
**ESTUDO E APLICAÇÕES DO MÉTODO DE AGRUPAMENTO
BASEADO NO MODELO**

PATRÍCIA BATISTA LEAL

Orientadora: Prof.^a Dra. Viviana Giampaoli
Área de concentração: Estatística Aplicada

Dissertação submetida como requerimento parcial para obtenção do grau de Mestre em Estatística pela Universidade Federal de Pernambuco

Recife, 2004

Universidade Federal de Pernambuco
Mestrado em Estatística

27 de Fevereiro de 2004

(16/4)

Nós recomendamos que a dissertação de mestrado de autoria de

Patrícia Estiana Leal

Intitulada

Estudo e aplicações do método de agrupamento hierárquico em matrizes

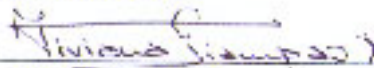
seja aceita como cumprimento parcial dos requerimentos para o grau de Mestre em Estatística.



Coordenador da Pós-Graduação em Estatística

Banca Examinadora:

 **Prof. Francisco Cabral Neto**
Coordenador de Mestrado
em Estatística - UFPE

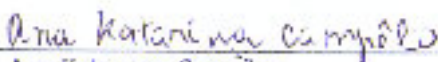


Viviana Araújo

orientador



Maria Cristina Patrício Raposo



Ana Katarina Campêlo

Este documento será anexado à versão final da dissertação.

Aos meus pais, Ivonete e Antônio.
As minhas irmãs, Rosana e Cristiane.

Agradecimentos

A Deus, por me conceder a graça de concluir mais uma etapa da minha vida.

A minha família que sempre me incentivou, em especial a minha mãe.

Agradeço, de forma especial, a minha orientadora Viviana Giampaoli pela segurança, disponibilidade, paciência e incentivo.

A professora Cristina, por estar sempre disposta a me ajudar, por sua alegria.

A Lindomberg, por seu amor, por acreditar em mim, pelo seu incentivo e compreensão.

A Gilmar, meu amigo, meu irmão, pela amizade incondicional.

A Michelli, minha querida amiga, pela amizade, pelo apoio, por saber que posso contar com ela sempre.

A Horácio, pela amizade e pelas vezes que sempre se dispôs a me ajudar quando precisei.

A Silvia, minha irmã de coração, pela amizade, pelo companheirismo e pelos momentos felizes que dividimos.

A Gilson, por sua amizade, por saber transmitir alegria e calma.

A Moisés, que nos meus momentos de dificuldade sempre aparecia disposto a me ajudar.

A minha turma de mestrado, que se tornou uma família: Sílvia, Gilson, Moisés, Bartolomeu, Felipe, Tatiene, Tarciana, Raydonal, João Marcelo e Keila. Juntos dividimos nossas saudades, angústias, apreensões, nosso sucesso, nossas experiências e principalmente nossas alegrias.

Aos meus colegas de mestrado Michelli, Diana, Carla, Patrícia Leone, Amanda, Heráclito, Sandra Pinheiro, Tatiane, Andreia, Sandra Rêgo, Lenaldo, Cherubino, Fernando, Sílvia, Gecy, André, Júnior e Cristina Moraes.

Aos professores do Programa de Mestrado em Estatística, pela competência e atenção.

A Valéria Bittencourt, pela competência, organização, atenção e amizade.

Aos funcionários do departamento de estatística, Adriana, Cícero e Antônio, pela atenção.

Aos colegas de graduação, Lindomberg, Michelli, Horácio, Diana e Juliana.

Aos professores do Departamento de Matemática e Estatística da Universidade Federal do Estado da Paraíba, em especial aos professores Francisco Moraes, Luiz Mendes e Vandik Estevam.

A Michelli, Horácio, Diana, Lenaldo, Renata, Avelita e Kátia, pela convivência pacífica e harmoniosa.

A CAPES, pelo apoio financeiro.

Resumo

Análise de agrupamento é um termo genérico que abrange vários métodos numéricos para estudar dados multivariados com o intuito de classificar as observações em grupos homogêneos.

Neste trabalho se apresenta uma síntese dos métodos de agrupamento existentes e se realiza uma descrição detalhada de um método particular de agrupamento hierárquico aglomerativo baseado em modelos que considera um critério bayesiano para a determinação do melhor modelo e dos possíveis grupos existentes.

Duas aplicações desta metodologia foram realizadas. A primeira utilizando o banco de dados do Instituto de Pesquisa Econômica e Aplicada que contém informações sobre finanças públicas. A segunda com o banco de dados do Sistema Único de Saúde relacionadas a área de saúde. Na realização das duas foram utilizados os municípios de Pernambuco visando uma classificação dos mesmos verificando os grupos mais homogêneos, os municípios com valores extremos para cada variável. É apresentada também uma discussão dos resultados obtidos.

Abstract

Cluster analysis are a generic term for a variety of numerical methods to study multivariate data with the purpose of classifying in specific groups homogeneous observations. In this work we present the synthesis of the existing methodologies. A particular method is detailed: model-based clustering for the determination of the best model and the possible existing groups is utilized an Bayesian criterion. Two different applications of this methodology are carried with data of the Instituto de Pesquisa Econômica Aplicada and of Sistema Único de Saúde. The first database contains information on public finances and health of the boroughs of Pernambuco.

Índice

| | | |
|-------|--|----|
| 1 | Introdução | 1 |
| 2 | Conceitos Básicos de Análise de Agrupamento..... | 3 |
| 2.1 | Medidas de Similaridade e Dissimilaridade para Variáveis Quantitativas | 5 |
| 2.1.1 | Para Variáveis de Escala Intervalar | 6 |
| 2.1.2 | Para Variáveis de Razão Intervalar | 7 |
| 2.2 | Medidas de Similaridade e Dissimilaridade para Variáveis Qualitativas | 7 |
| | Para Variáveis Binárias | 7 |
| | Medida de Coincidência Simples | 8 |
| | Medida de Concordância Positiva | 8 |
| 2.3 | Métodos de Agrupamento | 9 |
| 2.3.1 | Métodos de Partição | 9 |
| | Método das K-Médias | 9 |
| 2.3.2 | Métodos Hierárquicos | 10 |
| | Método do Centróide | 10 |
| | Método da Ligação Simples ou do Vizinho Mais Próximo | 10 |
| | Método da Ligação Completa ou do Vizinho Mais Distante | 10 |
| | Método de Ward | 11 |
| 3 | Método de Agrupamento Baseado no Modelo | 12 |
| 3.1 | Densidade de Mistura Finita | 12 |
| 3.2 | O Algoritmo EM para Modelos de Mistura..... | 13 |
| 3.3 | Seleção do Modelo | 15 |
| 3.4 | Agrupamentos Hierárquicos Baseados nos Modelos | 17 |
| 3.5 | Combinando Aglomeração Hierárquica, EM, e Fator de Bayes | 18 |
| 3.6 | Modelando as Perturbações | 18 |
| 3.7 | Vantagens e Limitações | 19 |
| 3.8 | Dados Não Gaussianos | 20 |
| 4 | Aplicação | 21 |
| 4.1 | Resultado do Agrupamento com os Dados do IPEA | 23 |

| | |
|--|----|
| 4.2 Resultado do agrupamento com os dados do SUS | 37 |
| 4.2 Conclusões Gerais dos Agrupamentos | 38 |
| ◇ Apêndice A | 45 |
| ◇ Apêndice B | 49 |
| ◇ Apêndice C | 53 |
| ◇ Referências | 57 |