**Pedro Rafael Diniz Marinho**

# *Detecção de Cluster através de um Modelo Híbrido Genético-Fuzzy*

João Pessoa - PB, Brasil

17 de dezembro de 2010

**Pedro Rafael Diniz Marinho**

# *Detecção de Cluster através de um Modelo Híbrido Genético-Fuzzy*

Monografia apresentada para obtenção do Grau de Bacharel em Estatística pela Universidade Federal da Paraíba - UFPB.

Orientador:
Joab de Oliveira Lima

Universidade Federal da Paraíba
Centro de Ciências Exatas e da Natureza
Departamento de Estatística

João Pessoa - PB, Brasil

17 de dezembro de 2010

Monografia de Projeto Final de Graduação sob o título *"Detecção de Cluster através de um Modelo Híbrido Genético-Fuzzy"*, defendida por Pedro Rafael Diniz Marinho e aprovada em 17 de dezembro de 2010, em João Pessoa, Estado da Paraíba, pela banca examinadora constituída pelos professores:

---

Prof. Dr. Joab de Oliveira Lima
Orientador

---

Prof. Me. Jozemar Pereira dos Santos
Universidade Federal da Paraíba

---

Prof. Me. Marcelo Rodrigo Portela Ferreira
Universidade Federal da Paraíba

# *Resumo*

O inter-relacionamento de dados epidemiológicos, geralmente, é tratado por métodos que levem em conta as suas características espaciais. Nesse sentido, vários métodos de decteção de cluster, tal como o algoritmo Scan, têm sido sugeridos na literatura. Normalmente, essa metodologia utiliza como centros espaciais os centróides dos elementos espaciais envolvidos e, além disso, se utiliza apenas de conceitos univariados. No entanto, o mesmo procedimento não consegue trabalhar bem quando os centros dos objetos em estudo são características qualitatitivas da população, como por exemplo, as causas externas de óbitos que acomentem uma certa região ou um município. Esse último fato motivou a criação de um modelo híbrido de agrupamento de dados que utiliza um mistura das filosofias dos Algoritmos Genéticos com a Lógica Nebulosa Fuzzy, chamado de Modelo Híbrido Genético-Fuzzy (MHGF). Essa proposta foi aplicada a um banco de dados, fornecido pela Secretaria de Saúde do Estado da Paraíba, que relacionou os 30 principais municípios paraibanos com as 8 principais causas externas de óbitos descritas no Capítulo XX da CID-10 para os anos de 2006 a 2010. Os resultados se mostraram bastante satisfatórios com relação aos agrupamentos formados dos municípios dentro das causas. Inclusive, em muitas situações, foi observado que muitos municípios participantes de um mesmo grupo (causa) mantinham, sobretudo, as suas características de associação espacial. Esse estudo aplicado sugere, portanto, uma boa qualidade e eficiência do modelo proposto e o aponta como uma metodologia alternativa de agrupamentos de dados.

# *Abstract*

The interrelationship of epidemiological data, is generally treated by methods that take into account their spatial characteristics. Accordingly, various methods of cluster detection, as the Scan algorithm, have been suggested in the literature. Typically, this methodology uses space centers as the centroids of the spatial elements involved and also makes use only of univariate concepts. However, the same procedure can not work well when the centers of the objects under study are the population characteristics of qualitative approaches, for example, external causes of deaths imposed on a certain region or a municipality. This latter fact prompted the creation of a hybrid model of grouping data using a mixture of the philosophies of Genetic Algorithms with Fuzzy Logic, called Genetic-Fuzzy Hybrid Model (GFHM). This proposal was applied to a database, provided by Health Department of the State of Paraíba, which listed the top 30 municipalities of Paraíba with 8 main causes of deaths described in Chapter XX of ICD-10 for the years 2006 to 2010 . The results were quite satisfactory with regard to groups formed within the municipalities within of causes. In fact, in many situations, it was noted that many municipalities participating in a same group (external causes) maintained, especially the characteristics of spatial association. This applied study therefore suggests a good quality and efficiency of the proposed model and shows as an alternative method of data clusters.

# *Dedicatória*

Dedico este trabalho a minha mãe Wilta Maria Diniz Américo Marinho e ao meu pai José Walter Marinho da Silva por serem pais presentes na minha vida dando total apoio aos meus sonhos. Estas duas pessoas em nenhum momento mediram esforços para realização dos meus desejos, me guiaram pelo caminho correto me ensinando dessa forma a fazer boas escolhas mostrando que a honestidade e o respeito são essenciais à vida e que sempre devo lutar para o que quero. Aos meus pais devo a pessoa que me tornei e tenho muito orgulho de chamá-los de pai e mãe.

# *Agradecimentos*

Agradeço aos meus pais, Wilta Maria Diniz Américo Marinho e José Walter Marinho da Silva por terem me dado incentivos e condições para concluir meus estudos, incentivos estes de grande importância para que eu pudesse chegar a conclusão do curso de bacharelado em estatística.

Agradeço aos meus demais familiares por terem mostrado interesses em minha caminhada na graduação e nos estudos mais elementares, me dando incentivos e forças para concluir o curso.

Ao meu orientador Joab de Oliveira Lima pela grande dedicação à esta monografia que com tanta presteza colaborou neste trabalho, em que as inúmeras reuniões realizadas para construção deste trabalho foram de grande importância para enriquecer meus conhecimentos em estatística e como pessoa.

Aos professores da iniciação científica Ronei Marcos de Moraes e Neir Antunes Paes pelos conhecimentos transmitidos durante toda minha graduação, experiências estas que acrescentaram bastante nos meus conhecimentos.

Agradeço aos demais professores da graduação em estatística pelos ensinamentos transmitidos, em especial aos professores Ulisses Umbelino dos Anjos, Eufrásio de Andrade Lima Neto, Hemílio Fernandes Campos Coêlho, Andréa Vanessa Rocha, Antonio Marcos de Moraes, Marcelo Rodrigo Portela Ferreira, Renata Patrícia Lima Jerônymo.

Ao meu amigo de infância Rafhael Lopes Pires Fernandes pela grande amizade. Agradeço à ele pelo bom convívio e várias discussões na área de Ciência da Computação, discussões estas saudáveis e de extrema importância para aprofundar meus conhecimentos de programação.

A minha amiga Surama Marjouri Campos da Fonsêca Maia por me dar bastante incentivos e acreditar no meu potencial, incentivos estes que ajudaram a me dedicar cada vez mais aos meus estudos.

Aos meus grandes amigos de graduação Josemir Ramos de Almeida, Marcílio Regis da Silva pelos bons momentos durante o bacharelado em estatística, pelos estudos em

conjunto e pelas madrugadas em claro estudando para a graduação e realizando trabalhos de consultoria.

Agradeço aos demais amigos do curso pelo bom convívio, trocas de informações e boas conversas, Jefferson, Natália Rodrigues Guedes Gondim, Rodrigo Cabral da Silva, Sadraque Enéias de Figueiredo Lucena, Telmo Cristiano da Silva.

# *Sumário*

# *Lista de Figuras*

# *Lista de Tabelas*

# 1 Introdução

*"Todo grande progresso da ciência resultou de uma nova audácia da imaginação."*
John Dewey

## 1.1 Evolução da violência

A violência, em seus mais variados contornos, é um fenômeno histórico na constituição da sociedade brasileira. Desde a escravidão, primeiro com os índios depois com mão de obra africana, a colonização mercantilista, o coronelismo, as oligarquias antes e depois da independência, tudo isso somado a um Estado caracterizado pelo autoritarismo burocrático, contribuiu fortemente para o aumento da violência que atravessa a história do Brasil.

Diversos fatores colaboram para aumentar a violência, tais como a urbanização acelerada, que traz um grande fluxo de pessoas para as áreas urbanas e assim contribui para um crescimento desordenado das cidades. Colaboram também para o aumento da violência as fortes aspirações de consumo, em parte frustradas pelas dificuldades de inserção no mercado de trabalho.

As causas da violência são associadas, em parte, a problemas sociais como miséria, fome e desemprego. Mas nem todos os tipos de violência derivam das condições econômicas. Além disso, um Estado ineficiente e sem programas de políticas públicas de segurança, contribui para aumentar a sensação de injustiça e impunidade, que é, talvez, a principal causa da violência.

Os acidentes de transportes também expressam uma forma de violência pois pôem em risco as pessoas, o meio ambiente, em que dessa forma representa uma forte ameaça à sociedade. Tais acidentes representam custos elevados para o Estado que só podem ser diminuídos com a aplicação de medidas públicas eficientes. Conforme a frota cresce e envelhece e o trânsito se torna mais perigoso, a rede pública de saúde precisa investir mais

em equipamentos e profissionais para o atendimento das vítimas. Tais investimentos não se devem unicamente à este tipo de violência mas a todas as formas de violência presentes em nosso país.

A Paraíba não se encontra fora do contexto nacional. Devido ao crescimento dos municípios paraibanos e a acelerada urbanização de alguns deles, é possível perceber um aumento da violência nessas regiões. Segundo a Secretaria de Saúde do Estado da Paraíba houve um aumento significativo nas mais variadas causas de violência que se deu principalmente na Grande João Pessoa nos últimos 5 anos. Tais problemas podem ser contornados com a aplicação de medidas públicas que visam a redução dessas violências e que consequentemente refletem a diminuição da mortalidade por causas não naturais.

## 1.2 Uma visão geral do problema abordado

Estudos sobre dados de estatísticas vitais de óbitos de uma região é de extremo interesse para o planejamento de políticas públicas de saúde. Estudos dessa natureza possibilitam às Secretarias de gestão dessas regiões a tomem decisões com base em métodos quantitativos que darão um embasamento maior ao gestor visando a soluções de um problema.

Utilizando os dados do Sistema de Informação de Mortalidade - (SIM) da Secretaria de Saúde do Estado da Paraíba, foi possível entender os relacionamentos de alguns municípios da Paraíba com as causas externas de óbitos. Pode-se entender por causas externas as causas de óbitos por motivos não naturais como homicídios, acidentes de transportes, afogamento, dentre outras.

As causas externas estão na Classificação Internacional das Doenças na versão 10 (CID-10) no Capítulo XX. Através da base de dados gerada pela Secretaria Estadual foram obtidas as informações de óbitos de 2006 a 07 de julho de 2010.

Foram consideradas 8 causas específicas de óbitos, em que a principal causa de óbito é aquela que apresentou o maior número de óbitos entre 01/01/2006 a 07/09/2010 . Dessa forma, foi possível, através das 8 causas de óbitos mais relevantes no Estado da Paraíba, selecionar 30 municípios mais significativos, segundo as causas consideradas. As causas estudadas neste trabalho estão listadas abaixo, bem como os municípios considerados.

**Causas externas de óbito**

1. V03 - Pedestre traumatizado em colisão com automóvel.
2. V09 - Pedestre traumatizado em acidentes de transporte não especificado;
3. V49 - Ocupante de automóvel traumatizado em outro acidente de transporte e em acidentes de transportes não especificados;
4. V89 - Acidente com um veículo a motor ou não-motorizado, tipo(s) de veículos(s) não especificado(s);
5. V99 - Acidente de transporte não especificado;
6. W19 - Queda sem especificação;
7. X70 - Lesão provocada intencionalmente por enforcamento, estrangulamento e sufocação;
8. X95 - Agressão por disparo de arma de fogo ou arma não especificada;

**Municípios**

Tabela 1: Codificação dos 30 municípios considerados paraibanos.

| Municípios | Códigos | Municípios | Códigos |
|---|---|---|---|
| João Pessoa | JP | Catolé do Rocha | CR |
| Campina Grande | CG | Lagoa Seca | LS |
| Santa Rita | SR | Pedras de Fogo | PE |
| Bayeux | BA | Alhandra | AH |
| Patos | PA | Alagoa Grande | AG |
| Cabedelo | CA | Solânea | SL |
| Sapé | SA | Remígio | RE |
| Sousa | SO | Monteiro | MO |
| Queimadas | QE | Conde | CO |
| Mamanguape | MA | Cuité | CU |
| Guarabira | GA | Soledade | SD |
| São Bento | SB | Alagoa Nova | AN |
| Esperança | ES | Areia | AR |
| Caaporã | CP | Aroeira | AO |
| cajazeiras | CJ | Boqueirão | BO |

Apesar dos dados serem provenientes e de responsabilidade da Secretaria de Saúde Estadual, é importante perceber que essas informações são bastante diversificadas, po-

dendo responder sobre as políticas de seguranças dessas regiões. A causa $X95$, por exemplo, reflete o grau de criminalidade de um dado município. Já a causa $W19$ (causa de óbito de queda sem especificação) refere-se a uma variável pouco informativa, no sentido de que um município associado à esta causa não reflete muito as condições sócio econômicas dessa localidade. Contudo, esta variável foi considerada no estudo por ser de interesses de algumas pessoas para o monitoramento da qualidade dos dados, que não será tratado neste trabalho.

Para construção desses relacionamentos foram consideradas as variáveis sociais sexo, escolaridade, faixa etária, raça/cor. Todas essas variáveis, sabidamente, influenciam os inter-relacionamentos entre municípios e causas.

## 1.3 Justificativa

Tradicionalmente os métodos de agrupamentos que envolvem elementos espaciais se baseiam em medidas e ponderações espaciais, como é o caso do método de Varredura *Scan* para detecção de conglomerados no espaço. Por outro lado, quando o objetivo são os métodos de agrupamentos de dados (não espaciais), as principais obras da literatura estatística apontam, quase sempre, para os procedimentos de agrupamentos rígidos, ou seja, que obrigam que cada item (objeto ou pessoa) pertença a um, e somente um, grupo. Mesmo essas últimas técnicas apresentam algumas limitações, como por exemplo, a manipulação de dados de atributos. Portanto, percebe-se que não existe um método de agrupamento, digamos, uniformemente melhor, mais eficiente e mais versátil que se aplique a todas as situações práticas. Mas foram as limitações dos métodos citados que motivaram a construção de um método de agrupamento capaz de combinar as características de algoritmos de agrupamento espaciais com a possibilidade de utilização de um plantel de variáveis categóricas multidimensionais. O método não faz uso da distribuição geográfica dos elementos espaciais, mas, ainda assim, consegue fornecer respostas muita próximas da técnica de Varredura *Scan*, com o atrativo de se basear em conjuntos de dados multivariados com atributos. O método não faz uso da distribuição geográfica dos elementos espaciais, mas, ainda assim, consegue fornecer respostas muita próximas da técnica de varredura Scan, com o atrativo de se basear em conjuntos de dados multivariados com atributos.

## 1.4 Motivação

Os métodos de classificação de dados, sejam eles métodos hierárquicos ou não-hierárquicos, possuem uma filosofia de classificação rígida, conhecida na literatura como classificação *crisp*. Na contramão dessa idéia estão os métodos de aprendizado supervisionado, como é o caso dos métodos nebulosos *Fuzzy* que relaxam essa filosofia de rigidez, permitindo assim, que um indivíduo possa pertencer de um *cluster* ao mesmo tempo.

Tendo em vista que uma das metas do trabalho é agrupar os municípios paraibanos com as causas de óbitos estudadas, parece ser mais realista permitir que um mesmo município esteja relacionado com mais de uma causa ao mesmo tempo, de acordo com alguma medida de proximidade. A necessidade de se construir modelos de agrupamentos mais flexíveis, no sentido de necessitar não utilizar parâmetros espaciais foi o grande fator motivador desse trabalho. Um estudo poderia está interessado em agrupar municípios, por exemplo, utilizando dimensões, ou seja, agrupar tais localidades (dentro das causas externas) para certas faixas de PIB (dimensão 1) ou de IDH (dimensão 2), problema este que pode ser difícil tratar com as metodologias clássias de agrupamento de dados.

## 1.5 Objetivo do trabalho

### 1.5.1 Objetivo Geral

O objetivo principal desse trabalho é desenvolver um modelo de agrupamento híbrido que utilize as técnicas da Lógica Nebulosa *Fuzzy* e de Algoritmo Genético para agrupar os municípios paraibanos dentro das principais causas externas de óbitos para os anos de 2006 a 2010.

### 1.5.2 Objetivos específicos

- Avaliar a evolução das inter-relações entre municípios e causas externas no período de 2006 a 2010;
- Estudar as características analíticas do processo de simulação/otimização;

## 1.6 Estrutura da monografia

A monografia apresenta dividida em oito curtos capítulos. Tal divisão foi feita visando um melhor entendimento das informações transmitidas por este trabalho, visando um melhor entendimento dos conteúdos aqui apresentados.

Este primeiro capítulo apresenta uma visão superficial do problema que irá ser abordado, apresentando os objetivos a serem alcançados bem como as justificativas da importância desse problema. O Capítulo 2 apresenta uma revisão da literatura em que serão listados os principais trabalhos referentes ao tema abordado, trabalhos estes nas áreas de teoria *Fuzzy* e Algoritmos Genéticos. Também será citado nesse capítulo alguns autores que trabalharam sobre um método misto utilizando algoritmo *Fuzzy Genético*. No Capítulo 3 serão abordados alguns conceitos de Análise Multivarida. Este capítulo não tem a pretensão de se aprofundar nesses conceitos, no entanto, serve como um embasamento para o entendimento dos demais capítulos. O Capítulo 4 apresenta uma introdução à lógica nebulosa também conhecida como lógica *Fuzzy* ou lógica difusa. Será feita uma comparação da lógica *Fuzzy* com a lógica booleana o que ajudara no entendimento das diferenças entre as lógicas. Após a introdução dos conhecimentos sobre lógica *Fuzzy* e entender as diferenças entre a lógica clássica para a lógica difusa. Será apresentado o método de agrupamento *Fuzzy C-Means*. No Capítulo 5 serão introduzidos os conceitos de algoritmos evolutivos em especial os algoritmos genéticos, em que serão apresentados alguns exemplos que facilitaram o entendimento da teoria evolucionária. No Capítulo 6 será apresentado uma proposta de um modelo *Fuzzy C-Means Genético* para agrupamento de dados. Neste Capítulo também será apresentado a descrição do problema a ser otimizado, a função objetivo a ser otimizada, bem como o passo a passo do algoritmo *Fuzzy C-Means Genético*. Já no O Capítulo 7 apresentará os resultados alcançados com o método proposto entre os anos de 2006 a 2010. Por fim, no Capítulo 8 serão apresentados as conclusões e perspectivas de trabalho futuros.

# 2 *Revisão da Literatura e Referencial Teórico*

Os primeiros trabalhos sobre teoria *Fuzzy* ocorreram nos na segunda metade dos anos 60, em que em 1965 Lotfali Askar Zadeh matemático do Azerbaijão nascido em 1921 propôs a lógica difusa, mais conhecida como lógica *Fuzzy*. O primeiro trabalho sobre a lógica *Fuzzy*, *Fuzzy Sets*, foi publicado no volume 8 da revista *Information and control* nas páginas 338 a 353 no ano de 1965 (43). Trata-se de um artigo que formaliza a teoria nebulosa em que há toda uma formalização matemática dos conceitos da teoria *Fuzzy*, em que são definidas a álgebra do conjunto *Fuzzy*. Ainda em 1965 Lotfali A. Zadeh publicou um outro artigo entitulado Fuzzy sets and systems no Jornal Fox de 10 páginas (29-39) (44). Até os dias de hoje Zadeh continua publicando trabalhos sobre teoria *Fuzzy*, trabalhos estes teóricos e aplicados. Atualmente Zadeh está com 89 anos e atua no Departamento de Computação da Universidade da Califórnia, Berkley.

Usando os conceitos introduzidos por Zaheh há vários artigos aplicados utilizando lógica difusa, como o artigo Publicado por Garcia e Filho com o título Desempenho energético de um sistema de refrigeração aplicando o controle adaptativo *Fuzzy* em que consideram um sistema de ar condicionado que minimizasse o custo de energia (13). Outros trabalhos que utilizam lógica *Fuzzy* são: HARRIS C. J., MOORE C. G., BROWN M., Intelligent control: Aspects of *Fuzzy* Logic and Neural Nets, World Scientific, 1993 (4); KOSKO, BART, Neural networks and *Fuzzy* systems: a dynamical systems approach to machine intelligence, Prentice-Hall International, 1992 (27), COX E., The *Fuzzy* Systems Handbook: a Practitioner's Guide to Building, Using and Maintaining *Fuzzy* Systems, Professional, 1994 (6); PEDRYCZ W., GOMIDE F., *Fuzzy* Systems Engineering: Toward Human-Centric Computing, Wiley/IEEE Press, 2007 (35).

O método de otimização conhecido como “Algoritmo Genético" foi criado por John Henry Holland. Existem na literatura várias aplicações da teoria de algoritmos genéticos, dentre esses trabalhos pode-se citar: GOLDBERG D. E. Genetic Algorithms in Search,

Optimization, and Machine Learning. EUA: Addison-Wesley, 1989, página 80 (16), em que discute formas de representação do espaço de busca, GOLDBERG D. E. Genetic Algorithms in Search, Optimization, and Machine Learning, EUA: Addison-Wesley, 1989, p. 121, que discute e compara diversas formas de seleção de indivíduos (14), GOLDBERG D. E. Genetic Algorithms in Search, Optimization, and Machine Learning. EUA: Addison-Wesley, 1989, p. 147 que discute operações que podem ser aplicadas nos indivíduos para a reprodução, dentre outros (15).

Há vários artigos sobre algoritmos evolutivos e em especial sobre os algoritmos genéticos em língua portuguesa, dentre esses artigos podemos citar: Algoritmos Genéticos: Aplicação à Robótica, dissertação defendida por Pires em 1998 na Faculdade de Engenharia da Universidade do Porto (36), GOLDBARG C. Algoritmos evolucionários na determinação da configuração de custo mínimo de sistemas de co-geração de energia com base no gás natural, Pesqui. Oper. 25(2): pp. 231-259, 2005 (17); LEITE. Aplicação de algoritmos genéticos na determinação da operação ótima de sistemas hidrotérmicos de potência, Sba Controle & Automação 17(1): pp. 81-88, 2006 (28).

Ainda assim na literatura há poucos artigos que combinam algoritmos genéticos com a teoria *Fuzzy*, dentre eles podemos citar KLAWONN F., KELLER A. *Fuzzy* Clustering with Evolutionary Algorithm, internacional of Intelligent Systems, volume 13,1998 (25). Este artigo apresenta o método uma abordagem da teoria *Fuzzy C-Means* para identificação de *clusters*. Uma proposta bem parecida com o técnica utilizada nessa monografia. Contundo, o artigo propõe um método em que os centróides não são definidos a priori. No nosso caso os centróides serão pre-estabelecidos antes mesmo da execução do método de agrupamento.

O autor também propõe uma abordagem para a construção de agrupamentos usando algoritmos genéticos, contudo segundo os autores essa abordabem não é promissora devido a característica do espaço paramétrico em que a função objetivo varia ser bastante complicado. Pode-se citar também LIU H. C., JEAG B. C.,YU Y. K. *Fuzzy* C-Means Algorithm Based on Standard Mahalanobis Distances, International Symposium on Information Processing, China, 2009, pp. 422-427, em que os autores trazem algumas modificações do algoritmos *Fuzzy C-Means* (29). Os autores propuseram uma metodologia de agrupamento *Fuzzy* utilizando a distância de Mahalanobis, ao invés das distâncias euclidianas. Um outro trabalho relevante é o trabalho publicado por NUOVO A. G., CATANIA V., PALESI M. The Hybrid Genetic *Fuzzy* C-Means: a Reasoned Implementation, International Conference on *Fuzzy* Systems, Cavtat, Croatia, 2006, pp 33-38 (32). Neste artigo

é apresentada uma abordagem híbrida que integram *Fuzzy C-Means* (FCM) e algoritmos Algoritmos Genéticos (AGs) para projetar um classificador ótimo para o problema de classificação específica. Essa integração permite a geração automática de um sistema de classificação, com um subconjunto de funcionalidades otimizadas, a partir de um banco de dados de exemplos. Segundo os autores o classificador gerado supera o algoritmo FCM clássico.

# 3 *Análise Multivariada de Dados*

"*A ciência se compõe de erros que, por sua vez, são os passos até a verdade*"
Jules Verne

## 3.1 Introdução

A estatística multivariada lida com problemas em que se observa várias medidas sobre um mesmo item (pessoa ou objeto). Na verdade, as principais principais idéias da Análise Multivariada surgiram a partir da generalização de técnicas univariadas. Segundo Furtado(2008) a necessidade de compreensão das relações entre as diversas variáveis, de maneira conjunta faz com que as análises multivariadas sejam complexas ou até mesmo difíceis. E, talvez, por isso, sejam, ainda, tão pouco utilizadas como ferramenta estatística. Em geral, os objetivos para os quais a análise multivariada se presta são:

- Compreensão analítica conjunta das inter-relações existentes entre as diversas variáveis-resposta;
- Redução da massa de dados conservando o máximo de informação original possível;
- Agrupamento de ítens similares baseados em dados amostrais ou experimentais;
- Investigação de inter-dependências entre as variáveis objetos de estudo, buscando, para isso, identificar relações estruturais;
- Predição com base no relacionamento das variáveis consideradas.

A análise multivariada de dados vem sendo amplamente utilizada por várias áreas do conhecimento. Em estudos da área médica, por exemplo, técnicas multivariadas podem ser implementadas para criar uma função que separe pessoas doentes e não doentes considerando algumas característica. Em estudos de pesquisa de mercado, muitas empresas

tentam traçar o perfil sócio-econômico de seus clientes em busca de identificar grupos diferenciados de consumo.

Os dados multivariados provêm de estudos em que são selecionadas $p \geq 1$ variáveis ou características para serem mensuradas. Entretanto, alguns pesquisadores definem análise multivariada como sendo exames de relações entre mais de duas variáveis. A representação das informações multivariadas é dada por $x_{mn}$ que indica o valor da $m$-ésima unidade amostral ou experimental de uma $n$-ésima variável mensurada.

Dessa forma as medidas das $p$ variáveis em $n$ unidades amostrais ou experimentais podem ser representadas pela matriz $\underset{\sim}{X}$ logo abaixo.

$$\underset{\sim}{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{1j} & \cdots & x_{1n} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mj} & \cdots & x_{mn} \end{bmatrix}$$

Outros autores preferem definir a análise multivariada como uma técnica adequada estudar problemas nos quais todas as múltiplas variáveis são assumidas como tendo uma distribuição normal multivariada, como apresentada pela equação 3.1.

$$f(x) = \frac{1}{(2\pi)^{\frac{p}{2}} |\underset{\sim}{\Sigma}|^{\frac{1}{2}}} \cdot e^{[-\frac{1}{2}(\underset{\sim}{X}-\underset{\sim}{\mu})^t \underset{\sim}{\Sigma}^{-1} (\underset{\sim}{X}-\underset{\sim}{\mu})]}, \tag{3.1}$$

onde $(\underset{\sim}{X} - \underset{\sim}{\mu})^t \underset{\sim}{\Sigma}^{-1} (\underset{\sim}{X} - \underset{\sim}{\mu})$ representa uma medida de similaridade ponderada pela covariância entre os elementos e será discutida rapidamente na Seção 3.2. Os valores $\underset{\sim}{\mu}$ e $\underset{\sim}{\Sigma}$ são o vetor de média e matriz de covariância populacional respectivamente. Os estimadores do vetor de média e matriz de covariância populacional serão representados da forma que segue (11).

$$\overline{\underset{\sim}{X}} = \begin{bmatrix} \overline{X}_1 \\ \overline{X}_2 \\ \vdots \\ \overline{X}_p \end{bmatrix}$$

$$S = \begin{bmatrix} S_{11} & S_{12} & \dots & S_{1p} \\ S_{21} & S_{22} & \dots & S_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ S_{p1} & S_{p2} & \dots & S_{pp} \end{bmatrix}$$

## 3.2 Medidas de similaridade

A maioria dos esforços dispendidos na produção de uma estrutura grupal simples, a partir de um conjunto de dados complexo, requer medidas de "proximidade" ou "similaridade". O que se sabe é que existe sempre um elevado grau de subjetividade no que tange à escolha de uma medida de similaridade. Considerações importantes como a natureza das variáveis (discreta, contínua, binária), as escalas de medida (nominal, ordinal, intervalo) e o conhecimento específico do assunto em questão devem ser considerados.

Quando ítens (unidades, casos ou indivíduos) são agrupados, sua proximidade é indicada por algum tipo de distância. Por outro lado, as variáveis são agrupadas baseadas no seu coeficiente de correlação ou outras medidas estatísticas de associação. Nas Seções 3.2.1 e 3.2.2 serão discutidos os conceitos das principais medidas de similaridades(22).

### 3.2.1 Covariância e Correlação

A partir da matriz de dados $\underset{\sim}{X}(m \times n)$ a matriz de covariância $C$, sendo esta uma medida de associação pode ser obtida pela expressão abaixo:

$$c_{jk} = \frac{1}{m-1} \sum_{i=1}^{m} (x_{jk} - \overline{x}_j)(x_{jk} - \overline{x}_j) \tag{3.2}$$

Através da matriz de covariância pode-se calcular a matriz de correlação $R$, em que cada elemento da matriz é dado por $r_{jk} = \frac{c_{jk}}{s_j s_k}$, em que $s_j$ $s_k$ são os desvios padrões das variáveis $j$ e $k$ (22).

### 3.2.2 Medidas de distâncias

Nos métodos de agrupamentos, a similaridade entre amostras são expressas, em geral, por uma função de distância entre dois pontos representativos da amostra no espaço n-dimensional. A forma mais usual para o cálculo das distância entre dois pontos é conhecida como distância Euclidiana, definida como segue para dois pontos $a$ e $b$ no

espaço n-dimensional.

$$X^2{}_{ab} = \sum_{j=1}^{n} (d_{aj} - d_{bj})^2 \tag{3.3}$$

Existem outras medidas de distância comumente utilizadas em diversos métodos de agrupamentos,entre elas estão as distâncias de Mahalanobis e de Manhattan.

**Distância Euclidiana**

A distância Euclidiana ou distância métrica é a distância entre dois pontos que pode ser demonstrada pela aplicação repetida do teorema de Pitágoras. Ao aplicar o teorema de Pitágoras como distância, o espaço euclidiano torna-se um espaço métrico. A distância Euclidiana, em sua forma mais geral, é definida como:

$$\sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2}$$

**Distância de Mahalanobis**

A distância de Mahalanobis é uma medida alternativa que agrega a informação das covariâncias entre as variáveis ao seu cálculo. O cálculo dessa distância é bastante simples e pode ser obtido pela expressão abaixo.

$$d(\underset{\sim}{X}_1, \underset{\sim}{X}_2) = \sqrt{(\underset{\sim}{X}_1 - \underset{\sim}{X}_2)^t S^{-1} (\underset{\sim}{X}_1 - \underset{\sim}{X}_2)},$$

em que $S^{-1}$ é a inversa da matriz se variância e covariância amostral.

## 3.3 Análise de Correspondência

A Análise de Correspondência (AC) é uma metodologia multivariada baseada numa abordagem composicional e que é normalmente utilizada para mapear as percepções observadas entre as categorias de uma tabela de contingência.

A expressão Mapa Perceptual designa uma técnica de *marketing* importada do campo da psicologia e que parte do princípio de que os consumidores constroem uma imagem do produto com base em características específicas, nos seus benefícios ou no seu preço. Dessa

forma, é possível posicionar os produtos em um mapa ou gráfico em que cruzam-se duas variáveis. Em outras palavras, um mapa perceptual reflete uma espécie de representação visual de percepções que um respondente tem sobre objetos em duas ou mais dimensões. Cada objeto tem uma posição especial no mapa perceptual que está associado a uma similaridade ou preferência relativa a outros objetos no que se refere às dimensões do mapa perceptual. A maioria das aplicações da AC envolve um conjunto de objetos e atributos em que os resultados são configurados em um *mapa perceptual* comum.

As primeiras considerações matemáticas a respeito da Análise de Correspondência foram feitas por Hirshfeld em 1935. Alguns autores definem Análise de Correspondência como um método de análise fatorial para variáveis categóricas. A AC foi primeiramente utilizada, por Fisher (1940) para a análise de tabelas de contingência posteriormente foi redescoberta na França por Benzecri em 1969. Atualmente a técnica Análise de Correspondência tem sido extremamente empregada em diversas áreas do conhecimento, com uma ênfase maior para as áreas de Psicologia, Psiquiatria, Neurociências e *Datamining*.

em que tem sido extremamente usado naquele país como um método estatístico no processo de *Data Mining*. A partir de 1975, a técnica de Análise de Correspondência vem sendo utilizada em diversas áreas do conhecimento, em publicações em diversos idiomas (20).

Uma aplicação comum da metodologia de Análise de Correspondência tem sido na redução dimensionalidade dos dados, obtidas através do mapeamento perceptual das relações de inter-dependência de informações amostrais que, na maioria das vezes, representam dados não-métricos. Tal método também é conhecido como escalonamento ou escore ótimo, média recíproca ou análise de homogeneidade. A proximidade entre os objetos no gráfico indica o nível de associação entre tais objetos no âmbito amostral.

O método AC difere de outras técnicas de escalonamento multidimensional em sua habilidade de acomodar tanto dados não-métricos quanto relações não lineares. As três principais características que diferem a Análise de Correspondência dos outros métodos de escalonamento multidimensional são:

- É uma técnica composicional, e não decomposicional, porque o mapa perceptual é baseado na associação entre objetos e um conjunto de características descritivas ou atributos especificados pelo pesquisador.

- Sua aplicação mais direta é na retratação da correspondência de categorias de variáveis, particularmente àquelas medidas em escalas nominais. Tal correspondência

é, desse modo, a base para o desenvolvimento de mapas perceptuais.

- Os principais benefícios da AC residem em sua habilidade para representar linhas e colunas em um espaço conjunto.

Na AC, uma decomposição dos dados é obtida para se estudar a estrutura dos mesmos sem que um modelo seja hipotetizado ou que uma distribuição de probabilidade tenha sido assumida, o que é uma grande vantagem. Outro aspecto, discutido por Von der Heizden et all (1989), é que a análise de correspondência, geralmente, é introduzida sem qualquer tratamento estatístico prévio, para dados categóricos, o que prova sua utilidade e flexibilidade. Em consequência disso, os testes estatísticos inferênciais clássicos, em sua maioria, não são aplicáveis, estando a solução sugerida apenas pela distribuição gráfica de seus resultados (1).

É também, um meio de criar configurações representando as linhas da tabela por pontos no espaço, tal que a distância Euclidiana entre os pontos na configuração seja igual a distância qui-quadrado calculadas entre as linhas da tabela.

O único pressuposto a ser respeitado é que os dados sejam uma matriz retangular de entradas não negativas. O tipo mais comum de matriz de entrada é uma tabela contingência com categorias específicas definindo as linhas e colunas. Para o caso de Análise de Correspondência Simples temos uma tabela de dupla entrada sendo essa a forma mais simples de AC. A Tabela 2 abaixo apresenta a estrutura do banco de dados em uma Análise de Correspondência Simples.

Tabela 2: Estrutura dos dados para Análise de Correspondência.

| | B | | | | | | |
|---|---|---|---|---|---|---|---|
| A | 1 | 2 | ... | j | ... | J | Total Linha |
| 1 | $n_{11}$ | $n_{12}$ | $\cdots$ | $n_{1j}$ | $\cdots$ | $n_{1J}$ | $n_{1+}$ |
| 2 | $n_{21}$ | $n_{22}$ | $\cdots$ | $n_{2j}$ | $\cdots$ | $n_{2J}$ | $n_{2+}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| i | $n_{i1}$ | $n_{i2}$ | $\cdots$ | $n_{ij}$ | $\cdots$ | $n_{iJ}$ | $n_{i+}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| I | $n_{I1}$ | $n_{I2}$ | $\cdots$ | $n_{Ij}$ | $\cdots$ | $n_{IJ}$ | $n_{I+}$ |
| Total Coluna | $n_{+1}$ | $n_{+2}$ | $\cdots$ | $n_{+j}$ | $\cdots$ | $n_{+J}$ | N |

onde temos que:

- $n_{ij}$ é a frequência observada pela intersecção da $i$-ésima categoria da variável A com a $j$-ésima categoria da variável B;

- $n_{i+}$ é a frequência total observada na $i$-ésima categoria de A;
- $n_{+j}$ é a frequência total observada na $j$-ésima categoria de B;
- $n$ é o total geral de frequências observadas.

Seja $N$ a matriz de frequências absolutas, ou seja, temos que $N = [n_{ij}]_{IxJ}$ gerada a partir da Tabela 2. A matriz de frequências relativas será $P = \frac{1}{n}N$ e é chamada de *matriz de correspondência*, em que cada linha ou coluna da matriz retangular $P$ pode ser considerada um vetor de proporções.

Tabela 3: Tabela de Correspondência.

| | B | | | | | | |
|---|---|---|---|---|---|---|---|
| A | 1 | 2 | ... | j | ... | J | Total Linha |
| 1 | $p_{11}$ | $p_{12}$ | $\cdots$ | $p_{1j}$ | $\cdots$ | $p_{1J}$ | $p_{1+}$ |
| 2 | $p_{21}$ | $p_{22}$ | $\cdots$ | $p_{2j}$ | $\cdots$ | $p_{2J}$ | $p_{2+}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| i | $p_{i1}$ | $p_{i2}$ | $\cdots$ | $p_{ij}$ | $\cdots$ | $p_{iJ}$ | $p_{i+}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| I | $p_{I1}$ | $p_{I2}$ | $\cdots$ | $p_{Ij}$ | $\cdots$ | $p_{IJ}$ | $p_{I+}$ |
| Total Coluna | $p_{+1}$ | $p_{+2}$ | $\cdots$ | $p_{+j}$ | $\cdots$ | $p_{+J}$ | 1 |

em que $p_{ij} = \frac{n_{ij}}{n}$, $p_{i+} = \frac{n_{i+}}{n}$ e $p_{+j} = \frac{n_{+j}}{n}$.

Da Tabela 3 os vetores de frequências relativas marginais são denominado *massas* e são calculado em relação ao total geral $n$. Dessa forma, temos que a massa da $i$-ésima linha é $\frac{n_{i+}}{n}$ e a massa da $j$-ésima coluna é definido por $\frac{n_{j+}}{n}$. Com as massas calculadas para linhas e colunas podemos definir os vetores de massas $r$ e $c$ para linha e coluna respectivamente, em que $r = [p_{1+}, p_{2+}, \cdots, p_{i+}, \cdots, p_{I+}]$ e $c = [p_{+1}, p_{+2}, \cdots, p_{+j}, \cdots, p_{+J}]$.

O vetor $a_i = [\frac{n_{i1}}{n_{i+}}, \frac{n_{i2}}{n_i}, \cdots, \frac{n_{iJ}}{n_{i+}}]^t$ é definido como *perfil de linha*. Em função da matriz de correspondência $P$, o $i$-ésimo perfil linha será $a_i = [\frac{p_{i1}}{p_{i+}}, \frac{p_{i2}}{p_i}, \cdots, \frac{p_{iJ}}{p_{i+}}]^t$, $\forall i = 1, 2, \cdots, I$. Cada vetor perfil $a_i$ representa uma realização da distribuição multinomial condicionada à $i$-ésima categoria da variável A.

De forma análoga também é definido um vetor $b_j = [\frac{n_{1j}}{n_{+j}}, \frac{n_{2j}}{n_{+j}}, \cdots, \frac{n_{Ij}}{n_{+j}}]^t = [\frac{p_{1j}}{p_{+j}}, \frac{p_{2j}}{p_{+j}}, \cdots, \frac{p_{Ij}}{p_{+j}}]^t$, $\forall j = 1, 2, \cdots, J$, chamado de *perfil coluna*. Com os valores $a_i$ são definidas as distâncias entre perfis linha utilizando a métrica Euclidiana ponderada, também conhecida como distância qui-quadrado, definida abaixo:

$$d_c(a_i, {a_i}^t) = (a_i, {a_i}^t){D_c}^{-1}(a_i, a_i) = \sum_{j=1}^{J} \frac{(\frac{n_{ij}}{n_{i+}} - \frac{n_{i^t j}}{n_{i^t +}})^2}{\frac{n_{+j}}{n}}, \tag{3.4}$$

ou seja, $d_c(a_i, {a_i}^t)$ trata-se da distância qui-quadrado entre os vetores $a_i$ e ${a_i}^t$ ponderada por $D_c$ sendo esta uma matriz diagonal de elementos $c_j = \frac{n_{+j}}{n}\ \forall j = 1, 2, \cdots, J$. O vetor formado pelas massas de coluna (proporções marginais de colunas) $c = [c_1, c_2, \cdots, c_J]^t$ é também chamado de perfil linha médio ou *centróide dos perfis linha*. De forma análoga, as distâncias qui-quadrado entre perfis colunas são obtidas na métrica $D_r$ que é uma matriz diagonal das proporções marginais de linha $r_i = \frac{n_{i+}}{n} \forall i = 1, 2, \cdots, I$, que compõe o vetor $r$ denominado *centróides dos perfis coluna ou vetor massa de linhas*(18).

## 3.4 Análise de Agrupamentos

A Análise de Agrupamento também conhecida como *cluster Analysis* constitui um conjunto de técnicas estatísticas voltadas para o agrupamento de observações (ou de variáveis), de modo que as observações contidas em um mesmo grupo sejam similares entre si (em algum sentido) e as observações pertencentes a grupos distintos sejam dissimilares entre si. A Análise de Agrupamento, nesse sentido, representa uma técnica de aprendizagem não supervisionada, ou seja, o próprio procedimento que gera os grupos aprende e decide sozinho (daí o termo não supervisionado) para quais conglomerados receberão as diversas observações. Por esse motivo, essa metodologia vem sendo aplicada nos mais variados campos da ciência, como nas áreas de mineração de dados, de aprendizagem de máquina, de reconhecimento de padrões, de análise de imagens, na bioinformática, dentre outras.

Na literatura, algoritmos são utilizados para formar os agrupamentos, cada um deles se vale dos mais variados critérios para construir os grupos (11). Ainda assim, os métodos de agrupamento, em sua essência, buscam agrupar uma massa de dados em classes de elementos similares seguindo regras específicas para tais classificações (3). Além disso, por ser uma técnica de aprendizagem não supervisionada a Análise de Agrupamentos não considera um número de grupos fixo e é realizada com base na similaridade ou dissimilaridade entre as observações. No geral, os métodos de agrupamento de dados utilizam como medidas de similaridades as mais variadas métricas de distâncias entre pontos, como distâncias Euclidianas, Mahalanobis, Manhattan, dentre outras (21).

Dentre os principais métodos de agrupamento de dados, se pode citar a classe dos métodos hierárquicos e não-hierárquicos. Os métodos de agrupamentos hierárquicos cria uma hierarquia de agrupamentos que podem ser representadas em uma estrutura de árvore, chamada de Dendograma. As raízes contém um único agrupamento com todas as

observações e as folhas correspondem às observações individuais. Os agrupamentos hierárquicos são realizados por sucessivas fusões ou sucessivas divisões. Dentro desse sistema de hierarquias, os procedimentos de agrupamentos aglomerativos iniciam-se com tantos grupos quanto forem os objetos, ou seja, cada objeto forma um grupo. Dessa forma, os objetos mais similares são agrupados e fundidos sequencialmente até formar um grupo único contendo todos os objetos. Já os métodos hierárquicos divisivos trabalham no sentido inverso, ou seja, o processo inicia com um único grupo contendo todos os objetos e, sucessivamente, esse grupo vai sendo subdividido em grupos cada vez menores até que haja tantos grupos quantos forem o número de objetos existentes na amostra. Por outro lado os métodos não hierárquicos têm o objetivo de particionar os $n$ objetos considerados em $k$ grupos distintos. Tais métodos necessitam de uma pré-fixação de critérios que produzam medidas sobre a qualidade das partições que serão originadas. Dessa forma, os métodos não hierárquicos produzem partições num número fixo de classes, ou seja, há que se escolher o número de agrupamentos à priori. Um dos métodos não hierárquicos mais utilizados é o método das *k-médias* (22).

Os métodos clássicos de agrupamento de dados são bastante rígidos na construção dos agrupamentos, ou seja, se são identificados $n$ ítens, cada um desses ítens deverá pertencer a um, e somente um, dos $k$ grupos preestabelecidos. Para situações práticas em que a noção de rigidez precisa ser relaxada, isto é, que se permita a um determinado item pertencer a um ou mais grupos ao mesmo tempo, os métodos que utilizam a lógica nebulosa *Fuzzy*são de especial interesse. Na prática, a idéia da lógica nebulosa *Fuzzy* consiste na possibilidade de permitir que, um anfíbio, por exemplo, seja classificado com um animal aquático e terrestre ao mesmo tempo, apresentando as características semelhantes à ambos os grupos.

em que dessa forma não perdemos a noção de intersecção que poderia ofuscar de alguma forma a realidade. Para problemas em que a noção dessas intersecção são necessárias, os métodos *Fuzzy* de agrupamentos de dados são interessantes para solução desses problemas. Na prática, a separação dos agrupamentos é uma noção *Fuzzy* (nebulosa), por exemplo, se um anfíbio é um animal que tem características de animais aquáticos e terrestres, portanto, vai apresentar características semelhantes à ambos os grupos. O conceito de conjunto *Fuzzy* proporciona a vantagem de expressar esse tipo de situação em que um indivíduo compartilha similaridade com vários grupos através da possibilidade de um algoritmo associar cada indivíduo parcialmente a todos os grupos (29)(12).Essa metodologia será discutida com mais detalhes na seção que segue.

# 4 *Introdução à Teoria Fuzzy*

## 4.1 Teoria Fuzzy

### 4.1.1 Histórico e conceitos iniciais

Na década de 60, um professor de engenharia elétrica e ciências da computação, chamado Lotfi Zadeh (1965) desenvolveu uma variação da tradicional teoria dos conjuntos e lógica booleana para tornar a análise e controle de sistemas complexos de controle mais tratáveis. Ele observou que as regras que as pessoas usavam para fazer inferências não eram consistentes, ou seja, não podiam ser explicadas pelas pessoas que as usavam. Por exemplo, podemos olhar uma pessoa e dizer “ela parece ter por volta de 40 anos de idade” mas não se está preparado para explicar como sabemos disso. A idéia de Zadeh o levou a desenvolver o que é conhecido como lógica *Fuzzy*. A pesar de ter sido criticada inicialmente, a lógica *Fuzzy* acabou sendo bem aceita por engenheiros e cientistas da computação, tornando-se comuns as suas aplicações.

Desde a publicação do conceito de lógica *Fuzzy* em 1965, houveram muitas aplicações do conceito de lógica nebulosa como em 1980, no controle *Fuzzy* de operação de um forno de cimento. Em seguida vieram várias outras aplicações como o controle *Fuzzy* de plantas nucleares, refinarias, processos biológicos e químicos, trocador de calor, máquina diesel, tratamento de água e sistemas de operação automática de trens.

Do começo da ciência moderna até o fim do século XX as incertezas eram, em geral, indesejáveis em qualquer estudo ou experimento, sendo, portanto, sempre evitada. Com o passar dos anos, essa atitude foi gradativamente mudando com o desenvolvimento da estatística e dos mecanismos estatísticos. Para lidar com grandes complexidades mecânicos, no nível molecular, diversos mecanismos estatísticos eram utilizados com sucesso em várias áreas da ciência e, essencialmente, empregavam o cálculo de médias e teoria de probabilidade. Contudo, a teoria probabilística não era capaz de agregar altos níveis de “certezas" em todas suas manifestações (8). Em particular a teoria probabilística é inca-

paz de tratar a incerteza resultante dos termos da linguagem natural vagos. [Mukaidono, 2001] deu um exemplo disso com a palavra "meia-idade". É comum classificar uma pessoa de meia idade ou não, apesar de não se saber exatamente quando se começa ou termina o período considerado "meia-idade". Considere, a título de exemplo, que o período de meia idade compreende o intervalo fechado $I = [35, 55]$ anos. Usando a lógica tradicional, uma pessoa de 34 anos de idade só pode pertencer ao intervalo $I$, ou seja, ao grupo de pessoa de meia-idade depois do dia de seu aniversário de 35 anos. De mesmo modo uma pessoa de 56 anos não pertence mais à esse grupo. Contudo, tamanha precisão não é desejada em relação a este conceito, dados que os limites do intervalo $I$ não podem ser definidos precisamente. Na verdade a única idéia que se tem é uma idéia vaga a respeito dos limites do intervalo de meia-idade (40).

Pelo exemplo de incerteza do conceito de meia idade foi possível perceber que a lógica tradicional impõe limites bruscos ao contrário da lógica *Fuzzy* que não impõe limites tão ríspidos, proporcionando graus de pertinências de elementos a uma determinada categoria. Dessa forma, a lógica *Fuzzy* é capaz de capturar informações vagas, em geral descrita em linguagem natural e convertê-las para um formato numérico e de fácil manipulação.

### 4.1.2 Comparação entre Lógica Fuzzy, Lógica Booleana e Probabilidade

O filósofo grego Aristóteles, aluno de Platão e professor de Alexandre o Grande (384 - 322 a.C.), foi o fundador da ciência da lógica onde estabeleceu um conjunto de regras rígidas para que conclusões pudessem ser tomadas utilizando de tais regras. O emprego da lógica de Aristóteles leva a uma linha de raciocínio lógico baseado em premissas e conclusões binárias, ou seja, uma conclusão ou é verdadeira ou falsa. A lógica nebulosa permite uma transição gradual de uma proposição dentre os conjuntos a que esta proposição pode pertencer através da associação de graus de pertinências desta aos conjuntos analisados. Dessa forma a lógica *Fuzzy* permite a dualidade estabelecendo que algo pode coexistir com seu oposto (44).

A teoria probabilista e lógica nebulosa podem ser utilizadas para mensurar incertezas. O que diferencia a probabilidade da lógica *Fuzzy* é dizendo que a teoria probabilista lida com a *expectativa* de eventos futuros, baseados em fatores conhecidos. O senso de incerteza se refere à predição de ocorrência de um evento. O senso de incerteza representado pela lógica *Fuzzy* se resulta da imprecisão do significado de um conceito expresso em lógica natural, em que dessa forma estamos comparando um indivíduo e um dado conceito

impreciso. (8)

### 4.1.3 Componente da Teoria dos Conjuntos Fuzzy

#### 4.1.3.1 Variáveis linguísticas

Uma grande vantagem do uso da lógica *Fuzzy* é a possibilidade de transformar a linguagem natural em conjuntos numéricos, permitindo dessa forma a manipulação computacional. Zedeh definiu variáveis linguísticas como "variáveis nas quais os valores são palavras ou sentenças em linguagem natural ou artificial". As variáveis linguísticas assumem valores denominados linguísticos, como por exemplo baixo, médio, alto referente a variável altura.

$$I_1(x) = \begin{cases} 1 & \text{para} \quad x \leq 20 \\ \frac{35-x}{15} & \text{para} \quad 20 < x < 35 \\ 0 & \text{para} \quad x \geq 35 \end{cases}$$

$$I_2(x) = \begin{cases} 0 & \text{para} \quad x \leq 20 \quad \text{ou} \quad x \geq 60 \\ \frac{x-20}{15} & \text{para} \quad 20 < x < 35 \\ \frac{60-x}{15} & \text{para} \quad 45 < x < 60 \\ 1 & \text{para} \quad 35 \leq x \leq 45 \end{cases}$$

$$I_3(x) = \begin{cases} 0 & \text{para} \quad x \leq 45 \\ \frac{x-45}{15} & \text{para} \quad 45 < x < 60 \\ 1 & \text{para} \quad x \geq 60 \end{cases}$$

As funções apresentadas acima podem ser observada na Figura 1. Podemos observar que a função em questão apresenta uma forma triangular, no entanto a função de pertinência poderia assumir outras estruturas como Linear, Trapezoidal, Formato S, Formato Z, Gaussiana, Irregular dentre outras, sendo essas formas padrões definidas na literatura que as função de pertinências podem assumir, podendo também ser definida pelo usuário.

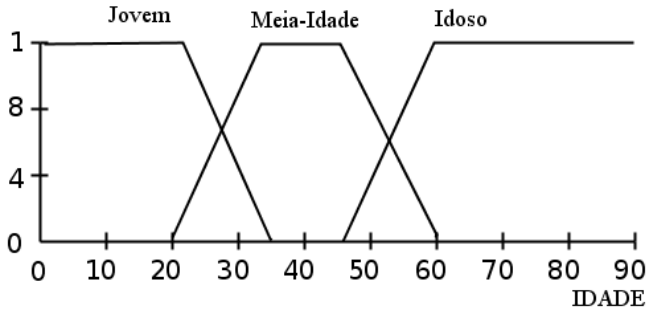



Figura 1: Gráfico da função de pertinência para a variável linguística meia-idade.

#### 4.1.3.2 Funções de Pertinências

Um subconjunto *Fuzzy* $A$ é definido em termos de relevância à um conjunto universal $U$, por uma função denominada de função de pertinência, associando a cada elemento $x$ um número, $\mu_A(x)$ no intervalo $[0,1]$. Essa função, portanto, caracteriza o grau de pertinência de $x$ em $A$. Dessa forma um elemento pertence a um conjunto em uma escala que varia no intervalo $[0,1]$. A função de pertinência é apresentada abaixo e em seguida será definido o conceito de conjunto *Fuzzy*(43).

$$\mu_A : U \to [0,1]$$

**Definição 4.1.1.** *Um subconjunto Fuzzy $F$ de $U$ é caracterizado por uma função $\mu : U \to [0,1]$, chamada função de pertinência do conjunto Fuzzy $F$. Dessa forma, o valor $\mu(x) \in [0,1]$ indica o grau com que o elemento $x$ de $U$ está no conjunto Fuzzy $F$, com $\mu(x) = 0$ e $\mu(x) = 1$ indicando, respectivamente, a não pertinência e pertinência completa de $x$ ao conjunto Fuzzy $F$.*

Do ponto de vista formal, a definição de um subconjunto *Fuzzy* foi obtida simplesmente aplicando-se o contra-domínio da função características, que é o conjunto $\{0,1\}$, para o intervalo $[0,1]$. Nesse sentido, podemos dizer que um conjunto clássico é um caso particular de conjunto *Fuzzy*. Por exemplo, considere o conjunto $P$ dos números pares. Esse conjunto, tem função característica $C_p(n) = 1$ se $n$ é par e $C_p(n) = 0$ se $n$ é ímpar. Portanto o conjunto dos números pares é um particular conjunto *Fuzzy* já que $C_p(n) \in [0,1]$ (44). A critério de exemplo considere o subconjunto *Fuzzy* $F$ dos números naturais pequenos, ou seja, $F$ é definido da forma que segue:

$$F = \{n \in N : n \text{ é pequeno}\}$$

O número 0(zero) pertence pertence a esse conjunto? E o número 1000? Usando a teoria *Fuzzy*, poderíamos dizer ambos pertencem à $F$ porém com diferentes graus de pertinência de acordo com a propriedade que o caracteriza. Ou seja, a função de pertinência de $F$ deve ser "construída" de forma coerente com o termo "pequeno" que caracteriza seus elementos no conjunto universo dos números naturais. Uma possível função de pertinência para o subconjunto *Fuzzy* $F$ é:

$$\mu(n) = \frac{1}{n+1}$$

Se a função de pertinência $\mu(n)$ for o caso, se pode dizer que o número 0(zero) pertence ao subconjunto $F$ com um grau de pertinência $\mu(0) = 1$, enquanto o número 1000 pertence ao subconjunto *Fuzzy* $F$ com um grau de pertinência $\mu(1000) = 0,0011$. Neste caso, a função de pertinência $\mu(n)$ foi escolhida de maneira totalmente arbitrária, podendo ser qualquer outra função definida pelo usuário.

## 4.2 Método *Fuzzy* C-Means

O algoritmo de agrupamento *Fuzzy* mais conhecido é o *Fuzzy C-Means* (FCM) que é uma modificação proposta por Jim Bezdek (1981) de uma metodologia de agrupamento *crisp*. O algoritmo nebuloso *Fuzzy c-means* é o método de agrupamento mais utilizado por ser simples de ser implementado, o mais rápida execução e por não infligir restrições ao conjunto de dados (39). O algoritmo é melhor que o método *k-means* que é um algoritmo *hard* pois o algoritmo *Fuzzy C-means* evita mínimos locais. O método trata-se de uma versão nebulosa do método de agrupamento rígido *k-means*, sendo usado classificar um universo de amostras em categorias nebulosas de acordo com sua disposição no espaço euclidiano. Bezdek introduziu a idéia de um parâmetro de fuzzificação ($m$) no intervalo $[1, n]$, que determina o grau de fuzificação (*fuziness*) nos agrupamentos, ou seja, para $m = 1$ o efeito é uma agrupamento *crisp* dos elementos aos *clusters*. Porém, quando temos $m > 1$, o grau de fuzzificação entre pontos no espaço de decisão aumenta (29)(3). O algoritmo do método *Fuzzy C-means* se divide em dois procedimentos onde o primeiro consiste em calcular os centros de cada grupo e a atribuição de pontos à estes centros utilizando uma forma de distância euclidiana. Este processo se repete até que os centros dos agrupamentos tenha se estabilizados.

Supondo que tem-se um conjuntos de indivíduos $X = \{x_1, x_2, \cdots, x_n\}$, onde cada indivíduo $x_k \in \Re^p, k = 1, \cdots, n$, e deseja-se organizá-los em $c$ grupos, $C = \{C_1, C_2, \cdots, C_n\}$.

O algoritmo de agrupamento FCM é um algoritmo não hierárquico cujo objetivo é fornecer uma partição *Fuzzy* de um conjunto de indivíduos homogêneos em $c$ agrupamentos. Para a construção dos $c$ grupos o FCM define e minimiza uma função objetivo, função esta que mede a adequação entre os indivíduos e agrupamentos. A função objetivo é definida abaixo:

$$J(U,G) = \sum_{k=1}^{n}\sum_{i=1}^{c}(u_{ik})^m \phi(x_k, g_i) = \sum_{k=1}^{n}\sum_{i=1}^{c}(u_{ik})^m \sum_{l=1}^{p}(x_{kl} - g_{il})^2 \tag{4.1}$$

$$\phi(x_k, g_i) = \sum_{l=1}^{p}(x_{kl} - g_{il})^2 \tag{4.2}$$

onde

$U \in M =$ matriz de pertinência $\{u_{ik}\}$ do indivíduo $k$ ao *cluster* $C_i$;

$G =$ vetor de protótipos de *clusters* ou centróides $(g_1, g_2, \cdots, g_k), g_i = (g_{i1}, \cdots, g_{ip})$;

$m \in ]1, +\infty[=$ é um parâmetro que controla a (*fuzzines*) da pertinência dos indivíduos;

$\phi(x_k, g_i) =$ é o quadrado da distância $L_2$ *Minkowsky*, ou Euclidiana, a qual mede a dissimilaridade entre um indivíduo $k$ e um protótipo de *cluster* $i$.

**Proposição 4.2.1.** *Sejam os protótipos $g_i = (g_{i1}, \cdots, g_{ip})$ do cluster $C_i (i = 1, \cdots, c)$, eles minimizam o critério de clustering $J(U,G)$ e são atualizados de acordo com a expressão que segue:*

$$g_{il} = \frac{\sum_{k=1}^{n}(u_{ik})^m x_{kl}}{\sum_{k=1}^{n}(u_{ik})^m}, l = 1, 2, \cdots, p \tag{4.3}$$

Prova: Uma vez que, no passo de representação, a pertinência de cada indivíduo $k$ no *cluster* $C_i$, o parâmetro $m$ estão fixos, pode-se reescrever o critério $J(U,G)$ como:

$$\sum_{i=1}^{c}\sum_{d=1}^{p} J_{il}(g_{il})$$

onde:

$$J_{il}(g_{il}) = \sum_{k=1}^{n} = (u_{ik})^m (x_{kl} - g_{il})^2$$

,

O critério $J(U,G)$ é aditivo e desta forma o problema torna-se achar o $g_{il}$ que minimize $J_{il}(g_{il})$. Logo, a solução deste problema é resolver a equação $\frac{dJ_{il}(g_{il})}{dg_{il}} = 0$. Dessa forma temos que:

$$\frac{dJ_{il}(g_{il})}{dg_{il}} = 0 \Rightarrow \frac{d}{dg_{il}}\left[\sum_{k=1}^{n}(u_{ik})^m(x_{kl}-g_{il})^2\right] \Rightarrow \sum_{k=1}^{n}(u_{ik})^m(x_{kl}-g_{il}) = 0$$

O resultado acima leva a

$$g_{il} = \frac{\sum_{k=1}^{n}(u_{ik})^m x_{kl}}{\sum_{k=1}^{n}(u_{ik})^m}$$

O resultado até agora mostra que atualizando-se $g_{il}$ pela Equação 4.3 obtém-se um ponto extremo de $J_{il}(g_{il})$. Para concluir se este é um ponto de mínimo usaremos o teste da segunda derivada.

$$\frac{d^2J_{il}(g_{il})}{d(g_{il})^2} = \frac{d}{d(g_{il})}\left[-2\sum_{k=1}^{n}(u_{ik})^m(x_{kl}-g_{il})\right] = 2\sum_{k=1}^{n}(u_{ik})^m > 0$$

Como a segunda derivada de $J_{il}(g_{il})$ é positiva, pode-se então concluir que calculando-se $g_{il}$ conforme a Equação 4.3 minimiza $J_{il}(g_{il})$.

**Proposição 4.2.2.** *Seja a pertinência $u_{ik}$ de cada indivíduo $k(k = 1, \cdots, n)$ ao cluster $i(i = 1, \cdots, c)$, ela minimiza o critério clustering $J(U,G)$ sob as seguintes restrições $u_{ik} \geq 0$ e $\sum_{i=1}^{c} u_{ik} = 1$ e é atualizada de acordo com expressão a seguir:*

$$u_{ik} = \left[\sum_{h=1}^{c}\left\{\frac{\sum_{l=1}^{p}(x_{kl}-g_{il})^2}{\sum_{l=1}^{p}(x_{kl}-g_{hl})^2}\right\}^{\frac{1}{m-1}}\right]^{-1}, i = 1, \cdots, c; k = 1, \cdots, n. \quad (4.4)$$

Prova: Considere que no passo de alocação, os protótipos $g_i$ do *cluster* $C_i$, $(i = 1, \cdots, c)$, e o parâmetro m estão fixos. Uma vez que $U$ é uma matriz degenerada, suas colunas são independentes e a minimização de $J(U,G)$ pode ser obtida aplicando-se o método dos multiplicadores de Lagrange a cada termo $u_k$ . Seja a função $\kappa(u_k)$:

$$\kappa(u_k) = \sum_{i=1}^{c}(u_{ik})^m \sum_{l=1}^{p}(x_{kl} - g_{il})^2$$

minimizar $J(U, G)$ implica em minimizar $\kappa(u_k)$ sob a restrição $\sum_{i=1}^{c} u_{ik} = 1$.

Seja o lagrangiano de $\kappa(u_k)$:

$$F(\mu, u_k) = \sum_{i=1}^{c}(u_{ik})^m \sum_{l=1}^{p}(x_{kl} - g_{il})^2 - \mu(\sum_{i=1}^{c}(u_{ik})^m - 1)$$

O par $(\mu, u_k)$ é estacionário para a função $F$ se o gradiente de $F$, $\Delta F$, for zero. Igualando o gradiente de $F$ igual a zero, tem-se:

$$\frac{\partial F}{\partial \mu}(\mu, u_k) = (\sum_{i=1}^{c}(u_{ik})^m - 1) = 0 \tag{4.5}$$

$$\frac{\partial F}{\partial u_{ik}}(\mu, u_k) = \left[m(u_{ik})^{m-1} \sum_{l=1}^{p}(x_{kl} - g_{il})^2 - \mu\right] = 0 \tag{4.6}$$

A partir da Equação 4.6, obtém-se:

$$u_{ik} = \left[\frac{\mu}{m \sum_{l=1}^{p}(x_{kl} - g_{il})^2}\right]^{\frac{1}{m-1}} \tag{4.7}$$

Usando o resultado da Equação 4.5 na Equação 4.6, obtém-se:

$$\sum_{h=1}^{c} u_{hk} = \sum_{h=1}^{c} \left(\frac{\mu}{m}\right)^{\frac{1}{m-1}} \left[\frac{1}{\sum_{l=1}^{p}(x_{kl} - g_{hl})^2}\right]^{\frac{1}{m-1}} =$$

$$\left(\frac{\mu}{m}\right)^{\frac{1}{m-1}} \left\{\sum_{h=1}^{c} \left[\frac{1}{\sum_{l=1}^{p}(x_{kl} - g_{hl})^2}\right]^{\frac{1}{m-1}}\right\} = 1$$

Deste resultado conclui-se:

$$\left(\frac{\mu}{m}\right)^{\frac{1}{m-1}} = \frac{1}{\sum_{h=1}^{c}\left[\frac{1}{(x_{kl}-g_{hl})^2}^{\frac{1}{m-1}}\right]}$$

Aplicando o resultado acima na Equação 4.7, tem-se:

$$u_{ik} = \left\{\frac{1}{\sum_{h=1}^{c}\left[\frac{1}{\sum_{l=1}^{p}(x_{kl}-g_{hl})^2}\right]}\right\}\left[\frac{1}{\sum_{l=1}^{p}(x_{kl}-g_{il})^2}\right]^{\frac{1}{m-1}}$$

$$= \frac{1}{\sum_{h=1}^{c}\left[\frac{\sum_{l=1}^{p}(x_{kl}-g_{il})^2}{\sum_{l=1}^{p}(x_{kl}-g_{hl})}\right]^{\frac{1}{m-1}}}$$

$$= \left[\sum_{h=1}^{c}\left\{\frac{\sum_{d=1}^{p}(x_{kl}-g_{il})^2}{\sum_{d=1}^{p}(x_{kl}-g_{hd})^2}\right\}^{\frac{1}{m-1}}\right]^{-1},$$

para $i = 1, \cdots, c; k = 1, \cdots, n$.

O resultado acima permite concluir que calculando-se $u_{ik}$ pela Equação 4.4leva a um valor extremo de $J(U, G)$. Usar-se-á o teste da segunda derivada para verificar se esta mesma expressão é um mínimo da função objetivo. Minimizar $\kappa(u_k)$ leva à minimização de $J(U, G)$. Já foi encontrado um valor extremo $\kappa(u_k)$ e consequentemente para $J(U, G)$ sob a restrição $\sum_{i=1}^{c} u_{ik} = 1$. Pelo teste da segunda derivada da função $\kappa(u_k)$ se este valor é mínimo.

$$\frac{\partial\kappa}{\partial u_{ik}}(u_k) = [m(u_{ik})^{m-1}\sum_{d=1}^{p}(x_{kl}-g_{il})^2]$$

$$\frac{\partial^2\kappa}{\partial(u_{ik})^2}(u_k) = [m(m-1)(u_{ik})^{m-2}\sum_{d=1}^{p}(x_{kl}-g_{il})^2]$$

Note que:

$$\frac{\partial^2\kappa}{\partial u_{zk}u_{wk}}(u_k) = 0, \forall z \neq w$$

Esse fato leva a matriz Hessiana de $\kappa(u_k)$ ser definida da seguinte forma:

$$H(\kappa(\mathbf{u}_k)) = \begin{bmatrix} [m(m-1)(u_{1k})^{m-2}\sum_{d=1}^{p}(x_{kl}-g_{1l})^2] & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & [m(m-1)(u_{ck})^{m-2}\sum_{d=1}^{p}(x_{kl}-g_{cl})^2] \end{bmatrix}$$

A matriz $H(\kappa(u_k))$ é uma matriz diagonal com todos elementos positivos. Este fato permite concluir que $H(\kappa(u_k))$ é definida positiva e portanto isto implica que atualizar $u_{ik}$ pela Equação 4.4 leva a um valor mínimo da função objetivo $J(U,G)$ (3).

# 5 *Introdução aos Algoritmos Genéticos*

## 5.1 Algoritmo Evolutivo

Algoritmos evolutivos são procedimentos baseados em mecanismos da evolução biológica e servem para originar conceitos um pouco mais recentes, como o dos Algoritmos Genéticos. A motivação para a construção de tais modelos computacionais surgiu de teorias através das quais a Natureza, por meio de seus recursos, resolveu problemas de grande complexidade. Assim pode-se dizer que a natureza otimiza seus mecanismos para resolver um ou mais problemas.

A partir de um problema de otimização, mesmo que se desconheça o que se está otimizando, é possível encontrar uma solução ótima, ou no mínimo, uma boa solução, através dos Algoritmos Evolutivos e suas variações. Em resumo, tais algoritmos podem trabalhar em cima de problemas, sem que exista um conhecimento explicito, sobre as suas soluções ótimas.

Os Algoritmos Evolutivos buscam tratar as estruturas de objetos abstratos de uma população, como, por exemplo, variáveis de um problema de otimização, através da manipulação de operadores inspirados na evolução biológica, chamados comumente de operadores genéticos.

### 5.1.1 Programação Evolutiva

A programação evolutiva (PE) foi proposta por Fogel em 1962 com o objetivo de utilizar os conceitos da evolução humana no desenvolvimento da Inteligência Artificial (IA). Na programação evolutiva cada indivíduo de uma população é representado por uma máquina de estados finitos ou autômatos finitos[1]. Durante a avaliação, os indivíduos

[1]Uma máquina de estados finitos ou autômato finito é uma modelagem de um comportamento composto por estados, transições e ações refletindo as mudanças desde a entrada num estado, no início do

são analisados por uma função de *payoff* (penalidade) de acordo com a saída da máquina. A reprodução desses indivíduos é feita apenas por operadores de mutação, sendo que todos indivíduos da população atual geram novos descendentes. Esse processo caracteriza o que se chama de *reprodução assexuada*[2]. Na seleção de novos indivíduos os descendentes (filhos) competem com os pais e somente os indivíduos com maior *fitness* sobrevivem(3).

Dessa forma, a PE garante que todos os indivíduos de uma tal população irão produzir novos descendentes e, somente, os melhores indivíduos entre os atuais e os descendentes irão sobreviver, em que o domínio dos melhores indivíduos é chamado de *elitismo total*. O elitismo mais utilizado garante a sobrevivência apenas dos $k$-melhores indivíduos em que $k < N$, sendo $N$ o tamanho da população (26).

### 5.1.2 Algoritmos Genéticos

Um algoritmo genético (AG) é uma técnica de busca utilizada na ciência da computação para achar soluções aproximadas em problemas de otimização e busca, fundamentado principalmente pelo americano John Henry Holland. Algoritmos genéticos são uma classe particular de algoritmos evolutivos que usam técnicas inspiradas pela biologia evolutiva como hereditariedade, mutação, seleção natural e recombinação (*crossing over*). Holland estudou a evolução natural considerando esta um processo robusto, simples e poderoso, que poderia ser adaptado para obtenção de soluções computacionais eficientes para problemas de otimização. O conceito de robustez refere-se ao fato de os AG's independentemente das escolhas dos parâmetros iniciais, em geral, reproduzem soluções de qualidade (26).

Algoritmos genéticos são implementados como uma simulação de computador em que uma população de representações abstratas de solução é selecionada em busca de soluções melhores. A evolução geralmente se inicia a partir de um conjunto de soluções criado aleatoriamente e é realizada por meio de gerações. A cada geração, a adaptação de cada solução na população é avaliada, alguns indivíduos são selecionados para a próxima geração, e recombinados ou mutados para formar uma nova população (37).

A grande vantagem dos algoritmos genéticos esta no fato de não precisarmos saber como funciona a função objetivo que queremos otimizar, apenas tê-la disponível para ser aplicada aos indivíduos e comparar os resultados é suficiente. A nova população então é utilizada como entrada para a próxima iteração do algoritmo. A seguir mostram-se os principais conceitos de algoritmos genéticos (26).

sistema, até o momento presente.

[2]Tipo de reprodução que ocorre sem a troca de material genético.

1. Função objetivo: Refere-se ao objeto de otimização. Pode ser um problema de otimização, um conjunto de regras e testes para identificar os indivíduos mais aptos, ou mesmo uma "caixa preta" em que sabemos apenas o formato das entradas;

2. Código genético: É uma representação do espaço de busca do problema a ser resolvido. O código genético deve ser uma representação capaz de representar todo o conjunto dos valores no espaço de busca e precisa ter tamanho finito;

3. Indivíduo: É meramente um portador do seu código genético;

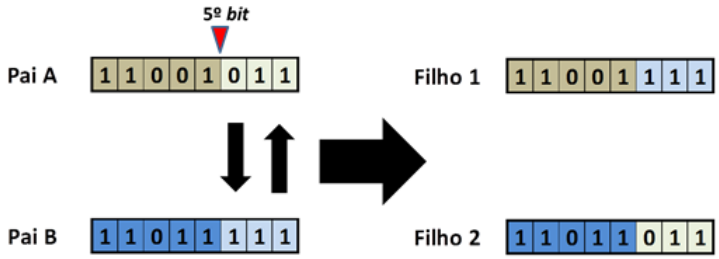
4. Seleção: Trata-se de uma parte chave do algoritmo genético. Em geral, usa-se o mecanismo de seleção por "roleta", em que os indivíduos são ordenados de acordo com a função-objetivo e lhes são atribuídas probabilidades decrescentes de serem escolhidos. A escolha é feita então aleatoriamente de acordo com essas probabilidades. Dessa forma conseguimos escolher como pais os mais bem adaptados, sem deixar de lado a diversidade dos menos adaptados. Outras formas de seleção podem ser aplicadas dependendo do problema a ser tratado;

5. Reprodução: A reprodução é se divide em três etapas sendo elas o acasalamento, a recombinação e a mutação.

    - Acasalamento: É a escolha de dois indivíduos para se reproduzirem em que em geral geral dois descendentes para manter o tamanho populacional;
    - Recombinação: Também conhecida por *crossing-over* a recombinação é um processo que imita o processo biológico homônimo na reprodução sexuada em que os descendentes carregam em seu código genético parte do código genético do do pai e parte do código da mãe. Dessa forma essa recombinação garante que os melhores indivíduos sejam capazes de trocar entre si as informações que os levam a ser mais aptos a sobreviver, e assim gerar descendentes ainda mais aptos.

      A critério de exemplo considere a representação binária com 8 *bits* de dois números em que a cadeia de bits (*cromossomos*) de um número $A = \mathbf{11001}011$ e do número $B = 11011\mathbf{111}$. Uma regra de *cross-over* pode ser sortear aleatoriamente a posição de 5 *bit* do número $A$ e 3 *bits* do número $B$ e concatenar esses elementos originando um novo indivíduo (filho) $C = \mathbf{11001111}$, dado que o evento *cross-over* ocorreu com uma certa probabilidade relativamente alta. Com essa informação, é possível, por exemplo, gerar dois filhos:

Figura 2: Troca de Material Genético para o Algoritmo Genético Binário.

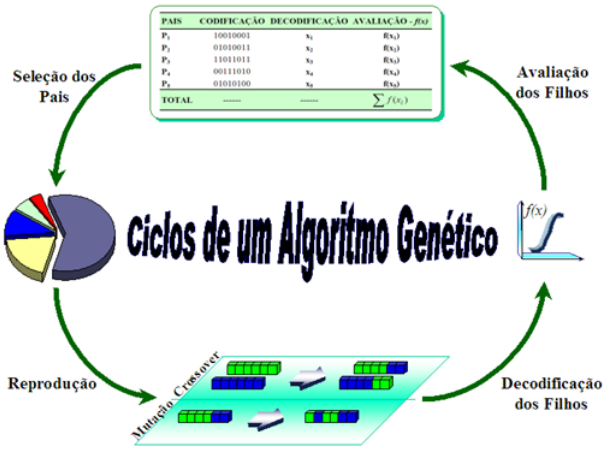



  - O Filho 1 com a parte genética inicial do Pai A e a parte genética final do Pai B;
  - O Filho 2 com a parte genética inicial do Pai B e a parte genética final do Pai A.

- Mutação: O processo de mutação é realizado com a probabilidade mais baixa possível no entanto essa probabilidade existe, e tem como objetivo permitir maior variabilidade genética na população, impedindo que a busca fique estagnada em um mínimo local. O processo de mutação não há troca de material genético como no exemplo anterior, ou seja, não há uma recombinação dos *cromossomos*. A única coisa que há é uma mudança na sequência da cadeia do *cromossomo* dos indivíduos. Uma regra de mutação poderia ser sortear aleatoriamente um *bit* de um indivíduo e trocar o seu valor. Dessa forma seja um indivíduo $A = 11001001$. Se o segundo elemento da cadeia de *bits* for selecionado de forma aleatória teremos um indivíduo mutado $A^* = 10001001$.

- O processo dos Algoritmos Genéticos passam pode ser resumido pela Figura abaixo:

Figura 3: Resumo dos Algoritmos Genéticos.



Os exemplos de *cross-over* e mutação apresentados acima pertencem à classe de algoritmos genéticos de representação binária , em que os materiais genéticos são representados por *cromossomos* que são uma sequência de zeros e uns de conjuntos de *bits* estabelecido. Assim, a representação binária tem dificuldades com múltiplas dimensões de variáveis contínuas, especialmente quando uma grande precisão é requerida. Para que a precisão desejada seja alcançada, um grande número de *bits* na cadeia dos *cromossomos* será necessário. A grande quantidades de *bits* utilizados implicam em *cromossomos* extremamente extensos, dificultando dessa forma, a operacionalização do algoritmo genético. Além disso, ainda existe a dificuldade de representar números (indivíduos) contínuos (com casas decimais) na escala binária, o que obriga o processo a discretizar (truncar) os números reais.

Percebe-se assim que algoritmos genéticos com representação real são mais adequados para a maioria dos problemas, principalmente quando se trabalha em um espaço paramétrico contínuo. Há inúmeros trabalhos na literatura que faz uso de algoritmos genéticos com representação real. Pode-se citar, por exemplo (Joab., 2009) que discutiu o uso de função de verossimilhança na estimação da proporção de ítens conformes na presença de erros de classificação e de classificações repetidas repetidas usando algoritmos genéticos num processo de otimização multiobjetivo (30).

A seguir, serão mostrados os principais passos de um algoritmo genético geral:

1. No tempo $t_0$, um conjunto de soluções iniciais é gerado e esse funcionará como uma população inicial para as futuras descendências. Durante toda a execução do processo evolutivo, o algoritmo genético mantém atualizada essa de populações pontenciais $P(gen) = \{x_1{}^{gen}, x_2{}^{gen}, \cdots, x_n{}^{gen}\}$;

2. Para essa população inicial cada indivíduo é avaliado através de uma medida de aptidão, ou *fitness*, geralmente, representada pela própria função objetivo;

3. Dois indivíduos (os pais) são escolhidos de acordo com os seus graus de adaptação, isto é, indivíduos mais adaptados (mais fortes) têm uma probabilidade maior de serem selecionados;

4. Os indivíduos selecionados passarão por um processo de reprodução que envolve não só o crossing-over, como também a mutação, produzindo filhos. Esses filhos caracterizam-se como dois novos candidatos a perpetuarem a espécie.

5. Esses filhos são adicionados à população inicial e todo o processo é repetido até que a população não possa ser mais “melhorada".

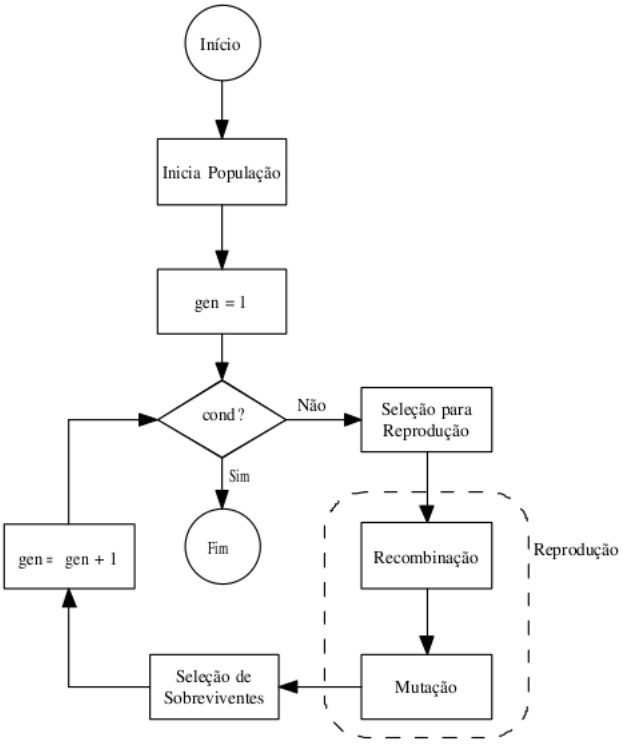



Figura 4: Diagrama do fluxo do algoritmo genético.

Para otimização de parâmetros contínuos a representação real é a mais adequada, e segue o mesmo fluxo descrito no diagrama apresentado na Figura 4. Em todo esse tópico foi discutido os conceitos mais relevantes da teoria sober os algoritmos genéticos mas, para tornar essa idéia mais clara e objetiva, considere o exemplo abaixo como parte da explicação da teoria de algoritmos genéticos. O objetivo do problema é encontrar o parâmetro $\theta$ que maximiza um função $f(\theta)$, com $\theta$ variando no intervalo $[-2,5; 2,5]$. A função $f(\theta)$ é a que segue:

$$f(\theta) = 6 + \theta^2 seno(14\theta), \tag{5.1}$$

com $-2,5 \leq \theta \leq 2,5$. O gráfico da função $f(\theta)$ segue logo abaixo:

Como podemos perceber o gráfico da função $f(\theta)$ apresenta vários pontos de mínimo é máximo global o que é um problema para muitos métodos de otimização, em que dessa forma muitos algoritmos não são capazes de localizar o ótimo global na presença de múltiplos ótimos locais. A exemplo o método de busca Hill Climbing sendo este um método de busca local.

Na Figura 5 pode-se observar um ponto vermelho, sendo este o ponto de máximo global da função $f(\theta)$ com $\theta$ variando no intervalo $[-2,5; 2,5]$ e valor ótimo nesse ponto de 11,561141. O $\theta$ que otimiza a função $f(\theta)$ é igual à 2,360511947, sendo este valor

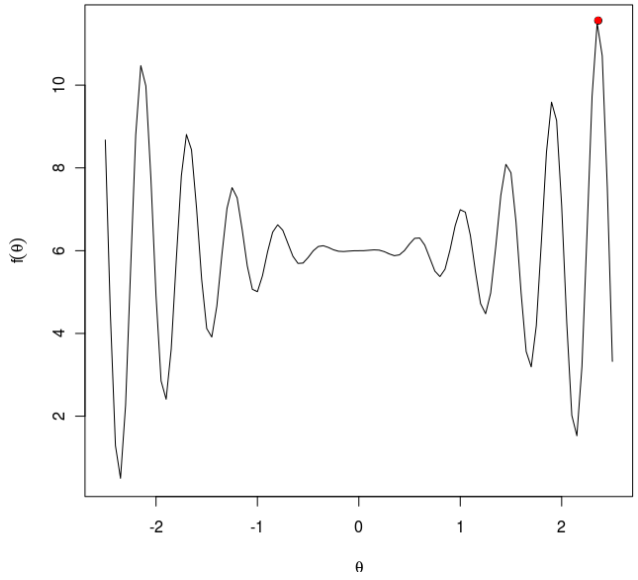



Figura 5: Função objetivo $f(\theta)$.

pertencente ao espaço paramétrico que maximiza a função objetivo. Assim, o objetivo do algoritmo genético é encontrar esse valor para $\theta$ para um certo número de gerações. Os passos do algoritmo genético para otimizar a função $f(\theta)$ seguem abaixo:

1. Definir os parâmetros de entrada do algoritmo genético. Esses parâmetros são:
   - A função objetivo $FO$;
   - Número de iterações ($NIT$): $1 \leq t \leq NIT$, em que $t$ indica as gerações;
   - Probabilidade de recombinação (*cross-over*) $PrR$: Sugestão $PrR = 0,90$;
   - Probabilidade de mutação ($PrM$): Sugestão $PrM = 0,01$;
   - Número de Soluções de Elite ($NSE$): Sugestão $NSE = 5$;
   - Nível de precisão ($\varepsilon$) para o critério de parada ($EPS$): Sugestão $EPS = 10^{-8}$;
2. (Geração $t = 0$): Gerar, aleatoriamente, digamos, 15 valores para $\theta$, chamadas de soluções candidatas;
3. Calcular a função objetivo ($f(\theta) = 6 + \theta^2 seno(14\theta)$) para cada uma das 15 soluções candidatas e ordenar esse vetor (com base na função objetivo) em rodem decrescente;
4. Calcular a proporções simples e as proporções acumuladas a partir da soma das funções objetivo obtidas. Abaixo segue um exemplo fictício da estrutura dessa.
5. Selecionar (usando a técnica da Roleta Russa) duas soluções candidatas (os dois indivíduos) dentre as 15 soluções apresentadas na Geração $t = 0$, de modo que essa

Tabela 4: Tabela com as Soluções Candidatas Iniciais e com as Proporções Calculadas com Base na Função Objetivo.

| $\theta_i$ | Soluções | $FO$ | Proporções | Acumulada ($P_{rA}$) |
|---|---|---|---|---|
| $\theta_1$ | 2,319054079 | 10,66723 | 0,10559 | 0,10559 |
| $\theta_2$ | -2,157887509 | 10,34924 | 0,10244 | 0,20803 |
| $\theta_3$ | -1,637852107 | 8,16439 | 0,08081 | 0,28884 |
| $\theta_4$ | 1,535309915 | 7,12349 | 0,07051 | 0,35935 |
| $\theta_5$ | -1,182515946 | 7,04802 | 0,06976 | 0,42911 |
| $\theta_6$ | -2,477416303 | 6,77317 | 0,06704 | 0,49615 |
| $\theta_7$ | -0,787606433 | 6,62003 | 0,06553 | 0,56168 |
| $\theta_8$ | -0,297479171 | 6,07555 | 0,06014 | 0,62182 |
| $\theta_9$ | -0,285729545 | 6,06180 | 0,06000 | 0,68182 |
| $\theta_{10}$ | -0,228049562 | 6,00266 | 0,05942 | 0,74124 |
| $\theta_{11}$ | 0,886791589 | 5,88148 | 0,05822 | 0,79946 |
| $\theta_{12}$ | -1,562166204 | 5,70587 | 0,05648 | 0,85594 |
| $\theta_{13}$ | -1,557588427 | 5,55394 | 0,05497 | 0,91091 |
| $\theta_{14}$ | -2,000869778 | 4,96244 | 0,04912 | 0,96003 |
| $\theta_{15}$ | 1,630069887 | 4,03953 | 0,03998 | 1,00000 |
| Total | * | 101,02883 | 1,00000 | * |

seleção seja proporcional à adaptação de cada indivíduo, ou seja, uma solução mais apta (com função objetivo maior) teria uma maior probabilidade de ser selecionada que uma solução menos apta (com função objetivo menor). Isso é feito através do seguinte procedimento:

**Para selecionar a primeira solução candidata**

- Gerar um número aleatório ($NUM_1$) de uma Distribuição Uniforme $(0; 1)$;
- Se $NUM_1 \leq PrA$ (Proporção acumulada da $j$-ésima solução candidata), então selecione a solução candidata $\theta_j$;

**Para selecionar a segunda solução candidata**

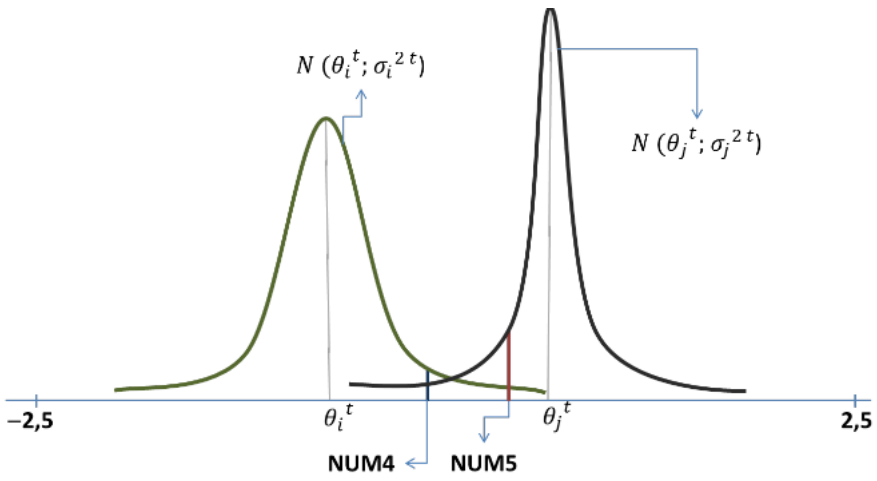
- Gerar um outro número aleatório ($NUM_2$) de uma distribuição uniforme $(0; 1)$;
- Se $NUM_2 \leq PrA$ (Proporção acumulada da $j$-ésima solução candidata), então selecione a solução candidata $\theta_j$. O processo de seleção é feito utilizando uma técnica de amostragem com reposição.

  O processo de seleção é feito utilizando uma técnica de amostragem com reposição. Em vista disso, uma mesma solução candidata pode ser selecionada duas vezes numa mesma geração;

6. Com as duas soluções candidatas selecionadas, juntamente com as suas respectivas funções objetivo calculadas, verificar se haverá (ou não) o processo de *crossover*

ou recombinação (troca de material genético dos dois candidatos). Isso poderá ser realizado a partir do seguinte procedimento:

- Gerar um número aleatório ($NUM_3$) de uma Distribuição de $Bernoulli(PrR)$, ou seja, gerar um dígito binário, de modo que a probabilidade de sair "1" é de $PrR$ (Probabilidade de recombinação definida no 1º Passo);
- Se $NUM_3 = 1$, então realizar o processo de recombinação da seguinte forma:
  - Sejam $\theta_i$, $\theta_j$, $FO(\theta)_{it}$ (função objetivo na geração $t$ relativa à solução candidata $\theta_i$) e $FO(\theta)_{jt}$ (função objetivo na geração $t$ relativa à solução candidata $\theta_j$) os valores das soluções candidatas selecionadas e suas respectivas funções objetivo;
  - Calcule $\sigma_{\theta it} = FO(\theta)_{it} - 1$ e $\sigma_{\theta jt} = FO(\theta)_{jt} - 1$ que funcionarão como uma espécie de uma medida de variabilidade para cada solução candidata e que serão inversamente proporcionais aos valores de suas funções objetivo;
  - Agora gere um outro número aleatório ($NUM_4$), atribuído ao indivíduo $\theta_i$, de uma Distribuição Normal com média $\theta_i$ e desvio-padrão $\sigma_{?it}$. Faça o mesmo para a outra solução candidata, isto é, gere um número aleatório ($NUM_5$), atribuído ao indivíduo $\sigma_j$, de uma Distribuição Normal com média $\sigma_j$ e desvio-padrão $\sigma_{?jt}$. O efeito dessa estratégia pode ser representado graficamente como segue:




  - Faça $\theta_i = NUM_5$ e $\theta_j = NUM_4$, ou seja, a solução candidata?$\theta_i$ recebe o valor aleatório gerado a partir da distribuição de probabilidade em torno de $\theta_j$, enquanto que a solução candidata?$\theta_j$ recebe o valor aleatório gerado a partir da distribuição de probabilidade em torno de $\theta_i$. Essa é uma maneira de criar soluções candidatas novas a partir das informações "genéticas" dos seus concorrentes. Se $NUM_4 > 2,5$ ou $NUM_4 < -2,5$, então faça $NUM_4$

igual ao limite do domínio da função (ou -2,5 ou 2,5). Proceda de forma análoga para $NUM_5$.

7. Com as duas novas soluções candidatas (depois de terem, ou não, realizadas o *crossover*), verificar se haverá (ou não) o processo de mutação (melhoramento do material genético interno de cada uma das soluções candidatas). Isso poderá ser realizado a partir do seguinte procedimento:

   - Gerar um número aleatório ($NUM_6$) de uma Distribuição de Bernoulli ($PrM$), ou seja, gerar um dígito binário, de modo que a probabilidade de sair "1" é de $PrM$ (Probabilidade de Mutação definida no 1º passo);
   - Se $NUM_6 = 1$, então realizar o processo de mutação da seguinte forma:
     - Gerar dois números aleatórios ($ERR_1$ e $ERR_2$) de duas Distribuições Uniformes distintas no intervalo $[\frac{-1}{t}; \frac{1}{t}]$. Perceba que esse intervalo vai diminuindo à medida que as gerações vão avançando, evitando assim, variações altas nas gerações mais adiantadas;
     - Faça $\theta_i = \theta_i + ERR_1$ e $\theta_j = \theta_j + ERR_2$. Se $\theta_i > 2,5$ ou $\theta_j < -2,5$, então faça $\theta_i$ igual ao limite do domínio da função (ou -2,5 ou 2,5). Proceda de forma análoga para $\theta_j$.

8. Calcule a função objetivo para cada uma das novas soluções candidatas e utilize o seguinte critério para incluí-las (ou não) na lista das 15 soluções candidatas da geração inicial:

   - Se pelo menos uma das soluções candidatas apresentar uma adaptação melhor (maior valor da função objetivo) que qualquer uma das 15 soluções candidatas geradas no 2º passo, então inclua tal solução (ou as duas) na listagem das 15 soluções candidatas (que ficaria nesse momento 16 soluções) e delete a última (ou as duas últimas);
   - Se pelo menos uma das soluções candidatas apresentar uma adaptação melhor (maior valor da função objetivo) que a melhor solução candidata (aquela com o maior valor da função objetivo, ou seja, a primeira candidata da lista ordenada) gerada no 2º passo, então inclua tal solução (ou as duas) na listagem das 15 soluções candidatas (que ficaria nesse momento 16 soluções) e delete a última (ou as duas últimas). E, além disso, verifique se é possível incluí-la também (ou as duas) na lista de Soluções de Elite (que receberá, no máximo, as $NSE = 5$ melhores soluções).

9. Ordene as 15 novas soluções candidatas e repita tudo novamente a partir do 4º passo até que tenha atingido o total de iterações ($NIT$) ou que a diferença, em módulo, entre a solução encontrada no 8º passo e a melhor solução da lista de Soluções de Elite seja menor que $EPS$.

Esses procedimentos serão repetidos até que o critério de parada seja atingido. Nos anexos desse trabalho podem ser encontrado duas implementações desse exemplo na linguagem R e no pacote estatístico SAS. O algoritmo implementado na linguagem R considerou uma população de 8000 indivíduos e foram feitas 300 interações, ou seja, foram consideradas 300 gerações que para "convergência" o algoritmo levou 4,265293 segundos. O valor da função $f(\theta)$ encontrado pelo algoritmo foi 11,5618064 com $\theta$ ótimo igual à 2,360763, valores estes bem próximos do real. Com base em tudo que foi visto, decidiu-se propor um método híbrido de otimização que combine as melhores propriedades da lógica nebulosa Fuzzy com as principais idéias da filosofia dos algoritmos genéticos. Essa proposta será discutida com mais detalhes a seguir.

# 6 *Proposta de um Modelo Fuzzy C-Means Genético para Agrupamento de Dados*

## 6.1 Introdução

A proposta de um modelo híbrido para agrupamentos de dados une conceitos de algoritmos genéticos com os da lógica de conjuntos nebulosos *Fuzzy*. Esse *mix* de ferramentas fornecerá um método de agrupamento mais flexível e eficiente na hora de agrupar os dados, representando assim melhor a realidade dos diversos problemas práticos. Modelos clássicos para agrupar dados, em geral, se utilizam de uma filosofia *crisp* em que um indivíduo pertence a um, e somente um agrupamento. Em geral, essa construção rígida de agrupamentos de dados não é coerente com a realidade. Para contornar esse problema, é comum a utilização de conceitos da lógica difusa (ver Capítulo 4), nos quais permitem que um indivíduo possa pertencer a todos os grupos, segundo uma medida de pertinência.

O modelo proposto também faz uso dos Algoritmos Genéticos como uma ferramenta fundamental na otimização dos parâmetros da metodologia *Fuzzy*, visto que esses algoritmos evolutivos não fazem uso de derivadas, inversões de matrizes ou qualquer outro cálculo que pode tornar o procedimento complexo computacionalmente.

Mas antes da discussão detalhada do método proposto, convém explanar como foi realizado todo o processo de definição, captação, crítica e delineamento do banco de dados que será utilizado nesse trabalho.

## 6.2 Delineamento do Processo Amostral

### 6.2.1 Base de dados utilizada

Esse trabalho fez uso da base de dados do Sistema de Informação sobre Mortalidade (SIM) da Secretaria de Saúde do Estado da Paraíba no período de 1 de janeiro de 2006 à 07 de julho de 2010, data esta da última geração do banco de dados. A base de dados foi fornecida pela Gerencia Operacional de Resposta Rápida da Secretaria de Saúde. A banco de dados inicial gerado pelo SIM apresentou todas as informações de óbitos de pessoas do Estado da Paraíba no período considerado, de forma que cada linha desse arquivo representava um indivíduo que veio à óbito em um dado ano, sendo a chave primária o número da declaração de óbito (DO) do indivíduo.

A operacionalização do Sistema é composta pelo preenchimento e coleta do documento padrão - a Declaração de Óbito (DO), sendo esse o documento de entrada do sistema nos estados e municípios. Os dados coletados são de grande importância para a vigilância sanitária e análise epidemiológica, além de estatísticas de saúde e demografia. Sendo assim, o SIM é um sistema de vigilância epidemiológica nacional, cujo objetivo é captar dados sobre os óbitos do país a fim de fornecer informações sobre mortalidade para todas as instâncias do sistema de saúde.

O banco de dados, inicialmente gerado, apresentou mais de um milhão de linhas, representando cerca de 700 MB em formato DBF. Os dados listavam os óbitos ocorridos por todas as causas de óbitos fornecidas pelo livro de Classificação Internacional de Doenças (CID-10). As informações de interesse (tais como nome do paciente, nomes dos pais, CRN do médico que constatou o óbito, nome do hospital, bem como o endereço do indivíduo que veio a óbito, dentre outras informações ) foram filtradas utilizando o *software* TabWin 3.61, fornecido pelo Ministério da Saúde.

A partir do carregamento da base de dados original, os dados passaram a ser tratados utilizando a linguagem R, em que foram filtrados os óbitos ocorridos por homicídios referente à causas externas pertencente ao capítulo XX da CID-10. Para a consulta das descrições das causas foi utilizado o *software* livre Classix 0.1. Este software permite a consulta das causas de óbitos da CID-10 por capítulo ou pela codificação das causas e está disponível para o sistema Linux.

A opção de se utilizar a linguagem R foi devido o TabWin esconder, de certa forma, algumas informações,uma vez que os cruzamentos de variáveis são feitos por alguns arquivos

de definições.

### 6.2.2 Linguagem R

Para a realização das simulações foi utilizado o pacote estatístico R em sua versão 11.1 de 31 de maio de 2010. O pacote R é uma linguagem e um ambiente de desenvolvimento integrado, para cálculos estatísticos e gráficos. A linguagem R foi criada inicialmente por Ross Ihaka e Robert Gentleman na Universidade de Auckland, Nova Zelândia e foi desenvolvido por um esforço colaborativo de pessoas em vários locais do mundo (38). O nome R provêm em parte das iniciais dos criadores e também de um jogo figurado com a linguagem S (da Bell Laboratories, antiga AT&T). No entanto, a linguagem R não deriva da linguagem S como muitos acreditam (9).

A linguagem resultante é muito similar em sintaxe ao S, mas a implementação subjacente e semântica são derivados da linguagem Scheme. O código fonte do R está disponível sob a licença GNU GPL (*GNU General Public License*) e as versões binárias pré-compiladas são fornecidas para Windows, Macintosh, e muitos sistemas operacionais Unix/Linux.

Apesar do seu caráter gratuito, o R é uma ferramenta bastante poderosa com boas capacidades ao nível de programação e um conjunto bastante vasto (e em crescimento) de *packages* que acrescentam bastante potencialidades à já poderosa versão base do R.

## 6.3 Descrição do Problema a ser Otimizado

A necessidade de inter-relacionar alguns municípios do Estado da Paraíba com as principais causas externas de óbito, descritas na Seção 1.2 é o grande objetivo de estudo e de aplicação desse trabalho.

Diante desse cenário, deseja-se encontrar a melhor função de pertinência da lógica nebulosa *Fuzzy*, de modo que essa venha representar, da melhor forma possível os inter-relacionamentos entre municípios e causas. Uma consequência natural da busca da melhor matriz de pertinência é a otimização do processo de construção de grupos homogêneos de municípios dentro das diversas causas. Obviamente, eEssas medidas de pertinência passam, obrigatoriamente, pelas idéias de similaridade vistas na Seção 3.2.

Dessa forma, como o problema é multidimensional, o que se quer é encontrar as melhores pertinências de todos os municípios para todas as causas. Como uma maneira de

simplificar o espaço de busca, decidiu-se trabalhar apenas com os 30 principais municípios paraibanos com registros nas e 10 8 principais causas de óbitos. Dessa forma, as matrizes de pertinências candidatas a soluções ótimas terão 30 linhas por 8 colunas. Com isso, ainda que considerando essas restrições, as soluções ótimas do subespaço de busca estarão contidas no espaço R30 8, de dimensão 240. A melhor maneira de transformar informações de contagem em medidas de distância (proximidade) é aplicando a técnica multivariada de Análise de Correspondência. É esse procedimento que será apresentado a seguir.

### 6.3.1 Transformando Tabelas de Contingências em Medidas de Distâncias Bidimensionais

Em busca de se verificar as interações entre municípios paraibanos com as causas de óbitos,foi empregada a metodologia de Análise de Correspondência (AC), descrita na Seção 3.3 do Capítulo 3, para transformar contagens em distâncias (similaridades). A Análise de Correspondência possibilitou a identificação das proximidades entre municípios e causas para os anos de 2006 a 2010.

A partir da tabela de contigência relacionou os municípios e as causas externas de óbito (ver Tabela 5), por exemplo para o ano de 2010, fazendo uso da AC é possível encontrar as coordenadas do plano cartesiano dos indivíduos (cada linha da tabela de contingência) e de todas as colunas da tabela. Desse modo, tem-se as coordenadas dos municípios e causas externas de óbitos consideradas neste trabalho que podem ser representadas no $\Re^2$. Os dados foram reduzidos ao espaço bidimensional devido este ser mais simples de visualizar os relacionamentos as categorias das variáveis envolvidas.

Para a escolha do número de dimensões que seriam adotadas utilizou-se a regra dos autovalores maiores que 1. A partir desse critério, decidiu-se por utilizar apenas 2 dimensões. Essas dimensões explicaram, para todos os anos em estudo, no mínimo 90% da variabilidade total dos dados originais. A AC em sua teoria utiliza a distância Euclidiana para calcular as dissimilaridades entre linhas e colunas da matriz de correspondência. Contudo, preferiu-se empregar as distância de Mahalanobis, uma vez que essa levam em consideração a matriz de variância-covariância dos pontos no $R^2$.

Tabela 5: Tabela de contigência a critério ilustrativo com o quantitativo de óbitos ocorridos em 2010.

| | Causas Externas de Óbito | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Municípios | V03 | V09 | V49 | V89 | V99 | W19 | X70 | X95 | Total |
| Alagora Grande | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Alagoa Nova | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 3 |
| Alhandra | 0 | 2 | 0 | 3 | 0 | 0 | 0 | 4 | 9 |
| Areia | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 2 | 4 |
| Aroeira | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 2 | 3 |
| Bayeux | 1 | 0 | 1 | 2 | 0 | 0 | 1 | 30 | 35 |
| Boqueirão | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 2 | 3 |
| Caaporã | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 4 | 6 |
| Cabedelo | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 15 | 18 |
| Cajazeiras | 1 | 0 | 1 | 2 | 0 | 0 | 1 | 2 | 7 |
| Campina Grande | 0 | 0 | 0 | 30 | 0 | 4 | 7 | 57 | 98 |
| Catolé do Rocha | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 2 |
| Conde | 0 | 0 | 1 | 2 | 0 | 0 | 1 | 2 | 6 |
| Cuité | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 2 | 7 |
| Esperança | 0 | 0 | 0 | 2 | 0 | 1 | 0 | 1 | 4 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| Sapé | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 15 | 18 |
| Sousa | 0 | 0 | 1 | 4 | 0 | 0 | 0 | 4 | 9 |
| Total | 17 | 9 | 15 | 103 | 1 | 18 | 24 | 381 | 568 |

A AC em sua teoria utiliza a distância euclidiana para calcular as distâncias entre linhas e colunas da matriz de correspondência. Contudo, utilizou-se da distância de Mahalanobis para o cálculo das distâncias por considerar a variabilidade dos pontos no espaço $\Re^2$, modificando um poucos a distribuição espacial dos pontos fornecidas pelo método AC. O gráfico bidimensional com os pontos de municípios e causas mostram a proximidade entre esses pontos, entretanto não é capaz de mostrar os agrupamentos dos pontos, pois não existe nenhuma medida eficiente para agrupar tais dados.

Os pares de coordenadas de cada ponto foram padronizados da forma que segue:

$$Q(\underset{\sim}{X^*}_1; \underset{\sim}{X^*}_2) = Q\left(\frac{\underset{\sim}{X}_1 - \mu_1}{\sigma_1}; \frac{\underset{\sim}{X}_2 - \mu_2}{\sigma_2}\right), \tag{6.1}$$

em que $\underset{\sim}{X}_1$ e $\underset{\sim}{X}_2$ são vetores das coordenadas obtidas pela análise de correspondência dos pontos, sejam eles os municípios ou causas, isto é, linha ou coluna da tabela de contingência. Os valores $\sigma_1$ e $\sigma_2$ se referem aos desvios padrões dos vetores $\underset{\sim}{X}_1$ e $\underset{\sim}{X}_2$ com médias $\mu_1$ e $\mu_2$ respectivamente. Com as coordenadas padronizadas, utilizou-se a Equação 3.4 para o cálculo das distâncias entre os pontos. Com tais coordenas estimadas é possível perceber de forma superficial o relacionamento entre os dados em gráficos perceptuais,

em que neste caso é bidimensional. A critério de exemplo, segue dois gráficos perceptuais com as coordenadas não padronizadas e com as coordenadas padronizadas, gráficos estes obtidos pela Análise de Correspondência para o ano de 2010, em que $D1$ e $D2$ refere-se as dimensões das linhas e colunas da tabela de contingência respectivamente, em que pontos são as causas e as estrelas os municípios.

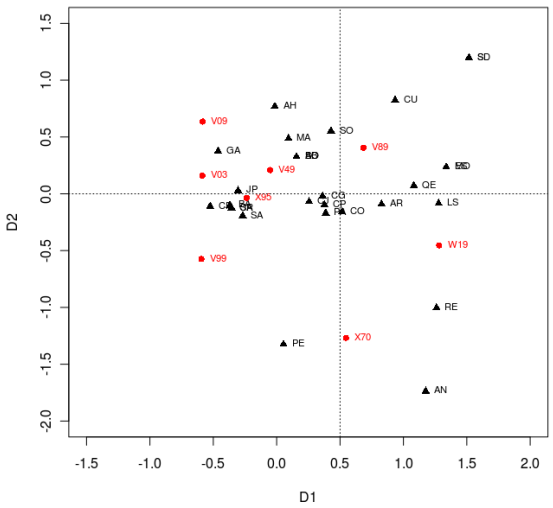



Figura 6: Distâncias Euclidiana.

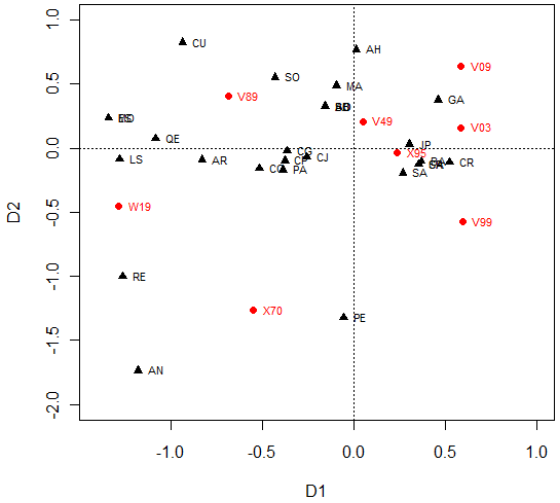



Figura 7: Distâncias de Mahalanobis.

Com a distribuição dos $n$ municípios e $c$ causas em um espaço $\Re^2$, buscou-se formar $c$ agrupamentos com base na medida de proximidade $d(M_i, C_{[1,\cdots,c]})$. Assim, se um município $M_1$ está mais próximo de um determinado grupo $c_{20}$ do que do conjunto $c_2$, por exemplo, o município $M_1$ terá maior chance de pertencer ao grupo $c_{20}$ do que ao $c_2$. Ainda assim, tal fato não impede que um indivíduo $M_1$ não venha a estar associado a mais de um agrupamento ao mesmo tempo, podendo o município $M_1$ pertencer aos agrupamentos $c_2$ e $c_{20}$ com mesma representatividade de pertinência para ambos agrupamentos. Dessa forma, percebe-se que os agrupamentos formados têm uma interpretação clara e bastante direta devido à característica inerente a cada centróide, onde todos os centróides que são os centros de cada um dos agrupamentos formados possuem uma expressão linguística associada à ele, como por exemplo, a causa Homicídio por Arma de Fogo referente a causa X95 do Capítulo XX da CID-10. Assim, o problema busca encontrar os melhores agrupamentos utilizando a filosofia de algoritmos genéticos em conjunto com a Lógica Nebulosa *Fuzzy*. Otimizando uma função objetivo que estabeleça os melhores agrupamentos, é possível tirar resultados bastante interessantes de cada agrupamento formado, resultados estes que são fáceis de serem observados porque, de forma bastante superficial, o problema apresenta uma contextualização bastante simplória que é agrupar regiões em conglomerados que possuem uma expressão linguística associada. Mas esse problema de otimização tem uma característica particular. Os centros (centróides) dos agrupamentos

que serão formados são fixos. Por isso, uma adaptação do método Fuzzy C-Means foi utilizada para tornar os centros dos grupos fixos.

## 6.4 Função Objetivo

A forma com que os agrupamentos são formados é dado através da minimização da função $J(U, D)$, onde $U$ é uma matriz retangular de dimensão $m \times c$ e D é a matriz de distâncias também $m \times c$. Essa função é apresentada na Equação 4.1, reescrita de outra forma.

$$J(U,D) = \sum_{i=1}^{n}\sum_{j=1}^{c}(u_{ij})^m \phi(x_i, d_j) = \sum_{i=1}^{n}\sum_{j=1}^{c}(u_{ij})^m \sum_{j=1}^{p}(x_{ij} - d_{ij})^2,$$

onde $u_{ij}$ é calculado da forma que segue com $m = 2$ que é o parâmetro de fuzzificação mais apropriado proposto pela literatura.

$$u_{ij} = \left[ \sum_{j=1}^{c} \left\{ \frac{\sum_{j=1}^{p}(x_{ij} - g_{ij})^2}{\sum_{j=1}^{p}(x_{ij} - g_{ij})^2} \right\}^{\frac{1}{m-1}} \right]^{-1} , i = 1, \cdots, n; \quad j = 1, \cdots, c.$$

Tal função objetivo do Método Híbrido Genético Fuzzy é a mesma função objetivo do método *Fuzzy C-Means* (FCM), no entanto os valores $g_{ij}$ não são atualizados, ou seja, os centróides não sofrem alterações em cada interação do algoritmo como ocorre no método FCM. No Método Híbrido Genético Fuzzy os $g_{ij}$ são pré-estabelecidos e obtidos pela análise de correspondência. Logo o que resta otimizar é a matriz de $U_{m\times c}$ que é a matriz de pertinências em que cada linha da matriz $U_{m\times c}$ apresenta as pertinências de um determinado indivíduo (município) pertencer a cada um dos $c$ agrupamentos formados. É importante observar que o espaço paramétrico da função $J(U, D)$ é bastante complicado, pois os valores que a função objetivo assumem são matrizes de dimensões $m \times c$. Dessa modo, tem-se que encontrar as melhores combinações entre linhas e colunas da matriz de pertinência $U_{m\times c}$ que melhor representem as pertinências de cada indivíduo pertencer a um dado agrupamento $c$.

É importante observar que as distâncias dos municípios para as causas (grupos) são fixas, o que não ocorre no *Fuzzy C-Means*, em que as distâncias dos indivíduos mudam a cada interação devido os centróides sofrerem um processo estocástico. A função objetivo

para o Modelo Híbrido Genético Fuzzy pode ser reescrita da forma que segue com $m = 2$.

$$J(U, D) = \sum_{j=1}^{c} \sum_{i=1}^{n} (u_{ij})^2 (d_{ij})^2, \tag{6.2}$$

onde $u_{ij}$ representam as pertinências de um indivíduo $j$ pertencer a um dado agrupamento $i$ e $(d_{ij})^2$ representam as distâncias de um indivíduo $j$ a um agrupamento $i$, distâncias estas fixas em todas as gerações.

Dessa forma, cada matriz $U$ fornecerá, por meio do cálculo de $J(U, D)$, uma medida de qualidade dos agrupamentos formados, de modo que as matrizes $U$ desejadas serão aquelas que tornarem a função objetivo com valor menor possível. Na Seção 6.5, logo abaixo, serão discutidos todos os detalhes da implementação do modelo de otimização proposto.

## 6.5 Implementação do Modelo Híbrido Genético Fuzzy

### 6.5.1 Passos do Modelo Híbrido Genético Fuzzy - MHGF

- **Passo 1**: Definir os parâmetros iniciais do modelo. Tais parâmetros são: População inicial $P_0$ em que é definido a quantidade de soluções candidatas que pertencerão à geração $t$, probabilidade de recombinação $P_r$, probabilidade de mutação $P_M$, número de gerações $n_g$, precisão de convergência $\epsilon$, $n_{p_0}$ que é o tamanho da população inicial $(P_0)$, $n_e$ que é o número de soluções de elite e $n_{ss}$ que é o critério de parada.

  - As probabilidades de *crossover* $(P_r)$ e mutação $(P_M)$ pode variar a depender do problema.
  - O parâmetro $n_{p_0}$ refere-se ao número de soluções de elites que irão iniciar a população inicial armazenadas no *array* $P_0$. Em geral escolhe-se um número grande de soluções iniciais de forma a garantir uma maior variabilidade nas gerações, possibilitando assim encontrar soluções mais diversificadas.

- **Passo 2**: Gerar aleatoriamente $n_{p_0}$ e abrigá-las no *array* $P_0$.

  - As soluções candidatas devem ser geradas de forma aleatória de forma que cada linha da matriz candidata $(U_{n\times c})$some 1.

- **Passo 3**: Calcule a função objetivo para todos os indivíduos da população $P_0$, obtendo-se um vetor de soluções $F_O$.
  - Para cada solução candidata será aplicado a função objetivo $J(U, D)$ e os valores serão guardados no vetor $F_O$.

- **Passo 4**: Ordene, de forma crescente, os valores de $F_O$ de modo que os primeiros indivíduos serão aqueles mais adaptados, ou seja, com menor função objetivo. A ordenação deve ser feita de modo que as posições iniciais antes da ordenação sejam guardadas.

- **Passo 5**: Guarde as melhores soluções em um vetor de elite E de tamanho $n_e$, bem como as matrizes de pertinências $U_{n\times c}$ que deram origem a essas soluções ótimas. Essa espécie de "cópia de segurança" das melhores soluções candidatas é uma estratégia de programação para não perder, ao longo da evolução do algoritmo, esses indivíduos.

- **Passo 6**: Selecione, aleatoriamente, 2 indivíduos (2 soluções candidatas) para avançarem para a fase da reprodução, de modo que tais indivíduos selecionados tenham probabilidades de seleção proporcionais à sua função objetivo, ou seja, indivíduos mais aptos (com valores menores para a função objetivo) tenham mais chances de serem selecionados.

- **Passo 7**: Gere um número com distribuição $Bernoulli(P_r)$, sendo 1 o sucesso e 0 o fracasso. Caso o número gerado seja 1 haverá recombinação (crossover) entre as soluções candidatas. Em havendo o processo de recombinação, siga os sub-passos abaixo:
  - Para cada solução candidata $(U_{n\times c})$ selecionados ${U^{(1)}}_{n\times c}$ e ${U^{(2)}}_{n\times c}$ da população $P_0$ faça $R_1 = ({U^{(1)}}_{n\times c})^2 \otimes (d_{n\times c})^2$ e $R_{2n\times c} = ({U^{(2)}}_{n\times c})^2 \otimes (d_{n\times c})^2$, ou seja, $R_{1n\times c}$ e $R_{2n\times c}$ é o produto entre os elementos da matriz $(U_{n\times c})^2$ pela matriz de distâncias ao quadrado $((d_{n\times c})^2)$.
  - Para cada matriz $R_{n\times c}$ calcula a soma de das $n$ linhas. Linhas que possuem uma soma baixa é denominada linha boa pois contribuem para minimização da função objetivo $J(U, D)$ e linhas com maiores somas são denominadas linhas ruins. Selecione a melhor e pior linha das matrizes ${R^{(1)}}_{n\times c}$ e ${R^{(2)}}_{n\times c}$ respectivamente e guarde suas posições. Considera as ilustrações do processo de recombinação apresentados abaixo.

– Multiplique os elementos da matriz $U^{(1)}{}_{n\times c}$ selecionada pela matriz $(d_{n\times c})^2$. Observe que não trata-se de produto matricial e que os elementos das matrizes estão elevados ao quadrado.

$$\mathbf{U^{(1)}} \qquad\qquad \mathbf{D}$$

$$\begin{bmatrix} \mu_{11}^{2(1)} & \mu_{12}^{2(1)} & \cdots & \mu_{1c}^{2(1)} \\ \mu_{21}^{2(1)} & \mu_{22}^{2(1)} & \cdots & \mu_{2c}^{2(1)} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{n1}^{2(1)} & \mu_{n2}^{2(1)} & \cdots & \mu_{nc}^{2(1)} \end{bmatrix} \otimes \begin{bmatrix} d_{11}^2 & d_{12}^2 & \cdots & d_{1c}^2 \\ d_{21}^2 & d_{22}^2 & \cdots & d_{2c}^2 \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1}^2 & d_{n2}^2 & \cdots & d_{nc}^2 \end{bmatrix}$$

$$\mathbf{U^{(1)} \otimes D} \qquad\qquad \mathbf{R_1}$$

$$= \begin{bmatrix} \mu_{11}^{2(1)} d_{11}^2 & \mu_{12}^{2(1)} d_{12}^2 & \cdots & \mu_{1c}^{2(1)} d_{1c}^2 \\ \mu_{21}^{2(1)} d_{21}^2 & \mu_{22}^{2(1)} d_{22}^2 & \cdots & \mu_{2c}^{2(1)} d_{2c}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{n1}^{2(1)} d_{n1}^2 & \mu_{n2}^{2(1)} d_{n2}^2 & \cdots & \mu_{nc}^{2(1)} d_{nc}^2 \end{bmatrix} \Rightarrow \begin{matrix} [1] \\ [2] \\ \vdots \\ [n] \end{matrix} \begin{bmatrix} \sum_{j=1}^{c} \mu_{1j}^{2(1)} d_{1j}^2 \\ \sum_{j=1}^{c} \mu_{2j}^{2(1)} d_{2j}^2 \\ \vdots \\ \sum_{j=1}^{c} \mu_{nj}^{2(1)} d_{nj}^2 \end{bmatrix}$$

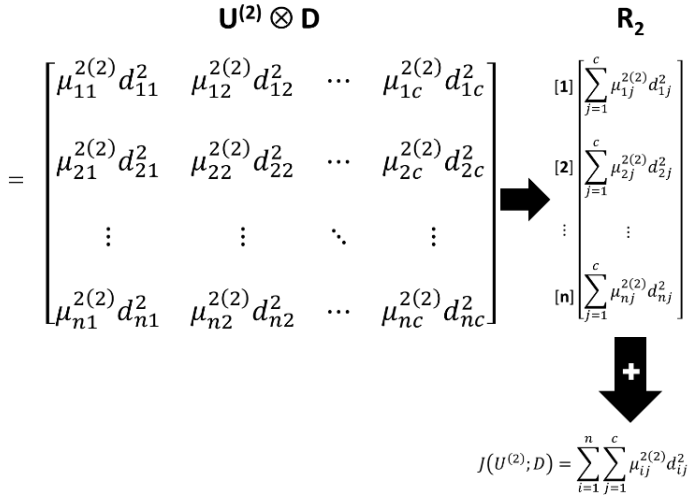
$$\Downarrow +$$

$$J(U^{(1)}; D) = \sum_{i=1}^{n} \sum_{j=1}^{c} \mu_{ij}^{2(1)} d_{ij}^2$$

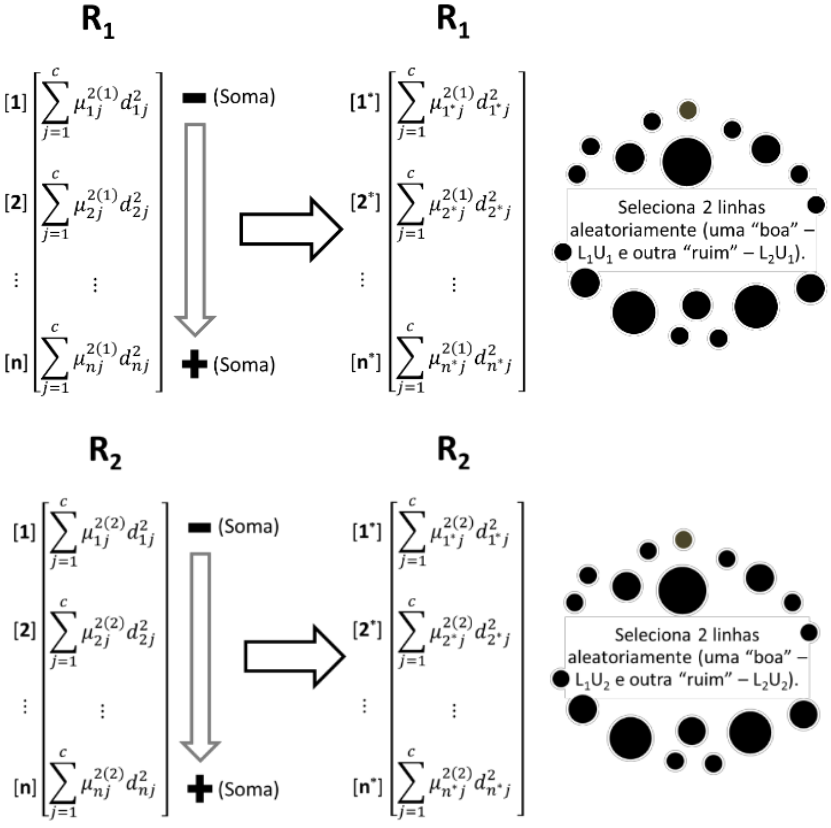
– Realize o mesmo procedimento acima para a matriz $(U_{n\times c})^2$, ou seja:

$$\mathbf{U^{(2)}} \qquad\qquad \mathbf{D}$$

$$\begin{bmatrix} \mu_{11}^{2(2)} & \mu_{12}^{2(2)} & \cdots & \mu_{1c}^{2(2)} \\ \mu_{21}^{2(2)} & \mu_{22}^{2(2)} & \cdots & \mu_{2c}^{2(2)} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{n1}^{2(2)} & \mu_{n2}^{2(2)} & \cdots & \mu_{nc}^{2(2)} \end{bmatrix} \otimes \begin{bmatrix} d_{11}^2 & d_{12}^2 & \cdots & d_{1c}^2 \\ d_{21}^2 & d_{22}^2 & \cdots & d_{2c}^2 \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1}^2 & d_{n2}^2 & \cdots & d_{nc}^2 \end{bmatrix}$$

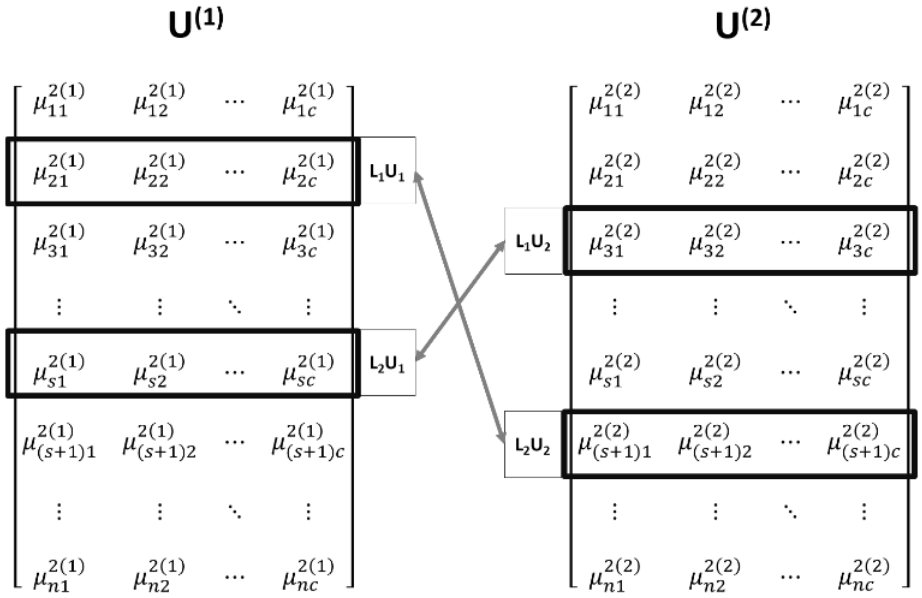



– Em seguida ordene as matrizes produtos $R^{(1)}{}_{n\times c}$ e $R^{(2)}{}_{n\times c}$ e posteriormente selecione uma linha boa e uma linha ruim utilizando os mesmos critérios de seleção das soluções candidatas.




– Seleciona uma linha boa e uma linha ruim para cada uma das matrizes $R^{(1)}{}_{n\times c}$ e $R^{(2)}{}_{n\times c}$. Troque a linha boa da matriz $R^{(1)}{}_{n\times c}$ pela linha ruim da matriz $R^{(2)}{}_{n\times c}$ e a linha ruim de $R^{(1)}{}_{n\times c}$ pela linha boa de $R^{(2)}{}_{n\times c}$. Com dessa mudança entre linhas de $R^{(1)}{}_{n\times c}$ e $R^{(2)}{}_{n\times c}$ é possível que haja troca de informações entre as soluções candidatas. A troca de uma linha boa por uma linha ruim pode vir

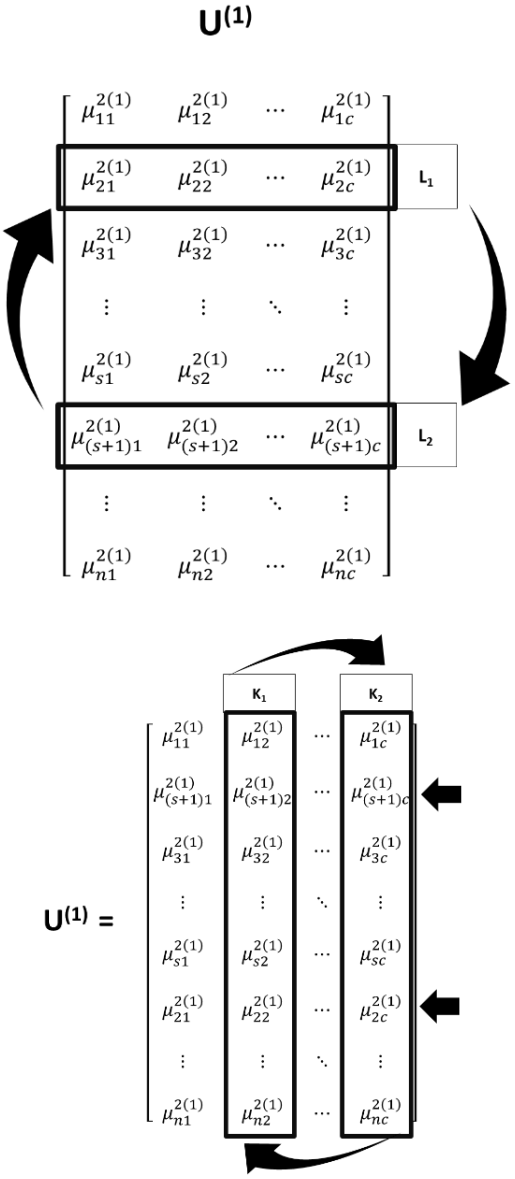
a melhorar as soluções candidatas, todavia isso não necessariamente ocorre. O que é importante é entender que a troca de material genético deve ocorrer, sendo isso suficiente para o algoritmo. Este procedimento pode ser resumido da seguinte forma.




– Para $Bernoulli(P_M) = 1$, pegue as posições das matrizes ${R^{(1)}}_{n\times c}$ e ${R^{(2)}}_{n\times c}$ obtidas no passo anterior e troque a linha ruim da matriz ${U^{(1)}}_{n\times c}$ com a linha boa da matriz $U^{(2)}$ e a linha boa da matriz ${U^{(1)}}_{n\times c}$ com a matriz linha ruim da matriz ${U^{(2)}}_{n\times c}$. Observe que a troca é feita nas matrizes $U_{n\times c}$ e não na matriz $R_{n\times c}$. A matriz $R_{n\times c}$ serve apenas para julgar quais linhas das matrizes $U_{n\times c}$ são boas ou ruim para minimização da função objetivo $J(U, D)$.

- **Passo 8**: Gere 2 números aleatórios independentes com distribuição $Bernoulli(P_M)$. Os números gerados referem-se à possibilidade de mutação entre as soluções candidatas.

  – Selecione duas linhas e duas colunas da matriz que sofrerá mutação para serem trocadas. Observe que a mutação se dá por uma desordenação dos elementos genéticos da própria solução candidata. O procedimento pode ser resumido da forma que segue.




- **Passo 19**: Para aqueles que irão sofrer mutação, ou seja, para aqueles que tiveram $Bernoulli(P_M) = 1$ o processo de mutação será iniciado.
- **Passo 20**: Gere 2 números aleatórios independentes com distribuição $Uniforme(1, n)$ de modo a selecionar duas das $n$ linhas das matrizes de pertinências. As linhas selecionadas serão trocadas.
- **Passo 21**: Gere mais dois números aleatórios independentes com distribuição $Uniforme(1, c)$. Os dois valores geradores referem-se as colunas das matrizes de pertinências que serão trocadas, finalizando assim o processo de mutação.

- **Passo 9**: Recalcule a função objetivo $J(U, D)$ para cada uma das matrizes de pertinências.

- **Passo 10**: Compare as funções objetivos, obtidas após os operadores genéticos de recombinação e/ou mutação com os valores das funções objetivo dos indivíduos da população $P_0$. Caso os valores das funções objetivos encontrados são melhores que ao menos uma das funções objetivo calculadas para $P_0$, as novas matrizes substituirão as "piores" soluções candidatas de $P_0$. Este mesmo processo também é aplicado para as soluções de elite.

- **Passo 11**: Repita este processo a partir do Passo 2 até que número número de gerações indicado no Passo 1 seja atingido ou até que o primeiro valor da solução de elite seja menor que $\epsilon$ para um número razoável de gerações consecutivas, ou seja, até que $n_{ss}$ seja maior que um valor pré-estabelecido.

- **Passo 12**: Ao final de todas as gerações, quando o critério de parada do for atingido, as soluções do grupo de elite serão as melhores soluções encontradas.

No capítulo que segue, uma aplicação da metodologia proposta será efetuada e os principais resultados serão discutidos.

# 7 Aplicações e Discussões

## 7.1 Delineamento do Processo de Amostral de Simulação

A base de dados considerada nesse trabalho refere-se à população de indivíduos que vieram à óbito dos 30 principais municípios paraibanos registrados entre os anos de 2006 a 2010. Vale salientar que os dados considerados refletem o universo completo de casos, não sendo necessário, portanto, se faz inferências no que diz respeito ao tamanho da amostra ou outras características amostrais. Como já comentado nesse texto, o objeto principal desse trabalho é agrupar os municípios dentro das causas estudadas para os anos de 2006 a 2010. Esses agrupamentos dependerão da escolha ótima da matriz de pertinência que relaciona os municípios e as causas. E para a obtenção dessa escolha ótima, será utilizado um Modelo Híbrido Genético Fuzzy proposto nesse trabalho.

Para a execução do processo de simulação, considerou-se uma população inicial de 1000 soluções candidatas, descritas aqui como as matrizes de pertinência $U_{n\times c}$. Além disso, foi considerado $n_g = 50.000$ evoluções genéticas, sendo preservadas, no grupo de soluções de elite, 20 soluções. Já para a aplicação dos operadores genéticos, foram considerados uma probabilidade de recombinação de 85% e de mutação de 5%. Como critério de parada, utilizou-se $n_{ss} = 100$ com um $\epsilon = 10^{-3}$.

O problema de simulação abordado nesse documento é bastante complexo, principalmente devido à alta dimensionalidade do subespaço de busca da função objetivo $J(U; D)$ que se encontra no espaço vetorial $R^{240}$ dimensional. Mesmo assim, os resultados obtidos foram bastante consistentes para a grande maioria dos municípios, não deixando de haver também, algumas ocorra eventuais inconsistências, o que não deixa de ser compreensível, haja vista a complexidade do problema a ser otimizado.

Por outro lado, em termos de desempenho, utilizou-se um computador com processador Intel(R) Core(TM) 2 Duo CPU, T5550, 1.83GHz e observou-se que, em média, o

algoritmo consumiu cerca de 4,45 horas para gerar os resultados para cada ano, sendo que em torno de 76% desse tempo, o algoritmo consumiu realizando as tarefas de recombinação, como mostra a Tabela 5.

Tabela 6: Tempo de convergência em horas do algoritmo MHGF.

| Tempo em Hora | | | |
|---|---|---|---|
| Anos | Total | Cross-over | Mutação |
| 2006 | 4,302651 | 3,011856 | 1,290795 |
| 2007 | 4,658944 | 3,727155 | 0,931789 |
| 2008 | 4,268316 | 2,987821 | 1,280495 |
| 2009 | 4,254411 | 2,935544 | 1,318867 |
| 2010 | 4,752415 | 4,277174 | 0,475242 |

## 7.2 Relação entre Municípios e Causas Externas de Óbitos

Para obter as proximidades entre os municípios e as causas externas consideradas, foi empregada a técnica de Análise de Correspondência Simples (AC). O método AC permitiu transformar medidas de contagem de variáveis categóricas em relacionamentos (distâncias) entre coordenadas contínuas dos municípios para as causas. Abaixo é apresentada uma tabela que ilustra as contingências observadas para municípios e causas externas para o ano de 2006.

Pelos dados constantes na Tabela 6 não é tão direta a análise conjunta de quais relações entre município e causa são mais significativas, pois não existe uma medida adequada de dissimilaridade. Mas após a construção do gráfico perceptual (Figura 6), tudo fica mais claro e direto. Observe nesse gráfico, por exemplo, que os municípios de Jõao Pessoa (JP), Santa Rita (SR), Bayeux (BA) e Cabedelo (CA) estão todos inter-relacionados com a causa X95, uma vez que todos esses municípios estão próximos dessa causa. Logo, o gráfico perceptual da AC fornece avaliações diretas de dissimilaridades entre os elementos estudados simplesmente analisando os seus padrões de distâncias.

Dessa forma, as coordenadas produzidas pela Análise de Correspondência para os municípios e causas foram utilizadas para se calcular as distâncias (expressão 6.1) entre esses dois conjuntos de elementos. A partir daí, essa matriz de distâncias fomentou a geração dos resultados simulados, através da execução do modelo de agrupamento proposto. As demais tabelas de contingências podem ser encontradas no Anexo A.4.

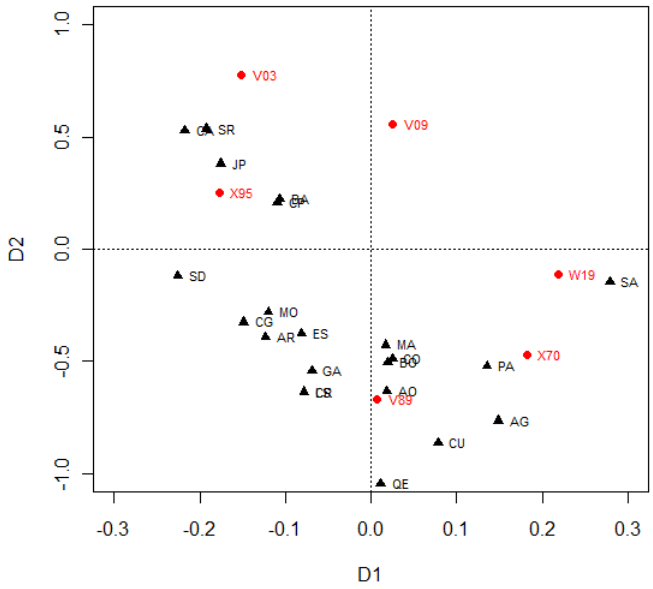
Tabela 7: Tabela de Contingência para o ano de 2006.

| | Causas | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Municipio | V03 | V09 | V49 | V89 | V99 | W19 | X70 | X95 |
| AG | 0 | 0 | 0 | 4 | 0 | 3 | 1 | 1 |
| AN | 0 | 0 | 0 | 7 | 0 | 0 | 0 | 0 |
| AH | 1 | 0 | 3 | 0 | 0 | 0 | 0 | 1 |
| AR | 0 | 0 | 0 | 3 | 0 | 0 | 1 | 4 |
| AO | 0 | 0 | 0 | 3 | 0 | 1 | 1 | 2 |
| BA | 4 | 1 | 1 | 10 | 0 | 2 | 0 | 32 |
| BO | 0 | 0 | 0 | 2 | 0 | 1 | 1 | 2 |
| CP | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 2 |
| CA | 3 | 2 | 0 | 2 | 0 | 1 | 0 | 11 |
| CJ | 0 | 0 | 3 | 2 | 0 | 3 | 3 | 0 |
| CG | 0 | 2 | 0 | 78 | 0 | 5 | 8 | 102 |
| CR | 0 | 0 | 0 | 4 | 0 | 0 | 1 | 3 |
| CO | 1 | 0 | 0 | 3 | 1 | 0 | 1 | 3 |
| CU | 0 | 0 | 0 | 2 | 0 | 0 | 2 | 1 |
| ES | 0 | 0 | 0 | 5 | 0 | 1 | 2 | 7 |
| GA | 0 | 0 | 0 | 4 | 0 | 1 | 0 | 3 |
| JP | 38 | 4 | 4 | 38 | 0 | 7 | 11 | 222 |
| LS | 0 | 0 | 0 | 4 | 0 | 0 | 1 | 3 |
| MA | 0 | 1 | 0 | 2 | 0 | 0 | 1 | 1 |
| MO | 0 | 0 | 0 | 3 | 0 | 0 | 2 | 6 |
| PA | 0 | 0 | 1 | 13 | 0 | 1 | 6 | 12 |
| PE | 1 | 0 | 1 | 1 | 1 | 1 | 2 | 1 |
| QE | 0 | 0 | 0 | 13 | 0 | 0 | 3 | 3 |
| RE | 0 | 1 | 0 | 9 | 2 | 0 | 0 | 1 |
| SR | 6 | 8 | 0 | 2 | 0 | 4 | 3 | 38 |
| SB | 0 | 2 | 1 | 1 | 0 | 0 | 2 | 7 |
| SA | 1 | 0 | 1 | 3 | 3 | 0 | 0 | 11 |
| SL | 0 | 0 | 1 | 7 | 0 | 0 | 0 | 3 |
| SD | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 2 |
| SO | 0 | 1 | 8 | 4 | 0 | 0 | 0 | 5 |

## 7.3 Resultados e Discussões

Com as distâncias entre municípios e causas externas de óbito foi possível obter as pertinências de cada município para com todas as causas de óbito através do Modelo Híbrido Genético Fuzzy (MHGF). Com essas matrizes de pertinência, basta escolher um ponto de corte ($P_c$) para que se tenha os grupos de municípios formados para cada uma das causas. Convém informar que esses Os agrupamentos são bastantes sensíveis ao ponto de corte escolhido, em que diferenças por menor que sejam podem favorecer a entrada ou saída de um município dentro de um *cluster*.

Figura 8: Gráfico Perceptual para o Ano de 2006.




Nesse trabalho decidiu-se definir os pontos de corte 0,10, 0,20, 0,30 e 0,50, no sentido de facilitar o estudo de possíveis impactos dessas escolhas sobre os agrupamentos formados. Por isso, resolveu-se apresentar os resultados comparando os agrupamentos formados para cada um dos pontos de corte considerados em cada ano. A essas análises chamou-se Análise de Sensibilidade e será discutida na Seção 7.3.1. Por outro lado, pensando na evolução temporal das inter-relações entre causas e os principais municípios paraibanos, para um ponto de corte específico, tais municípios foram avaliados. A esse conjunto de resultados, denominou-se Análise Longitudinal que será apresentada na Seção 7.3.2.

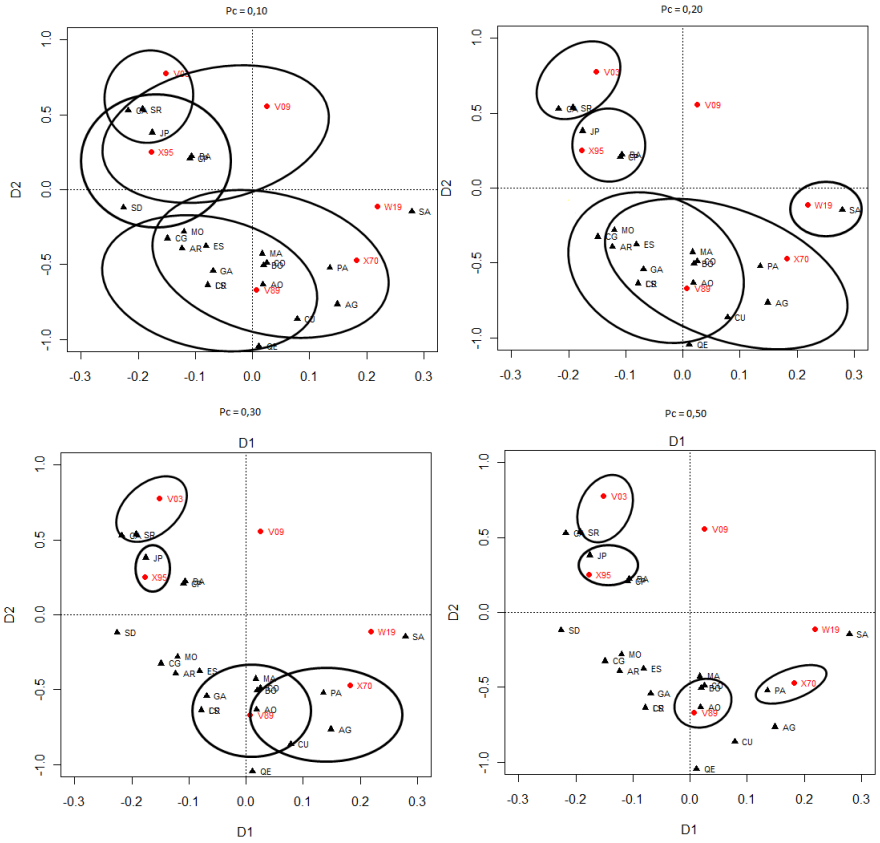
### 7.3.1 Análise de Sensibilidade dos Agrupamentos Formados

A análise de sensibilidade é importante para esclarecer o quanto a escolha de um ponto de corte influência os agrupamentos formados. A análise será feita de forma gráfica, em que para cada ano foi construído gráficos perceptuais para cada ponto de corte sugerido. Mais adiante seguem os gráficos para os anos de 2006 e 2007 segundo os 4 pontos de corte analisados. Observa-se pelos resultados apresentados que à medida que o ponto de corte cresce o número de municípios alocados à cada uma das 8 causas decresce, o que é muito natural, uma vez que aumenta-se a exigência para que um determinado município venha a pertencer a uma certa causa, o número de municípios propensos diminui. Os gráficos para os demais anos estão apresentados nos anexos desse trabalho. Ainda com relação aos gráficos ilustrados nas Figuras 7 e 8, alguns municípios ou causas não aparecem pois as escalas foram diminuídas para que fosse possível visualizar os agrupamentos formados

numa escala mais extendida. Também não foram circulados todos os agrupamentos para que a visualização não fosse prejudicada. De qualquer forma os agrupamentos formados podem ser observados nas Tabelas de Agrupamentos apresentadas no Anexo A.12, em que o código "1" indica que o município pertence à causa considerada com base no ponto de corte escolhido. Claramente nota-se que a escolha do ponto de corte impacta, e muito, os agrupamentos formados.

Com o ponto de corte $P_c = 0,10$ , por exemplo, o cluster formado pela causa $V09$ (Pedestre traumatizado em acidentes de transporte não especificado) engloba, também, todos os municípios pertencentes ao grupo formado pela causa $V03$ (Pedestre traumatizado em colisão com automóvel). Já para $P_c = 0,20$ os agrupamentos gerados parecem ser mais consistentes com um número de municípios participantes um pouco menor. Para $P_c = 0,30$, $P_c = 0,50$ os agrupamentos formados se mostram mais rigorosos, aproximando-se dos métodos tradicionais que utilizam a filosofia "Hard" (um município é alocado para uma, e somente uma, causa).

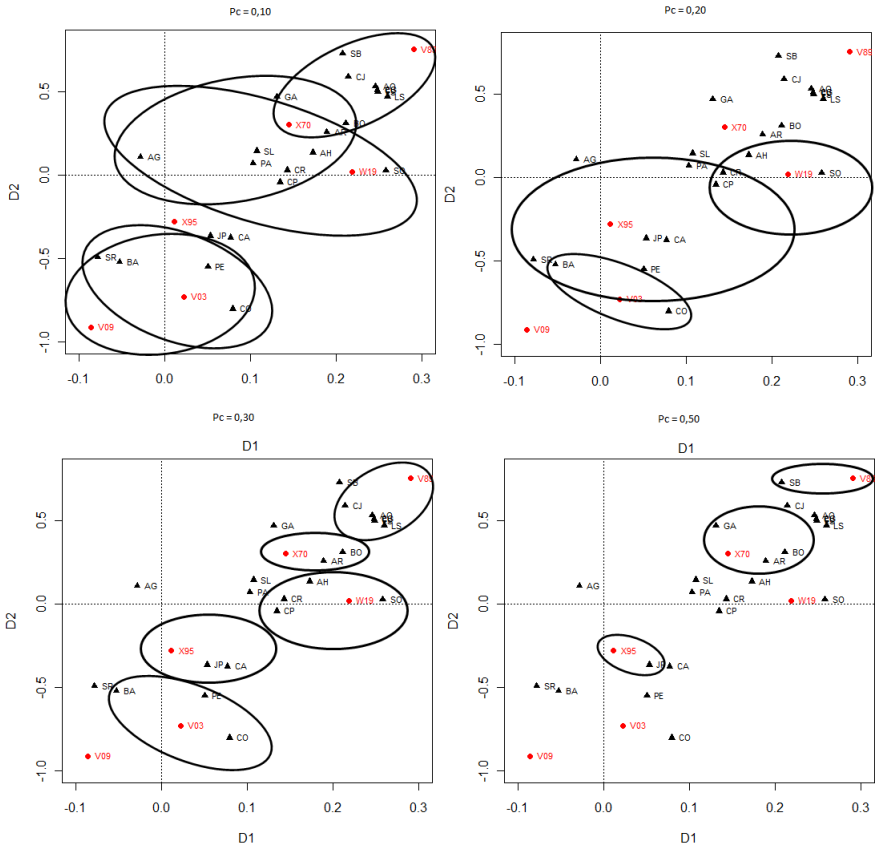
Figura 9: Gráficos para os 4 pontos de cortes considerados, 2006.




De um modo geral, o aumento da exigência, em relação ao ponto de corte, faz com que todos os municípios que pertencem aos grupos considerados de causas externas apresentam, no mínimo, uma inter-relação bastante forte com as respectivas causas. Ainda

assim, eventualmente, alguns municípios que apresentam uma distância razoavelmente pequena de uma dada causa não foi incorporado à essa causa, o que era para acontecer. Mas esses resultados são compreensíveis dada a tamanha complexidade do objeto a ser otimizado, que são as matrizes de pertinência.

Figura 10: Gráficos para os 4 pontos de cortes considerados, 2007.




Para o ano de 2007 também se observa que à medida que o ponto de corte vai aumentando os contornos definem os agrupamentos agrupamentos em torno das causas externas de óbitos vão diminuindo apresentando *clusters* mais coerentes com os pontos de cortes $P_c = 0,20$ ou $P_c = 0,30$. A Tabela de pertinências para o no de 2007, bem como para os demais anos estão no Anexo A.18.

Tabela 8: Pertinências Obtidas pelo Modelo Híbrido Genético Fuzzy (MHGF) para o ano de 2006.

| | PERTINÊNCIAS | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Município** | **V03** | **V09** | **V49** | **V89** | **V99** | **W19** | **X70** | **X95** |
| **AG** | 0,03075 | 0,02635 | 0,02021 | 0,55357 | 0,05922 | 0,05031 | 0,22818 | 0,03142 |
| **AN** | 0,04658 | 0,08117 | 0,03310 | 0,23504 | 0,21898 | 0,08080 | 0,25291 | 0,05142 |
| **AH** | 0,05378 | 0,05760 | 0,61335 | 0,04038 | 0,08655 | 0,04880 | 0,04882 | 0,05072 |
| **AR** | 0,03536 | 0,10576 | 0,02969 | 0,35887 | 0,06967 | 0,13547 | 0,16913 | 0,09605 |
| **AO** | 0,02061 | 0,01140 | 0,01527 | 0,55480 | 0,01804 | 0,02773 | 0,25502 | 0,09713 |
| **BA** | 0,08763 | 0,06264 | 0,01013 | 0,02384 | 0,01492 | 0,05345 | 0,02861 | 0,71878 |
| **BO** | 0,02750 | 0,03849 | 0,01527 | 0,60296 | 0,04116 | 0,06094 | 0,16120 | 0,05248 |
| **CP** | 0,04369 | 0,14816 | 0,00997 | 0,11362 | 0,07855 | 0,27135 | 0,03696 | 0,29771 |
| **CA** | 0,32223 | 0,18362 | 0,02528 | 0,06495 | 0,04611 | 0,04426 | 0,08140 | 0,23217 |
| **CJ** | 0,08965 | 0,13236 | 0,10119 | 0,16368 | 0,13638 | 0,09120 | 0,18195 | 0,10359 |
| **CG** | 0,04673 | 0,04368 | 0,02302 | 0,29971 | 0,03607 | 0,09600 | 0,30469 | 0,15009 |
| **CR** | 0,02465 | 0,03437 | 0,01447 | 0,71602 | 0,03994 | 0,04197 | 0,08208 | 0,04650 |
| **CO** | 0,02128 | 0,02589 | 0,00881 | 0,15026 | 0,12619 | 0,08342 | 0,52909 | 0,05505 |
| **CU** | 0,04140 | 0,03373 | 0,01985 | 0,23480 | 0,10536 | 0,05189 | 0,41416 | 0,09881 |
| **ES** | 0,04228 | 0,07992 | 0,02255 | 0,24708 | 0,13588 | 0,11156 | 0,29931 | 0,06141 |
| **GA** | 0,02761 | 0,03705 | 0,01020 | 0,30467 | 0,05290 | 0,23571 | 0,29226 | 0,03960 |
| **JP** | 0,11684 | 0,10335 | 0,01172 | 0,02530 | 0,01673 | 0,05030 | 0,02928 | 0,64648 |
| **LS** | 0,02883 | 0,02099 | 0,01864 | 0,46794 | 0,03033 | 0,04520 | 0,27780 | 0,11028 |
| **MA** | 0,02863 | 0,05567 | 0,02261 | 0,16756 | 0,11140 | 0,17754 | 0,32501 | 0,11158 |
| **MO** | 0,04106 | 0,07134 | 0,02524 | 0,14215 | 0,11403 | 0,20136 | 0,26264 | 0,14218 |
| **PA** | 0,03203 | 0,02368 | 0,01402 | 0,14074 | 0,06289 | 0,04597 | 0,59336 | 0,08731 |
| **PE** | 0,10977 | 0,09367 | 0,04740 | 0,10280 | 0,22012 | 0,12355 | 0,14408 | 0,15862 |
| **QE** | 0,04101 | 0,07576 | 0,02744 | 0,27527 | 0,19233 | 0,07885 | 0,26268 | 0,04666 |
| **RE** | 0,04149 | 0,07610 | 0,02962 | 0,23819 | 0,23811 | 0,07712 | 0,25374 | 0,04564 |
| **SR** | 0,53184 | 0,17564 | 0,00987 | 0,03101 | 0,03862 | 0,04121 | 0,02938 | 0,14244 |
| **SB** | 0,06763 | 0,19531 | 0,01882 | 0,13083 | 0,09361 | 0,32485 | 0,06205 | 0,10690 |
| **SA** | 0,08659 | 0,04020 | 0,00904 | 0,15430 | 0,06876 | 0,29913 | 0,21862 | 0,12338 |
| **SL** | 0,02775 | 0,03345 | 0,01537 | 0,09941 | 0,18158 | 0,09242 | 0,49118 | 0,05884 |
| **SD** | 0,05081 | 0,14617 | 0,01545 | 0,15790 | 0,09584 | 0,31653 | 0,08399 | 0,13332 |
| **SO** | 0,05198 | 0,05716 | 0,60973 | 0,04095 | 0,08886 | 0,05077 | 0,05044 | 0,05012 |

## 7.4 Análise Longitudinal das Inter-Relações entre Municípios e Causas

As causas externas de óbito vêm evoluindo não necessariamente de forma crescente ao longo dos anos. A cada ano, novas estruturas de inter-relacionamento entre municípios e causas externas são formadas. Devido o comportamento dinâmico das causas externas, faz-se necessário avaliar tais inter-relacionamentos segundos os anos. A análise se conterá em avaliar avaliar evolução temporal das causas que mais acomentem os principais

municípios da Paraíba. Para a análise não ficar muito tediosa e nem repetitiva, e ao mesmo tempo, para torna-la representativa do universo estudado, decidiu-se escolher os municípios de João Pessoa, Campina Grande, Cabedelo e Bayeux. Eventuais municípios que merecerem algum destaque poderão ser citados na análise. O ponto de corte escolhido para formar os agrupamentos foi $P_c = 0,20$. Toda a análise que será feita em cima desses quatro municípios pode ser naturalmente reproduzidas para os demais municípios paraibanos considerados neste trabalho.

Desse modo, foi observado que para o ponto de corte adotado, o município de João Pessoa sempre esteve associado à causa $X95$ referente a homicídios por arma de fogo, para todos os anos. Ainda para o ano de 2006, o município que apresentou inter-relacionamento com os óbitos por homicídios provocados por arma de fogo, além de João Pessoa e Bayeux foi o município de Caaporã. Todos os resultado aqui comentados podem ser encontrados na tabela abaixo para os 5 anos considerados para $P_c = 0,20$.

Tabela 9: Evolução das inter-relações entre municípios e causas externas para o ponto de corte 0,20.

| Ano | Município | | | |
|---|---|---|---|---|
| | JOÃO PESSOA | CAMPINA GRANDE | CABEDELO | BAYEUX |
| 2006 | X95 | X70; V89 | X95; V03 | X95 |
| 2007 | X95 | V89 | W19; X95 | V03 |
| 2008 | X95 | W19; X95 | X95 | X95 |
| 2009 | X95 | V89 | V03 | V09 |
| 2010 | X95 | V89; X95 | V03; V99; X95 | X95 |

Ainda em 2006, observa-se que o município de Campina Grande (CG) apresenta associação com a causa $V89$ referente à acidentes com veículos a motor ou não motorizados de tipo não especificado. O município de Cabedelo (CA), apesar de relativamente próxima da causa $X95$, foi incluído no *cluster* formado pela causa $V03$ referente à pedestre traumatizado em colisão com automóvel, juntamente com o município de Santa Rita (SR). No ano de 2007 observou-se que mais municípios foram incorporados ao *cluster* formado pela causa de homicídio provocado por arma de fogo. Os municípios que estão contidos nesse agrupamento são os mesmos do ano de 2006 (João Pessoa, Bayeux, Santa Rita e Caaporã) mais os municípios de Pedras de Fogo (PE), Patos (PA) e Catolé do Rocha. Para esse ano o município de Campina Grande (CG) apresentou mais relação com acidentes acidentes de transporte.

No ano de 2008, João Pessoa, Santa Rita, Cabedelo e o município de Patos (ver no anexo), continuam apresentando relações significativas com a causa de óbito por homi-

cídios por arma de fogo e o município Bayeux não foi mais incluído nesse agrupamento, mostrando possivelmente uma melhora nas políticas de seguranças públicas desse município. O município de Campina Grande continua apresentando inter-relacionamento mais forte com a causa $V89$. Já no ano de 2009 foi observado que apenas a cidade de João Pessoa apresentou um relacionamento com a causa de óbito $X95$ e os municípios de Santa Rita (SR), Caaporã (CA) e Bayeux ficaram bastante próximos mas não chegaram a pertencer a este agrupamento. O município de Campina Grande para este ano não apresentou inter-relacionamento com nenhuma causa de externa de óbito.

No ano de 2010, observou-se que muitos municípios que em 2009 não apresentaram inter-relacionamentos voltaram a refletir as mesmas relações dos anos anteriores à 2009. O município de Campina Grande voltou a apresentar um inter-relacionamento considerável com a causa externa de óbito $V89$ causa esta voltada aos acidentes de transporte. Observa-se que em nenhum momento João Pessoa esteve longe da causa de óbito de homicídio por arma de fogo ($X95$) mostrando sua forte ligação para esta prática de homicídio.

É importante verificar que uma característica muito peculiar do método MHGF é são as intersecções enter os agrupamentos mostrando a flexibilidade do modelo. Vale lembrar que interpretações espaciais podem ser tomadas partindo do pressuposto que os municípios dentro do mesmo agrupamento são vizinhos geograficamente, como é o caso dos municípios de João Pessoa, Bayeux, Cabedelo e Santa Rita que possivelmente formam um conglomerado espacial em torno da causa de óbito por homicídio provocados por arma de fogo.

# 8 Conclusões e Sugestões de Trabalhos Futuros

Este trabalho apresentou um novo método para agrupamento de dados utilizando a Lógica Nebulosa Fuzzy em paralelo com Algoritmos Genéticos. O método, em geral, apresentou resultados bastante satisfatórios, nos remetendo a agrupamentos que já eram esperados subjetivamente devido a vivência com os dados e com proximidades espaciais. Tais resultados também estão de acordo com os levantamentos feito pela Secretaria de Saúde da Paraíba, mostrando assim, uma coerência nos agrupamentos encontrados. Isso mostra que o modelo proposto pode ser também indicado quando se deseja identificar agrupamentos espaciais, em que não se faz necessário utilizar shapes de mapas geográficos, mesmo não sendo uma especialidade do metodologia sugerida. O modelo híbrido ainda pode ser recomendado para os casos em que se deseje identificar agrupamentos em torno de variáveis específicas, tais como PIB ou IDH, o que os demais métodos de agrupamentos não conseguem tratar. Em resumo, pode-se afirmar que a modelagem proposta nesse trabalho servirá como uma alternativa para estudos em que se necessite investigar padrões multivariados de inter-relacionamentos. Mas ainda não foi feito tudo em relação ao Modelo Híbrido Genético Fuzzy. Uma sugestão de trabalho futuro seria a otimização do parâmetro m de fuzzificação, que nesse trabalho, utilizou-se $m = 2$. Uma outra possibilidade seria a otimização do ponto de corte ($P_c$), no sentido de gerar grupos o mais homogêneo possível ou até mesmo a implementação de outros modelos híbridos que envolvam, por exemplo, (a) Algoritmo Scan com Algoritmo Genético; (b) Estimação de densidades Kernel com Algoritmos de Colônia de Formigas ou (c) Fuzzy C-Means com Redes Neurais.

# *ANEXO A – Códidos de Programação*

## A.1 Código em R do exemplo de otimização da função $f(\theta)$ do Capítulo 5

```
rm(list = ls(all = TRUE))
tempo = Sys.time()
require(Rlab) # Carrega o pacote para geracao de numeros pseudo-aleatorios Bernoulli
f <- function(teta,...){
6 + (teta^(2))*sin(14*teta)
}
# Contruindo a crusa da funcao f(teta)
#curve(6 + (x^2)*sin(14*x), -2.5, 2.5, xlab = expression(paste(theta)),
ylab = expression(f(paste(theta))))
geracao_inicial = teta = runif(8000,-2.5,2.5)
i <- 1 # Contador das geracoes
for(i in 1:300){
F0 = f(teta)
F0_D = sort(F0, decreasing = TRUE, index.return  = TRUE) # Funcao objetivo
#organizada de forma decrescente
teta_D = teta[F0_D$ix] # Valores de teta organizados de forma
#decrescente segundo a funcao objetivo no ponto teta.
Prop = abs(F0_D$x)/sum(abs(F0_D$x)) # Proporcoes
acumulado = cumsum(Prop) # Proporcao acumulada
Roleta = runif(1, min = min(acumulado), max = 1) # Numero pseudo-aleatorio no intervalo [0,1]
teta1 = min(acumulado[acumulado >= Roleta])
pteta1 = (1:length(acumulado))[acumulado == teta1] # Posicao do valor de teta
#selecionado pela roleta para o teta1
Roleta = runif(1, min = min(acumulado), max = 1)
teta2 = min(acumulado[acumulado >= Roleta])
pteta2 = (1:length(acumulado))[acumulado == teta2] # Posicao do valor de teta
#selecionado pela roleta para o teta2

##### CROSSOVER
cross = rbern(1, 0.8) # Se o valor e' 1 havera' crossover.
if(cross == 1){
F_teta1 = f(teta1)^(-1) # Funcao objetivo no valor de teta1 - 1
# Permite que o numero esteja no dominio
repeat{
teta1 = rnorm(1,teta1, F_teta1) # NUM4 vai ser o novo teta2
if(teta1 >= -2.5 && teta1 <= 2.5) break
}
```

```
F_teta2 = f(teta2)^(-1) # Funcao objetivo no valor de teta2 - 1

# Permite que o numero esteja no dominio
repeat{
teta2 = rnorm(1,teta2, F_teta2) # NUM5 vai ser o novo teta1
if(teta2 >= -2.5 && teta2 <= 2.5) break
}
#teta_D[c(pteta1,pteta2)] = c(teta2,teta1)   # Os tetas sao trocados como uma
#forma de troca de material genetico
#teta_D = teta = sort(teta_D, decreasing = TRUE, index.return  = F)
#F_tetas = f(teta_D[c(pteta1,pteta2)]) ################ Funcao objetivo dos
#dois novos valores otimos
F_tetas = f(c(teta2,teta1)) # Funcoes objetivos
F1 = F_tetas[1] # Funcao objetivo de teta1 caso exista um crossover
F2 = F_tetas[2] # Funcao objetivo de teta2 caso exista um crossover

}
#}
if(cross == 0){
F1 = F_teta1 = f(teta1) # Funcao objetivo de teta1 caso nao exista um crossover
F2 = F_teta2 = f(teta2) # Funcao objetivo de teta2 caso nao exista um crossover
}
#}
##### MUTACAO
mut = rbern(1, 0.1) # Se o valor e' 1 havera' mutacao.
if(mut == 1){
M1 = runif(1,-1/i,1/i) # Obs: O intervalo vai diminuindo a cada interacao
M2 = runif(1,-1/i,1/i)
teta1 = teta1 + M1
teta2 = teta2 + M2
if(teta1 < -2.5 || teta1 > 2.5){
teta1 = sample(size = 1, x = c(-2.5,2.5))
}
if(teta2 < -2.5 || teta2 > 2.5){
teta2 = sample(size = 1, x = c(-2.5,2.5))
}
F_tetas = f(c(teta2,teta1)) # Funcoes objetivos
F1 = F_tetas[1] # Funcao objetivo de teta1 caso exista uma mutacao
F2 = F_tetas[2] # Funcao objetivo de teta2 caso exista uma mutacao
}

FO = f(c(teta1,teta2)) # Vetor com as duas funcoes objetivos
PFO_MAX = which(FO == max(FO)) # Posicao da FO maxima
if(length(PFO_MAX) > 1) PFO_MAX = 1
PFO_MIN = which(FO == min(FO)) # Posicao da FO minima
if(length(PFO_MIN) > 1) PFO_MIN = 1
FO_MAX = FO[PFO_MAX]
logico = FO_MAX > F0_D$x
logico = length(logico[logico == TRUE]) # Caso maior ou igual a 1 existe ao
#menos um valor que seja menor que FO_MAX
TETA = c(teta1,teta2)
#Guardando as funcoes objetivos mais adaptadas e colocando na populacao
#para fazer parte do processo de roleta.
if(logico > 0){
```

```
#teta_D[c(pteta1,pteta2)] = c(teta2,teta1)   # Os tetas sao trocados como
#uma forma de troca de material genetico
#teta_D = teta = sort(teta_D, decreasing = TRUE, index.return  = F)
#F_tetas = f(teta_D[c(pteta1,pteta2)]) ############### Funcao objetivo
#dos dois novos valores otimos
F0_D$x[length(F0_D$x)] = FO_MAX
teta_D[length(F0_D$x)] = TETA[PFO_MAX]
F0_D$x = sort(F0_D$x, decreasing = TRUE, index.return  = TRUE)
teta_D = teta = teta[F0_D$ix]
}

}
Prop
acumulado
teta
F0_D$x
#f(2.360511947) =  10.71240
tempo =   Sys.time() - tempo
tempo
```

## A.2 Código em SAS do exemplo de otimização da função $f(\theta)$ do Capítulo 5

```
/*
================================================
=
=  ALGORITMO ALGORITMO GENÉTICO
=
=   AUTOR:   JOAB DE OLIVEIRA
=   CRIADO:   29/10/2010
=                ATUALIZADO: 02/11/2010
=
=
================================================

PARÂMETROS DE ENTRADA:

NIT;  * NÚMERO DE ITERAÇÕES;
PRR;  * PROBABILIDADE DE CROSSOVER(RECOMBINAÇÃO);
PRM;  * PROBABILIDADE DE MUTAÇÃO;
NSE; * NÚMERO DE SOLUÇÕES DE ELITE;
EPS; * EPSLON (ERRO ADMISSÍVEL DE PARADA);
TAM_POP; * TAMANHO DA POPULAÇÃO INICIAL;

*/

%MACRO GA_SIMPLES(NIT,PRR,PRM,NSE,EPS,TAM_POP);


DATA POP_INICIAL;
DO I=1 TO &TAM_POP;
```

```
TETA=-2.5+5*RANUNI(0);
FO=6+(TETA**2)*SIN(14*TETA);
OUTPUT;
END;
DROP I;
RUN;

PROC IML SYMSIZE=300 WORKSIZE=400;

USE POP_INICIAL;
READ ALL INTO POP_INICIAL;
NIT=&NIT;
PRR=&PRR;
PRM=&PRM;
NSE=&NSE;
EPS=&EPS;
TAM_POP=&TAM_POP;

LISTA_SOLUCOES=J(TAM_POP,4,.);

SOMA_FO=POP_INICIAL[+,2];

DO I=1 TO TAM_POP;
LISTA_SOLUCOES[I,1]=POP_INICIAL[I,1];
LISTA_SOLUCOES[I,2]=POP_INICIAL[I,2];
LISTA_SOLUCOES[I,3]=POP_INICIAL[I,2]/SOMA_FO;

IF I=1 THEN LISTA_SOLUCOES[I,4]=LISTA_SOLUCOES[I,3];
ELSE LISTA_SOLUCOES[I,4]=LISTA_SOLUCOES[I-1,4]+LISTA_SOLUCOES[I,3];
END;

/* ORDENA A COLUNA DE PROPORÇÕES DECRESCENTEMENTE */
CALL SORT (LISTA_SOLUCOES,{2},{2});


/* INICIA DE EVOLUÇÃO DO ALGORITMO GENÉTICO */
DO IT=1 TO NIT;

*RESET PRINT;
IF IT=1 THEN SOLUCOES_ELITE=LISTA_SOLUCOES[1:NSE,1:2];
NUM1=RANUNI(0);
NUM2=RANUNI(0);

DO I=1 TO TAM_POP;
IF NUM1<=LISTA_SOLUCOES[I,4] THEN
DO;
SOL1=LISTA_SOLUCOES[I,1];
FO1=LISTA_SOLUCOES[I,2];
I=16;
END;
END;

DO I=1 TO TAM_POP;
```

```
IF NUM2<=LISTA_SOLUCOES[I,4] THEN
DO;
SOL2=LISTA_SOLUCOES[I,1];
FO2=LISTA_SOLUCOES[I,2];
I=16;
END;
END;

OCOR_CROSS=RANBIN(0,1,PRR);
OCOR_MUT=RANBIN(0,1,PRM);

IF OCOR_CROSS=1 THEN
DO;
SIG_SOL1=1/FO1;
SIG_SOL2=1/FO2;
NUM4=SOL1+SIG_SOL1*RANNOR(0);
NUM5=SOL2+SIG_SOL2*RANNOR(0);

IF NUM5>2.5 THEN SOL1_=2.5;
ELSE
DO;
IF NUM5<-2.5 THEN SOL1_=-2.5;
ELSE
DO;
SOL1_=NUM5;
END;
END;

IF NUM4>2.5 THEN SOL2_=2.5;
ELSE
DO;
IF NUM4<-2.5 THEN SOL2_=-2.5;
ELSE
DO;
SOL2_=NUM4;
END;
END;

END;
ELSE
DO;
SOL1_=SOL1;
FO1_=FO1;
SOL2_=SOL2;
FO2_=FO2;
END;

IF OCOR_MUT=1 THEN
DO;
ERR1=(-1/IT)+(2/IT)*RANUNI(0);
ERR2=(-1/IT)+(2/IT)*RANUNI(0);
SOL1_=SOL1_+ERR1;
SOL2_=SOL2_+ERR2;
IF SOL1_>2.5 THEN SOL1_=2.5;
```

```
ELSE
DO;
IF SOL1_<-2.5 THEN SOL1_=-2.5;
END;

IF SOL2_>2.5 THEN SOL2_=2.5;
ELSE
DO;
IF SOL2_<-2.5 THEN SOL2_=-2.5;
END;

END;

FO1_=6+(SOL1_**2)*SIN(14*SOL1_);
FO2_=6+(SOL2_**2)*SIN(14*SOL2_);

DO I=1 TO TAM_POP;
IF FO1_>LISTA_SOLUCOES[I,2] THEN
DO;
NOVAS_SOLUCOES[TAM_POP,1]=SOL1_;
NOVAS_SOLUCOES[TAM_POP,2]=FO1_;
CALL SORT (NOVAS_SOLUCOES,{2},{2});
IF I<=5 THEN
DO;
IF FO1_>SOLUCOES_ELITE[I,2] THEN
DO;
SOLUCOES_ELITE[5,1]=SOL1_;
SOLUCOES_ELITE[5,2]=FO1_;
END;
END;
I=TAM_POP+1;
END;
END;

DO I=1 TO TAM_POP;
IF FO2_>NOVAS_SOLUCOES[I,2] THEN
DO;
NOVAS_SOLUCOES[TAM_POP,1]=SOL2_;
NOVAS_SOLUCOES[TAM_POP,2]=FO2_;
CALL SORT (NOVAS_SOLUCOES,{2},{2});

IF I<=5 THEN
DO;
IF FO2_>SOLUCOES_ELITE[I,2] THEN
DO;
SOLUCOES_ELITE[I,1]=SOL2_;
SOLUCOES_ELITE[I,2]=FO2_;
CALL SORT (SOLUCOES_ELITE,{2},{2});
END;
END;

I=TAM_POP+1;
END;
END;
```

```
NA MATRIZ NOVAS_SOLUCOES */
LISTA_SOLUCOES=NOVAS_SOLUCOES;

LISTA_SOLUCOES[,3]=LISTA_SOLUCOES[,2]/LISTA_SOLUCOES[+,2];

CALL SORT (LISTA_SOLUCOES,{2},{2});

DO I=1 TO TAM_POP;
IF I=1 THEN LISTA_SOLUCOES[I,4]=LISTA_SOLUCOES[I,3];
ELSE LISTA_SOLUCOES[I,4]=LISTA_SOLUCOES[I-1,4]+LISTA_SOLUCOES[I,3];
END;
END;

QUIT;

%MEND GA_SIMPLES;

/* EXECUTA A MACRO GA_SIMPLES*/

%GA_SIMPLES(50,0.85,0.10,5,0.000001,200);
```

## A.3 Código da Análise de Correspondência e Método Fuzzy-C-Means

```
########################################
#                                      #
# MÉTODO HÍBRIDO GENÉTICO FUZZY        #
#                                      #
########################################
#Criado por: Pedro Rafael Diniz Marinho
#Ter 07 Dez 2010 23:26:15 BRT
#rm(list = ls(all = TRUE)) # Apagar todos os objetos
require(Rlab) # Carrega o pacote para geracao de numeros pseudo-aleatorios Bernoulli
require(MASS) # Permite salvar a matriz de pertinencia otima usando o objeto write.matrix()
# A funcao MP gera uma matriz de pertinencia com as dimensoes desejadas.
# A matriz e' gerada aleatoriamente por uma distribuicao de Poisson com parametro tambem aleatorio
# com distribuicao uniforme no intervalo [1,100].
MP <- function(row, col){
U = matrix(1:row*col,row,col)
i = 1
while(i<=row){
linha = c(prop.table(rpois(col,runif(1,1,100))^10))
U[i,] = linha
i = i + 1
}
return(U)
}
# A funcao NMP gera o numero de matrizes de pertinencias U com as dimensoes
# desejadas.
NMP <- function(n,row,col){
```

```
i = 1
U <- NULL
while(i<=n){
U = cbind(U,MP(row,col))
i = i+1
}
return(array(U,dim=c(row,col,n)))
}

# A funcao R calcula o produto entre as matrizes de
#pertinencias e a matriz de distancias ao quadrado.
# Devese entrar com a matriz de pertinencias, matrizes de distancias e o numero
# POP -- Matrizes de pertinencias geradas pela funcao NMP;
# d   -- Matrizes de distancias gerada pela Analise de Correspondencia.
R <- function(POP,d){
i = 1
R = array(, dim = c(dim(POP[,,1])[1], dim(POP[,,2])[2],
length(POP)/(dim(POP[,,1])[1]*dim(POP[,,1])[2])))
while(i<=length(POP)/(dim(POP[,,1])[1]*dim(POP[,,1])[2])){
R[,,i] = ((POP[,,i]^(2))*(d^2))
i = i + 1
}
return(R)
}
# A funcao J.1 e' a funcao objetivo que queremos otimizar J(x).
# Ou seja, e' calculado as somas de cada uma das matrizes R.
J.1 <- function(R){
     i = 1
     SR <- NULL
     while(i<=length(R)/(dim(R[,,1])[1]*dim(R[,,1])[2])){
      SR = c(SR,sum(R[,,i]))
i = i + 1
     }
     return(SR)
}

############# Metodo GA
#d <- matrix(c(runif(15,0,10)),5,3) # Matriz de distancias. Esta matriz e' fixa.
i = 1
T = Sys.time() # Tempo inicial.
POP = NMP(1000,29,8) # Populacao inicial com 15 matrizes de pertinencias 5 por 3.
for(i in 1:50000){
Matrizes_R = R(POP,d)
FO = J.1(Matrizes_R) # Funcoes objetivos calculadas para cada matriz R.
FO_D = sort(FO, decreasing = FALSE, index.return  = TRUE) # Ordenando de forma crescente
Prop = abs(FO_D$x)/sum(abs(FO_D$x)) # Proporcoes
acumulado = cumsum(Prop) # Proporcao acumulada ordenado
#acumulado = sort(prop.table(rpois(length(FO_D$ix),runif(1,1,100))^10),decreasing = TRUE)
Roleta = runif(2, min(acumulado), max(acumulado)) # Gera dois numeros aleatorios no intervalo [0,1]
PR1 = PU1 = which(acumulado == min(acumulado[acumulado >= Roleta[1]])) # Posicao da primeira matriz U / R
PR2 = PU2 = which(acumulado == min(acumulado[acumulado >= Roleta[2]])) # Posicao da segunda matriz U / R
n = length(POP)/(dim(POP[,,1])[1]*dim(POP[,,1])[2]) # Numero de matrizes iniciais.
PR1 = n - PR1 + 1
PR2 = n - PR2 + 1
```

```
R1 = Matrizes_R[,,PR1] # Matriz R1
R2 = Matrizes_R[,,PR2] # Matriz R2
U1 = as.matrix((POP[,,PU1])) # Matriz U1
U2 = as.matrix((POP[,,PU2])) # Matriz U2
#Crossover
cross = rbern(1, 0.85) # Se o valor e' 1 havera' crossover.
if(cross == 1){
SR1 = apply(R1,1,sum) # Soma das linhas da matriz R1
SR2 = apply(R2,1,sum) # Soma das linhas da matriz R2
linha_ord1 = sort(SR1, decreasing = FALSE, index.return  = TRUE) # Ordenando as linhas com menores
#somas da matriz R1
linha_ord2 = sort(SR2, decreasing = FALSE, index.return  = TRUE) # Ordenando as linhas com
#menores somas da matriz R2
acumulado1 = sort(prop.table(rpois(dim(R1)[1],runif(1,1,100))^10),decreasing = TRUE)
acumulado2 = sort(prop.table(rpois(dim(R2)[1],runif(1,1,100))^10),decreasing = TRUE)
Roleta1 = runif(1,0,max(acumulado1))
Roleta2 = runif(1,0,max(acumulado2))
Roleta = c(Roleta1,Roleta2)
#Prop1 = abs(linha_ord1$x)/sum(abs(linha_ord1$x))
#acumulado1 = cumsum(Prop1)
#Prop2 = abs(linha_ord2$x)/sum(abs(linha_ord2$x))
#acumulado2 = cumsum(Prop2)
PL1 = min(which(acumulado1 == min(acumulado1[acumulado1 >= Roleta[1]]))) # Posicao da matriz
#R1 que ira sofrer crossover
PL2 = min(which(acumulado2 == min(acumulado2[acumulado2 >= Roleta[2]])))# Posicao da matriz
#R2 que ira sofrer crossover
L1U = U1[PL1,] # Elementos da linha da matriz U1 que sera jogado na matriz U2
L2U = U2[PL2,] # Elementos da linha da matriz U2 que sera jogado na matriz U1
U1[PL1,] = L2U  # Filho 1
U2[PL2,] = L1U  # Filho 2
}
#################
# Mutacao
mut = rbern(1, .05) # Se o valor e' 1 havera' mutacao.
if(mut == 1){
PL1 = trunc(runif(1,1,dim(POP[,,1]))[1])
PL2 = trunc(runif(1,1,dim(POP[,,1]))[1])
# Troca duas linhas da matriz U1
L1 = U1[PL1,]
L2 = U1[PL2,]
U1[PL2,] = L1
U1[PL1,] = L2
# Seleciona uma linha da matriz U1 e gera uma nova linha
linha_troca = round(runif(1,1,dim(U1)[1]),digits = 0)
linha = c(prop.table(rpois(dim(U1)[2],runif(1,1,100))^round(runif(1,10,30),digits = 0) ))
U1[linha_troca,] = linha
}
FU1 = sum((U1*(d^2))) # Funcao objetivo para o valor de U1
FU2 = sum((U2*(d^2))) # Funcao objetivo para o valor de U2
logico1 = FU1 > FO_D$x
logico2 = FU2 > FO_D$x
logico1 = length(logico1[logico1 == TRUE]) # A funcao objetivo em U1 eh maior
#que ao menos uma funcao objetivo da minha geracao anterior?
logico2 = length(logico2[logico2 == TRUE]) # A funcao objetivo em U1 eh maior
```

```
#que ao menos uma funcao objetivo da minha geracao anterior?
substituir = FO_D$ix[c(length(acumulado)-1,length(acumulado))]

if(logico1>=1){
POP[,,substituir[1]] = U2
}
if(logico2>=1){
 POP[,,substituir[2]] = U1
}
Matrizes_R = R(POP,d)
FO = J.1(Matrizes_R) # Funcoes objetivos calculadas para cada matriz R.
FO_D = sort(FO, decreasing = FALSE, index.return  = TRUE)
}
T1 = Sys.time()
Tempo = T1 - T
FO_D$x[1:3]
FO_D$ix[1:3]
POP[,,FO_D$ix[1]]
d
write.matrix(d,"Distancias_2010.csv")
write.matrix(as.matrix(POP[,,FO_D$ix[1]]),"Pertinencias_2010.csv")
```

# A.4 Tabelas de Contingências

## A.4.1 Ano de 2006

Tabela 10: Tabela de Contingência para o Ano de 2006.

| Municipio | V03 | V09 | V49 | V89 | V99 | W19 | X70 | X95 |
|---|---|---|---|---|---|---|---|---|
| AG | 0 | 0 | 0 | 4 | 0 | 3 | 1 | 1 |
| AN | 0 | 0 | 0 | 7 | 0 | 0 | 0 | 0 |
| AH | 1 | 0 | 3 | 0 | 0 | 0 | 0 | 1 |
| AR | 0 | 0 | 0 | 3 | 0 | 0 | 1 | 4 |
| AO | 0 | 0 | 0 | 3 | 0 | 1 | 1 | 2 |
| BA | 4 | 1 | 1 | 10 | 0 | 2 | 0 | 32 |
| BO | 0 | 0 | 0 | 2 | 0 | 1 | 1 | 2 |
| CP | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 2 |
| CA | 3 | 2 | 0 | 2 | 0 | 1 | 0 | 11 |
| CJ | 0 | 0 | 3 | 2 | 0 | 3 | 3 | 0 |
| CG | 0 | 2 | 0 | 78 | 0 | 5 | 8 | 102 |
| CR | 0 | 0 | 0 | 4 | 0 | 0 | 1 | 3 |
| CO | 1 | 0 | 0 | 3 | 1 | 0 | 1 | 3 |
| CU | 0 | 0 | 0 | 2 | 0 | 0 | 2 | 1 |
| ES | 0 | 0 | 0 | 5 | 0 | 1 | 2 | 7 |
| GA | 0 | 0 | 0 | 4 | 0 | 1 | 0 | 3 |
| JP | 38 | 4 | 4 | 38 | 0 | 7 | 11 | 222 |
| LS | 0 | 0 | 0 | 4 | 0 | 0 | 1 | 3 |
| MA | 0 | 1 | 0 | 2 | 0 | 0 | 1 | 1 |
| MO | 0 | 0 | 0 | 3 | 0 | 0 | 2 | 6 |
| PA | 0 | 0 | 1 | 13 | 0 | 1 | 6 | 12 |
| PE | 1 | 0 | 1 | 1 | 1 | 1 | 2 | 1 |
| QE | 0 | 0 | 0 | 13 | 0 | 0 | 3 | 3 |
| RE | 0 | 1 | 0 | 9 | 2 | 0 | 0 | 1 |
| SR | 6 | 8 | 0 | 2 | 0 | 4 | 3 | 38 |
| SB | 0 | 2 | 1 | 1 | 0 | 0 | 2 | 7 |
| SA | 1 | 0 | 1 | 3 | 3 | 0 | 0 | 11 |
| SL | 0 | 0 | 1 | 7 | 0 | 0 | 0 | 3 |
| SD | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 2 |
| SO | 0 | 1 | 8 | 4 | 0 | 0 | 0 | 5 |

### A.4.2 Ano de 2007

Tabela 11: Tabela de Contingência para o Ano de 2007.

| Municipio | V03 | V09 | V49 | V89 | V99 | W19 | X70 | X95 |
|---|---|---|---|---|---|---|---|---|
| AG | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 |
| AN | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 |
| AH | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 2 |
| AR | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 2 |
| AO | 0 | 0 | 0 | 2 | 0 | 0 | 2 | 2 |
| BA | 6 | 7 | 1 | 5 | 1 | 2 | 3 | 41 |
| BO | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 7 |
| CP | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 6 |
| CA | 2 | 2 | 1 | 3 | 0 | 2 | 2 | 20 |
| CJ | 0 | 0 | 1 | 2 | 0 | 1 | 3 | 1 |
| CG | 0 | 0 | 3 | 103 | 0 | 4 | 15 | 98 |
| CR | 0 | 0 | 0 | 2 | 0 | 0 | 2 | 7 |
| CO | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 |
| CU | 0 | 0 | 0 | 3 | 0 | 0 | 1 | 3 |
| ES | 0 | 0 | 0 | 6 | 0 | 0 | 1 | 6 |
| GA | 1 | 0 | 2 | 3 | 0 | 0 | 1 | 1 |
| JP | 21 | 8 | 6 | 38 | 1 | 7 | 12 | 273 |
| LS | 0 | 0 | 0 | 3 | 0 | 1 | 2 | 3 |
| MA | 1 | 2 | 3 | 2 | 1 | 0 | 0 | 1 |
| MO | 0 | 0 | 0 | 7 | 0 | 1 | 1 | 1 |
| PA | 0 | 0 | 3 | 6 | 0 | 0 | 5 | 19 |
| PE | 0 | 3 | 0 | 2 | 0 | 0 | 0 | 9 |
| QE | 0 | 0 | 0 | 9 | 0 | 2 | 3 | 4 |
| RE | 0 | 0 | 0 | 6 | 1 | 0 | 0 | 0 |
| SR | 2 | 7 | 1 | 3 | 1 | 1 | 5 | 38 |
| SB | 0 | 0 | 2 | 6 | 0 | 0 | 1 | 3 |
| SA | 0 | 0 | 1 | 0 | 14 | 0 | 1 | 13 |
| SL | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 2 |
| SD | 0 | 0 | 0 | 6 | 0 | 0 | 0 | 1 |
| SO | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |

### A.4.3 Ano de 2008

Tabela 12: Tabela de Contingência para o Ano de 2008.

| Municipio | V03 | V09 | V49 | V89 | V99 | W19 | X70 | X95 |
|---|---|---|---|---|---|---|---|---|
| AG | 1 | 1 | 1 | 0 | 1 | 2 | 4 | 0 |
| AN | 0 | 0 | 0 | 3 | 0 | 1 | 4 | 3 |
| AH | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 4 |
| AR | 0 | 0 | 0 | 2 | 0 | 1 | 0 | 1 |
| AO | 0 | 0 | 0 | 3 | 0 | 1 | 0 | 0 |
| BA | 7 | 3 | 4 | 7 | 2 | 2 | 5 | 43 |
| BO | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 3 |
| CP | 2 | 0 | 0 | 1 | 0 | 0 | 0 | 2 |
| CA | 6 | 1 | 0 | 3 | 0 | 1 | 0 | 16 |
| CJ | 2 | 0 | 1 | 3 | 2 | 0 | 1 | 0 |
| CG | 1 | 0 | 4 | 74 | 0 | 8 | 11 | 88 |
| CR | 0 | 0 | 0 | 2 | 0 | 0 | 1 | 4 |
| CO | 1 | 1 | 0 | 3 | 0 | 0 | 0 | 1 |
| CU | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 1 |
| ES | 0 | 0 | 0 | 4 | 0 | 1 | 1 | 5 |
| GA | 1 | 0 | 0 | 3 | 0 | 0 | 2 | 2 |
| JP | 28 | 6 | 4 | 49 | 2 | 16 | 13 | 296 |
| LS | 0 | 0 | 0 | 3 | 0 | 0 | 4 | 3 |
| MA | 1 | 0 | 1 | 4 | 0 | 2 | 0 | 5 |
| MO | 0 | 0 | 0 | 3 | 0 | 0 | 1 | 3 |
| PA | 2 | 1 | 2 | 8 | 0 | 0 | 4 | 49 |
| PE | 2 | 0 | 1 | 3 | 0 | 0 | 0 | 2 |
| QE | 0 | 0 | 0 | 11 | 0 | 2 | 3 | 4 |
| RE | 0 | 0 | 0 | 5 | 0 | 0 | 1 | 2 |
| SR | 6 | 4 | 1 | 6 | 2 | 0 | 1 | 62 |
| SB | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 2 |
| SA | 0 | 1 | 2 | 2 | 4 | 2 | 3 | 12 |
| SL | 0 | 1 | 0 | 2 | 0 | 0 | 0 | 2 |
| SD | 0 | 0 | 0 | 5 | 0 | 0 | 1 | 3 |
| SO | 0 | 1 | 0 | 4 | 0 | 0 | 4 | 3 |

### A.4.4 Ano de 2009

Tabela 13: Tabela de Contingência para o Ano de 2009.

| Municipio | V03 | V09 | V49 | V89 | V99 | W19 | X70 | X95 |
|---|---|---|---|---|---|---|---|---|
| AG | 0 | 1 | 0 | 1 | 5 | 1 | 4 | 1 |
| AN | 0 | 0 | 0 | 3 | 0 | 3 | 1 | 1 |
| AH | 0 | 0 | 0 | 2 | 0 | 1 | 3 | 10 |
| AR | 0 | 0 | 0 | 3 | 0 | 1 | 1 | 0 |
| AO | 0 | 0 | 0 | 8 | 0 | 0 | 1 | 8 |
| BA | 3 | 2 | 0 | 5 | 1 | 2 | 1 | 74 |
| BO | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 |
| CP | 0 | 2 | 1 | 4 | 0 | 0 | 0 | 7 |
| CA | 0 | 0 | 1 | 6 | 0 | 0 | 0 | 41 |
| CJ | 2 | 1 | 2 | 1 | 0 | 1 | 2 | 4 |
| CG | 0 | 0 | 1 | 39 | 0 | 19 | 10 | 124 |
| CR | 0 | 0 | 1 | 5 | 0 | 0 | 0 | 12 |
| CO | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 8 |
| CU | 0 | 0 | 0 | 5 | 0 | 1 | 2 | 2 |
| ES | 0 | 0 | 0 | 4 | 0 | 0 | 4 | 7 |
| GA | 0 | 0 | 0 | 4 | 2 | 0 | 2 | 2 |
| JP | 13 | 8 | 8 | 47 | 0 | 7 | 13 | 358 |
| LS | 0 | 0 | 0 | 4 | 0 | 2 | 2 | 7 |
| MA | 1 | 0 | 0 | 7 | 1 | 0 | 2 | 12 |
| MO | 0 | 0 | 0 | 12 | 0 | 1 | 0 | 0 |
| PA | 0 | 0 | 1 | 19 | 0 | 4 | 5 | 47 |
| PE | 1 | 0 | 0 | 2 | 0 | 0 | 2 | 10 |
| QE | 0 | 0 | 0 | 11 | 0 | 1 | 3 | 11 |
| RE | 0 | 0 | 0 | 6 | 0 | 1 | 0 | 1 |
| SR | 4 | 2 | 0 | 13 | 0 | 3 | 1 | 57 |
| SB | 0 | 0 | 2 | 6 | 0 | 0 | 2 | 15 |
| SA | 0 | 0 | 4 | 2 | 1 | 1 | 2 | 25 |
| SL | 1 | 0 | 2 | 2 | 3 | 0 | 0 | 1 |
| SD | 0 | 0 | 0 | 2 | 0 | 2 | 0 | 7 |
| SO | 0 | 0 | 1 | 6 | 0 | 1 | 2 | 7 |

### A.4.5 Ano de 2010

Tabela 14: Tabela de Contingência para o Ano de 2010.

| Municipio | V03 | V09 | V49 | V89 | V99 | W19 | X70 | X95 |
|---|---|---|---|---|---|---|---|---|
| AG | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| AN | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| AH | 0 | 2 | 0 | 3 | 0 | 0 | 0 | 4 |
| AR | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 2 |
| AO | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 2 |
| BA | 1 | 0 | 1 | 2 | 0 | 0 | 1 | 30 |
| BO | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 2 |
| CP | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 4 |
| CA | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 15 |
| CJ | 1 | 0 | 1 | 2 | 0 | 0 | 1 | 2 |
| CG | 0 | 0 | 0 | 30 | 0 | 4 | 7 | 57 |
| CR | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| CO | 0 | 0 | 1 | 2 | 0 | 0 | 1 | 2 |
| CU | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 2 |
| ES | 0 | 0 | 0 | 2 | 0 | 1 | 0 | 1 |
| GA | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 5 |
| JP | 12 | 5 | 7 | 21 | 0 | 3 | 4 | 174 |
| LS | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 |
| MA | 1 | 0 | 2 | 3 | 0 | 0 | 0 | 4 |
| MO | 0 | 0 | 0 | 2 | 0 | 1 | 0 | 1 |
| PA | 0 | 0 | 2 | 7 | 0 | 1 | 3 | 13 |
| PE | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 4 |
| QE | 0 | 0 | 0 | 3 | 0 | 2 | 0 | 3 |
| RE | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 |
| SR | 1 | 0 | 0 | 2 | 0 | 0 | 1 | 26 |
| SB | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 4 |
| SA | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 15 |
| SL | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| SD | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 |
| SO | 0 | 0 | 1 | 4 | 0 | 0 | 0 | 4 |

# A.5 Distâncias entre Municípios e Causas Externas de Óbitos.

## A.5.1 Ano de 2006

Tabela 15: Distâncias Calculadas entre Municípios e Causas Externas de óbito, 2006.

| | DISTÂNCIAS | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Município** | **V03** | **V09** | **V49** | **V89** | **V99** | **W19** | **X70** | **X95** |
| **AG** | 1,56922 | 1,32748 | 3,14922 | 0,17017 | 0,57096 | 0,65489 | 0,29480 | 1,06789 |
| **AN** | 2,20388 | 1,97920 | 3,58575 | 0,75182 | 0,65194 | 1,32449 | 0,96508 | 1,68524 |
| **AH** | 3,56480 | 3,41819 | 0,69537 | 3,83400 | 3,59144 | 3,40997 | 3,58834 | 3,67786 |
| **AR** | 1,16771 | 0,96040 | 3,27351 | 0,30750 | 1,02372 | 0,44043 | 0,31489 | 0,64634 |
| **AO** | 1,42047 | 1,19177 | 3,21846 | 0,03740 | 0,75417 | 0,55814 | 0,23043 | 0,90830 |
| **BA** | 0,55237 | 0,35711 | 3,13686 | 0,90251 | 1,51880 | 0,46886 | 0,75378 | 0,07591 |
| **BO** | 1,29037 | 1,06053 | 3,17215 | 0,16730 | 0,84457 | 0,43775 | 0,16517 | 0,78085 |
| **CP** | 0,56817 | 0,37321 | 3,14156 | 0,88695 | 1,50596 | 0,45955 | 0,74012 | 0,08046 |
| **CA** | 0,25452 | 0,24421 | 3,23211 | 1,22088 | 1,84091 | 0,77680 | 1,07791 | 0,28037 |
| **CJ** | 2,23608 | 1,97980 | 1,45506 | 1,85022 | 1,48296 | 1,60795 | 1,64605 | 2,06520 |
| **CG** | 1,10066 | 0,89921 | 3,28010 | 0,37893 | 1,08986 | 0,42384 | 0,36154 | 0,57805 |
| **CR** | 1,41368 | 1,19777 | 3,30936 | 0,09201 | 0,82945 | 0,60097 | 0,30745 | 0,89398 |
| **CO** | 1,27742 | 1,04674 | 3,16277 | 0,18147 | 0,85128 | 0,42314 | 0,15796 | 0,76886 |
| **CU** | 1,65418 | 1,42067 | 3,25488 | 0,20492 | 0,57740 | 0,76219 | 0,40443 | 1,14371 |
| **ES** | 1,15252 | 0,93789 | 3,22821 | 0,30838 | 1,00722 | 0,39720 | 0,27970 | 0,63440 |
| **GA** | 1,31944 | 1,10245 | 3,26836 | 0,14998 | 0,88173 | 0,51548 | 0,26025 | 0,80088 |
| **JP** | 0,39560 | 0,26698 | 3,19386 | 1,06669 | 1,68871 | 0,63156 | 0,92394 | 0,12841 |
| **LS** | 1,41368 | 1,19777 | 3,30936 | 0,09201 | 0,82945 | 0,60097 | 0,30745 | 0,89398 |
| **MA** | 1,21389 | 0,98367 | 3,14984 | 0,24394 | 0,90394 | 0,37212 | 0,17044 | 0,70607 |
| **MO** | 1,05503 | 0,84862 | 3,24030 | 0,41152 | 1,10520 | 0,37668 | 0,35791 | 0,53432 |
| **PA** | 1,32784 | 1,08362 | 3,06907 | 0,19666 | 0,75658 | 0,41588 | 0,06763 | 0,83391 |
| **PE** | 1,56275 | 1,28325 | 2,25249 | 0,99110 | 0,85636 | 0,75869 | 0,77168 | 1,26656 |
| **QE** | 1,82967 | 1,60402 | 3,39763 | 0,37660 | 0,58736 | 0,95618 | 0,59998 | 1,31269 |
| **RE** | 2,01969 | 1,78175 | 3,34965 | 0,57106 | 0,43919 | 1,10705 | 0,74773 | 1,51068 |
| **SR** | 0,24230 | 0,21809 | 3,20656 | 1,22327 | 1,83606 | 0,76862 | 1,07521 | 0,28476 |
| **SB** | 0,68086 | 0,40237 | 2,69704 | 1,01586 | 1,44073 | 0,42151 | 0,77899 | 0,50779 |
| **SA** | 1,01621 | 0,74652 | 2,82085 | 0,59109 | 1,02730 | 0,06819 | 0,34062 | 0,60462 |
| **SL** | 1,55181 | 1,28435 | 2,83110 | 0,44182 | 0,48944 | 0,58901 | 0,32440 | 1,10258 |
| **SD** | 0,89715 | 0,72062 | 3,30654 | 0,59881 | 1,29591 | 0,44364 | 0,53875 | 0,37395 |
| **SO** | 2,69756 | 2,49998 | 0,51193 | 2,73445 | 2,46060 | 2,36050 | 2,49866 | 2,70220 |

### A.5.2 Ano de 2007

Tabela 16: Distâncias Calculadas entre Municípios e Causas Externas de óbito, 2007.

| | DISTÂNCIAS | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Município** | **V03** | **V09** | **V49** | **V89** | **V99** | **W19** | **X70** | **X95** |
| **AG** | 0,84024 | 1,02277 | 0,18092 | 0,72269 | 4,22977 | 0,26191 | 0,26051 | 0,38973 |
| **AN** | 2,33255 | 2,53370 | 1,57999 | 0,82212 | 4,78746 | 1,55755 | 1,29779 | 1,89492 |
| **AH** | 0,87918 | 1,08013 | 0,38035 | 0,63208 | 4,42610 | 0,12326 | 0,16973 | 0,44531 |
| **AR** | 1,00214 | 1,20253 | 0,41558 | 0,50925 | 4,42977 | 0,23819 | 0,06381 | 0,56567 |
| **AO** | 1,28522 | 1,48564 | 0,60534 | 0,22604 | 4,47002 | 0,51459 | 0,25329 | 0,84757 |
| **BA** | 0,22291 | 0,39379 | 0,67278 | 1,32258 | 4,33306 | 0,60592 | 0,84688 | 0,24957 |
| **BO** | 1,06078 | 1,26180 | 0,45547 | 0,45043 | 4,44695 | 0,29217 | 0,06754 | 0,62544 |
| **CP** | 0,69764 | 0,89850 | 0,38461 | 0,81376 | 4,41264 | 0,10524 | 0,34512 | 0,26734 |
| **CA** | 0,36042 | 0,56287 | 0,58271 | 1,15094 | 4,42172 | 0,42005 | 0,68066 | 0,11525 |
| **CJ** | 1,33543 | 1,53370 | 0,62178 | 0,18237 | 4,43702 | 0,56984 | 0,29634 | 0,89375 |
| **CG** | 1,26367 | 1,46442 | 0,59244 | 0,24754 | 4,47278 | 0,49248 | 0,23399 | 0,82679 |
| **CR** | 0,77060 | 0,97094 | 0,36541 | 0,74097 | 4,41012 | 0,07625 | 0,27245 | 0,33666 |
| **CO** | 0,09227 | 0,19865 | 0,98008 | 1,57397 | 4,54305 | 0,83594 | 1,10798 | 0,52793 |
| **CU** | 1,26075 | 1,46153 | 0,59053 | 0,25047 | 4,47286 | 0,48951 | 0,23132 | 0,82394 |
| **ES** | 1,25229 | 1,45319 | 0,58557 | 0,25897 | 4,47388 | 0,48084 | 0,22388 | 0,81579 |
| **GA** | 1,20642 | 1,40088 | 0,47758 | 0,32719 | 4,35737 | 0,45821 | 0,16850 | 0,75987 |
| **JP** | 0,36691 | 0,56548 | 0,56304 | 1,14644 | 4,39622 | 0,42006 | 0,67414 | 0,09499 |
| **LS** | 1,22441 | 1,42617 | 0,57573 | 0,28770 | 4,48586 | 0,45131 | 0,20338 | 0,79004 |
| **MA** | 1,01048 | 1,07253 | 0,55349 | 1,30169 | 3,56599 | 0,94968 | 0,94273 | 0,77348 |
| **MO** | 1,86759 | 2,06928 | 1,13902 | 0,35868 | 4,64127 | 1,09250 | 0,83527 | 1,43146 |
| **PA** | 0,80622 | 1,00263 | 0,31704 | 0,71019 | 4,36521 | 0,12582 | 0,23504 | 0,36283 |
| **PE** | 0,18312 | 0,38818 | 0,73073 | 1,32816 | 4,44012 | 0,59506 | 0,85777 | 0,27287 |
| **QE** | 1,52016 | 1,72192 | 0,82076 | 0,03389 | 4,54768 | 0,74565 | 0,49046 | 1,08481 |
| **RE** | 2,33876 | 2,49137 | 1,44410 | 1,16733 | 3,73887 | 1,70792 | 1,42603 | 1,90349 |
| **SR** | 0,25878 | 0,42107 | 0,63935 | 1,30190 | 4,30094 | 0,59273 | 0,82553 | 0,23066 |
| **SB** | 1,47478 | 1,67153 | 0,72827 | 0,08638 | 4,43046 | 0,71144 | 0,43416 | 1,03081 |
| **SA** | 3,58036 | 3,54284 | 3,17181 | 3,67103 | 0,88717 | 3,60726 | 3,51243 | 3,45059 |
| **SL** | 0,87998 | 1,07583 | 0,31501 | 0,63821 | 4,35988 | 0,16671 | 0,16200 | 0,43543 |
| **SD** | 2,02106 | 2,22219 | 1,27928 | 0,51069 | 4,67445 | 1,24640 | 0,98673 | 1,58364 |
| **SO** | 0,79345 | 1,00110 | 0,47808 | 0,73043 | 4,52496 | 0,04022 | 0,29863 | 0,39341 |

### A.5.3 Ano de 2008

Tabela 17: Distâncias Calculadas entre Municípios e Causas Externas de óbito, 2008.

| | DISTÂNCIAS | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Município** | **V03** | **V09** | **V49** | **V89** | **V99** | **W19** | **X70** | **X95** |
| **AG** | 3,82821 | 3,63008 | 3,40307 | 3,93408 | 1,09690 | 3,97177 | 3,41327 | 4,07241 |
| **AN** | 2,00861 | 1,98450 | 1,89846 | 0,72616 | 5,16659 | 0,65191 | 0,99999 | 1,62207 |
| **AH** | 0,84286 | 0,85614 | 0,85202 | 0,45023 | 4,96544 | 0,52660 | 0,59248 | 0,46146 |
| **AR** | 2,21286 | 2,17695 | 2,07464 | 0,94339 | 5,13084 | 0,87297 | 1,14635 | 1,84301 |
| **AO** | 1,23110 | 1,23989 | 1,21034 | 0,11682 | 5,09677 | 0,16126 | 0,60382 | 0,81704 |
| **BA** | 0,31635 | 0,45669 | 0,63048 | 1,06574 | 5,03117 | 1,13979 | 1,08552 | 0,19094 |
| **BO** | 1,65644 | 1,62490 | 1,53554 | 0,39768 | 4,97289 | 0,33803 | 0,65065 | 1,29449 |
| **CP** | 0,54196 | 0,57392 | 0,62237 | 0,75037 | 4,91323 | 0,82773 | 0,76404 | 0,23235 |
| **CA** | 0,49130 | 0,62248 | 0,77196 | 0,96421 | 5,15375 | 1,03331 | 1,06610 | 0,07636 |
| **CJ** | 0,42233 | 0,33990 | 0,30486 | 0,95017 | 4,62282 | 1,03083 | 0,76603 | 0,47757 |
| **CG** | 0,88663 | 0,92646 | 0,94997 | 0,45985 | 5,09722 | 0,52652 | 0,70434 | 0,44743 |
| **CR** | 0,70701 | 0,76593 | 0,82374 | 0,64174 | 5,07825 | 0,71186 | 0,79512 | 0,25899 |
| **CO** | 0,87314 | 0,92695 | 0,96731 | 0,51195 | 5,14980 | 0,57465 | 0,76956 | 0,41305 |
| **CU** | 1,75837 | 1,72924 | 1,64091 | 0,48830 | 5,02674 | 0,42142 | 0,75060 | 1,38803 |
| **ES** | 1,12990 | 1,10798 | 1,04547 | 0,18571 | 4,88544 | 0,26407 | 0,39668 | 0,78282 |
| **GA** | 2,20780 | 2,02336 | 1,77811 | 1,98221 | 3,06980 | 2,01278 | 1,47869 | 2,29551 |
| **JP** | 0,44663 | 0,56741 | 0,71105 | 0,95512 | 5,09200 | 1,02666 | 1,02753 | 0,05821 |
| **LS** | 1,22286 | 1,22135 | 1,17970 | 0,07396 | 5,03279 | 0,14475 | 0,53967 | 0,82749 |
| **MA** | 0,85917 | 0,78964 | 0,68483 | 0,56479 | 4,61317 | 0,64182 | 0,33260 | 0,67151 |
| **MO** | 2,28161 | 2,26153 | 2,17719 | 0,99510 | 5,32646 | 0,91816 | 1,27326 | 1,88287 |
| **PA** | 0,89253 | 0,92503 | 0,93969 | 0,43641 | 5,06826 | 0,50545 | 0,67014 | 0,46554 |
| **PE** | 0,56759 | 0,61940 | 0,68405 | 0,74327 | 4,98205 | 0,81833 | 0,80306 | 0,18691 |
| **QE** | 1,30686 | 1,30285 | 1,25530 | 0,03336 | 5,04904 | 0,06030 | 0,56107 | 0,91071 |
| **RE** | 2,02248 | 2,00868 | 1,93475 | 0,73485 | 5,26940 | 0,65664 | 1,05917 | 1,61825 |
| **SR** | 0,55853 | 0,65160 | 0,75811 | 0,82327 | 5,09778 | 0,89371 | 0,93191 | 0,07742 |
| **SB** | 0,71242 | 0,73772 | 0,75835 | 0,58443 | 4,96119 | 0,66050 | 0,67407 | 0,33544 |
| **SA** | 0,30969 | 0,26377 | 0,31880 | 1,01717 | 4,69730 | 1,09748 | 0,87438 | 0,41690 |
| **SL** | 3,21163 | 3,01646 | 2,80675 | 3,52315 | 1,60684 | 3,57297 | 2,99498 | 3,51475 |
| **SD** | 0,96810 | 1,01313 | 1,03806 | 0,42210 | 5,16215 | 0,48041 | 0,73668 | 0,51373 |
| **SO** | 1,16440 | 1,15360 | 1,10305 | 0,12629 | 4,95855 | 0,20694 | 0,46716 | 0,79123 |

### A.5.4 Ano de 2009

Tabela 18: Distâncias Calculadas entre Municípios e Causas Externas de óbito, 2009.

| | DISTÂNCIAS | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Município** | **V03** | **V09** | **V49** | **V89** | **V99** | **W19** | **X70** | **X95** |
| **AG** | 2,01340 | 1,67528 | 1,26967 | 2,15696 | 0,83038 | 1,67111 | 1,24141 | 2,18569 |
| **AN** | 1,45753 | 1,28666 | 0,98686 | 0,75608 | 2,12681 | 0,61681 | 0,29267 | 1,32028 |
| **AH** | 1,05622 | 1,18312 | 1,28179 | 0,41839 | 2,78926 | 0,71842 | 1,21806 | 0,66776 |
| **AR** | 1,39331 | 1,39320 | 1,30989 | 0,14633 | 2,71262 | 0,67967 | 0,92688 | 1,07882 |
| **AO** | 1,93260 | 1,90608 | 1,75586 | 0,68843 | 3,03139 | 1,16119 | 1,19081 | 1,62227 |
| **BA** | 0,28284 | 0,24063 | 0,52899 | 1,05872 | 1,98100 | 0,59112 | 1,04036 | 0,41686 |
| **BO** | 1,00253 | 1,14055 | 1,25745 | 0,46668 | 2,76861 | 0,71364 | 1,23055 | 0,61041 |
| **CP** | 0,25065 | 0,54142 | 0,89960 | 1,09996 | 2,32056 | 0,81033 | 1,34174 | 0,17678 |
| **CA** | 0,25983 | 0,57011 | 0,94044 | 1,13954 | 2,34813 | 0,86003 | 1,39163 | 0,20837 |
| **CJ** | 1,73422 | 1,39407 | 1,00764 | 2,00399 | 0,74406 | 1,47673 | 1,12643 | 1,93298 |
| **CG** | 1,04707 | 1,09262 | 1,10215 | 0,21948 | 2,58489 | 0,48943 | 0,94848 | 0,71482 |
| **CR** | 0,88999 | 0,92199 | 0,94323 | 0,36622 | 2,43891 | 0,35765 | 0,88011 | 0,58665 |
| **CO** | 0,87329 | 0,89089 | 0,90055 | 0,38763 | 2,39448 | 0,31292 | 0,84137 | 0,58700 |
| **CU** | 1,87398 | 1,90569 | 1,82998 | 0,63241 | 3,19950 | 1,20146 | 1,37144 | 1,52142 |
| **ES** | 1,07292 | 1,08992 | 1,06361 | 0,18520 | 2,53025 | 0,43652 | 0,86400 | 0,76381 |
| **GA** | 1,19617 | 1,09250 | 0,90425 | 0,44752 | 2,25311 | 0,34495 | 0,47873 | 0,99982 |
| **JP** | 0,39077 | 0,62155 | 0,91697 | 0,94311 | 2,38634 | 0,71025 | 1,26572 | 0,01404 |
| **LS** | 1,46748 | 1,30157 | 1,00723 | 0,74034 | 2,15303 | 0,62278 | 0,31847 | 1,32329 |
| **MA** | 0,80094 | 0,81360 | 0,83501 | 0,46422 | 2,33611 | 0,27807 | 0,83301 | 0,53291 |
| **MO** | 1,19703 | 1,19702 | 1,13438 | 0,09256 | 2,57108 | 0,50030 | 0,84252 | 0,89377 |
| **PA** | 0,41653 | 0,62113 | 0,89409 | 0,89462 | 2,37480 | 0,66140 | 1,21972 | 0,04840 |
| **PE** | 0,63912 | 0,64478 | 0,70956 | 0,63491 | 2,22257 | 0,28692 | 0,85482 | 0,42542 |
| **QE** | 1,58434 | 1,53114 | 1,36779 | 0,41442 | 2,66988 | 0,78067 | 0,83295 | 1,30911 |
| **RE** | 1,59363 | 1,57968 | 1,46380 | 0,34938 | 2,81646 | 0,84683 | 0,99140 | 1,28265 |
| **SR** | 0,17716 | 0,49199 | 0,87524 | 1,16765 | 2,26816 | 0,84589 | 1,36065 | 0,25598 |
| **SB** | 1,21516 | 0,96676 | 0,58694 | 1,00181 | 1,70508 | 0,55055 | 0,16785 | 1,20556 |
| **SA** | 1,15332 | 0,81278 | 0,47409 | 1,64557 | 1,04274 | 1,06397 | 0,94210 | 1,38636 |
| **SL** | 0,77423 | 0,83642 | 0,91231 | 0,48479 | 2,42250 | 0,39305 | 0,95095 | 0,45874 |
| **SD** | 1,42377 | 1,42783 | 1,34690 | 0,17062 | 2,74847 | 0,71645 | 0,95886 | 1,10440 |
| **SO** | 1,36664 | 1,20246 | 0,91823 | 0,69676 | 2,11400 | 0,52386 | 0,27402 | 1,22749 |

### A.5.5 Ano de 2010

Tabela 19: Distâncias Calculadas entre Municípios e Causas Externas de óbito, 2010.

| | DISTÂNCIAS | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Município** | **V03** | **V09** | **V49** | **V89** | **V99** | **W19** | **X70** | **X95** |
| **AG** | 4,30862 | 4,74535 | 4,16172 | 4,19413 | 3,65904 | 3,30418 | 2,57646 | 3,99173 |
| **AN** | 2,58781 | 2,95386 | 2,29983 | 2,19633 | 2,11733 | 1,28760 | 0,78476 | 2,20993 |
| **AH** | 0,83567 | 0,58468 | 0,56164 | 0,78930 | 1,46301 | 1,78236 | 2,11305 | 0,83641 |
| **AR** | 1,43494 | 1,58750 | 0,92767 | 0,51533 | 1,50092 | 0,58125 | 1,20948 | 1,06529 |
| **AO** | 0,76051 | 0,80230 | 0,23775 | 0,53425 | 1,17192 | 1,36932 | 1,64110 | 0,53514 |
| **BA** | 0,33928 | 0,76811 | 0,44434 | 1,16827 | 0,52362 | 1,68736 | 1,48230 | 0,14671 |
| **BO** | 0,76051 | 0,80230 | 0,23775 | 0,53425 | 1,17192 | 1,36932 | 1,64110 | 0,53514 |
| **CP** | 0,99596 | 1,20843 | 0,52563 | 0,58841 | 1,08137 | 0,97266 | 1,18265 | 0,61632 |
| **CA** | 0,36739 | 0,79731 | 0,45616 | 1,17035 | 0,50437 | 1,67097 | 1,45408 | 0,15150 |
| **CJ** | 0,87164 | 1,09652 | 0,41383 | 0,63947 | 0,98768 | 1,09580 | 1,23276 | 0,49311 |
| **CG** | 0,96446 | 1,15113 | 0,47184 | 0,53306 | 1,10474 | 1,01679 | 1,26109 | 0,59886 |
| **CR** | 0,27545 | 0,74754 | 0,57088 | 1,31368 | 0,46985 | 1,83867 | 1,57751 | 0,29671 |
| **CO** | 1,14529 | 1,35628 | 0,67462 | 0,58935 | 1,18291 | 0,82173 | 1,10805 | 0,76090 |
| **CU** | 1,65895 | 1,52976 | 1,16109 | 0,48866 | 2,07127 | 1,32429 | 2,12758 | 1,45345 |
| **ES** | 1,92492 | 1,96296 | 1,38850 | 0,67431 | 2,09443 | 0,69137 | 1,69875 | 1,59781 |
| **GA** | 0,24955 | 0,28703 | 0,44413 | 1,14719 | 0,95912 | 1,93105 | 1,92839 | 0,47045 |
| **JP** | 0,30987 | 0,66828 | 0,31153 | 1,05820 | 0,66859 | 1,65805 | 1,55081 | 0,09451 |
| **LS** | 1,87943 | 1,99652 | 1,36049 | 0,76874 | 1,93449 | 0,36838 | 1,39036 | 1,51523 |
| **MA** | 0,75387 | 0,69354 | 0,31279 | 0,59723 | 1,26392 | 1,51540 | 1,81196 | 0,61899 |
| **MO** | 1,92492 | 1,96296 | 1,38850 | 0,67431 | 2,09443 | 0,69137 | 1,69875 | 1,59781 |
| **PA** | 1,02652 | 1,26141 | 0,57843 | 0,64657 | 1,06070 | 0,93896 | 1,10951 | 0,63736 |
| **PE** | 1,61640 | 2,06222 | 1,53775 | 1,84070 | 0,99202 | 1,50569 | 0,49580 | 1,32064 |
| **QE** | 1,67043 | 1,75921 | 1,14123 | 0,51796 | 1,79662 | 0,56221 | 1,44307 | 1,32366 |
| **RE** | 2,18209 | 2,46788 | 1,78591 | 1,52059 | 1,90397 | 0,54997 | 0,76199 | 1,78234 |
| **SR** | 0,36481 | 0,79361 | 0,45233 | 1,16747 | 0,50858 | 1,67071 | 1,45667 | 0,14812 |
| **SB** | 0,76051 | 0,80230 | 0,23775 | 0,53425 | 1,17192 | 1,36932 | 1,64110 | 0,53514 |
| **SA** | 0,47485 | 0,88770 | 0,45771 | 1,12482 | 0,50081 | 1,57140 | 1,34785 | 0,15989 |
| **SL** | 2,34565 | 2,17535 | 1,85377 | 1,15040 | 2,75633 | 1,66805 | 2,64976 | 2,14517 |
| **SD** | 2,34565 | 2,17535 | 1,85377 | 1,15040 | 2,75633 | 1,66805 | 2,64976 | 2,14517 |
| **SO** | 1,08822 | 1,01635 | 0,58988 | 0,29586 | 1,52167 | 1,31893 | 1,82404 | 0,88917 |

## A.6 Tabelas de Pertinências Segundo o Método Genético *Fuzzy* (MHGF)

## A.7 Pertinências para o Ano de 2006

Tabela 20: Pertinências dos inter-relacionamentos entre municípios e causas externas de óbito para o ano de 2006.

| | PERTINÊNCIAS | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Município** | **V03** | **V09** | **V49** | **V89** | **V99** | **W19** | **X70** | **X95** |
| **AG** | 0,03075 | 0,02635 | 0,02021 | 0,55357 | 0,05922 | 0,05031 | 0,22818 | 0,03142 |
| **AN** | 0,04658 | 0,08117 | 0,03310 | 0,23504 | 0,21898 | 0,08080 | 0,25291 | 0,05142 |
| **AH** | 0,05378 | 0,05760 | 0,61335 | 0,04038 | 0,08655 | 0,04880 | 0,04882 | 0,05072 |
| **AR** | 0,03536 | 0,10576 | 0,02969 | 0,35887 | 0,06967 | 0,13547 | 0,16913 | 0,09605 |
| **AO** | 0,02061 | 0,01140 | 0,01527 | 0,55480 | 0,01804 | 0,02773 | 0,25502 | 0,09713 |
| **BA** | 0,08763 | 0,06264 | 0,01013 | 0,02384 | 0,01492 | 0,05345 | 0,02861 | 0,71878 |
| **BO** | 0,02750 | 0,03849 | 0,01527 | 0,60296 | 0,04116 | 0,06094 | 0,16120 | 0,05248 |
| **CP** | 0,04369 | 0,14816 | 0,00997 | 0,11362 | 0,07855 | 0,27135 | 0,03696 | 0,29771 |
| **CA** | 0,32223 | 0,18362 | 0,02528 | 0,06495 | 0,04611 | 0,04426 | 0,08140 | 0,23217 |
| **CJ** | 0,08965 | 0,13236 | 0,10119 | 0,16368 | 0,13638 | 0,09120 | 0,18195 | 0,10359 |
| **CG** | 0,04673 | 0,04368 | 0,02302 | 0,29971 | 0,03607 | 0,09600 | 0,30469 | 0,15009 |
| **CR** | 0,02465 | 0,03437 | 0,01447 | 0,71602 | 0,03994 | 0,04197 | 0,08208 | 0,04650 |
| **CO** | 0,02128 | 0,02589 | 0,00881 | 0,15026 | 0,12619 | 0,08342 | 0,52909 | 0,05505 |
| **CU** | 0,04140 | 0,03373 | 0,01985 | 0,23480 | 0,10536 | 0,05189 | 0,41416 | 0,09881 |
| **ES** | 0,04228 | 0,07992 | 0,02255 | 0,24708 | 0,13588 | 0,11156 | 0,29931 | 0,06141 |
| **GA** | 0,02761 | 0,03705 | 0,01020 | 0,30467 | 0,05290 | 0,23571 | 0,29226 | 0,03960 |
| **JP** | 0,11684 | 0,10335 | 0,01172 | 0,02530 | 0,01673 | 0,05030 | 0,02928 | 0,64648 |
| **LS** | 0,02883 | 0,02099 | 0,01864 | 0,46794 | 0,03033 | 0,04520 | 0,27780 | 0,11028 |
| **MA** | 0,02863 | 0,05567 | 0,02261 | 0,16756 | 0,11140 | 0,17754 | 0,32501 | 0,11158 |
| **MO** | 0,04106 | 0,07134 | 0,02524 | 0,14215 | 0,11403 | 0,20136 | 0,26264 | 0,14218 |
| **PA** | 0,03203 | 0,02368 | 0,01402 | 0,14074 | 0,06289 | 0,04597 | 0,59336 | 0,08731 |
| **PE** | 0,10977 | 0,09367 | 0,04740 | 0,10280 | 0,22012 | 0,12355 | 0,14408 | 0,15862 |
| **QE** | 0,04101 | 0,07576 | 0,02744 | 0,27527 | 0,19233 | 0,07885 | 0,26268 | 0,04666 |
| **RE** | 0,04149 | 0,07610 | 0,02962 | 0,23819 | 0,23811 | 0,07712 | 0,25374 | 0,04564 |
| **SR** | 0,53184 | 0,17564 | 0,00987 | 0,03101 | 0,03862 | 0,04121 | 0,02938 | 0,14244 |
| **SB** | 0,06763 | 0,19531 | 0,01882 | 0,13083 | 0,09361 | 0,32485 | 0,06205 | 0,10690 |
| **SA** | 0,08659 | 0,04020 | 0,00904 | 0,15430 | 0,06876 | 0,29913 | 0,21862 | 0,12338 |
| **SL** | 0,02775 | 0,03345 | 0,01537 | 0,09941 | 0,18158 | 0,09242 | 0,49118 | 0,05884 |
| **SD** | 0,05081 | 0,14617 | 0,01545 | 0,15790 | 0,09584 | 0,31653 | 0,08399 | 0,13332 |
| **SO** | 0,05198 | 0,05716 | 0,60973 | 0,04095 | 0,08886 | 0,05077 | 0,05044 | 0,05012 |

## A.8 Pertinências para o Ano de 2007

Tabela 21: Pertinências dos inter-relacionamentos entre municípios e causas externas de óbito para o ano de 2007.

| | PERTINÊNCIAS | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Município** | **V03** | **V09** | **V49** | **V89** | **V99** | **W19** | **X70** | **X95** |
| **AG** | 0,04468 | 0,05211 | 0,39328 | 0,06251 | 0,00670 | 0,18555 | 0,10162 | 0,15355 |
| **AN** | 0,06203 | 0,08353 | 0,10823 | 0,43810 | 0,02605 | 0,09675 | 0,08020 | 0,10512 |
| **AH** | 0,03924 | 0,04744 | 0,30982 | 0,06189 | 0,00562 | 0,26589 | 0,13283 | 0,13728 |
| **AR** | 0,07969 | 0,01924 | 0,04739 | 0,05252 | 0,01288 | 0,18724 | 0,56676 | 0,03429 |
| **AO** | 0,04385 | 0,06533 | 0,09836 | 0,46112 | 0,01194 | 0,09576 | 0,13342 | 0,09023 |
| **BA** | 0,53959 | 0,09232 | 0,07304 | 0,04330 | 0,01225 | 0,06099 | 0,05075 | 0,12775 |
| **BO** | 0,07989 | 0,01954 | 0,04640 | 0,05991 | 0,01315 | 0,17773 | 0,57010 | 0,03329 |
| **CP** | 0,05476 | 0,05326 | 0,06067 | 0,03441 | 0,01233 | 0,48779 | 0,09926 | 0,19751 |
| **CA** | 0,09300 | 0,07347 | 0,04664 | 0,02922 | 0,01303 | 0,33922 | 0,07247 | 0,33293 |
| **CJ** | 0,04262 | 0,06441 | 0,09637 | 0,49499 | 0,01192 | 0,08884 | 0,11304 | 0,08781 |
| **CG** | 0,04408 | 0,06547 | 0,09912 | 0,44713 | 0,01188 | 0,09816 | 0,14325 | 0,09091 |
| **CR** | 0,05354 | 0,09623 | 0,09958 | 0,04792 | 0,02137 | 0,34423 | 0,13026 | 0,20687 |
| **CO** | 0,65116 | 0,13046 | 0,05105 | 0,03479 | 0,01026 | 0,03804 | 0,03540 | 0,04885 |
| **CU** | 0,09109 | 0,03215 | 0,17096 | 0,23273 | 0,02309 | 0,18263 | 0,21904 | 0,04832 |
| **ES** | 0,09112 | 0,03215 | 0,17111 | 0,22764 | 0,02305 | 0,18346 | 0,22298 | 0,04849 |
| **GA** | 0,02547 | 0,02195 | 0,08069 | 0,09355 | 0,00703 | 0,06855 | 0,66238 | 0,04037 |
| **JP** | 0,06273 | 0,04072 | 0,08116 | 0,02022 | 0,00792 | 0,05480 | 0,03684 | 0,69560 |
| **LS** | 0,04420 | 0,06546 | 0,09920 | 0,42518 | 0,01165 | 0,10236 | 0,16030 | 0,09165 |
| **MA** | 0,08934 | 0,13363 | 0,15796 | 0,06921 | 0,03420 | 0,16081 | 0,12540 | 0,22945 |
| **MO** | 0,05231 | 0,07394 | 0,09954 | 0,49266 | 0,01836 | 0,08954 | 0,07795 | 0,09570 |
| **PA** | 0,09008 | 0,02707 | 0,07613 | 0,04946 | 0,01406 | 0,29267 | 0,38513 | 0,06540 |
| **PE** | 0,18463 | 0,10501 | 0,04397 | 0,02965 | 0,01401 | 0,32907 | 0,07002 | 0,22365 |
| **QE** | 0,01750 | 0,02402 | 0,16564 | 0,64558 | 0,01143 | 0,03456 | 0,08011 | 0,02116 |
| **RE** | 0,11253 | 0,05429 | 0,19202 | 0,19550 | 0,04562 | 0,18002 | 0,15336 | 0,06667 |
| **SR** | 0,14360 | 0,10280 | 0,05223 | 0,03131 | 0,01459 | 0,33192 | 0,07328 | 0,25027 |
| **SB** | 0,01778 | 0,01569 | 0,03602 | 0,80194 | 0,00592 | 0,03685 | 0,06036 | 0,02544 |
| **SA** | 0,04686 | 0,04519 | 0,05041 | 0,07052 | 0,65071 | 0,04432 | 0,04566 | 0,04633 |
| **SL** | 0,04965 | 0,05039 | 0,07729 | 0,04382 | 0,01268 | 0,41865 | 0,17887 | 0,16866 |
| **SD** | 0,05657 | 0,07806 | 0,10419 | 0,46622 | 0,02125 | 0,09340 | 0,08016 | 0,10016 |
| **SO** | 0,05246 | 0,04065 | 0,07430 | 0,11443 | 0,02351 | 0,52818 | 0,07288 | 0,09360 |

## A.9 Pertinências para o Ano de 2008

Tabela 22: Pertinências dos inter-relacionamentos entre municípios e causas externas de óbito para o ano de 2008.

| | PERTINÊNCIAS | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Município** | **V03** | **V09** | **V49** | **V89** | **V99** | **W19** | **X70** | **X95** |
| **AG** | 0,04873 | 0,07495 | 0,06419 | 0,05708 | 0,53676 | 0,05664 | 0,10612 | 0,05553 |
| **AN** | 0,04973 | 0,05731 | 0,11575 | 0,28584 | 0,02538 | 0,29837 | 0,10838 | 0,05923 |
| **AH** | 0,06059 | 0,06690 | 0,12375 | 0,27054 | 0,01850 | 0,25680 | 0,10048 | 0,10245 |
| **AR** | 0,05328 | 0,06109 | 0,11987 | 0,27742 | 0,02861 | 0,28563 | 0,11211 | 0,06199 |
| **AO** | 0,05302 | 0,01996 | 0,02045 | 0,36462 | 0,00485 | 0,30673 | 0,04623 | 0,18413 |
| **BA** | 0,17963 | 0,10673 | 0,11454 | 0,09310 | 0,01825 | 0,11952 | 0,03979 | 0,32844 |
| **BO** | 0,04055 | 0,04824 | 0,10666 | 0,30324 | 0,02011 | 0,32572 | 0,10638 | 0,04911 |
| **CP** | 0,25720 | 0,06493 | 0,07729 | 0,18543 | 0,01137 | 0,10167 | 0,11665 | 0,18547 |
| **CA** | 0,12326 | 0,07231 | 0,09272 | 0,08660 | 0,01627 | 0,11339 | 0,03151 | 0,46393 |
| **CJ** | 0,08576 | 0,45517 | 0,22356 | 0,03724 | 0,02119 | 0,05161 | 0,04727 | 0,07820 |
| **CG** | 0,06355 | 0,05459 | 0,11757 | 0,13535 | 0,02862 | 0,30199 | 0,08618 | 0,21214 |
| **CR** | 0,06782 | 0,07046 | 0,12332 | 0,23950 | 0,01796 | 0,23322 | 0,07989 | 0,16784 |
| **CO** | 0,13113 | 0,06232 | 0,07578 | 0,12805 | 0,02161 | 0,09501 | 0,14982 | 0,33627 |
| **CU** | 0,09268 | 0,05704 | 0,03618 | 0,18038 | 0,01309 | 0,45391 | 0,11605 | 0,05068 |
| **ES** | 0,03773 | 0,04537 | 0,10367 | 0,34464 | 0,01633 | 0,29460 | 0,10759 | 0,05006 |
| **GA** | 0,09201 | 0,15248 | 0,11558 | 0,13531 | 0,11268 | 0,08982 | 0,17552 | 0,12660 |
| **JP** | 0,09592 | 0,10214 | 0,03543 | 0,05537 | 0,03428 | 0,09599 | 0,10629 | 0,47458 |
| **LS** | 0,02558 | 0,03269 | 0,08985 | 0,43341 | 0,01371 | 0,30779 | 0,06390 | 0,03307 |
| **MA** | 0,08173 | 0,05312 | 0,06123 | 0,22809 | 0,00909 | 0,21919 | 0,13124 | 0,21631 |
| **MO** | 0,04582 | 0,06525 | 0,04652 | 0,30902 | 0,02372 | 0,26182 | 0,13679 | 0,11105 |
| **PA** | 0,12430 | 0,08177 | 0,10882 | 0,16273 | 0,02035 | 0,17742 | 0,07488 | 0,24973 |
| **PE** | 0,14048 | 0,06954 | 0,07970 | 0,08559 | 0,01974 | 0,05800 | 0,13390 | 0,41304 |
| **QE** | 0,12029 | 0,05414 | 0,01276 | 0,29466 | 0,01686 | 0,41405 | 0,02988 | 0,05735 |
| **RE** | 0,05904 | 0,10081 | 0,07817 | 0,37670 | 0,02599 | 0,18418 | 0,09563 | 0,07947 |
| **SR** | 0,04298 | 0,03782 | 0,17323 | 0,12690 | 0,03542 | 0,14466 | 0,03642 | 0,40258 |
| **SB** | 0,07745 | 0,11693 | 0,09120 | 0,33747 | 0,01897 | 0,12355 | 0,08078 | 0,15364 |
| **SA** | 0,30018 | 0,16675 | 0,09921 | 0,03116 | 0,00701 | 0,03864 | 0,23427 | 0,12278 |
| **SL** | 0,09019 | 0,14902 | 0,10790 | 0,12205 | 0,18240 | 0,07682 | 0,14944 | 0,12217 |
| **SD** | 0,05715 | 0,06215 | 0,11767 | 0,28400 | 0,01876 | 0,27087 | 0,09047 | 0,09892 |
| **SO** | 0,07336 | 0,05400 | 0,03005 | 0,33412 | 0,03645 | 0,31998 | 0,08757 | 0,06447 |

## A.10 Pertinências para o Ano de 2009

Tabela 23: Pertinências dos inter-relacionamentos entre municípios e causas externas de óbito para o ano de 2009.

| | PERTINÊNCIAS | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Município** | **V03** | **V09** | **V49** | **V89** | **V99** | **W19** | **X70** | **X95** |
| **AG** | 0,08672 | 0,10949 | 0,14450 | 0,08367 | 0,31414 | 0,07777 | 0,10061 | 0,08311 |
| **AN** | 0,05101 | 0,10289 | 0,16818 | 0,07378 | 0,04028 | 0,11425 | 0,37827 | 0,07135 |
| **AH** | 0,11915 | 0,06792 | 0,06432 | 0,28585 | 0,05802 | 0,19033 | 0,08362 | 0,13080 |
| **AR** | 0,09463 | 0,04902 | 0,05080 | 0,41855 | 0,05249 | 0,17046 | 0,08011 | 0,08394 |
| **AO** | 0,11300 | 0,07775 | 0,07596 | 0,28872 | 0,06070 | 0,11402 | 0,21011 | 0,05975 |
| **BA** | 0,12444 | 0,25569 | 0,11182 | 0,10087 | 0,05469 | 0,14813 | 0,10138 | 0,10297 |
| **BO** | 0,15451 | 0,07149 | 0,10639 | 0,13200 | 0,04972 | 0,12661 | 0,16928 | 0,18999 |
| **CP** | 0,54662 | 0,06230 | 0,04442 | 0,03036 | 0,03343 | 0,04305 | 0,05509 | 0,18473 |
| **CA** | 0,59930 | 0,05993 | 0,03633 | 0,05634 | 0,01501 | 0,04018 | 0,02501 | 0,16791 |
| **CJ** | 0,10002 | 0,17347 | 0,09481 | 0,07581 | 0,21036 | 0,14464 | 0,11237 | 0,08852 |
| **CG** | 0,10584 | 0,05860 | 0,05827 | 0,33987 | 0,05401 | 0,19690 | 0,08134 | 0,10518 |
| **CR** | 0,10059 | 0,18891 | 0,06439 | 0,23243 | 0,02838 | 0,13881 | 0,13633 | 0,11015 |
| **CO** | 0,07633 | 0,06705 | 0,19640 | 0,15967 | 0,03005 | 0,13929 | 0,08921 | 0,24199 |
| **CU** | 0,11339 | 0,06701 | 0,06891 | 0,29183 | 0,06550 | 0,18605 | 0,10193 | 0,10538 |
| **ES** | 0,10168 | 0,05553 | 0,05638 | 0,35570 | 0,05299 | 0,19915 | 0,08157 | 0,09700 |
| **GA** | 0,14200 | 0,08255 | 0,11170 | 0,12949 | 0,03512 | 0,18942 | 0,18398 | 0,12574 |
| **JP** | 0,02319 | 0,02969 | 0,08530 | 0,06966 | 0,00379 | 0,07189 | 0,00716 | 0,70932 |
| **LS** | 0,07209 | 0,09990 | 0,05631 | 0,09806 | 0,02454 | 0,23258 | 0,28141 | 0,13510 |
| **MA** | 0,07482 | 0,17936 | 0,08692 | 0,15772 | 0,02725 | 0,17955 | 0,06345 | 0,23093 |
| **MO** | 0,09090 | 0,05502 | 0,05059 | 0,46030 | 0,04262 | 0,09414 | 0,16956 | 0,03686 |
| **PA** | 0,14643 | 0,07013 | 0,08269 | 0,05152 | 0,02300 | 0,08936 | 0,10727 | 0,42959 |
| **PE** | 0,05991 | 0,09603 | 0,13842 | 0,07437 | 0,01880 | 0,21488 | 0,22448 | 0,17311 |
| **QE** | 0,10594 | 0,06170 | 0,06655 | 0,30342 | 0,06144 | 0,19284 | 0,11275 | 0,09536 |
| **RE** | 0,10533 | 0,06006 | 0,06314 | 0,32950 | 0,06000 | 0,18590 | 0,10021 | 0,09586 |
| **SR** | 0,59825 | 0,06806 | 0,04535 | 0,02868 | 0,03373 | 0,04138 | 0,05474 | 0,12981 |
| **SB** | 0,05070 | 0,13085 | 0,09711 | 0,04824 | 0,02796 | 0,10121 | 0,46851 | 0,07542 |
| **SA** | 0,14800 | 0,13498 | 0,23582 | 0,04619 | 0,07123 | 0,10153 | 0,19098 | 0,07127 |
| **SL** | 0,15767 | 0,10641 | 0,05775 | 0,33958 | 0,03923 | 0,14840 | 0,05587 | 0,09509 |
| **SD** | 0,03151 | 0,03397 | 0,03825 | 0,67067 | 0,03597 | 0,06052 | 0,07000 | 0,05912 |
| **SO** | 0,05422 | 0,13125 | 0,08493 | 0,07755 | 0,02825 | 0,12240 | 0,41890 | 0,08250 |

## A.11 Pertinências para o Ano de 2010

Tabela 24: Pertinências dos inter-relacionamentos entre municípios e causas externas de óbito para o ano de 2010.

| | PERTINÊNCIAS | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Município** | **V03** | **V09** | **V49** | **V89** | **V99** | **W19** | **X70** | **X95** |
| **AG** | 0,10268 | 0,08755 | 0,12438 | 0,08584 | 0,11825 | 0,16576 | 0,20854 | 0,10700 |
| **AN** | 0,07128 | 0,05960 | 0,07655 | 0,10271 | 0,11107 | 0,17805 | 0,29832 | 0,10241 |
| **AH** | 0,09338 | 0,20476 | 0,30694 | 0,14863 | 0,06132 | 0,05755 | 0,04005 | 0,08739 |
| **AR** | 0,08937 | 0,07530 | 0,13157 | 0,14418 | 0,09350 | 0,19613 | 0,16635 | 0,10360 |
| **AO** | 0,07769 | 0,15743 | 0,37201 | 0,15089 | 0,05638 | 0,05467 | 0,03736 | 0,09356 |
| **BA** | 0,17952 | 0,11725 | 0,14459 | 0,05705 | 0,13477 | 0,05610 | 0,05847 | 0,25225 |
| **BO** | 0,07769 | 0,15743 | 0,37201 | 0,15089 | 0,05638 | 0,05467 | 0,03736 | 0,09356 |
| **CP** | 0,05637 | 0,08399 | 0,30330 | 0,20664 | 0,11493 | 0,05464 | 0,06512 | 0,11502 |
| **CA** | 0,24586 | 0,03742 | 0,06504 | 0,02543 | 0,22399 | 0,01894 | 0,02296 | 0,36035 |
| **CJ** | 0,07901 | 0,15056 | 0,31836 | 0,14956 | 0,06964 | 0,06815 | 0,05091 | 0,11381 |
| **CG** | 0,07470 | 0,04614 | 0,13186 | 0,22478 | 0,05847 | 0,06506 | 0,08517 | 0,31381 |
| **CR** | 0,29340 | 0,04752 | 0,06199 | 0,02701 | 0,24043 | 0,02043 | 0,02498 | 0,28423 |
| **CO** | 0,09651 | 0,07888 | 0,14985 | 0,12357 | 0,10090 | 0,16202 | 0,16746 | 0,12081 |
| **CU** | 0,08299 | 0,16689 | 0,12529 | 0,23095 | 0,05008 | 0,13261 | 0,06477 | 0,14641 |
| **ES** | 0,08914 | 0,07821 | 0,12467 | 0,14640 | 0,09177 | 0,20842 | 0,16391 | 0,09749 |
| **GA** | 0,22965 | 0,19718 | 0,15567 | 0,06178 | 0,11489 | 0,05649 | 0,05651 | 0,12784 |
| **JP** | 0,16732 | 0,11346 | 0,15208 | 0,05368 | 0,11344 | 0,05261 | 0,05356 | 0,29386 |
| **LS** | 0,04644 | 0,03701 | 0,09201 | 0,13360 | 0,07346 | 0,44066 | 0,06771 | 0,10910 |
| **MA** | 0,07702 | 0,09584 | 0,16959 | 0,12193 | 0,04864 | 0,09099 | 0,03613 | 0,35985 |
| **MO** | 0,09321 | 0,04562 | 0,06216 | 0,44100 | 0,04781 | 0,11913 | 0,13161 | 0,05945 |
| **PA** | 0,07707 | 0,14999 | 0,29735 | 0,15841 | 0,07229 | 0,08184 | 0,06070 | 0,10236 |
| **PE** | 0,04830 | 0,04457 | 0,12106 | 0,11312 | 0,18096 | 0,13736 | 0,27709 | 0,07753 |
| **QE** | 0,07865 | 0,05896 | 0,12608 | 0,14416 | 0,12199 | 0,32748 | 0,07650 | 0,06619 |
| **RE** | 0,08300 | 0,05728 | 0,10127 | 0,06253 | 0,09852 | 0,24491 | 0,27694 | 0,07554 |
| **SR** | 0,13902 | 0,12523 | 0,09794 | 0,05370 | 0,08622 | 0,03866 | 0,07117 | 0,38805 |
| **SB** | 0,07769 | 0,15743 | 0,37201 | 0,15089 | 0,05638 | 0,05467 | 0,03736 | 0,09356 |
| **SA** | 0,16065 | 0,11507 | 0,14827 | 0,06028 | 0,14238 | 0,05899 | 0,06230 | 0,25206 |
| **SL** | 0,14207 | 0,09484 | 0,08539 | 0,20558 | 0,09789 | 0,10783 | 0,13610 | 0,13030 |
| **SD** | 0,08651 | 0,07616 | 0,13107 | 0,22483 | 0,14997 | 0,11245 | 0,11131 | 0,10769 |
| **SO** | 0,07157 | 0,15723 | 0,29079 | 0,24495 | 0,05543 | 0,06358 | 0,04060 | 0,07586 |

## A.12 Tabelas de Agrupamentos pelo Método Híbrido Genético Fuzzy (MHGF)

## A.13 Ano de 2006

### A.13.1 Ponto de Corte igual à 0,10

Tabela 25: Agrupamento dos Municípios para com as Causas Externas de Óbito.

| | AGRUPAMENTOS | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **V89** | **X70** | **X95** | **V99** | **W19** | **V09** | **V03** | **V49** |
| **CJ** | 1 | 1 | 1 | 1 | | 1 | | 1 |
| **CG** | 1 | 1 | 1 | | | | | |
| **LS** | 1 | 1 | 1 | | | | | |
| **MA** | 1 | 1 | 1 | 1 | 1 | | | |
| **MO** | 1 | 1 | 1 | 1 | 1 | | | |
| **PE** | 1 | 1 | 1 | 1 | 1 | | 1 | |
| **SA** | 1 | 1 | 1 | | 1 | | | |
| **CO** | 1 | 1 | | 1 | | | | |
| **ES** | 1 | 1 | | 1 | 1 | | | |
| **CU** | 1 | 1 | | 1 | | | | |
| **QE** | 1 | 1 | | 1 | | | | |
| **RE** | 1 | 1 | | 1 | | | | |
| **AN** | 1 | 1 | | 1 | | | | |
| **AG** | 1 | 1 | | | | | | |
| **AR** | 1 | 1 | | | 1 | 1 | | |
| **AO** | 1 | 1 | | | | | | |
| **BO** | 1 | 1 | | | | | | |
| **GA** | 1 | 1 | | | 1 | | | |
| **PA** | 1 | 1 | | | | | | |
| **SD** | 1 | | 1 | | 1 | 1 | | |
| **CP** | 1 | | 1 | | 1 | 1 | | |
| **SB** | 1 | | 1 | | 1 | 1 | | |
| **CA** | | | 1 | | | 1 | 1 | |
| **JP** | | | 1 | | | 1 | 1 | |
| **SR** | | | 1 | | | 1 | 1 | |
| **BA** | | | 1 | | | | | |
| **AH** | | | | | | | | 1 |
| **SO** | | | | | | | | 1 |
| **SL** | | 1 | | 1 | | | | |
| **CR** | 1 | | | | | | | |

### A.13.2 Ponto de Corte igual à 0,20

Tabela 26: Agrupamento dos Municípios para com as Causas Externas de Óbito.

| | AGRUPAMENTOS | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **X70** | **V89** | **W19** | **X95** | **V99** | **V03** | **V49** | **V09** |
| **AG** | 1 | 1 | | | | | | |
| **AN** | 1 | 1 | | | 1 | | | |
| **RE** | 1 | 1 | | | 1 | | | |
| **CU** | 1 | 1 | | | | | | |
| **ES** | 1 | 1 | | | | | | |
| **GA** | 1 | 1 | 1 | | | | | |
| **LS** | 1 | 1 | | | | | | |
| **QE** | 1 | 1 | | | | | | |
| **AO** | 1 | 1 | | | | | | |
| **CG** | 1 | 1 | | | | | | |
| **MA** | 1 | | | | | | | |
| **MO** | 1 | | 1 | | | | | |
| **SA** | 1 | | 1 | | | | | |
| **PA** | 1 | | | | | | | |
| **SL** | 1 | | | | | | | |
| **CO** | 1 | | | | | | | |
| **AR** | | 1 | | | | | | |
| **BO** | | 1 | | | | | | |
| **CR** | | 1 | | | | | | |
| **CP** | | | 1 | 1 | | | | |
| **CA** | | | | 1 | | 1 | | |
| **BA** | | | | 1 | | | | |
| **SR** | | | | | | 1 | | |
| **AH** | | | | | | | 1 | |
| **CJ** | | | | | | | | |
| **JP** | | | | 1 | | | | |
| **PE** | | | | | 1 | | | |
| **SB** | | | 1 | | | | | |
| **SD** | | | 1 | | | | | |
| **SO** | | | | | | | 1 | |

### A.13.3 Ponto de Corte igual à 0,30

Tabela 27: Agrupamento dos Municípios para com as Causas Externas de Óbito.

| | AGRUPAMENTOS | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **V89** | **X70** | **W19** | **X95** | **V03** | **V49** | **V09** | **V99** |
| **AG** | 1 | | | | | | | |
| **AR** | 1 | | | | | | | |
| **AO** | 1 | | | | | | | |
| **BO** | 1 | | | | | | | |
| **CR** | 1 | | | | | | | |
| **GA** | 1 | | | | | | | |
| **LS** | 1 | | | | | | | |
| **CG** | | 1 | | | | | | |
| **CO** | | 1 | | | | | | |
| **CU** | | 1 | | | | | | |
| **MA** | | 1 | | | | | | |
| **PA** | | 1 | | | | | | |
| **SL** | | 1 | | | | | | |
| **SB** | | | 1 | | | | | |
| **SD** | | | 1 | | | | | |
| **JP** | | | | 1 | | | | |
| **BA** | | | | 1 | | | | |
| **CA** | | | | | 1 | | | |
| **SR** | | | | | 1 | | | |
| **AH** | | | | | | 1 | | |
| **SO** | | | | | | 1 | | |
| **AN** | | | | | | | | |
| **CP** | | | | | | | | |
| **CJ** | | | | | | | | |
| **MO** | | | | | | | | |
| **ES** | | | | | | | | |
| **PE** | | | | | | | | |
| **QE** | | | | | | | | |
| **RE** | | | | | | | | |
| **SA** | | | | | | | | |

### A.13.4 Ponto de Corte igual à 0,50

Tabela 28: Agrupamento dos Municípios para com as Causas Externas de Óbito.

| | AGRUPAMENTOS | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **V89** | **X70** | **X95** | **V49** | **V03** | **V09** | **V99** | **W19** |
| **AG** | 1 | | | | | | | |
| **AO** | 1 | | | | | | | |
| **BO** | 1 | | | | | | | |
| **CR** | 1 | | | | | | | |
| **CO** | | 1 | | | | | | |
| **PA** | | 1 | | | | | | |
| **JP** | | | 1 | | | | | |
| **BA** | | | 1 | | | | | |
| **AH** | | | | 1 | | | | |
| **SO** | | | | 1 | | | | |
| **SR** | | | | | 1 | | | |
| **CU** | | | | | | | | |
| **ES** | | | | | | | | |
| **GA** | | | | | | | | |
| **AR** | | | | | | | | |
| **PE** | | | | | | | | |
| **QE** | | | | | | | | |
| **RE** | | | | | | | | |
| **SB** | | | | | | | | |
| **SA** | | | | | | | | |
| **SL** | | | | | | | | |
| **SD** | | | | | | | | |
| **LS** | | | | | | | | |
| **MA** | | | | | | | | |
| **MO** | | | | | | | | |
| **CP** | | | | | | | | |
| **CA** | | | | | | | | |
| **CJ** | | | | | | | | |
| **CG** | | | | | | | | |
| **AN** | | | | | | | | |

# A.14 Ano de 2007

## A.14.1 Ponto de Corte igual à 0,10

Tabela 29: Agrupamento dos Municípios para com as Causas Externas de Óbito.

| | AGRUPAMENTOS | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **W19** | **X70** | **V89** | **X95** | **V49** | **V03** | **V09** | **V99** |
| **AG** | 1 | 1 | | 1 | 1 | | | |
| **AH** | 1 | 1 | | 1 | 1 | | | |
| **AR** | 1 | 1 | | | | | | |
| **BO** | 1 | 1 | | | | | | |
| **CR** | 1 | 1 | | 1 | | | | |
| **LS** | 1 | 1 | 1 | | | | | |
| **RE** | 1 | 1 | 1 | | 1 | 1 | | |
| **CU** | 1 | 1 | 1 | | 1 | | | |
| **ES** | 1 | 1 | 1 | | 1 | | | |
| **MA** | 1 | 1 | | 1 | 1 | | 1 | |
| **PA** | 1 | 1 | | | | | | |
| **SL** | 1 | 1 | | 1 | | | | |
| **CP** | 1 | | | 1 | | | | |
| **CA** | 1 | | | 1 | | | | |
| **PE** | 1 | | | 1 | | 1 | 1 | |
| **SR** | 1 | | | 1 | | 1 | 1 | |
| **SO** | 1 | | 1 | | | | | |
| **AO** | | 1 | 1 | | | | | |
| **CJ** | | 1 | 1 | | | | | |
| **CG** | | 1 | 1 | | | | | |
| **SD** | | | 1 | 1 | 1 | | | |
| **AN** | | | 1 | 1 | 1 | | | |
| **QE** | | | 1 | | 1 | | | |
| **MO** | | | 1 | | | | | |
| **SB** | | | 1 | | | | | |
| **GA** | | 1 | | | | | | |
| **BA** | | | | 1 | | 1 | | |
| **JP** | | | | 1 | | | | |
| **CO** | | | | | | 1 | 1 | |
| **SA** | | | | | | | | 1 |

### A.14.2 Ponto de Corte igual à 0,20

Tabela 30: Agrupamento dos Municípios para com as Causas Externas de Óbito.

| | AGRUPAMENTOS | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **V89** | **W19** | **X70** | **X95** | **V03** | **V49** | **V99** | **V09** |
| **CU** | 1 | | 1 | | | | | |
| **ES** | 1 | | 1 | | | | | |
| **AN** | 1 | | | | | | | |
| **AO** | 1 | | | | | | | |
| **CJ** | 1 | | | | | | | |
| **CG** | 1 | | | | | | | |
| **LS** | 1 | | | | | | | |
| **MO** | 1 | | | | | | | |
| **QE** | 1 | | | | | | | |
| **SB** | 1 | | | | | | | |
| **SD** | 1 | | | | | | | |
| **PA** | | 1 | 1 | | | | | |
| **SL** | | 1 | | | | | | |
| **PE** | | 1 | | 1 | | | | |
| **CA** | | 1 | | 1 | | | | |
| **CR** | | 1 | | 1 | | | | |
| **SR** | | 1 | | 1 | | | | |
| **CP** | | 1 | | | | | | |
| **AH** | | 1 | | | | 1 | | |
| **SO** | | 1 | | | | | | |
| **GA** | | | 1 | | | | | |
| **AR** | | | 1 | | | | | |
| **BO** | | | 1 | | | | | |
| **MA** | | | | 1 | | | | |
| **JP** | | | | 1 | | | | |
| **BA** | | | | | 1 | | | |
| **CO** | | | | | 1 | | | |
| **AG** | | | | | | 1 | | |
| **SA** | | | | | | | 1 | |
| **RE** | | | | | | | | |

### A.14.3 Ponto de Corte igual à 0,30

Tabela 31: Agrupamento dos Municípios para com as Causas Externas de Óbito.

| | AGRUPAMENTOS | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **V89** | **W19** | **X70** | **V49** | **V03** | **X95** | **V99** | **V09** |
| **CJ** | 1 | | | | | | | |
| **CG** | 1 | | | | | | | |
| **AN** | 1 | | | | | | | |
| **AO** | 1 | | | | | | | |
| **LS** | 1 | | | | | | | |
| **MO** | 1 | | | | | | | |
| **QE** | 1 | | | | | | | |
| **SB** | 1 | | | | | | | |
| **SD** | 1 | | | | | | | |
| **CP** | | 1 | | | | | | |
| **CA** | | 1 | | | | 1 | | |
| **CR** | | 1 | | | | | | |
| **SR** | | 1 | | | | | | |
| **SO** | | 1 | | | | | | |
| **SL** | | 1 | | | | | | |
| **PE** | | 1 | | | | | | |
| **PA** | | | 1 | | | | | |
| **AR** | | | 1 | | | | | |
| **BO** | | | 1 | | | | | |
| **GA** | | | 1 | | | | | |
| **AG** | | | | 1 | | | | |
| **AH** | | | | 1 | | | | |
| **CO** | | | | | 1 | | | |
| **BA** | | | | | 1 | | | |
| **JP** | | | | | | 1 | | |
| **SA** | | | | | | | 1 | |
| **CU** | | | | | | | | |
| **ES** | | | | | | | | |
| **MA** | | | | | | | | |
| **RE** | | | | | | | | |

### A.14.4 Ponto de Corte igual à 0,50

Tabela 32: Agrupamento dos Municípios para com as Causas Externas de Óbito.

| | AGRUPAMENTOS | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **X70** | **V03** | **V89** | **V99** | **W19** | **X95** | **V09** | **V49** |
| **AR** | 1 | | | | | | | |
| **BO** | 1 | | | | | | | |
| **GA** | 1 | | | | | | | |
| **CO** | | 1 | | | | | | |
| **BA** | | 1 | | | | | | |
| **QE** | | | 1 | | | | | |
| **SB** | | | 1 | | | | | |
| **SA** | | | | 1 | | | | |
| **SO** | | | | | 1 | | | |
| **JP** | | | | | | 1 | | |
| **CU** | | | | | | | | |
| **ES** | | | | | | | | |
| **AO** | | | | | | | | |
| **AG** | | | | | | | | |
| **AN** | | | | | | | | |
| **AH** | | | | | | | | |
| **CP** | | | | | | | | |
| **CA** | | | | | | | | |
| **CJ** | | | | | | | | |
| **CG** | | | | | | | | |
| **CR** | | | | | | | | |
| **LS** | | | | | | | | |
| **MA** | | | | | | | | |
| **MO** | | | | | | | | |
| **PA** | | | | | | | | |
| **PE** | | | | | | | | |
| **RE** | | | | | | | | |
| **SR** | | | | | | | | |
| **SL** | | | | | | | | |
| **SD** | | | | | | | | |

# A.15 Ano de 2008

## A.15.1 Ponto de Corte igual à 0,10

Tabela 33: Agrupamento dos Municípios para com as Causas Externas de Óbito.

| | AGRUPAMENTOS | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **V89** | **W19** | **X95** | **X70** | **V49** | **V03** | **V09** | **V99** |
| **CP** | 1 | 1 | 1 | 1 | | 1 | | |
| **MA** | 1 | 1 | 1 | 1 | | | | |
| **MO** | 1 | 1 | 1 | 1 | | | | |
| **AH** | 1 | 1 | 1 | 1 | 1 | | | |
| **CG** | 1 | 1 | 1 | | 1 | | | |
| **CR** | 1 | 1 | 1 | | 1 | | | |
| **AO** | 1 | 1 | 1 | | | | | |
| **PA** | 1 | 1 | 1 | | 1 | 1 | | |
| **SR** | 1 | 1 | 1 | | 1 | | | |
| **SB** | 1 | 1 | 1 | | | | 1 | |
| **AN** | 1 | 1 | | 1 | 1 | | | |
| **AR** | 1 | 1 | | 1 | 1 | | | |
| **BO** | 1 | 1 | | 1 | 1 | | | |
| **ES** | 1 | 1 | | 1 | 1 | | | |
| **CU** | 1 | 1 | | 1 | | | | |
| **LS** | 1 | 1 | | | | | | |
| **QE** | 1 | 1 | | | | 1 | | |
| **RE** | 1 | 1 | | | | | 1 | |
| **SD** | 1 | 1 | | | 1 | | | |
| **SO** | 1 | 1 | | | | | | |
| **CO** | 1 | | 1 | 1 | | 1 | | |
| **GA** | 1 | | 1 | 1 | 1 | | 1 | 1 |
| **SL** | 1 | | 1 | 1 | 1 | | 1 | 1 |
| **AG** | | | | 1 | | | | 1 |
| **BA** | | 1 | 1 | | 1 | 1 | 1 | |
| **CA** | | 1 | 1 | | | 1 | | |
| **JP** | | | 1 | 1 | | | 1 | |
| **PE** | | | 1 | 1 | | 1 | | |
| **SA** | | | 1 | 1 | | 1 | 1 | |
| **CJ** | | | | | 1 | | 1 | |

## A.15.2 Ponto de Corte igual à 0,20

Tabela 34: Agrupamento dos Municípios para com as Causas Externas de Óbito.

| | AGRUPAMENTOS | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **V89** | **W19** | **X95** | **V03** | **V09** | **V49** | **V99** | **X70** |
| **MA** | 1 | 1 | 1 | | | | | |
| **AN** | 1 | 1 | | | | | | |
| **AH** | 1 | 1 | | | | | | |
| **AR** | 1 | 1 | | | | | | |
| **AO** | 1 | 1 | | | | | | |
| **LS** | 1 | 1 | | | | | | |
| **MO** | 1 | 1 | | | | | | |
| **QE** | 1 | 1 | | | | | | |
| **BO** | 1 | 1 | | | | | | |
| **CR** | 1 | 1 | | | | | | |
| **ES** | 1 | 1 | | | | | | |
| **SD** | 1 | 1 | | | | | | |
| **SO** | 1 | 1 | | | | | | |
| **SB** | 1 | | | | | | | |
| **RE** | 1 | | | | | | | |
| **CG** | | 1 | 1 | | | | | |
| **CU** | | 1 | | | | | | |
| **JP** | | | 1 | | | | | |
| **SR** | | | 1 | | | | | |
| **CO** | | | 1 | | | | | |
| **PA** | | | 1 | | | | | |
| **PE** | | | 1 | | | | | |
| **BA** | | | 1 | | | | | |
| **CA** | | | 1 | | | | | |
| **CP** | | | | 1 | | | | |
| **SA** | | | | 1 | | | | 1 |
| **CJ** | | | | | 1 | 1 | | |
| **AG** | | | | | | | 1 | |
| **GA** | | | | | | | | |
| **SL** | | | | | | | | |

### A.15.3 Ponto de Corte igual à 0,30

Tabela 35: Agrupamento dos Municípios para com as Causas Externas de Óbito.

| | AGRUPAMENTOS | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **V89** | **W19** | **X95** | **V03** | **V09** | **V99** | **V49** | **X70** |
| **SO** | 1 | 1 | | | | | | |
| **AO** | 1 | 1 | | | | | | |
| **BO** | 1 | 1 | | | | | | |
| **LS** | 1 | 1 | | | | | | |
| **RE** | 1 | | | | | | | |
| **ES** | 1 | | | | | | | |
| **MO** | 1 | | | | | | | |
| **SB** | 1 | | | | | | | |
| **CG** | | 1 | | | | | | |
| **CU** | | 1 | | | | | | |
| **QE** | | 1 | | | | | | |
| **BA** | | | 1 | | | | | |
| **CA** | | | 1 | | | | | |
| **CO** | | | 1 | | | | | |
| **PE** | | | 1 | | | | | |
| **SR** | | | 1 | | | | | |
| **JP** | | | 1 | | | | | |
| **SA** | | | | 1 | | | | |
| **CJ** | | | | | 1 | | | |
| **AG** | | | | | | 1 | | |
| **AN** | | | | | | | | |
| **AH** | | | | | | | | |
| **AR** | | | | | | | | |
| **CP** | | | | | | | | |
| **CR** | | | | | | | | |
| **GA** | | | | | | | | |
| **MA** | | | | | | | | |
| **PA** | | | | | | | | |
| **SL** | | | | | | | | |
| **SD** | | | | | | | | |

### A.15.4 Ponto de Corte igual à 0,50

Tabela 36: Agrupamento dos Municípios para com as Causas Externas de Óbito.

| | AGRUPAMENTOS | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **V99** | **V03** | **V09** | **V49** | **V89** | **W19** | **X70** | **X95** |
| **AG** | 1 | | | | | | | |
| **AN** | | | | | | | | |
| **AH** | | | | | | | | |
| **AR** | | | | | | | | |
| **AO** | | | | | | | | |
| **BA** | | | | | | | | |
| **BO** | | | | | | | | |
| **CP** | | | | | | | | |
| **CA** | | | | | | | | |
| **CJ** | | | | | | | | |
| **CG** | | | | | | | | |
| **CR** | | | | | | | | |
| **CO** | | | | | | | | |
| **CU** | | | | | | | | |
| **ES** | | | | | | | | |
| **GA** | | | | | | | | |
| **JP** | | | | | | | | |
| **LS** | | | | | | | | |
| **MA** | | | | | | | | |
| **MO** | | | | | | | | |
| **PA** | | | | | | | | |
| **PE** | | | | | | | | |
| **QE** | | | | | | | | |
| **RE** | | | | | | | | |
| **SR** | | | | | | | | |
| **SB** | | | | | | | | |
| **SA** | | | | | | | | |
| **SL** | | | | | | | | |
| **SD** | | | | | | | | |
| **SO** | | | | | | | | |

# A.16 Ano de 2009

## A.16.1 Ponto de Corte igual à 0,10

Tabela 37: Agrupamento dos Municípios para com as Causas Externas de Óbito.

| | AGRUPAMENTOS | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **W19** | **V03** | **X70** | **V89** | **X95** | **V09** | **V49** | **V99** |
| **CR** | 1 | 1 | 1 | 1 | 1 | 1 | | |
| **BA** | 1 | 1 | 1 | 1 | 1 | 1 | 1 | |
| **GA** | 1 | 1 | 1 | 1 | 1 | | 1 | |
| **BO** | 1 | 1 | 1 | 1 | 1 | | 1 | |
| **CU** | 1 | 1 | 1 | 1 | 1 | | | |
| **QE** | 1 | 1 | 1 | 1 | | | | |
| **AO** | 1 | 1 | 1 | 1 | | | | |
| **RE** | 1 | 1 | 1 | 1 | | | | |
| **SA** | 1 | 1 | 1 | | | 1 | 1 | |
| **CJ** | 1 | 1 | 1 | | | 1 | | 1 |
| **CG** | 1 | 1 | | 1 | 1 | | | |
| **AH** | 1 | 1 | | 1 | 1 | | | |
| **ES** | 1 | 1 | | 1 | | | | |
| **SL** | 1 | 1 | | 1 | | 1 | | |
| **AN** | 1 | | 1 | | | 1 | 1 | |
| **LS** | 1 | | 1 | | 1 | | | |
| **PE** | 1 | | 1 | | 1 | | 1 | |
| **SB** | 1 | | 1 | | | 1 | | |
| **SO** | 1 | | 1 | | | 1 | | |
| **AR** | 1 | | | 1 | | | | |
| **CO** | 1 | | | 1 | 1 | | 1 | |
| **MA** | 1 | | | 1 | 1 | 1 | | |
| **CP** | | 1 | | | 1 | | | |
| **CA** | | 1 | | | 1 | | | |
| **PA** | | 1 | 1 | | 1 | | | |
| **SR** | | 1 | | | 1 | | | |
| **AG** | | | 1 | | | 1 | 1 | 1 |
| **MO** | | | 1 | 1 | | | | |
| **SD** | | | | 1 | | | | |
| **JP** | | | | | 1 | | | |

## A.16.2 Ponto de Corte igual à 0,20

Tabela 38: Agrupamento dos Municípios para com as Causas Externas de Óbito.

| | AGRUPAMENTOS | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **V89** | **X70** | **X95** | **V03** | **V99** | **W19** | **V09** | **V49** |
| **AO** | 1 | 1 | | | | | | |
| **AH** | 1 | | | | | | | |
| **AR** | 1 | | | | | | | |
| **CG** | 1 | | | | | | | |
| **CR** | 1 | | | | | | | |
| **CU** | 1 | | | | | | | |
| **ES** | 1 | | | | | | | |
| **MO** | 1 | | | | | | | |
| **QE** | 1 | | | | | | | |
| **RE** | 1 | | | | | | | |
| **SL** | 1 | | | | | | | |
| **SD** | 1 | | | | | | | |
| **AN** | | 1 | | | | | | |
| **PE** | | 1 | | | | 1 | | |
| **LS** | | 1 | | | | 1 | | |
| **SB** | | 1 | | | | | | |
| **SO** | | 1 | | | | | | |
| **PA** | | | 1 | | | | | |
| **CO** | | | 1 | | | | | |
| **MA** | | | 1 | | | | | |
| **JP** | | | 1 | | | | | |
| **CP** | | | | 1 | | | | |
| **CA** | | | | 1 | | | | |
| **SR** | | | | 1 | | | | |
| **AG** | | | | | 1 | | | |
| **CJ** | | | | | 1 | | | |
| **BA** | | | | | | | 1 | |
| **SA** | | | | | | | | 1 |
| **BO** | | | | | | | | |
| **GA** | | | | | | | | |

### A.16.3 Ponto de Corte igual à 0,30

Tabela 39: Agrupamento dos Municípios para com as Causas Externas de Óbito.

| | AGRUPAMENTOS | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **V89** | **V03** | **X70** | **X95** | **V99** | **V09** | **V49** | **W19** |
| **QE** | 1 | | | | | | | |
| **RE** | 1 | | | | | | | |
| **SL** | 1 | | | | | | | |
| **SD** | 1 | | | | | | | |
| **AR** | 1 | | | | | | | |
| **CG** | 1 | | | | | | | |
| **ES** | 1 | | | | | | | |
| **MO** | 1 | | | | | | | |
| **CP** | | 1 | | | | | | |
| **CA** | | 1 | | | | | | |
| **SR** | | 1 | | | | | | |
| **AN** | | | 1 | | | | | |
| **SB** | | | 1 | | | | | |
| **SO** | | | 1 | | | | | |
| **JP** | | | | 1 | | | | |
| **PA** | | | | 1 | | | | |
| **AG** | | | | | 1 | | | |
| **CR** | | | | | | | | |
| **CO** | | | | | | | | |
| **CU** | | | | | | | | |
| **GA** | | | | | | | | |
| **LS** | | | | | | | | |
| **MA** | | | | | | | | |
| **AO** | | | | | | | | |
| **BA** | | | | | | | | |
| **BO** | | | | | | | | |
| **CJ** | | | | | | | | |
| **AH** | | | | | | | | |
| **PE** | | | | | | | | |
| **SA** | | | | | | | | |

### A.16.4 Ponto de Corte igual à 0,50

Tabela 40: Agrupamento dos Municípios para com as Causas Externas de Óbito.

| | AGRUPAMENTOS | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **V03** | **V89** | **X95** | **V09** | **V49** | **V99** | **W19** | **X70** |
| **CP** | 1 | | | | | | | |
| **CA** | 1 | | | | | | | |
| **SR** | 1 | | | | | | | |
| **SD** | | 1 | | | | | | |
| **JP** | | | 1 | | | | | |
| **AG** | | | | | | | | |
| **AN** | | | | | | | | |
| **AH** | | | | | | | | |
| **AR** | | | | | | | | |
| **AO** | | | | | | | | |
| **BA** | | | | | | | | |
| **BO** | | | | | | | | |
| **CJ** | | | | | | | | |
| **CG** | | | | | | | | |
| **CR** | | | | | | | | |
| **CO** | | | | | | | | |
| **CU** | | | | | | | | |
| **ES** | | | | | | | | |
| **GA** | | | | | | | | |
| **LS** | | | | | | | | |
| **MA** | | | | | | | | |
| **MO** | | | | | | | | |
| **PA** | | | | | | | | |
| **PE** | | | | | | | | |
| **QE** | | | | | | | | |
| **RE** | | | | | | | | |
| **SB** | | | | | | | | |
| **SA** | | | | | | | | |
| **SL** | | | | | | | | |
| **SO** | | | | | | | | |

# A.17 Ano de 2010

## A.17.1 Ponto de Corte igual à 0,10

Tabela 41: Agrupamento dos Municípios para com as Causas Externas de Óbito.

| | AGRUPAMENTOS | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **V49** | **V89** | **X95** | **V09** | **V99** | **W19** | **X70** | **V03** |
| **CU** | 1 | 1 | 1 | 1 | | 1 | | |
| **PA** | 1 | 1 | 1 | 1 | | | | |
| **CJ** | 1 | 1 | 1 | 1 | | | | |
| **MA** | 1 | 1 | 1 | | | | | |
| **CG** | 1 | 1 | 1 | | | | | |
| **CO** | 1 | 1 | 1 | | 1 | 1 | 1 | |
| **CP** | 1 | 1 | 1 | | 1 | | | |
| **AR** | 1 | 1 | 1 | | | 1 | 1 | |
| **SD** | 1 | 1 | 1 | | 1 | 1 | 1 | |
| **PE** | 1 | 1 | | | 1 | 1 | 1 | |
| **QE** | 1 | 1 | | | 1 | 1 | | |
| **SO** | 1 | 1 | | 1 | | | | |
| **SB** | 1 | 1 | | 1 | | | | |
| **AH** | 1 | 1 | | 1 | | | | |
| **AO** | 1 | 1 | | 1 | | | | |
| **BO** | 1 | 1 | | 1 | | | | |
| **ES** | 1 | 1 | | | | 1 | 1 | |
| **BA** | 1 | | 1 | 1 | 1 | | | 1 |
| **GA** | 1 | | 1 | 1 | 1 | | | 1 |
| **JP** | 1 | | 1 | 1 | 1 | | | 1 |
| **AG** | 1 | | 1 | | 1 | 1 | 1 | 1 |
| **RE** | 1 | | | | | 1 | 1 | |
| **SA** | 1 | | 1 | 1 | 1 | | | 1 |
| **AN** | | 1 | 1 | | 1 | 1 | 1 | |
| **LS** | | 1 | 1 | | | 1 | | |
| **SL** | | 1 | 1 | | | 1 | 1 | 1 |
| **MO** | | 1 | | | | 1 | 1 | |
| **SR** | | | 1 | 1 | | | | 1 |
| **CA** | | | 1 | | 1 | | | 1 |
| **CR** | | | 1 | | 1 | | | 1 |

## A.17.2 Ponto de Corte igual à 0,20

Tabela 42: Agrupamento dos Municípios para com as Causas Externas de Óbito.

| | AGRUPAMENTOS | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **V49** | **X95** | **V89** | **W19** | **X70** | **V03** | **V99** | **V09** |
| **SO** | 1 | | 1 | | | | | |
| **AH** | 1 | | | | | | | 1 |
| **AO** | 1 | | | | | | | |
| **BO** | 1 | | | | | | | |
| **CP** | 1 | | 1 | | | | | |
| **CJ** | 1 | | | | | | | |
| **PA** | 1 | | | | | | | |
| **SB** | 1 | | | | | | | |
| **CG** | | 1 | 1 | | | | | |
| **CA** | | 1 | | | | 1 | 1 | |
| **CR** | | 1 | | | | 1 | 1 | |
| **BA** | | 1 | | | | | | |
| **JP** | | 1 | | | | | | |
| **MA** | | 1 | | | | | | |
| **SR** | | 1 | | | | | | |
| **SA** | | 1 | | | | | | |
| **MO** | | | 1 | | | | | |
| **SL** | | | 1 | | | | | |
| **SD** | | | 1 | | | | | |
| **QE** | | | | 1 | | | | |
| **RE** | | | | 1 | 1 | | | |
| **LS** | | | | 1 | | | | |
| **PE** | | | | | 1 | | | |
| **AG** | | | | | 1 | | | |
| **AN** | | | | | 1 | | | |
| **CU** | | | 1 | | | | | |
| **ES** | | | | 1 | | | | |
| **GA** | | | | | | 1 | | |
| **AR** | | | | | | | | |
| **CO** | | | | | | | | |

### A.17.3 Ponto de Corte igual à 0,30

Tabela 43: Agrupamento dos Municípios para com as Causas Externas de Óbito.

| | AGRUPAMENTOS | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **V49** | **X95** | **W19** | **V89** | **V03** | **V09** | **V99** | **X70** |
| **AH** | 1 | | | | | | | |
| **AO** | 1 | | | | | | | |
| **BO** | 1 | | | | | | | |
| **CP** | 1 | | | | | | | |
| **CJ** | 1 | | | | | | | |
| **SB** | 1 | | | | | | | |
| **CA** | | 1 | | | | | | |
| **CG** | | 1 | | | | | | |
| **MA** | | 1 | | | | | | |
| **SR** | | 1 | | | | | | |
| **LS** | | | 1 | | | | | |
| **QE** | | | 1 | | | | | |
| **MO** | | | | 1 | | | | |
| **BA** | | | | | | | | |
| **AR** | | | | | | | | |
| **AG** | | | | | | | | |
| **AN** | | | | | | | | |
| **CR** | | | | | | | | |
| **CO** | | | | | | | | |
| **CU** | | | | | | | | |
| **ES** | | | | | | | | |
| **GA** | | | | | | | | |
| **JP** | | | | | | | | |
| **PA** | | | | | | | | |
| **PE** | | | | | | | | |
| **RE** | | | | | | | | |
| **SA** | | | | | | | | |
| **SL** | | | | | | | | |
| **SD** | | | | | | | | |
| **SO** | | | | | | | | |

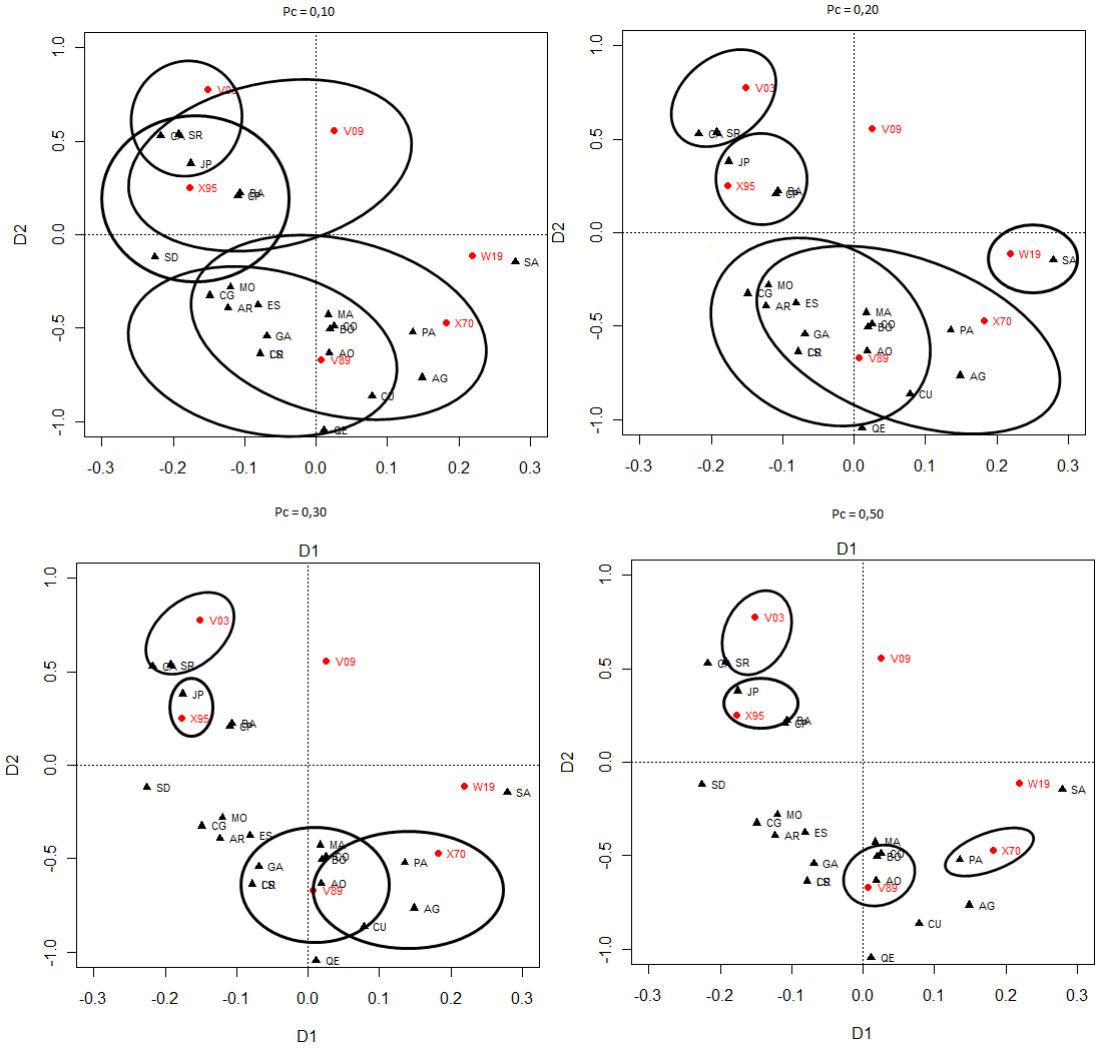
### A.17.4 Ponto de Corte igual à 0,50

Tabela 44: Agrupamento dos Municípios para com as Causas Externas de Óbito.

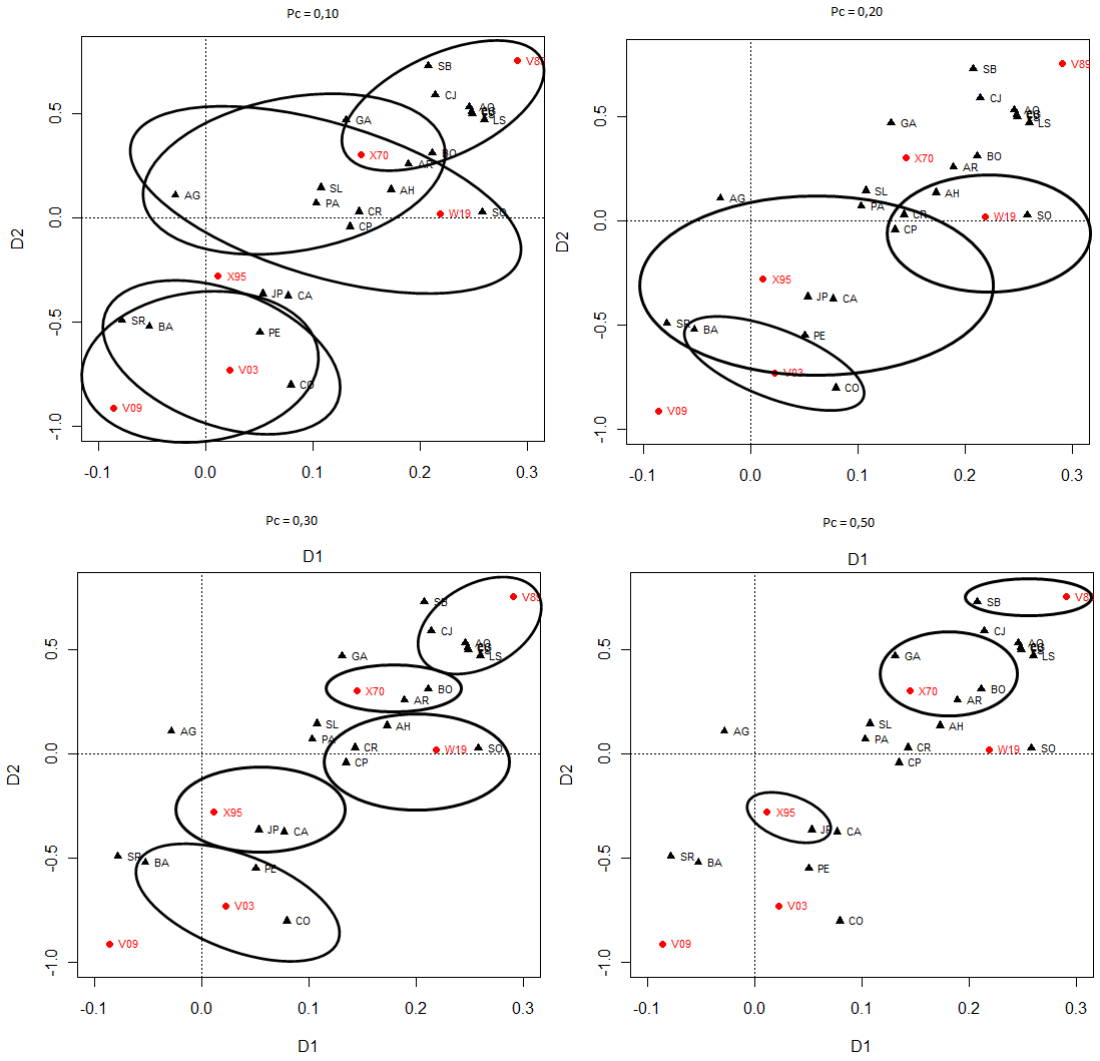
| | AGRUPAMENTOS | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **V03** | **V09** | **V49** | **V89** | **V99** | **W19** | **X70** | **X95** |
| **AG** | | | | | | | | |
| **AN** | | | | | | | | |
| **AH** | | | | | | | | |
| **AR** | | | | | | | | |
| **AO** | | | | | | | | |
| **BA** | | | | | | | | |
| **BO** | | | | | | | | |
| **CP** | | | | | | | | |
| **CA** | | | | | | | | |
| **CJ** | | | | | | | | |
| **CG** | NÃO HOUVE NENHUM AGRUPAMENTO | | | | | | | |
| **CR** | | | | | | | | |
| **CO** | | | | | | | | |
| **CU** | | | | | | | | |
| **ES** | | | | | | | | |
| **GA** | | | | | | | | |
| **JP** | | | | | | | | |
| **LS** | | | | | | | | |
| **MA** | | | | | | | | |
| **MO** | | | | | | | | |
| **PA** | | | | | | | | |
| **PE** | | | | | | | | |
| **QE** | | | | | | | | |
| **RE** | | | | | | | | |
| **SR** | | | | | | | | |
| **SB** | | | | | | | | |
| **SA** | | | | | | | | |
| **SL** | | | | | | | | |
| **SD** | | | | | | | | |
| **SO** | | | | | | | | |

# A.18 Gráficos de Sensibilidade para Diferentes Pontos de Corte por Ano

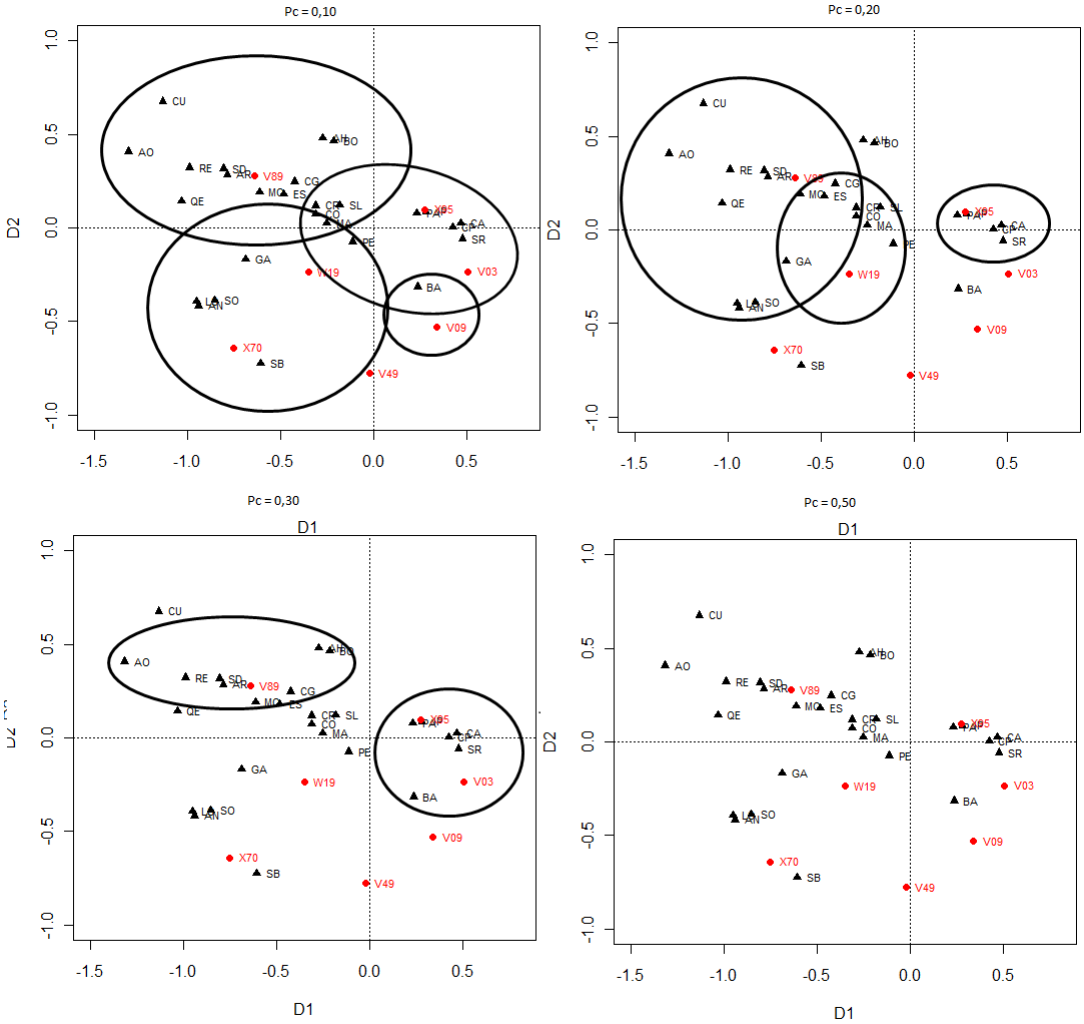
## A.18.1 Para o ano de 2006



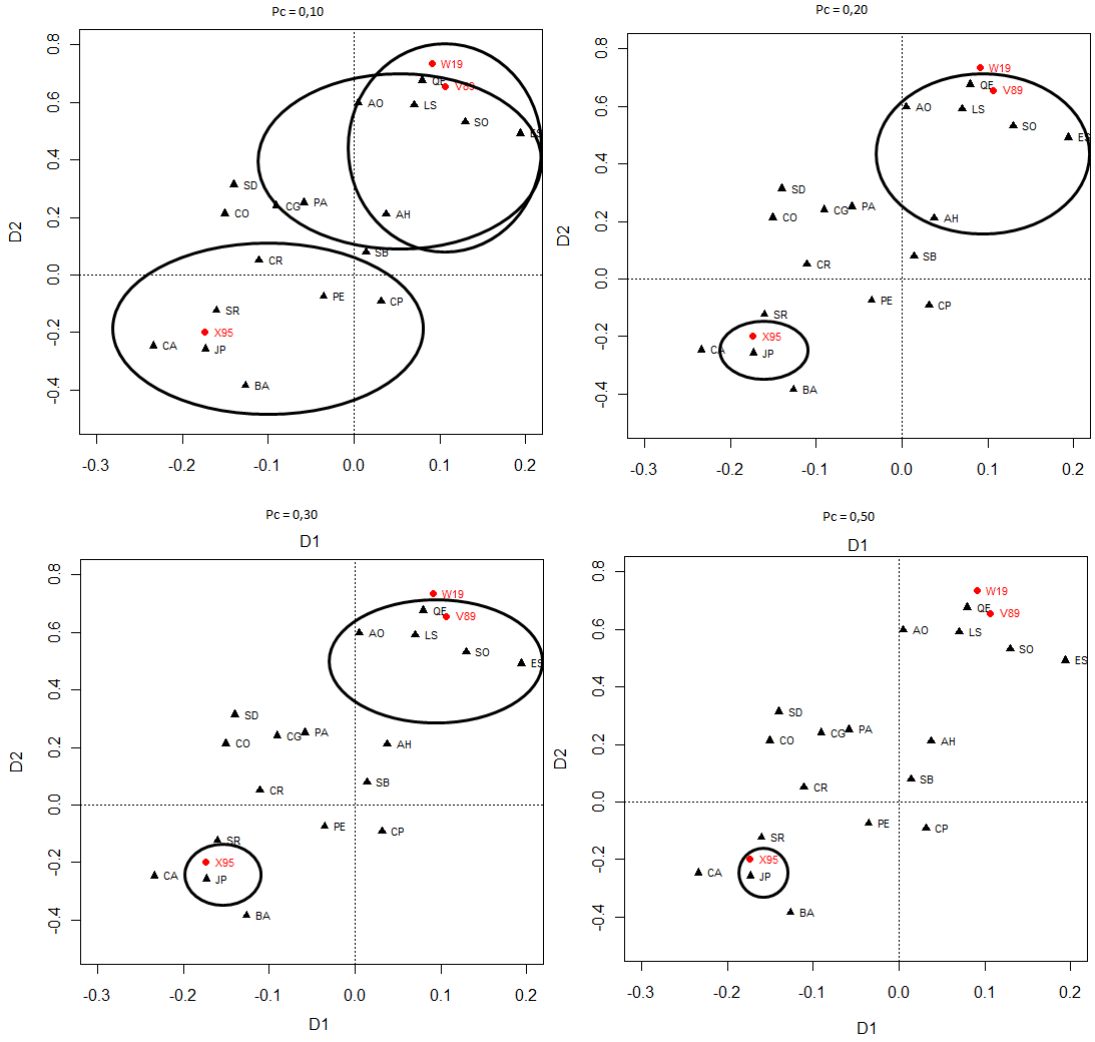
## A.18.2 Para o ano de 2007



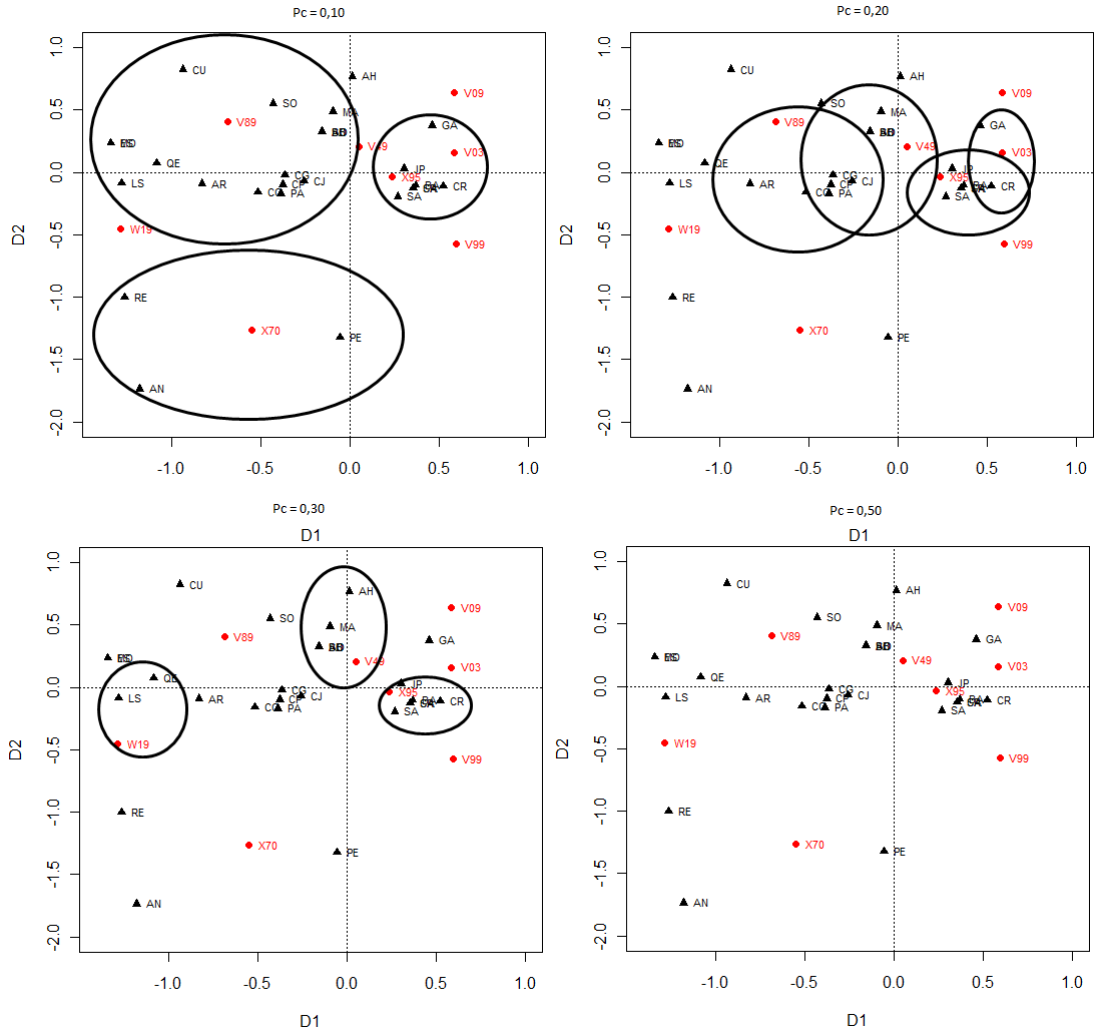
## A.18.3 Para o ano de 2008



## A.18.4 Para o ano de 2009



## A.18.5 Para o ano de 2010



# *Referências*

1 BENZÉCRI J.P. Correspondence Analysis Handbook. New York: Marcel, 1992.

2 BEZDEK J. C., CORAY C., GUNDERSON C. , WATSON J. Detection and Characterization of Cluster Substructure. SIAM Journal of Applied Mathematics, 40: pp. 339-372, 1981.

3 CAVALCANTI N. L. Clusterização baseada em algoritmos Fuzzy, Universidade Federal de Pernambuco, 2006.

4 BROWN M.,HARRIS C. J., MOORE C. G., Intelligent control: Aspects of Fuzzy Logic and Neural Nets, World Scientific, 1993.

5 CINTULA, LIBOR B. P. Fuzzy class theory, Academy of Sciences of the Czech Republic 154, volume 154, pp. 34–55, 2005.

6 COX E. The Fuzzy Systems Handbook: a Practitioner's Guide to Building, Using and Maintaining Fuzzy Systems, Eds: Professional,1, 1994;

7 CORREA, S. S. Lógica Nebulosa, XVIII Escola de Redes Neurais , volume 18, pp. 73–90, 1999.

8 CLAIR U. S., YUAN B. Fuzzy Set Theory: Foundations and Applications, Prentice Hall PTR, Upper Saddle River, NJ, 1997.

9 CRAWLEY M. J. The R Book, Wiley, 1, 2007.

10 DANTAS, GOLDBARG M. Algoritmos evolucionários na determinação da configuração de custo mínimo de sistemas de co-geração de energia com base no gás natural, Pesquisa Operacional 25, volume 25, pp. 231–259, 2005.

11 FERREIRA, F. D. Estatística Multivariada, 1, Eds: UFLA, pp. 341-391 ,2008.

12 FANCO C. R. Novos Métodos de Classificação Nebulosa e de Validação de Categorias e suas Aplicações a Problemas de Reconhecimento de Padrões, Universidade Federal do Rio de Janeiro, 2002.

13 GARCIA F. Desempenho energético de um sistema de refrigeração aplicando o controle adaptativo Fuzzy, Simpósio de Pós-Graduação em Engenharia Mecânica, MG, 2002.

14 GOLDBERG E. D. Genetic Algorithms in Search, Optimization, and Machine Learning, EUA: Addison-Wesley, 121, 1989.

15 GOLDBERG E. D. Genetic Algorithms in Search, Optimization, and Machine Learning, EUA: Addison-Wesley, 147, 1989.

16 GOLDBERG E. D. Genetic Algorithms in Search, Optimization, and Machine Learning. EUA: Addison-Wesley, 80, 1989.

17 GOLDBARG. Algoritmos evolucionários na determinação da configuração de custo mínimo de sistemas de co-geração de energia com base no gás natural, Pesqui. Oper. 25(2): 231-259, 2005.

18 GREENACRE M., HASTIE T. The geometric interpretation of correspondence analysis, Journal of the American Statistical Association, volume 82, pp. 437-447,1987.

19 HAIR J. F. BLACK W. C., BARRY J., BABIN R. E., ANERSON R. L.. Análise Multivariada de Dados, Eds: bookman, 2009.

20 HAIR O. F., BLACK W. C., BARRY J. B., ROLPH A. E., TATHAM R. L. Análise Multivariada de Dados, 6, Eds: Bookman,2006.

21 HEIJDEN P. et al. A combined approach to contingency table analysis using correspondence analysis and log-linear analysis. Applied statistics 38, 1989, pp. 249-292.

22 JOHNSON R. A., WICHERN D. W. Applied Multivariate Statistical Analysis, Ed. 6, Eds: Hardcover, 2007.

23 JOHNSON R. A., WICHERN D. Applied Multivariate statistical Analysis, 6ª, Eds: Hardcover, 2008.

24 KELLER, ANNETTE F. K. Fuzzy Clustering with Evolutionary Algorithms, International Journal of Intelligent Systems 13, volume 13, pp. 975–991, 1998.

25 KLAWONN F., KELLER A. Fuzzy Clustering with Evolutionary Algorithm,nternacional of Intelligent Systems, volume 13,1998.

26 KOZA, J.R. Genetic Programming: On the Programming of Computers by Means of Natural Selection. [S.l.]: MIT Press, 1992.

27 KOSKO B. Neural networks and Fuzzy systems: a dynamical systems approach to machine intelligence, Prentice-Hall International, 1992.

28 LEITE, TEIXEIRA. Aplicação de algoritmos genéticos na determinação da operação ótima de sistemas hidrotérmicos de potência, SBA Controle & Automação 17(1): pp. 81-88, 2006.

29 LIU H. C., JEAG B. C.,YU Y. K. Fuzzy C-Means Algorithm Based on Standard Mahalanobis Distances, International Symposium on Information Processing, China, 2009, pp. 422-427.

30 LIMA O. J. Classificações Repetidas nos Aspectos Inferenciais em Experimento de Bernoulli com Erros de Diagnóstico, 2009. Tese (Doutorado em Estatística) - Departamento de Estatística, Universidade Federal de Minas Gerais, MG.

31 MICHALEWICZ Z., NAZHIYATH G., MICHALEWICZ M. A Note on Usefulness of Geometrical Crossover for Numerical Optimization Problems, Proceedings of the 5th Annual Conference on Evolutionary Programming, San Diego, CA, 29 February - 3 March. MIT Press, Cambridge, MA, 1996, pp. 305-312.

32 NUOVO A. G., CATANIA V., PALESI M. The Hybrid Genetic Fuzzy C-Means: a Reasoned Implementation, International Conference on Fuzzy Systems, Cavtat, Croatia, 2006, pp. 33-38.

33 PALESI A., NUOVO V., MAURIZIO C. The Hybrid Genetic Fuzzy C-Means: a Reasoned Implementation, International Conference on Fuzzy Systems 7, volume 7, pp. 33–38, 2006.

34 PEDRYCZ W., GOMIDE F. Fuzzy Systems Engineering: Toward Human-Centric Computing, Wiley/IEEE Press, 2007.

35 PEDZYCZ W., GOMIDE F. Fuzzy Systems Engineering: Toward Human-Centric Computing, Wiley/IEEE Press, 2007.

36 PIRES, Algoritmos Genéticos. Aplicação à Robótica, Faculdade de Engenharia da Universidade do Porto, 1998.

37 POLI R., LANGDON W. B., MCPHEE N. F. A Field Guide to Genetic Programming, 2008.

38 R: A Language and Environment for Statistical Computing, R Development Core Team, R Foundation for Statistical Computing, Vienna, Austria, 2010, ISBN, 3-900051-07-0, http://www.R-project.org.

39 ROSS T. J. Fuzzy logic with engineering applications 1995, NY McGraw-Hill.

40 ROLLY I., MUKAIDONO M. Redundant Object and Dependency of Domain Attributes in - Goverings of the Universe, FUZZ-IEEE, 2001, pp. 1444-1447.

41 SCHMIDT, MIKLOS F. Fuzzy Clustering as a Means of Selecting Representative Conformers and Molecular Alignments, J. Chem. Inf. Comput. Sci 43, volume 43, pp. 810–818, 2003.

42 YU Y. K., LIU H. C., JENG B. C., YIH J. M. Fuzzy C-Means Algorithm Based on Standard Mahalanobis Distances, International Symposium on Information Processing 23, volume 23, pp. 422–427, 2009.

43 ZADEH L. A. Fuzzy Sets, Rev. Information and control, volume 8, pp. 338-353, 1965.

44 ZADEH L. A. Fuzzy sets and systems, Fox J, Ed. System Theory, Brooklyn, NY: Polytechnic Press, 1965: pp. 29–39.