se na verdade de outra forma de organizar esses dados e obter, a partir deles, outra forma de apresentação, segundo as necessidades de um pesquisador.

O aumento na disponibilização de dados e os recentes avanços nas técnicas metodológicas exigem formas de organização cada vez melhores.

## A Análise de Agrupamentos

A Análise de Agrupamentos é um conjunto variado de técnicas e algoritmos que objetivam identificar e agrupar objetos segundo a similaridade sobre algum atributo ou característica particular que possuem. Ela vem sendo utilizada em muitas áreas do conhecimento, principalmente para identificação de padrões de comportamento nos dados de observações. Também na Geografia, uma organização feita por meio da divisão desses dados em grupos, ou classes, pode possibilitar uma melhor compreensão de um fenômeno, facilitando sua interpretação e, mais tarde, sua representação.

Separar objetos em grupos similares, principalmente considerando apenas uma característica em particular, é uma atividade comum e intuitiva, e está presente no cotidiano do homem em qualquer coisa que faça e que requeira algum tipo de

organização. Mas quando existe a necessidade de se analisar mais de uma característica simultaneamente, identificar grupos de objetos passa a ser trabalhoso, exigindo conceitos mais sofisticados de semelhança e procedimentos mais “científicos” para se criar os agrupamentos (BASSAB *et al*, 1990 [3]).

Segundo EVERITT, 1974 [4], o problema básico que a Análise de Agrupamentos pretende resolver é, “dada uma amostra de “n” objetos (ou indivíduos), cada um deles segundo “p” variáveis, procurar um esquema de classificação que agrupe os objetos em “g” grupos. Devem ser determinados também o número e as características desses grupos”.

HAIR *et al*, 2005 [5], de forma análoga, dizem que a Análise de Agrupamentos é uma técnica analítica para a criação de grupos de indivíduos que objetiva classificar uma amostra de objetos em um pequeno número de grupos mutuamente excludentes, com base nas similaridades entre eles.

De fato, os algoritmos voltados para a Análise de Agrupamentos buscam particionar um conjunto de dados formando subconjuntos, ou grupos, de tal forma que os indivíduos, ou observações, presentes em um grupo, tenham alto grau de similaridade entre eles, enquanto que indivíduos pertencentes a diferentes grupos tenham alto grau de dissimilaridade.

Existem diversas etapas que precisam ser seguidas na realização da Análise de Agrupamentos sobre um conjunto de dados. SNEATH, 1967 [7], apontou essa necessidade. BASSAB *et al*, 1990 [3], também propõem uma estrutura apropriada para a aplicação das técnicas de agrupamento, podendo ser decomposta nas etapas seguintes:

1. Definição de objetivos, critérios, escolha de variáveis e objetos;
2. Obtenção dos dados;
3. Tratamento dos dados;
4. Escolha dos critérios de similaridade ou dissimilaridade;
5. Adoção e execução de um algoritmo para o agrupamento;
6. Apresentação dos resultados;
7. Avaliação e interpretação dos resultados.

Nessa estrutura, as etapas não são independentes. Ao contrário, pode ser necessário voltar a etapas anteriores para se corrigir ou aprimorar etapas posteriores.

Nas primeiras etapas são definidos os objetivos que se pretendem alcançar com a utilização da Análise de Agrupamentos. Aqui também são conhecidas as variáveis de análise e suas características no contexto do estudo. As variáveis também são consideradas quanto à escala em que se encontram e se são variáveis quantitativas ou qualitativas, derivadas, compostas, discretas ou contínuas, nominais ou ordinais. Se necessário as variáveis podem ser padronizadas para que se tornem adimensionais.

Na quarta etapa são feitas as principais opções da Análise de Agrupamentos. Se não for um dos objetivos a criação de um número já determinado de agrupamentos, opta-se, na maioria das vezes, pelas Técnicas Hierárquicas. Em contrapartida, se um determinado número de agrupamentos se faz necessário, opta-se pelo uso das Técnicas de Partição.

Uma vez escolhida qual técnica de partição que será utilizada é necessário definir a medida de distância, ou seja, o Coeficiente de Similaridade ou de Dissimilaridade.

Na quinta etapa aplicam-se diferentes algoritmos com o objetivo de se criarem agrupamentos preliminares. Diversos algoritmos podem ser utilizados até que se defina qual o mais adequado. Nesses casos, a experiência do pesquisador influi na definição da técnica a ser utilizada e isso é muito comum. Na criação dos agrupamentos, é importante que se diga, o que se quer garantir é que elementos pertencentes a um mesmo grupo apresentem comportamentos semelhantes e que elementos pertencentes a grupos diferentes apresentem comportamentos distintos.

Para se avaliar os agrupamentos, as demais etapas se utilizam de Dendogramas, Matrizes Cofenéticas e Gráficos, por exemplo.

Um exemplo detalhado de cálculo de agrupamentos demonstrando essa estrutura pode ser encontrado em ALENCAR, 2009. [1]

### Coeficientes de Similaridade e de Dissimilaridade

Um conceito fundamental na Análise de Agrupamento reside na escolha de um critério para medir a distância entre objetos ou quantificar sua similaridade. Segundo SNEATH & SOKAL, 1973 [8], agrupar é constatar uma distância mínima ou de alta correlação dentro de um padrão. Como afirma AZAMBUJA, 2005 [2], para se construir um simples grupo a partir de um conjunto de elementos é necessário utilizar algum critério de proximidade ou tipo de medida que possibilite a comparação entre os componentes desse conjunto, tornando possível verificar se um dado elemento A é mais parecido com B do que com C.

É, portanto, necessária a definição de um coeficiente de mensuração que quantifique a distância entre os objetos da análise e mostre o quanto dois elementos de um conjunto são similares.

Essa medida é chamada, como já citado anteriormente, de Coeficiente de Similaridade ou Coeficiente de Parecência. É importante observar que é possível se estabelecer também o conceito e o termo Coeficiente de Dissimilaridade. O Coeficiente de Similaridade é aquele cujo maior valor observado represente a maior proximidade, como é o caso do Coeficiente de Correlação. O Coeficiente de Dissimilaridade indica que quanto maior for o valor observado menor é a proximidade e menos parecidos são aqueles objetos, como é o caso da Distância Euclidiana.

Alguns coeficientes se adaptam melhor a determinadas situações de estudo e análise. Também, como já citado anteriormente, os coeficientes de Dissimilaridade são mais adequados para as variáveis quantitativas, e os de Similaridade para as variáveis qualitativas. Escolhido o coeficiente, torna-se possível construir a Matriz de Distâncias, que é de onde surgirão os grupos.

Uma variável Quantitativa é aquela que apresenta como possíveis realizações números oriundos de uma contagem, de uma medição ou de uma apuração. As variáveis quantitativas podem ser Discretas, se o conjunto de valores que ela pode assumir for um conjunto finito ou infinito enumerável, ou podem ser Contínuas, se o conjunto de valores que ela pode assumir for um conjunto infinito não enumerável (BUSSAB & MORETTIN, 2003, *apud* AZAMBUJA, 2005 [2]).

A Distância Euclidiana, por exemplo, é o Coeficiente de Dissimilaridade mais conhecido e mais utilizado para indicar a distância entre objetos de análise. Trata-se, objetivamente, da distância geométrica entre dois pontos no espaço e pode matematicamente ser expressa como:

$$d_{(A)(B)} = \sqrt{(x_{1(A)} - x_{2(B)})^2 + (y_{1(A)} - y_{2(B)})^2} \quad (1)$$

Onde:

$d_{(A)(B)}$  é a Distância Euclidiana entre os pontos A e B,

$x_{1(A)}, y_{1(A)}$  as coordenadas do ponto A e

$x_{2(B)}, y_{2(B)}$  as coordenadas do ponto B.

Generalizando, tem-se:

$$d_{(A)(B)} = \sqrt{\sum_{i=1}^p (x_{i(A)} - x_{i(B)})^2} \quad (2)$$

Onde:

$d_{(A)(B)}$ , é a distância euclidiana e

$x_{i(A)}$  e  $x_{i(B)}$  são os pontos considerados

A Distância Euclidiana apresenta simplicidade de cálculo e a distância entre quaisquer dois objetos não é afetada pela inserção de outros objetos ao conjunto de dados de análise. No entanto é conveniente que os dados sejam padronizados para que se evitem erros provocados pelas diferenças de escala associadas a dimensões.

Outras medidas também podem ser definidas, como a Distância Euclidiana Média, a Distância Euclidiana Quadrática, a Distância de Manhattan, o Coeficiente de Gower, de Cattel e outros. Maiores informações sobre esses coeficientes, incluindo aqueles utilizados para variáveis qualitativas, podem ser encontrados em ALENCAR, 2009 [1], em BASSAB *et al*, 1990 [3] e SNEATH & SOKAL, 1973 [8].

### Técnicas de Formação de Agrupamentos

O número de técnicas e algoritmos para a formação de agrupamentos é grande e diversificado. Um tipo de proposta de organização dessas técnicas é citada em EVERITT, 1974 [4], e CORMACK, 1971, *apud* BASSAB *et al*, 1990 [3]. Segundo os autores, pode-se dizer que existem três grandes famílias de técnicas voltadas para a formação de agrupamentos:

- Técnicas Hierárquicas, nas quais os objetos, ou observações, são classificados em grupos em diferentes etapas produzindo uma árvore de classificação;
- Técnicas de Partição, nos quais os grupos obtidos produzem uma partição no conjunto de objetos;
- Técnicas de Cobertura, nos quais os grupos formados recobrem o conjunto de objetos embora possam também se sobrepor.

Esse artigo se detém nas duas primeiras técnicas. Na verdade elas se diferem, basicamente, pela metodologia utilizada para se construir os agrupamentos. Escolher uma técnica em particular exige não somente o conhecimento de suas propriedades particulares como também sua adequação aos objetivos do pesquisador (BASSAB *et al*, 1990 [3]).

### Técnicas Hierárquicas para Análise de Agrupamento

As Técnicas Hierárquicas são utilizadas na Análise de Agrupamento logo depois que ocorre a formação de um grupo, no momento de reconstrução da Matriz de Distâncias. Como cita AZAMBUJA, 2005 [2], cada um dos elementos que participam do processo é considerado como um grupo e estará se juntando a outro elemento ou outro grupo de acordo com os valores calculados presentes na Matriz de Distâncias, que é atualizada cada vez que um agrupamento é feito. Ao final do processo todos os elementos são reunidos em um único grupo.

Nas técnicas hierárquicas não se sabe, preliminarmente, a quantidade de grupos que serão formados. Diferentes técnicas podem formar diferentes quantidades de grupos. Isso pode ser útil como levantamento preliminar, feito na fase exploratória da análise.

No contexto desse estudo serão trabalhadas as seguintes técnicas, também chamadas na literatura de métodos hierárquicos:

- Método das Médias das Distâncias (*Average Linkage*)
- Método da Ligação Simples (*Single Linkage*)
- Método da Ligação Completa (*Complete Linkage*)
- Método do Centróide (*Centroid Distance*)

O *Average Linkage* utiliza a média das distâncias entre todos os pares de objetos da matriz de dados para se criar a matriz de distâncias. Segundo JOHNSON & WICHERN, 1998 [6], esse método pode ser afetado se ocorrer mudança no coeficiente utilizado para os cálculos da matriz de distâncias, ainda que esse novo coeficiente mantenha a ordem dessas distâncias.

Esse método possui facilidade e rapidez de cálculo, principalmente se comparada às outras técnicas.

O *Single Linkage* também é conhecido como o Método do Vizinho mais Próximo ou Método da Distância Mínima. Esse método une grupos segundo uma distância mínima entre eles e os grupos formados tendem a ser menos homogêneos se comparados ao Método das Médias das Distâncias. Isso significa que em um mesmo grupo podem ser encontrados elementos bem distintos. Esse método tende a formar grupos com vários elementos enquanto isola outros elementos ainda não anexados (AZAMBUJA, 2005 [2]).

O Complete Linkage é também conhecido como o Método do Vizinho mais Distante. Nesse caso a similaridade é definida pelos objetos de cada grupo que menos se parecem.

Esse método, como cita AZAMBUJA, 2005 [2], *apud* Krzanowski & Marriot (1995), é indicado para a formação de grupos com tamanhos semelhantes e em geral forma grupos mais homogêneos que os formados por meio do Método das Médias das Distâncias e do Método da Ligação Simples. Além disso, como ressalta JOHNSON & WICHERN, 1998 [6], é um método que produz grupos que não se modificam mesmo quando outro coeficiente é adotado para o cálculo das distâncias, ou seja, ele mantém a ordenação das distâncias. O Método da Ligação Simples também possui essa característica.

O *Centroid Distance* é o mais direto no cálculo da distância entre um elemento de um grupo ou entre dois grupos. Para cada novo grupo formado uma nova distância é calculada e representa o centro médio dos elementos do grupo.

Esse método é bastante similar ao Método das Médias das Distâncias e seus resultados, em virtude disso, também são similares.

A maior dificuldade do método, no entanto, como ressalta ALENCAR, 2009 e BASSAB *et al*, 1990 [3], é a necessidade de se recuperar os dados originais, a cada grupo criado, para que seja possível recalcular o valor das distâncias. Quando muitas variáveis e objetos estão presentes o processo pode se tornar muito demorado.

Um detalhamento maior sobre essas técnicas pode ser encontrado em ALENCAR, 2009 [1].

### O Método K-Médias para Formação de Partições

As Técnicas de Partição, como o próprio nome diz, buscam produzir agrupamentos através de partições do conjunto original de elementos de análise. Cada partição deve mostrar coesão interna dentro de um mesmo grupo e isolamento entre os demais grupos. Além disso, as técnicas de partição já precisam definir o número total de grupos que serão criados antes de sua aplicação.

Um método de partição bastante conhecido e talvez o mais usado em Análise de Agrupamentos quando se tem muitos objetos (BASSAB *et al*, 1990 [3]) é o K-Médias, ou K-Means.

O K-Médias é uma técnica de partição que procura alocar os elementos de dados em grupos previamente definidos. É um método que minimiza a soma dos quadrados residuais dentro de cada grupo formado, aumentando a homogeneidade dentro dele ao mesmo tempo em que aumenta a diferença entre eles. Segundo JOHNSON & WICHERN, 1998 [6], o método foi introduzido por J. B. MacQueen em 1967.

O processo algorítmico do método é relativamente simples. Inicialmente é distribuído um elemento do conjunto de dados para cada grupo definido. Essa distribuição pode ser feita aleatoriamente ou através dos elementos que apresentem valores mais distantes de uma variável escolhida (como a média, por exemplo). Essa última opção é a mais utilizada e mais adequada.

Cada um desses K elementos se torna, então, o elemento central do grupo a que pertence e representam as sementes dos agrupamentos no momento inicial do processo. Quando, no decorrer do processo, novos elementos entrarem no grupo, o elemento central passa a ser a média entre eles. O processo segue designando cada novo elemento para um determinado grupo, especificamente aquele que apresente o elemento central mais próximo deste.

Depois que todos os elementos forem distribuídos nos K grupos definidos a soma dos quadrados residuais de cada grupo é calculada. Quanto menor for o valor dessa soma, mais homogêneos estarão os grupos formados.

No decorrer do processo os elementos de um grupo podem ser movidos para outros grupos buscando a formação de grupos os mais homogêneos possíveis, o que implicará em um novo cálculo. Se a soma dos quadrados residuais diminui, indicando aumento da homogeneidade, a movimentação é mantida e caso contrário ela é desfeita e o objeto movido retorna ao seu grupo original ou é movido para algum outro grupo, quando o ciclo se inicia novamente. Quando a soma não diminui mais ao longo das iterações (que podem, inclusive, ter um número previamente determinado), o processo termina. Os grupos são apresentados.

Um aprofundamento sobre outras características desse método pode ser encontrado em JOHNSON & WICHERN, 1998 [6], e ALENCAR, 2009 [1].

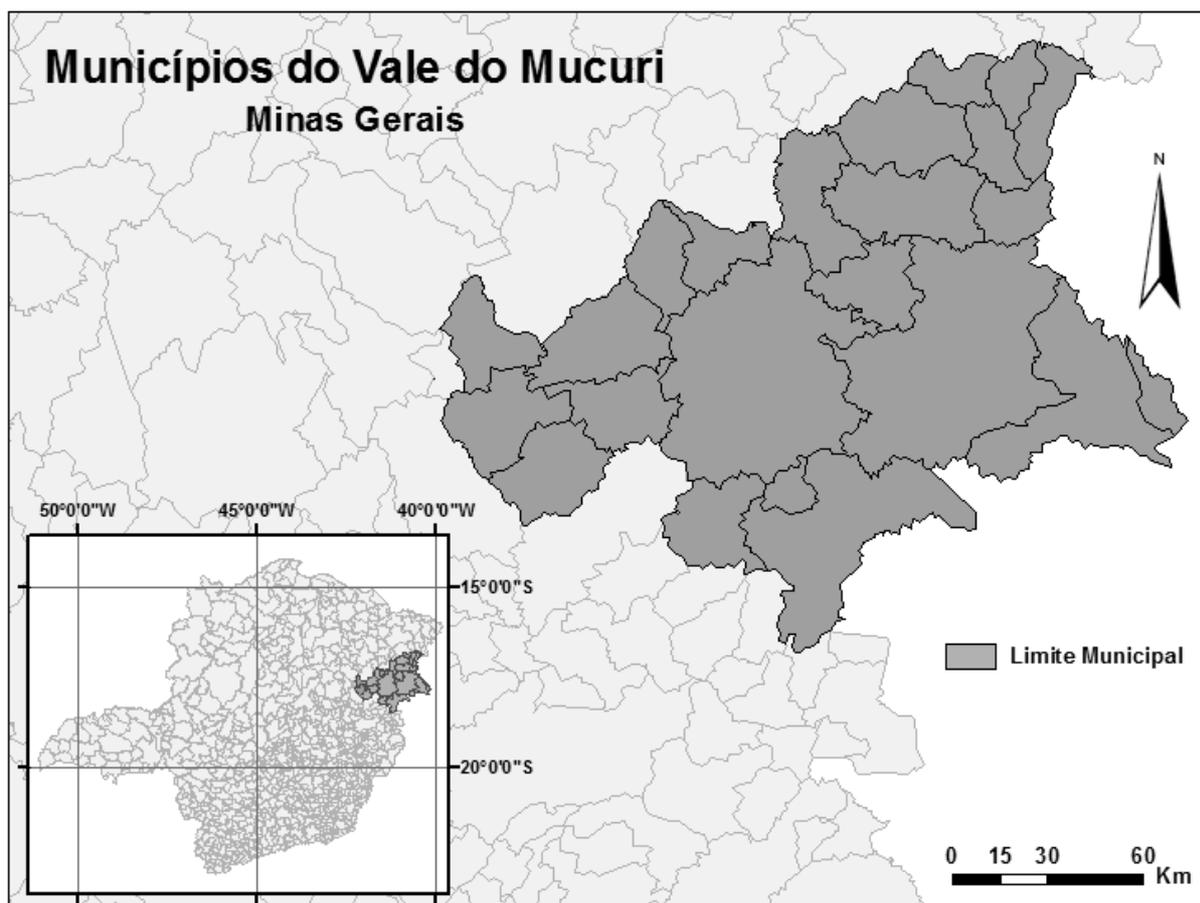
O que se conclui é que a escolha de um tipo particular de algoritmo para se realizar a Análise de Agrupamentos, deve ser função dos objetivos de um pesquisador frente aos dados de trabalho que ele possui. A utilização de computadores para tornar mais ágil os processos de cálculo envolvidos pode se tornar muito útil para o teste de diferentes métodos. A escolha de um que seja eficaz diante da necessidade de explicação de uma realidade deve ser critério desse pesquisador.

### Exemplo de Uso da Técnica

O exemplo a seguir considera alguns dados sócio-econômicos de 23 municípios pertencentes ao Vale do Mucuri, em Minas Gerais. O Mapa 1 mostra sua localização geográfica.

Foram trabalhadas 23 observações, ou casos, correspondentes aos municípios, e 16 variáveis censitárias descritas a seguir. O ano trabalho é o de 1991 e as diferenças de valores relatadas são entre 1991 e 2000.

O software utilizado para o cálculo foi o NinnaCluster, versão *Desktop*, disponível no site [www.globogeo.com/ninna](http://www.globogeo.com/ninna). A descrição completa do software e instruções sobre sua disponibilização está em ALENCAR, 2009 [1].



Mapa 1

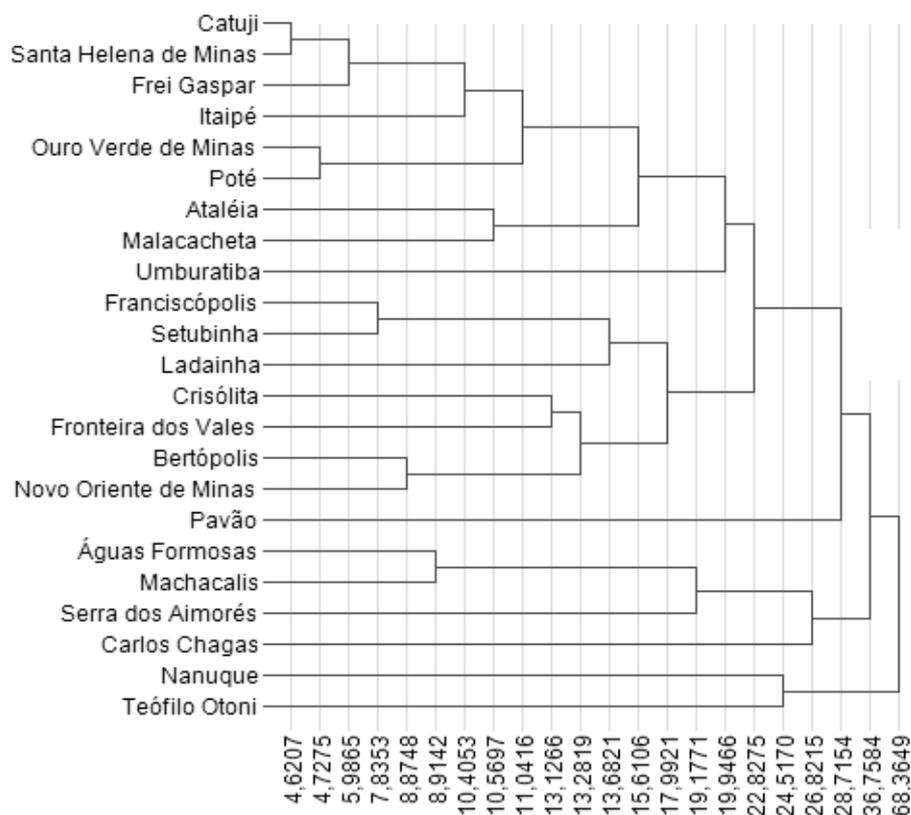
Vale do Mucuri - Localização Geográfica

São as seguintes as variáveis levantadas para estudo:

#	Nome	Descrição
1	EspVidaN	Esperança de Vida ao Nascer
2	TxAlfAdultos	Taxa de Alfabetização de Adultos
3	TxFreqEscola	Taxa Bruta de Freqüência à Escola
4	RendaPC	Renda per Capita
5	IDHM-M	Índice de Desenvolvimento Humano Municipal - Geral
6	IDHM-L	Índice de Desenvolvimento Humano Municipal - Longevidade
7	IDHM-E	Índice de Desenvolvimento Humano Municipal - Educação
8	IDHM-R	Índice de Desenvolvimento Humano Municipal - Renda

9	DifEspVida	Diferença da Esperança de Vida ao Nascer
10	DifTxAlfab	Diferença da Taxa de Alfabetização de Adultos
11	DifTxFreqE	Diferença da Taxa de Freqüência à Escola
12	DifRendaPC	Diferença da Renda per Capita
13	DifIDHM-M	Diferença do IDHM-M
14	DifIDHM-L	Diferença do IDHM-L
15	DifIDHM-E	Diferença do IDHM-E
16	DifIDHM-R	Diferença do IDHM-R

A figura a seguir mostra um exemplo de dendograma criado pelo software através de um dos métodos descritos.



Dendograma 1  
Método das Médias das Distâncias

A aplicação do método K-Médias dividindo-se o conjunto de municípios da região em cinco grupos revelou os seguintes agrupamentos:

Grupo I
Teófilo Otoni
Nanuque
Carlos Chagas

Grupo II
Águas Formosas
Machacalis
Serra dos Aimorés

Grupo III
Ataléia
Itaipé
Ouro Verde de Minas
Santa Helena de Minas
Frei Gaspar
Poté
Catuji

Grupo IV
Malacacheta
Pavão
Umburatiba

Grupo V
Bertópolis
Franciscópolis
Fronteira dos Vales
Ladainha
Setubinha
Novo Oriente de Minas
Crisólita

## Considerações Finais

A técnica pode ser aplicada em pesquisas nas mais diversas áreas do conhecimento. Em especial na Geografia ela simplifica muito a análise das inúmeras informações inerentes ao estudo de um fenômeno geográfico. Destaca-se, no entanto, a importância da presença de um geógrafo na análise e na validação dos resultados encontrados por meio da aplicação dessa e de outras técnicas matemáticas e estatísticas. O uso puro e simples de uma técnica não assegura, por si só, melhoria na explicação de um fenômeno espacial. Cabe ao geógrafo analisar e verificar se os resultados encontrados se adaptam à realidade do problema geográfico e servem para compor uma nova e mais rica representação ou explicação do fenômeno que se estuda.

## Agradecimentos

Os autores do presente artigo agradecem à FAPEMIG – Fundação de Amparo à Pesquisa do Estado de Minas Gerais – pelo apoio na apresentação desse trabalho.

## Bibliografia

- [1] ALENCAR, B. J., **A Análise Multivariada no Tratamento da Informação Espacial – Uma Abordagem Matemático-Computacional em Análise de Agrupamentos e Análise de Componentes Principais**. Tese de Doutorado em Geografia (PUC Minas). Orientadores: Leônidas Conceição Barroso e João Francisco de Abreu. 2009, 198p.
- [2] AZAMBUJA, S., **Estudo e Implementação da Análise de Agrupamento em Ambientes Virtuais de Aprendizagem** – Dissertação de Mestrado em Informática – UFRJ. Orientadores: Cláudia Lage Rebello da Motta e Marcos da Fonseca Elia. Rio de Janeiro, 2005
- [3] BASSAB, W. O., MIAZAKI, É. S. e ANDRADE, D. F., **Introdução à Análise de Agrupamentos**. Associação Brasileira de Estatística – ABE. 9º Simpósio Nacional de Probabilidade e Estatística. São Paulo – Julho de 1990.
- [4] EVERITT, B., **Cluster Analysis**. Heinemann Educational Books, London, 1974.
- [5] HAIR JR., J. F.; ANDERSON, R.; TATHAM, R. L.; BLACK, W. C., **Análise Multivariada de Dados** – 5ª Edição – Bookman – 2005 – 593p.
- [6] JOHNSON, R. A.; WICHERN, D. W., **Applied Multivariate Statistical Analysis**. Prentice Hall, New Jersey, USA, 1998, 816p.
- [7] SNEATH, P. H. A., **Some Statistical Problems in Numerical Taxonomy**. The Statistician, Vol. 17, Nr. 1 – 1967 – pp 1 – 12 – Blackwell Publishing for the Royal Statistical Society
- [8] SNEATH, P. H. A. e SOKAL, R. R., **Numerical Taxonomy – The Principles and Practice of Numerical Classification**. W. H. Freeman and Company – 1973 – 573p.