

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

FACULDADE DE ARQUITETURA

PROGRAMA DE PÓS-GRADUAÇÃO EM PLANEJAMENTO URBANO E REGIONAL

DISSERTAÇÃO DE MESTRADO

**MODELOS DE REGIONALIZAÇÃO ADMINISTRATIVA
ATRAVÉS DE UM SISTEMA ESPACIAL DE SUPORTE À DECISÃO:**

Estudo de Caso para o Estado do Rio Grande do Sul

ALEXANDER GURGEL MARQUES
Mestrando

PORTO ALEGRE, 2000.

RESUMO

MODELOS DE REGIONALIZAÇÃO ADMINISTRATIVA ATRAVÉS DE UM SISTEMA ESPACIAL DE SUPORTE À DECISÃO: Estudo de Caso para o Estado do Rio Grande do Sul

ALEXANDER GURGEL MARQUES

Julho de 2000

A distribuição de recursos públicos de modo equilibrado e bem aplicado é questão de suma importância para administradores públicos e planejadores, especialmente em países como o Brasil que, a cada ano, vê sua capacidade de investimento mais reduzida e onde mais se acentuam os desníveis sociais.

A metodologia, aqui empregada, busca incorporar ao modelo a característica de dinamismo própria da realidade regional e por isso permite grande abertura na fase de seleção de variáveis, tratamento dos dados e cálculos de correlações. Descarta de saída a possibilidade de ser resposta única para a questão da regionalização, e sobretudo, visa ser um modelo heurístico que se realimenta via interações, permitindo inúmeras soluções, tanto melhores, quanto mais forem as tentativas de otimização do método.

Este trabalho trata da questão da regionalização sob a ótica do estabelecimento de similaridades através de uma análise centrada nos dados sócio-econômicos que guardam melhor correlação com a estrutura espacial, utilizando a técnica de análise de agrupamentos e estatística multivariada, com o fim de facilitar o planejamento regional e munir os administradores públicos com instrumentos de decisão para uma distribuição melhor dimensionada de recursos.

O tratamento dos dados se desenvolve a partir de matrizes que relacionam cada objeto unitário com todos os demais, neste caso, cada unidade municipal do estado do Rio Grande do Sul com todos os demais municípios. Utiliza-se o cálculo de variadas formas de distâncias euclidianas e não euclidianas para estabelecer a similaridade entre os objetos, o que é medido através do Teste de Mantel que relaciona as matrizes de dados com a matriz de distância. Posteriormente, via uso de técnicas de análise de agrupamento obtém-se os padrões regionais atrelados à estrutura espacial.

As regionalizações geradas pelo método têm a vantagem de apresentarem-se em vários níveis hierárquicos, direcionando-se para uma maior diferenciação à medida que os níveis se aprofundam. Permitem uma visualização dos resultados em qualquer um dos níveis de diferenciação, o que proporciona ampla gama de possibilidades comparativas. Possuem um bom grau de isenção técnica, porque são o resultado de uma análise cujos principais referenciais são de domínio coletivo e definidores do espaço, em que pese o índice de correlação entre as matrizes de distâncias ter sido baixo, para esta partida de dados selecionada e aplicada ao estado do Rio Grande do Sul.

ABSTRACT

MODELS OF ADMINISTRATIVE REGIONALIZATION THROUGH A SPATIAL DECISION SUPPORT SYSTEM

Study Case to the Rio Grande do Sul State

ALEXANDER GURGEL MARQUES

July 2000

The equilibrated and well implemented distribution of public resources is a question of the greatest importance to public administrators and planners, specially in countries like Brazil which, year after year, sees its investment capacity more weakened, and where the social imbalances are much more accentuated.

The methodology here employed tries to incorporate to the model the characteristic of dynamism which is part of the regional reality, and which for this reason permits a great broadness in the phase of variables selection, data manipulation and calculation of correlation. It is discarded from the beginning the possibility of having a single answer to the question of regionalization, and moreover aims to be a heuristic model which feeds back through the interactions, allowing several solutions, which improve according to the number of optimization attempts of the method.

This study faces the question of regionalization under the perspective of establishing similarities by means of an analysis focussed on the socio-economic data which keep better correlation with the spatial structure, utilizing the technique of cluster analysis and multi-variate statistics, with the purpose of facilitating regional planning and of equipping public administrators with instruments of decision making oriented to a well measured distribution of resources.

The treatment of the data develops itself from the matrix that relate each unitarian object with all the others, which in this case was each municipal unit of Rio Grande do Sul

State with all the other municipalities. The calculation was used in several forms of Euclidean and non Euclidean distances in order to establish the similarity between the objects, which is tested through Mantel's Test that relates the data matrix with the distance matrix. After that, through the technique of cluster analysis we arrive to the regional patterns linked to the spatial structure.

The regionalizations generated by the method have the advantage of presenting themselves in various hierarchical levels, driving themselves towards a greater differentiation as the levels deepen. They permit a visualization of the results in each and everyone of the levels of differentiation, which offers a great scope of comparative possibilities. They have a good degree of technical exemption, because they are the result of an analysis which the main referential are of collective domain and spatial definers, in spite of the fact that the correlation index between the distance matrix have been low to this selected data sample to Rio Grande do Sul State.

*Ao Max, meu filho,
por sua alegria contagiante,
capaz de libertar e criar.*

*Aos meus pais,
a quem eu escolheria
se pudesse.*

AGRADECIMENTO

O que vale a pena ser feito não é necessariamente fácil ou desimpedido. Do esforço despendido deriva o valor do que é realizado. Uma dissertação não foge a estas circunstâncias e, por esta razão, o prazer de vencer etapas, superando obstáculo por obstáculo, é único e compensa fartamente os desafios assumidos.

No entanto, felizmente o isolamento do trabalho criativo não é total, o percurso envolve terceiros. Assim, aos que facilitaram ou possibilitaram em alguma medida a execução deste trabalho ou que, velada ou explicitamente, torceram por mim, dedico o meu agradecimento mais profundo.

Ao meu orientador, Carlos André Bulhões Mendes, pela inspiração e orientação pertinente, aos professores, funcionários, amigos e familiares pelos apoios de toda ordem, meus sinceros agradecimentos. E para alguns terei prazer em fazer uma menção especial:

Arlete Erbert, pela capacidade de trabalho de forma coletiva
Arnaldo Sisson Filho, pelo pronto apoio de última hora
Décio Rigatti, pelo apoio e tolerância aos prazos
Juçara Nunes e Silva, pela amizade suave e benevolente
Júlio César Berleze, pelo seu caráter diferenciado e cumplicidade propiciadora
Loribel Dias da Silva, por tantas fichas de empréstimo bibliotecário
Maria Aparecida Ramalho Forni, por descobrir formas de facilitação
Maria de Lourdes T. Jardim, pelo suprimento intelectual
Maria Izabel Mallmann, por sua supervisão amorosa e clara
Maria Luiza M. M. Rodrigues, por assumir a atenção de quem amamos
Marilene Dias Bandeira, pela atenção e capacidade de compreensão temática
Marines Zandavali Grando, pelo despertar de uma idéia
Naia Geila I. de Oliveira, pelo incentivo das conversas iniciais
Neiva Maria Pastorino, pela simpatia e capacidade executiva
Romulo Krafta, pela criação de um foco norteador
Rosane Margaréte Tormes Ballejos, pela presteza e cordialidade
Rosetta Mammarella, pela afinidade temática
Salette do Prado Oliveira, por sua proteção administrativa
Wrana Maria Panizzi, pelo marco que representa.

SUMÁRIO

Capa	i
Resumo	ii
Abstract	iv
Dedicatória	vi
Agradecimento	vii
Sumário	viii
Lista de Figuras	xii
Lista de Tabelas	xiv
Lista de Quadros	xv
Lista de Siglas	xv
Apresentação	xvi
CAPÍTULO 1 - INTRODUÇÃO	2
1.1 DEFINIÇÃO DE REGIÃO	3
1.2 RELEVÂNCIA	6
1.3 OBJETIVOS	7
1.4 ESTRUTURA DO TRABALHO	8
CAPÍTULO 2 – REVISÃO BIBLIOGRÁFICA	10
2.1 ASPECTOS CONCEITUAIS DA REGIONALIZAÇÃO	11
2.1.1 <i>Divisões regionais e critérios de regionalização</i>	11
2.1.2 <i>Divisões regionais para fins estatísticos</i>	15
2.1.3 <i>Divisões regionais administrativas</i>	15
2.1.3.1 <i>Conselhos Regionais de Desenvolvimento</i>	16
2.1.3.2 <i>Regiões do Orçamento Participativo</i>	17

2.2	ASPECTOS ESTRUTURAIS	20
2.2.1	<i>Índice de Desenvolvimento Humano</i>	20
2.2.2	<i>Índice de Carência Social e Índice de Desenvolvimento Social ..</i>	23
2.2.3	<i>Índice de Desenvolvimento Urbano</i>	24
2.3	ASPECTOS ANALÍTICOS	25
2.3.1	<i>Estatística Multivariada</i>	25
2.3.2	<i>Análise Espacial</i>	26
2.3.2.1	<i>Medidas de Análise Espacial</i>	33
2.3.2.2	<i>GIS e Autocorrelação espacial</i>	33
2.3.2.3	<i>Autocorrelação espacial, integração espacial e medida de segregação: onde a Sociologia encontra a Análise estatística espacial e a Geografia</i>	35
2.3.2.4	<i>Taxonomia Regional: Método de Análise de Agrupamento</i>	39
2.3.2.5	<i>Análise Regional</i>	40
2.3.2.6	<i>Modelos de Localização-Alocação</i>	42
2.4	ESTRUTURA DE DADOS CENSITÁRIOS	44
2.5	GEOPROCESSAMENTO APLICADO À ANÁLISE ESPACIAL	45
2.5.1	<i>Noções Gerais sobre Geoprocessamento</i>	45
2.5.2	<i>O Processo de Modelagem</i>	46
2.5.3	<i>O Universo do Mundo Real</i>	46
2.5.3.1	<i>Mapas Temáticos</i>	46
2.5.3.2	<i>Mapas Cadastrais</i>	47
2.5.3.3	<i>Modelos Numéricos de Terreno</i>	48
2.5.3.4	<i>Redes e Imagens</i>	48
2.5.4	<i>O Universo Conceitual</i>	48
2.5.5	<i>O Universo de Representação</i>	49
2.5.6	<i>Sistemas de Informação Geográfica</i>	51
CAPÍTULO 3 – METODOLOGIA		55
3.1	TRATAMENTO DOS DADOS	57
3.1.1	<i>Pré Processamento dos Dados</i>	57
3.1.2	<i>Análise Exploratória</i>	58

3.2	ANÁLISE DA SIMILARIDADE: ESPACIAL E DE ATRIBUTOS	67
3.2.1	<i>Coordenadas das Sedes Municipais e Distâncias Euclidianas..</i>	67
3.2.2	<i>Cálculo de Outros tipos de Distâncias entre os Dados</i>	68
3.2.2.1	<i>Distância Estatística</i>	68
3.2.2.2	<i>Outros Tipos de Distâncias entre os Dados</i>	70
3.2.2.3	<i>Monte Carlo</i>	75
3.3	TESTE DE MANTEL	75
3.4	ANÁLISE DE AGRUPAMENTOS	79
3.4.1	<i>Estatística Multivariada</i>	79
3.4.2	<i>Dendograma</i>	79
3.4.3	<i>Cluster – Agrupamentos Hierárquicos</i>	80
3.4.4	<i>Interpretação do Método de Agrupamento Hierárquico</i>	88
3.5	ANÁLISE COMPARATIVA COM OUTRAS DIVISÕES REGIONAIS	90
CAPÍTULO 4 - ESTUDO DE CASO: ESTADO DO RIO GRANDE DO SUL		92
4.1	FLUXOGRAMA DE APLICAÇÃO AO ESTUDO DE CASO	92
4.2	CONFIGURAÇÃO DA BASE DE APLICAÇÃO	95
4.3	MATRIZ DE “DISTÂNCIAS EUCLIDIANAS”	97
4.4	DETERMINAÇÃO DAS MATRIZES DE DISTÂNCIA DOS DADOS COM OITO COEFICIENTES DE SIMILARIDADE	98
4.5	TESTE DE MANTEL PARA OITO DISTÂNCIAS	99
4.6	ANÁLISE DE AGRUPAMENTO PARA A MATRIZ COM MAIS SIMILARIDADE	100
4.7	VISUALIZAÇÃO DOS RESULTADOS: MAPAS REGIONAIS	103
4.7.1	<i>Análise do Nível Hierárquico 2</i>	104
4.7.2	<i>Análise do Nível Hierárquico 3</i>	105
4.7.3	<i>Análise do Nível Hierárquico 4</i>	106
4.7.4	<i>Análise do Nível Hierárquico 5,6 e 7.....</i>	107
4.7.5	<i>Análise do Nível Hierárquico 8</i>	108
4.7.6	<i>Análise do Nível Hierárquico 9</i>	109
4.7.7	<i>Análise do Nível Hierárquico 10, 11, 12 e 13</i>	110
4.7.8	<i>Análise do Nível Hierárquico 14</i>	111

4.7.9	<i>Análise do Nível Hierárquico 15</i>	112
4.8	Análise Comparativa com as Divisões Regionais Existentes	113
CAPÍTULO 5 - CONCLUSÕES		118
5.1	ANÁLISE FINAL	118
5.1.1	<i>Dos Objetivos</i>	118
5.1.2	<i>Da Questão Espacial</i>	119
5.1.3	<i>Do Enfoque Metodológico</i>	119
5.1.4	<i>Dos Resultados Apresentados</i>	120
5.2	ASPECTOS PROBLEMÁTICOS	121
5.3	ASPECTOS DE RELEVÂNCIA	121
5.4	ABORDAGENS ESPECÍFICAS	122
5.5	INDICAÇÕES DE DESAFIOS POSSÍVEIS	122
5.6	FECHAMENTO	124
REFERÊNCIAS BIBLIOGRÁFICAS		126
ANEXOS		131
	ANEXO I – TabWin	132
	ANEXO II - ADE-4 Win	133
	ANEXO III - Fonte de Dados	136
	• CD-ROM BIM (IBGE)	136
	• CD-ROM FEE	136
	• DATASUS <i>Home Page</i>	136

LISTA DE FIGURAS

2.1	Mapa dos Conselhos Regionais de Desenvolvimento e do Orçamento Participativo.	18
2.2	Mapa das Microrregiões do Rio Grande do Sul – 1994	19
2.3	Mapa das Regionais de saúde para o estado Rio Grande do Sul – 1994.....	19
2.4	Mapa das Mesorregiões do Rio Grande do Sul – 1994	20
2.5	Comparativo das duas metodologias do IDH-renda até 1998 e de 1999.....	22
2.6	Índices de segregação da pobreza em Nova York	36
2.7	Índices de segregação da população negra em Nova York	37
2.8	Índices de segregação de habitantes sem-teto em Nova York	37
2.9	Exemplo de mapa temático: classificação de solos da Planície Dirol (Mauritânia) ...	47
2.10	Exemplo de mapa cadastral extraído do Sistema de Informações Estatísticas e Geográficas (SIEG) do IBGE (versão beta II de demonstração).....	47
2.11	Representação de dados nos formatos vetorial e matricial ou raster	50
3.1	Quadro metodológico	56
3.2	Representação esquemática de um objeto espacial.....	57
3.3	Representação dos polígonos municipais e seu código identificador.....	57
3.4	Origem e seleção de variáveis (22).....	64
3.5	Primeira fase da metodologia, pré processamento dos dados.....	66
3.6	Segunda fase da metodologia, análise da similaridade espacial	68
3.7	Terceira fase da metodologia, análise da similaridade dos atributos.....	75
3.8	Quarta fase da metodologia, análise da similaridade dos atributos.....	78
3.9	Dendograma de conexão única para cinco municípios gaúchos.....	82
3.10	Dendograma de conexão completa para cinco municípios gaúchos.....	83
3.11	Dendograma de conexão média para cinco municípios gaúchos.....	84
3.12	Dendograma do método de Ward para cinco municípios gaúchos.....	87
3.13	Quinta fase metodológica, agrupamento de regiões similares.....	89
3.14	Sexta fase metodológica, análise comparativa com outras regionalizações.....	90
4.1	Fluxograma esquemático de aplicação ao estudo de caso.....	94
4.1.1	Pré processamento.....	93
4.1.2	Similaridade espacial	93

4.1.3	Análise da similaridade das características medidas (atributos).....	93
4.1.4	Análise das distâncias de agrupamento com correlação espacial – Mantel.	93
4.1.5	Agrupamento de regiões similares mantendo a estrutura espacial	94
4.1.6	Análise comparativa com as demais formas de regionalização.....	94
4.2	Mapa político do Rio Grande do Sul – malha de 1994 (427 municípios).....	96
4.3	Mapa político do Rio Grande do Sul – malha de 1997 (467 municípios).....	96
4.4	Dendograma da Análise de agrupamento: método aglomerativo	102
4.5	Dendograma da Análise de agrupamento: método divisivo	102
4.6	Mapa do Rio Grande do Sul – Nível 1 – (427 + 40 municípios novos) – 1997	103
4.7	Regionalizações do RS: Nível hierárquico 2	104
4.8	Regionalizações do RS: Nível hierárquico 3 Método dos agrupamentos divisivos – Escala 1:250.000	105
4.9	Regionalizações do RS: Nível hierárquico 4 Método dos agrupamentos divisivos – Escala 1:250.000	106
4.10	Regionalizações do RS: Nível hierárquico 5, 6 e 7 Método dos agrupamentos divisivos – Escala 1:250.000	107
4.11	Regionalizações do RS: Nível hierárquico 8 Método dos agrupamentos divisivos – Escala 1:250.000	108
4.12	Regionalizações do RS: Nível hierárquico 9 Método dos agrupamentos divisivos – Escala 1:250.000	109
4.13	Regionalizações do RS: Nível hierárquico 10, 11, 12 e 13 Método dos agrupamentos divisivos – Escala 1:250.000	110
4.14	Regionalizações do RS: Nível hierárquico 14 Método dos agrupamentos divisivos – Escala 1:250.000	111
4.15	Regionalizações do RS: Nível hierárquico 15 Método dos agrupamentos divisivos – Escala 1:250.000	112
4.16	Conselhos Regionais de Desenvolvimento do RS (COREDES) Escala 1:250.000.	114
4.17	Microrregiões Homogêneas do RS (IBGE) Escala 1:250.000.....	115
4.18	Mesorregiões do RS (IBGE)	115
4.19	Regionais de Saúde do RS (IBGE) Escala 1:250.000	116

LISTA DE TABELAS

2.1 Tabela de Mínimos e Máximos para os indicadores do IDH	21
2.2 Distribuição dos registros bons e maus, para Brasil e Região Sul, segundo as aplicações	45
3.1 Extrato da matriz de indicadores para análise espacial (10 municípios).....	58
3.2 Matriz Normalizada (valores entre 0 e 1)	66
3.3 Matriz de correlação (valores entre 0 e 1)	66
3.4 Matriz de distância euclidiana	67
3.5 Latitude e longitude das sedes dos primeiros dez (10) municípios gaúchos (ordem alfabética)	69
3.6 Cálculo de distância “euclidiana” entre os dez primeiros municípios gaúchos	70
3.7 Cálculo de distância estatística para duas variáveis entre os dez (10) primeiros municípios gaúchos (Taxa de Migrantes X Capacidade de atendimento hospitalar) - 1996	70
3.8 Tabela de distâncias para cinco municípios gaúchos (1ª parte da tabela 3.6).....	81
4.1 Matriz de 427 X 2 de latitude e longitude para os municípios do RS.....	97
4.2 Matriz de distâncias normalizadas para os primeiros vinte municípios RS.....	98
4.3 Resultados do teste de Mantel para 8 distâncias.....	100
4.4 Descrição de 14 categorias para 15 níveis de agrupamento (gerada a partir do ADE-4).....	101

LISTA DE QUADROS

3.1 Testes de agrupamentos sob a forma de variações do produto gama (Γ)	77
--	----

LISTA DE SIGLAS

ADE-4 – *Analyses des Donées Écologiques*
AEC – Área estatisticamente comparável
BIM – Base de Informações Municipais
COREDES – Conselhos Regionais de Desenvolvimento
DATASUS – Sistema de Dados do Ministério da Saúde
DE – Delegacia de Educação
DLSEP – *Generalised Least Square Estimation Procedure*
DRS – Delegacia Regional de Saúde
DTA – Detecção e Imputação Automática
FEE – Fundação de Economia e Estatística Siegfried Emanuel Heuser
IBGE – Instituto Brasileiro de Geografia e Estatística
IDH – Índice de Desenvolvimento Humano
IDM – Índice de Desenvolvimento Municipal
IDR – Índice de Desenvolvimento Regional
IDS – Índice de Desenvolvimento Social
MNT – Modelo numérico do terreno
N.Y. – Nova York
ONU – Organização das Nações Unidas
PEA – População economicamente Ativa
PED – Pesquisa Emprego Desemprego
PEO – População economicamente Ocupada
PIA – População em Idade Ativa
PIB – Produto Interno Bruto
PNUD – Programa das Nações Unidas para o Desenvolvimento
POSS – População ocupada no setor secundário
POST – População ocupada no setor terciário
PPC – *Power parity currency* (Paridade do poder de compra em dólares)
RMPA – Região Metropolitana de Porto Alegre
SAA/RS – Secretaria da Agricultura e Abastecimento do Rio Grande do Sul
SCP/RS – Secretaria da Coordenação e Planejamento do Rio Grande do Sul
SDO/RS – Secretaria do Desenvolvimento Regional e Obras Públicas do Rio Grande do Sul
SGBD – Sistema Gerenciador de Banco de Dados
SIEG – Sistema de Informações Estatísticas e Geográficas
SIG – Sistema de Informação Geográfica
TabWin – Sistema de Tabulação para o Windows do DATASUS
U.S.A. – *United States of America*
UTM – *Universal Transverse Mercator Coordinate System*

APRESENTAÇÃO

O trabalho está constituído por cinco capítulos que objetivam conduzir o tema desde de sua definição e relevância, passando por uma apresentação de conteúdos afins, chegando à criação de um método de desenvolvimento do modelo, aplicado ao estudo de caso e finalizando com uma análise de resultados e indicações de novas propostas.

O Capítulo 1, Introdução, enfatiza a importância da questão colocada, define o tema principal e estabelece os objetivos a serem atingidos.

O Capítulo 2, Revisão Bibliográfica, enfoca a questão da regionalização do ponto de vista teórico enfatizando, contudo, o seu caráter prático. Percorre vários autores que desenvolveram temáticas semelhantes e apresenta um extenso embasamento estatístico.

O Capítulo 3, Metodologia, resume todo o processo metodológico em um único quadro composto por seis fases nas quais se definem a apreensão dos dados, a execução de análises espaciais e de atributos sob o prisma da similaridade utilizando matrizes de distâncias e a técnica de análise de agrupamentos, terminando por uma análise comparativa dos resultados.

O Capítulo 4, Estudo de Caso: Estado do Rio Grande do Sul, faz a aplicação do método desenvolvido aos dados sócio-econômicos do estado gerando quinze níveis hierárquicos de regionalização.

O Capítulo 5, Conclusões, realiza uma análise final dos objetivos, da questão espacial, do enfoque metodológico e dos resultados apresentados. Levanta os principais problemas encontrados, sugere novos caminhos de pesquisa e enfatiza a validade do modelo.



CAPÍTULO 1

INTRODUÇÃO

INTRODUÇÃO

O sucesso da aplicação de políticas regionais depende do conhecimento dos potenciais e das limitações das diversas áreas que integram uma região. A negligência em relação a esse aspecto acarreta ou aprofunda desigualdades que impedem o desenvolvimento integral e harmônico da região e oneram os cofres públicos. A metade sul do Rio Grande do Sul, por exemplo, encontra-se numa situação de estagnação econômica que, na opinião de muitos, a coloca para o Rio Grande do Sul como o nordeste está para o Brasil. Embora seja bem variado o leque de explicações evocadas para este isolamento socioeconômico, a superação desse quadro torna imprescindível aos responsáveis pelas decisões desenvolverem instrumentos que mensurem o potencial e as limitações dessa região para a aplicação de políticas adequadas que revertam essa situação e melhor distribuam os recursos existentes.

A função de transparência do gestor público, especialmente no que tange à aplicação dos escassos recursos públicos, em países como o Brasil, assume importância suprema para uma administração responsável e de gerência eficiente, além de garantir vida mais longa para propostas de administração alicerçadas em participação popular e compartilhamento de decisões. Se associado a este *modus faciendi* estiverem presentes técnicas de avaliação de custos e benefícios, e métodos capazes de medir as deficiências ou mesmo as excelências de áreas bem definidas e caracterizadas, grande será a oportunidade de ser realizado um trabalho de justiça social através de uma distribuição bem dosada de recursos, fundamentada em dados socioeconômicos e em aspectos espaciais da realidade enfocada. Ainda que, a alternativa de alocação de recursos como resposta à pressão de grupos de atividade política ou mesmo pela habilidade de atuação da cidadania em defender interesses de vizinhança sejam, também, de todo válidas (Talen, 1998), não substituem mas complementam os instrumentais técnicos de medida.

Neste trabalho, pretende-se desenvolver uma metodologia que dê suporte técnico à tomada de decisões de governos estaduais no que concerne ao planejamento regional, utilizando o estado do Rio Grande do Sul como exemplo para validação das hipóteses aqui apresentadas.

Propõe-se criar um modelo heurístico¹ que, a partir da Análise estatística multivariada e da Autocorrelação espacial de variáveis (Ding, 1992 e Shen, 1994), permita conceber regiões de modo a ser possível uma automática reconfiguração regional em função da atualização dos dados, seja pela utilização de fontes diferenciadas de dados, ou mesmo, pela simples eleição de variáveis tidas como pertinentes.

1.1 DEFINIÇÃO DE REGIÃO

Faz-se necessário inicialmente definir o que se entende por região, no contexto do presente estudo. Não é objetivo do trabalho conceituar região, nem entrar no âmbito das inúmeras discussões teórico-práticas desenvolvidas através de décadas sobre o tema (Mesquita, 1984)².

Observe-se que a questão de dividir o espaço para melhor estudá-lo é ponto de fundamento para toda e qualquer aproximação da realidade regional. Tem sido a preocupação de vários pesquisadores desta área, como consta, por exemplo, na parte IV, “Considerações Metodológicas sobre a Divisão Regional” do trabalho de Alonso, Benetti e Bandeira (1994), intitulado “Crescimento Econômico da Região Sul do Rio Grande do Sul – Causas e perspectivas”:

“A escolha de uma partição do espaço que proporcione uma base adequada para a organização de informações estatísticas é uma tarefa decisiva quando se pretende estudar problemas regionais, sejam eles de natureza econômica, social ou política.” (Alonso, 1994, p.215).

Contudo, como salienta o autor, a utilização de “divisões regionais já disponíveis”, como é frequentemente realizado, pode revelar-se imprópria. Isto porque a definição obtida em uma dada época não será a mesma ao longo do tempo. Pois as realidades regionais sofrem mudanças contínuas, seja pela implementação de planos do governo, ou pela

¹ “Conjunto de regras e métodos que conduzem à descoberta, à invenção e à resolução de problemas.” (Ferreira, 1986).

² Para um maior aprofundamento na questão conceitual de região consulte: Palacios (1983), Corrêa (1987), Duarte (1988), entre outros.

inexistência de planos, ou pela ação de agentes privados em sua própria exploração de demandas detectadas, na busca por mercados emergentes. Há, paralelamente, a ascensão e declínio de regiões relacionados aos ciclos econômicos ligados à exploração de recursos e produtos naturais que se esgotam, mudando substancialmente a configuração do espaço bem como seu uso e importância. No entanto, essa dimensão propriamente dinâmica não tem sido corretamente incorporada como critério de definição das divisões regionais. Exemplo disso é a divisão regional proposta pela Secretaria do Desenvolvimento Regional e Obras Públicas (SDO), publicado em 1974 e que utiliza critérios de centralidade e de polarização do espaço no Rio Grande do Sul (SDO/RS, 1974). Esta mesma divisão foi adaptada e adotada por Bandeira e Gründling em seu estudo sobre o crescimento industrial do estado na década de 70 (Bandeira, 1988). Já mais recentemente, há o trabalho de zoneamento do estado em regiões agroecológicas e ecológicas, respondendo ao anseio do setor primário do estado e estabelecendo uma política agrícola integrada à vocação ecológica regional (SAA/RS, 1994). Estes trabalhos ainda que de abrangência restrita, se justificam pela sua utilidade episódica e, alguns, por seu rigor metodológico como por exemplo, o trabalho que aborda o crescimento econômico da Região Sul do estado, já citado anteriormente (Alonso, op.cit.) e que também gera uma proposta de divisão regional.

Entretanto, o que aqui se pretende, como foi dito acima, é criar um modelo de regionalização administrativa que incorpore variáveis correlacionadas às transformações do meio, de modo que se possa obter, através da periódica realimentação dos dados, divisões regionais atualizadas e consistentes. Esta abordagem consiste na criação de um sistema que combine informações estatísticas sobrepostas às informações geográficas, através do uso de modelos de interação.

Para tratar do processo de modelagem regional foram selecionados quatro aspectos componentes da definição de região que, se pretende, seja própria e suficiente (Palacios, 1983).

Tais aspectos são amplamente conhecidos em sua maioria. O primeiro diz respeito a alguma característica dominante que imprima singularidade a uma dada parcela do território, designada por homogênea, seja social, física, climática ou política (Palacios, op.cit.). O segundo refere-se à questão da relação entre centro e periferia, clássica abordagem do esquema hierárquico na organização do território desenvolvida por Christaller (1933) e Lösch (1954) na Teoria do Lugar Central (Christaller, 1966). Como exemplos pode-se tomar os fluxos migratórios diários de trabalhadores de uma área para

outra, ou o volume de serviços especializados oferecidos exclusivamente por um centro regional para as áreas adjacentes. O terceiro engloba as relações que se estabelecem a partir das atividades humanas, o que confere o dinamismo ou cristalização de uma dada parcela territorial, passíveis de quantificação e qualificação através de uma análise regional. Exemplos seriam: volume de obras de infra-estrutura, capacidade de produção de energia, fluxo de capitais etc.

Por último, resta o aspecto responsável pelo caráter dinâmico da definição de região com que se trabalha. Esse aspecto traz possivelmente um acréscimo qualitativo na definição de região, a idéia de fluência, movimento e processo em curso, característica das relações dinâmicas e transformadoras por que passa qualquer parcela do território permanentemente alterada pelos agentes.

Tanto a delimitação espacial como as relações entre os municípios e a sede municipal, as próprias atividades humanas e alocação de recursos naturais, são de natureza dinâmica, significando, com isto, que o modelo deve ser sensível às variações contínuas do espaço regional em sua totalidade. Portanto, o modelo de regionalização deve ser capaz de detectar todas as transformações, sejam desmembramentos de municípios, alterações no equilíbrio de forças econômicas e de produção, bem como, escassez ou oferta de recursos naturais, fatores de atração do capital e do trabalho, ou até mesmo, mudanças e implementações da legislação de proteção do meio ambiente ou incentivos e isenções fiscais.

Com relação especificamente à questão da mútua influência centro-periferia relatada na literatura técnica como “*spread and backwash effects*” (Barkley, 1996) e Hughes (1994), cabe salientar uma idéia recorrente destes autores, que dizem ser absolutamente necessário um profundo entendimento dos elos entre as economias urbana e rural como modo dos planejadores detectarem os problemas inter-relacionados, por exemplo, o declínio das oportunidades econômicas em algumas áreas rurais e a diminuição da qualidade de vida em áreas urbanas com alta taxa de crescimento econômico.

Assim, o modelo a ser desenvolvido será calcado sobre uma definição de região que, por todos os aspectos considerados anteriormente, apresentará um enquadramento orgânico, no sentido de ser uma realidade intimamente associada aos processos socioeconômicos e políticos em andamento, caso se admita que a realidade regional possa ser expressa por indicadores de tais processos.

Por estas razões, **região** é aqui definida como uma porção relativamente similar do território, mas em constante transformação, constituída por um número determinado de municípios (divisão política), ocupada por uma parcela da população que realiza atividades sociais, econômicas e políticas inter-relacionadas vinculadas a um dado centro urbano proeminente e com autonomia dentro da sua área de abrangência.

1.2 RELEVÂNCIA

A alocação dos recursos aplicados em políticas públicas é normalmente pautada por critérios políticos, sobretudo em países com tradição política autárquica, como é o caso do Brasil. O estado do Rio Grande do Sul, como unidade da Federação, não foge à regra. Os desdobramentos negativos dessa orientação, ou em alguns casos extremos, desorientação, variam desde obras frustradas até demandas não atendidas. A aplicação politicamente orientada dos recursos públicos ocorre seja pela própria natureza do processo que favorece práticas clientelistas, seja pela inexistência de estudos que subsidiem tecnicamente a tomada de decisões. Em países como Estados Unidos (Ballard, 1980), Grã-Bretanha (Martin, 1994) e França, há décadas são utilizados recursos técnicos para otimizar a aplicação de recursos de modo a balancear o peso do fator político na tomada de decisões.

Quanto ao Brasil, o contexto atual desfavorece a utilização dos recursos públicos com critérios exclusivamente políticos ou aleatórios. A acelerada diversificação da nossa economia recomenda a racionalização das decisões públicas de modo a assegurar sua maior eficiência. Além disso, o aprofundamento do processo democrático expõe o poder público a demandas crescentes, cujo atendimento satisfatório depende, na maioria das vezes, da aplicação racional dos recursos. Porque uma distribuição justa e própria dos recursos associada a uma eficiência político-administrativa e tecnológica poderá ser um objetivo absolutamente necessário no futuro próximo, o que parece ser plenamente respaldado pelas palavras de David Harvey quando refere que *“... no longo prazo será muito benéfico se eficiência e distribuição forem exploradas conjuntamente, visto ser contraproducente no longo prazo criar uma distribuição socialmente justa se o tamanho do produto a ser distribuído encolher acentuadamente por causa do uso ineficiente dos*

recursos escassos. No longuíssimo prazo, contudo, justiça social e eficiência são muito a mesma coisa“ (Harvey, 1973).

Portanto, a escassez de estudos que subsidiem tecnicamente a tomada de decisões justifica o presente esforço na medida em que se pretende fornecer aporte técnico que agilize a obtenção de informações atualizadas sobre as diferentes regiões administrativas e contribua para a aplicação mais racional e eficiente dos recursos públicos.

1.3 OBJETIVOS

1.3.1 OBJETIVO GERAL

- Desenvolver um modelo capaz de detectar com clareza a similaridade de áreas regionais, baseada em características socioeconômicas municipais, a fim de servir de suporte para os agentes públicos na distribuição dos recursos públicos através da análise de dados espaciais e sua visualização espacial.

1.3.2 OBJETIVOS ESPECÍFICOS

- Escolher as variáveis a serem utilizadas que representem o processo de desenvolvimento, montando uma matriz de dados.
- Classificar, a partir dos dados selecionados, áreas relativamente homogêneas, significa dizer, áreas com características de similaridade, utilizando técnicas estatísticas multivariadas (análise de agrupamentos).
- Determinar a autocorrelação espacial dos vários agrupamentos de dados.
- Obter uma divisão regional preliminar e verificar sua consistência através de métodos de análise estatística espacial.
- Aplicar o modelo ao estado do Rio Grande do Sul com vistas à compreensão da realidade espacial e socioeconômica gaúcha.
- Obter soluções gráficas que sintetizem informações relevantes.
- Comparar as regiões administrativas obtidas através do modelo com as várias regionalizações adotadas oficialmente no Rio Grande do Sul.
- Análise crítica dos resultados obtidos.

1.4 ESTRUTURA DO TRABALHO

O trabalho está estruturado em cinco capítulos: introdução, revisão bibliográfica, metodologia, estudo de caso e conclusões. A introdução visa assinalar a importância e utilidade do modelo de regionalização desenvolvido definindo seus objetivos. A revisão bibliográfica que trata dos aspectos conceituais de divisão regional, regionalização e seus critérios; dos aspectos estruturais onde apresenta-se alguns índices de desenvolvimento; dos aspectos analíticos relativos à análise estatística e espacial; da questão da estrutura dos dados censitários e, por fim, traça um perfil da matéria geoprocessamento, enquanto ferramenta de aplicação no estudo. O terceiro capítulo apresenta a metodologia desenvolvida centrada em quatro blocos a saber: obtenção dos dados, análise estatística multivariada, validação dos resultados e apresentação dos resultados. No quarto capítulo são apresentados, sob a forma de mapas, vários níveis de regionalização obtidos pelo modelo e são tecidas considerações comparativas. O último capítulo aborda os aspectos finais de análise no que diz respeito à crítica dos resultados, potencial a ser explorado pelo método e validade do modelo proposto.



CAPÍTULO 2

REVISÃO BIBLIOGRÁFICA

2. REVISÃO BIBLIOGRÁFICA

O planejamento aplicado na execução de políticas regionais pressupõe uma visão o mais exata possível da área em que se vai intervir, caso haja uma preocupação efetiva com a eficiência. É preciso cumprir uma série de requisitos para que a aplicação dos recursos seja a menos onerosa para os estados e a mais útil para a população. Dito de um modo bastante reincidente, deve-se maximizar os benefícios e minimizar os investimentos compatíveis. Portanto, é preciso uma avaliação dos dados regionais relevantes de forma a obter um agrupamento regional adequado, o que fornecerá um panorama claro dos aspectos relevantes da estrutura regional e do uso dos instrumentos disponíveis (Fischer, 1979). Deste modo haverá uma contribuição efetiva para a locação dos recursos das diferentes secretarias, órgãos ou agências.

Certamente que estatísticas sociais ou métodos de análise espacial e de correlação entre variáveis socioeconômicas não são imprescindíveis para a adoção de programas sociais, entretanto a questão a ser colocada é: quão efetivos são estes programas e que mecanismos de avaliação possuem? A resposta à esta pergunta não é de natureza tão imediata, assim como é difícil de avaliar, à primeira vista, se, como e quanto os métodos estatísticos podem ser úteis na elaboração de políticas públicas (Cabello, 1960). Contudo, inumeráveis trabalhos nesta linha de desenvolvimento já foram publicados, alguns de caráter mais prático outros mais acadêmicos como será visto ao longo deste capítulo.

Quando se trabalha com um número menor de unidades espaciais básicas, obtidas por modelos regionais, obtém-se paralelamente uma organização e armazenamento de dados regionais mais eficiente, ainda que por este meio não seja possível explicar sistemas reais empíricos (Fischer, op.cit.). Mas, por originarem uma visualização mais definida entre

as similaridades e as diferenças relativas às regiões, criam condições de lançamento de hipóteses espaciais mais significativas, e de serem feitas explorações analíticas mais efetivas.

A estrutura deste segundo capítulo segue a seguinte ordem: aspectos conceituais da regionalização, onde procura-se diferenciar regionalização de divisão regional e apresenta-se tipos de divisões regionais existentes; aspectos estruturais que revisa a importância dos índices através de alguns exemplos; aspectos analíticos que consideram a questão da estatística multivariada e da análise espacial sendo apresentadas diversas medidas de correlação espacial e introduz-se o método de análise de agrupamento; o item sobre a estrutura de dados censitários faz uma análise a respeito da origem da informação censitária quanto à qualidade do dado; e por fim apresenta-se de modo genérico a aplicação do Geoprocessamento à Análise espacial.

2.1 ASPECTOS CONCEITUAIS DA REGIONALIZAÇÃO

2.1.1 DIVISÕES REGIONAIS E CRITÉRIOS DE REGIONALIZAÇÃO

A razão fundamental para a adoção de uma regionalização administrativa é a distribuição de recursos públicos da forma mais eficiente possível. Isto é, busca-se a utilização do recurso público em situações de oferta de serviços à população de forma que haja o maior retorno em termos de custo-benefício.

Cabe demarcar que alguns autores ao abordarem a questão das regionalizações do ponto de vista da Geografia, como Duarte (1980) ou Mesquita (1984), consideram serem diferentes os conceitos de divisões regionais e regionalizações, o que está sendo tomado no presente trabalho quase como o mesmo conceito. Contudo, faz-se necessário mostrar a diferença apresentada por estes autores como modo de aproveitar uma nuance que poderia, de outra forma, passar despercebida.

Os autores consideram como divisão regional o ato de particionar o espaço a partir de critérios arbitrariamente escolhidos que visam atender a determinados fins. O que teria sido largamente empregado no Brasil nas décadas de 60 e 70 em programas de planejamento regional, seja no âmbito federal ou estadual (Mesquita, op.cit.). Já regionalização seria, para eles, antes um processo que tem lugar na realidade e resulta na

formação de regiões e transformação de regiões como culminação de um relacionamento econômico e social definido por uma inter-relação entre a sociedade ou grupos sociais sobre o espaço em termos de organização, direção e ações (Duarte, op.cit.).

É neste ponto que a presente abordagem faz a convergência entre divisão regional e regionalização, porque busca-se de fato uma metodologia de divisão espacial segundo critérios de homogeneidade, mas que apreenda, a partir dos dados regionais, sociais e econômicos (atributos quantificáveis), a forma e a relação entre estes dados de modo que a regionalização obtida contemple e clareie os processos de transformação a que é submetido o espaço continuamente.

Nesta abordagem, que enfoca aspectos socioeconômicos, a escolha de variáveis reveste-se de grande importância, e deverá atender às demandas prioritárias de serviços públicos, levando em consideração as áreas social e econômica de natureza essencial, tais como saúde, educação, alimentação, habitação, níveis de emprego e renda e capacidade de produção, além de contemplar a área ambiental de alguma forma que demarque a importância do ambiente, enquanto estrutura de sustentação de recursos. Variáveis demográficas também são levadas em consideração, porque representam a quantificação das demandas por serviços públicos e privados, uma vez que toda oferta destes serviços é direcionada à população como um todo.

Analisando um conjunto de trabalhos (Lolonis, 1993), (Fischer, 1980), (Coombes, 1994), (Barkley, 1996), (Anson, 1991), (Pfeiffer, 1980) e (Shefer, 1997) que fazem estudos de regionalização e/ou criação de índices e realizam uma eleição de variáveis pertinentes ao contexto, observamos a reincidência de fatores comuns, o que apontaria para a importância de considerar estes atributos em trabalho de abordagem semelhante, como é o caso do presente estudo:

Demográficos

- Mortalidade infantil
- Taxa de migração
- Densidade populacional
- População urbana

Emprego

- População empregada no setor primário
- População empregada no setor secundário

- População empregada no setor terciário
- População economicamente ativa

Educação

- Número de estudantes de 1º grau / 1000 hab.
- Número de estudantes de 2º grau / 1000 hab.
- Número de estudantes de 3º grau / 1000 hab.
- Taxa de alfabetização de adultos

Habitação

- Habitações com instalações hidro-sanitárias
- Habitações com energia elétrica

Saúde

- Número de leitos hospitalares / 1000 hab.
- Número de médicos / 100.000 hab.

Econômicas

- PIB per capita municipal
- Estrutura fundiária
- Ramos Industriais

Ambientais

- Percentual de indústrias poluentes
- Cargas industriais remanescentes
- Volume de utilização de agrotóxicos

Geofísicas

- Área
- Altitude
- Cobertura vegetal
- Clima
- Temperatura média anual
- Precipitação média anual

Administração

- Tempo de emancipação

A maior parte das variáveis foram pré-selecionadas em função de já terem sido eleitas em outros trabalhos de temática similar, como já foi mencionado; por serem dados disponíveis e abertos por município em alguma das fontes de informação consultadas: IBGE (Instituto Brasileiro de Geografia e Estatística) e FEE (Fundação de Economia e Estatística – RS); e por que respondem à necessidade de se trabalhar com dados que

façam sentido se especializados, de modo a ser possível a aplicação de análise estatística espacial, uma das principais ferramentas deste trabalho. As variáveis que, mesmo sendo relevantes, não foram incluídas é devido ao fato de não estarem disponíveis, nestas fontes, para a maioria dos municípios gaúchos.

O trabalho da geógrafa Zilá Mesquita (1984), que faz a revisão das divisões regionais no estado, traz importante aporte para este trabalho na questão de localizar e identificar a origem das diversas divisões regionais adotadas no Rio Grande do Sul. Além de focar o assunto por ângulos geográficos e metodológicos significativos, aprofunda na caracterização das diferentes divisões regionais existentes segundo cinco categorias: divisões baseadas em regiões naturais, ligadas à ocupação do território, criadas para objetivos estatísticos, de caráter administrativo e para estudos específicos.

É diretamente pertinente ao enfoque deste trabalho as divisões regionais para fins estatísticos e as administrativas, as quais serão consideradas detidamente. As demais serão apenas mencionadas e o leitor interessado poderá encontrar o trabalho original na Fundação de Economia e Estatística (FEE), com extenso e minucioso detalhamento.

O conceito de regiões naturais, em oposição à forma arbitrária como são estabelecidas as regiões administrativas, surge da verificação sobre a instabilidade destas regiões em função dos desmembramentos e emancipações municipais que ao longo do tempo ocorrem desconhecendo a unidade de áreas territoriais homogêneas (Guimarães, 1963). Portanto, as regiões naturais são calcadas sobre as características fisiográficas da própria natureza.

Dois trabalhos são importantes mencionar como matrizes originárias de divisões regionais fundamentadas na ocupação do território (Mesquita, op. cit.): Bases Geográficas do povoamento do Estado do Rio Grande do Sul de Nilo Bernardes (1963) e A colonização alemã e o Rio Grande do Sul de Jean Roche (1969). No primeiro Bernardes afirma haver uma sobreposição entre as “zonas geográficas naturais” e “as características culturais homogêneas” rompendo com o padrão aceito até então que dissociava região natural de região humana. O segundo, de Roche, faz um mapeamento das diversas zonas de colonização do estado: açoriana, alemã, italiana e mista.

2.1.2 DIVISÕES REGIONAIS PARA FINS ESTATÍSTICOS

Desde de 1968, quando o IBGE criou uma nova divisão regional para o Brasil, que substituiria a divisão em zonas fisiográficas a qual vigorava desde 1940 (Mesquita, op. cit.), vem sendo utilizada a idéia de microrregiões homogêneas com origem na metodologia do geógrafo Michel Rochefort. Contudo, o número de microrregiões e mesorregiões (agrupamentos de microrregiões) vem aumentando, o que indica uma tendência de maior subdivisão do espaço como objeto de estudos estatísticos e sociais. No Rio Grande do Sul, tinham sido estipuladas, originalmente, 24 microrregiões que passaram, posteriormente, a ser 35, enquanto que as 6 mesorregiões tornaram-se 7³.

Contemporâneo à divisão regional do IBGE, surge o trabalho “Regiões polarizadas e homogêneas” de Neves e Abrantes cujos critérios fundamentais se vinculam à terra: tamanho médio das propriedades rurais, formas de utilização e posse da terra, distribuição da população rural (Mesquita, op.cit.).

Já em 1984 o trabalho “Áreas estatisticamente comparáveis no Rio Grande do Sul” de Alonso, Valente, Bandeira e Mesquita (Alonso, 1984) traz em si a preocupação de estabelecer critérios metodológicos que permitissem algum tipo de análise comparativa ao longo do tempo (1940 a 1975) tendo por base as 58 AECs (Áreas estatisticamente comparáveis) para o Rio Grande do Sul agrupadas conforme as datas das emancipações e origem dos desmembramentos municipais.

2.1.3 DIVISÕES REGIONAIS ADMINISTRATIVAS

Várias Secretarias de Estado fazem uso de sistemas de divisão regional para a administração e implantação de políticas públicas. Entre elas a Secretaria da Coordenação e Planejamento (SCP) com os Conselhos Regionais de Desenvolvimento (COREDES), a Secretaria da Educação com suas Delegacias de Educação⁴ (DE) e Distritos Geoeducacionais⁵ no âmbito federal (Ministério de Educação e do Desporto), e a Secretaria da Saúde com as Delegacias Regionais de Saúde (DRS). É importante comparar os resultados destas regionalizações, enquanto mapeamentos resultantes, para

³ Mapas das micro e mesorregiões vigentes nas páginas 18 e 19, respectivamente.

⁴ Órgão local e de representação da Secretaria da Educação, abrangendo município ou grupo de municípios.

⁵ Distrito Geoeducacional – Unidade Geográfica, cujos limites coincidem com limites políticos de um conjunto de municípios da mesma unidade da Federação, estabelecidos com a finalidade de embasar o processo de planejamento e tomada de decisões no âmbito da política educacional do 3º grau.

verificar que os critérios de subdivisão são muito variáveis. A subdivisão coincide completamente em apenas uma região que é a região Norte (classificação da SCP – COREDES). Para esta área de 27 municípios há uma coincidência nas três Secretarias de estado. Isto é, os municípios englobados por esta região são os mesmos para a DE e DRS, correspondendo à 15ª Delegacia de Educação e à 11ª Delegacia Regional de Saúde, e também à parte do 38º Distrito Geoeducacional. Existe, também, a coincidência de classificação para mais duas regiões no estado que são: a região Centro-sul (COREDES) correspondente à 2ª Delegacia Regional de Saúde e à 12ª Delegacia de Educação; e a região Litoral (COREDES) correspondente à 18ª Delegacia Regional de Saúde e à 11ª Delegacia de Educação. Depois verifica-se que a região da Fronteira Noroeste (COREDES) corresponde à 17ª Delegacia de Educação e a região do Médio Alto Uruguai (COREDES) corresponde à 15ª Delegacia Regional de Saúde. Para as demais regiões não existe correspondência direta, o que expõe uma dificuldade de entrosamento para estudos que, por exemplo, relacionassem a área de saúde com a de educação a partir de uma regionalização oficial destas secretarias.

As DRS a partir de janeiro de 1999 foram transformadas em coordenadorias regionais de saúde, tendo sido criados sete Pólos Regionais de saúde: Norte, Serra, Missioneira, Vales, Centro-oeste, Metropolitana e Sul.

Como os COREDES têm por objetivo ser “um fórum de discussão e decisão a respeito de políticas e ações que visem o desenvolvimento regional” cabe um detalhamento maior sobre este organismo.

2.1.3.1 CONSELHOS REGIONAIS DE DESENVOLVIMENTO (COREDES)⁶

Os objetivos dos Conselhos Regionais de Desenvolvimento criados pela Lei n.º. 10.283 de 17/10/94 e regulamentados pelo Decreto n.º. 35.764 de 28/12/94 são:

- a integração dos recursos e das ações do governo na região;
- a promoção do desenvolvimento regional harmônico e sustentável;
- a preservação e recuperação do meio ambiente;
- a melhoria da qualidade de vida da população;
- a distribuição eqüitativa da riqueza produzida;
- o estímulo à permanência do homem em sua região (Página da Internet, 1999)⁷.

⁶ Em abril de 1999, eram 22 COREDES, 29 DE e 18 DRS no estado do Rio Grande do Sul.

⁷ Home page: <http://www.scp.rs.gov.br/coredes/paginas/htm>

Através de uma análise primária dos objetivos dos COREDES, conforme a lei, é possível verificar a preocupação socioeconômica do legislador uma vez que, com exceção do primeiro objetivo, que é de natureza político-econômica, todos os demais apresentam um caráter social revestido de uma natureza econômica e permeados por uma tônica ecológica e ambiental. Como atribuições dos COREDES são listadas competências de natureza político-participativa, de análise e planejamento, e de inserção no âmbito federal, assim formuladas:

- promoção da participação dos segmentos sociais de cada região na definição de prioridades e potenciais, visando a elaboração e implementação de políticas de desenvolvimento regional;
- formulação de planos de desenvolvimento regional;
- valorização da ação política a partir da busca de participação da cidadania;
- regionalização do orçamento do estado;
- apoio e acompanhamento da eficácia das ações estaduais e federais nas regiões;
- respaldo do governo estadual na participação federativa.

São atividades dos COREDES colaborar na preparação, alteração e atualização de Planos Estaduais e Regionais junto aos órgãos estaduais de planejamento, universidades e organismos públicos ou privados, através da elaboração e discussão de diagnósticos regionais. Priorizar as necessidades de serviços públicos, fiscalizar a qualidade dos serviços públicos estaduais, reivindicar atendimento dos interesses regionais junto ao Governo Federal e buscar a participação da comunidade regional em programas públicos.

Para tanto, torna-se fundamental a utilização de uma forma técnica e prática de viabilizar estes estudos e planos. É nesta direção que aparece o geoprocessamento, como tecnologia crescentemente empregada em muitos países e que no Brasil começa a tomar corpo.

2.1.3.2 REGIÕES DO ORÇAMENTO PARTICIPATIVO

As regiões do orçamento participativo são as mesmas adotadas pelos COREDES e aproveitam a estrutura existente para agilizar a sua funcionalidade. São listadas a seguir:

1. Alto Jacuí com 18 municípios.
2. Campanha com 6 municípios.

3. Central com 34 municípios.
 4. Centro-sul com 16 municípios.
 5. Fronteira Noroeste com 20 municípios.
 6. Fronteira Oeste com 11 municípios.
 7. Hortênsias com 8 municípios.
 8. Litoral com 21 municípios.
 9. Médio Alto Uruguai com 30 municípios.
 10. Missões com 24 municípios.
 11. Nordeste com 20 municípios.
 12. Noroeste Colonial com 31 municípios.
 13. Norte com 27 municípios.
 14. Paranhana – Encosta da Serra com 11 municípios.
 15. Produção com 36 municípios.
 16. Serra com 32 municípios.
 17. Sul com 21 municípios.
 18. Vale do Caí com 18 municípios.
 19. Vale do Rio dos Sinos com 14 municípios.
 20. Vale do Rio Pardo com 24 municípios.
 21. Vale do Taquari com 36 municípios.
 22. Metropolitana Delta do Jacuí com 9 municípios.
- Total de 467 municípios.

A seguir exemplifica-se com as figuras 2.1, 2.2, 2.3 e 2.4 algumas das divisões regionais oficiais válidas para o Rio Grande do Sul.

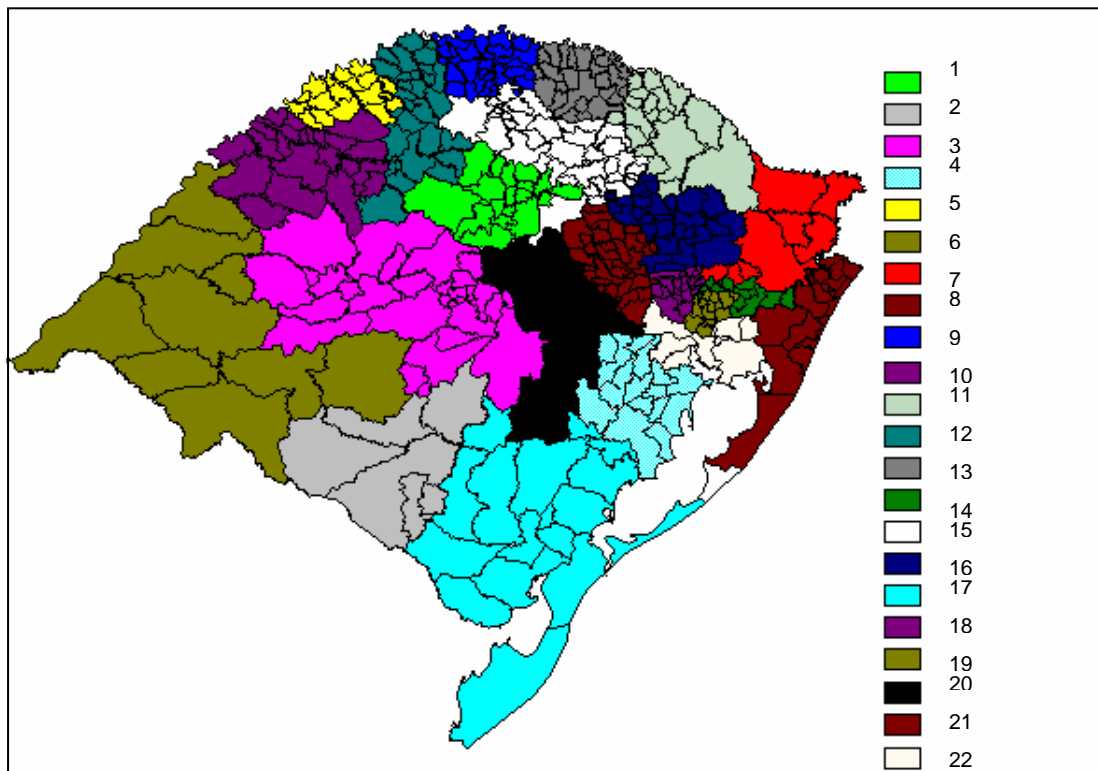


Figura 2.1 - Mapa dos Conselhos Regionais de Desenvolvimento e do Orçamento Participativo (Fonte: Ministério da Saúde, 1999).

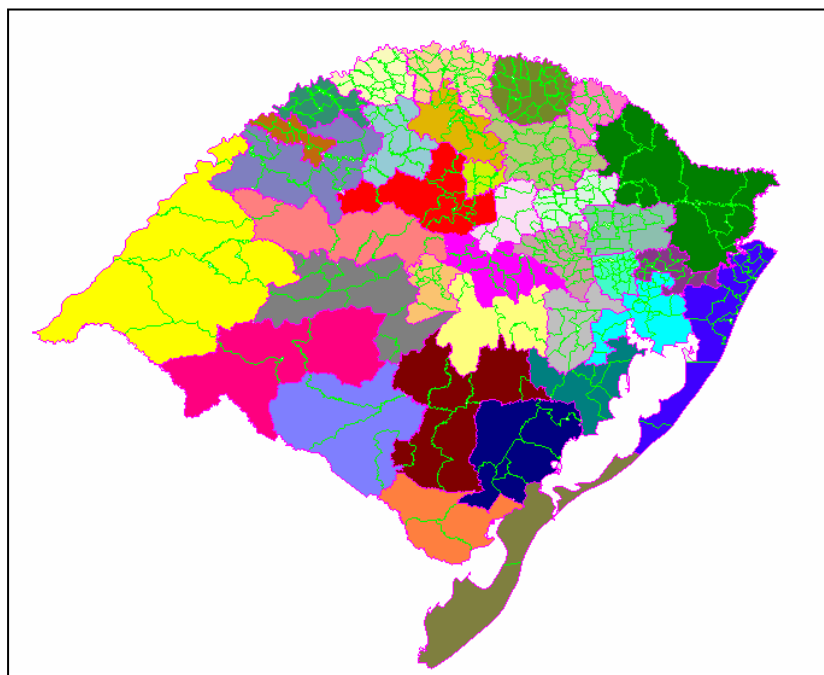


Figura 2.2 - Mapa das Microrregiões do Rio Grande do Sul – 1994
(Fonte: Ministério da Saúde, 1999).

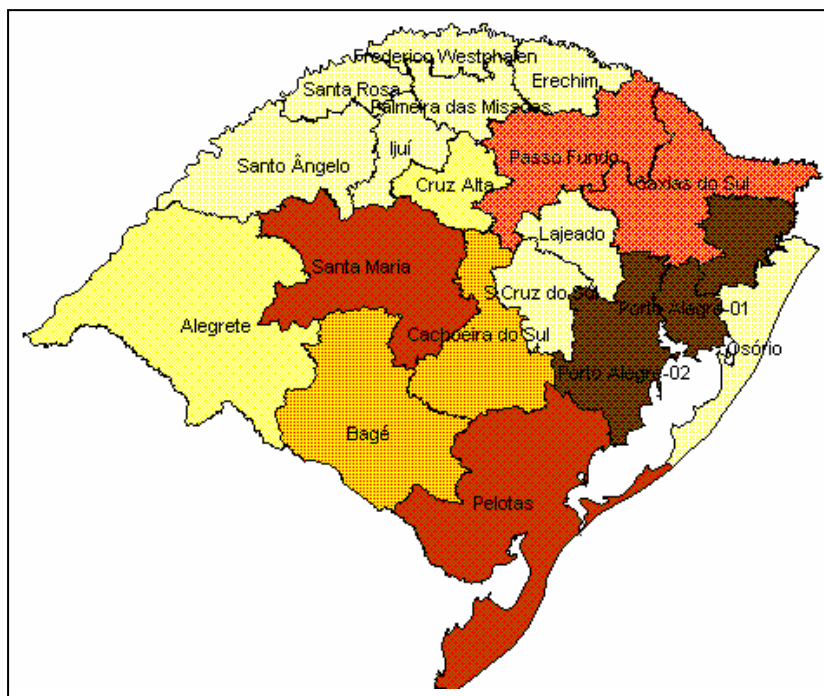


Figura 2.3- Mapa das Regionais de saúde para o estado Rio Grande do Sul – 1994
(Fonte: Ministério da Saúde, 1999).

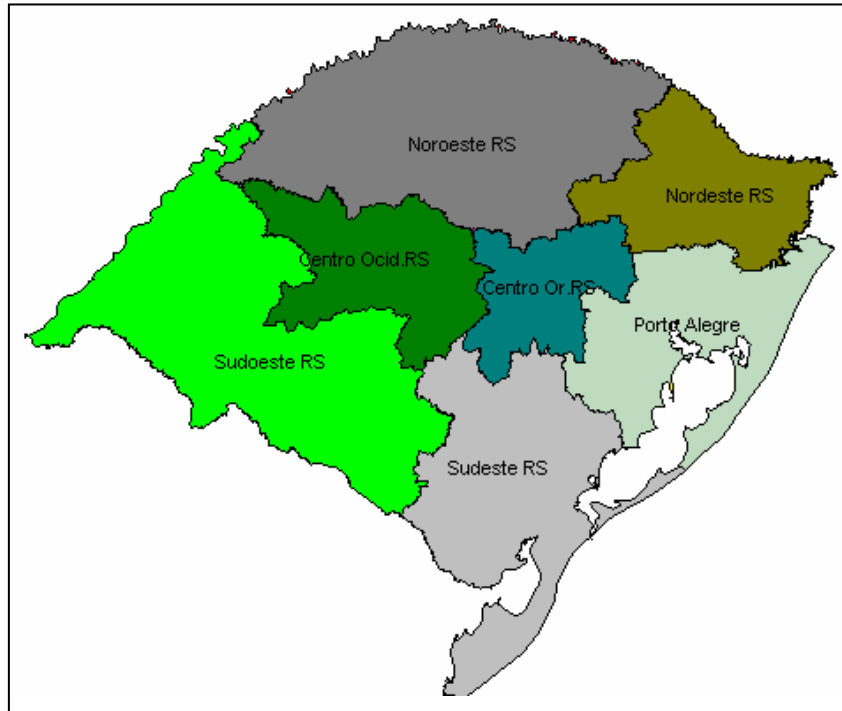


Figura 2.4 – Mapa das Mesorregiões do Rio Grande do Sul – 1994
(Fonte: Ministério da Saúde, 1999).

Observe-se nas quatro figuras acima a diversidade de traçados das áreas regionais.

2.2 ASPECTOS ESTRUTURAIS

2.2.1 ÍNDICE DE DESENVOLVIMENTO HUMANO (IDH)

A seguir exemplifica-se a utilidade dos índices com a apresentação sucinta da metodologia utilizada pelo Programa das Nações Unidas para o Desenvolvimento⁸. Trata-se do IDH (Índice de Desenvolvimento Humano) que tem sido largamente utilizado como critério de classificação do grau de desenvolvimento dos países.

O IDH, conforme o Relatório do Desenvolvimento Humano da ONU de 1998, resulta da combinação de três índices: Longevidade, Educação e Renda.

A longevidade é medida pela esperança de vida ao nascer. A educação é medida pela combinação da taxa de alfabetização de adultos e da taxa combinada de matrícula

⁸ <http://www.undp.org.br>

nos três níveis de ensino. A renda é medida pelo PIB per capita expresso pela paridade do poder de compra em dólares PPC.

A fórmula geral dos índices é:

$$\text{Índice} = \frac{\text{valor observado} - \text{valor mínimo}}{\text{valor máximo} - \text{valor mínimo}} \quad [2.1]$$

Tabela 2.1 - Tabela de Mínimos e Máximos para os indicadores do IDH

Indicador	Mínimo	Máximo	Unidade
Esperança de vida ao nascer	25	85	Anos
Taxa de alfabetização	0	100	%
Taxa combinada de matrícula	0	100	%
PIB per capita	100	40.000	US\$ PPC

(Fonte: Programa das Nações Unidas para o Desenvolvimento, 1999)

Assim o índice de Longevidade (IL_i) do país i , que possui esperança de vida ao nascer V_i é obtido pela fórmula:

$$IL_i = \frac{V_i - 25}{85 - 25} \quad [2.2]$$

O índice de Educação (IE_i) do país i , com taxa de alfabetização de adultos A_i e taxa combinada de matrícula M_i será dado por:

$$IE_i = 2/3 IA_i + 1/3 IM_i \quad [2.3]$$

onde

$$IA_i = A_i / 100$$

e

$$IM_i = M_i / 100$$

O índice de Renda (IR_i) do país i assume a premissa de que a contribuição da renda para o desenvolvimento humano apresenta rendimentos decrescentes, através da incorporação de um redutor para o PIB per capita observado, Y_i , pela fórmula de Atkinson, gerando W_i , PIB per capita ajustado para o país i :

$$W_i = Y^* + 2(Y^*)^{1/2} + 3(Y^*)^{1/3} + \dots + n[Y - (n - 1) Y^*]^{1/n} \text{ para } (n - 1)Y^* \leq Y \leq nY^* \quad [2.4]$$

onde Y^* é o PIB per capita médio mundial, definido como o patamar a partir do qual a renda apresenta rendimentos decrescentes.

Para $0 < Y_i < Y^*$, $W_i = Y^*$ e para $Y^* \leq Y_i \leq 2Y^*$, $W_i = Y^* + 2(Y_i - Y^*)^{1/2}$ e assim por diante até Y_i cair no intervalo $(n - 1)Y^* \leq Y_i \leq nY^*$ quando vale a fórmula geral de Atkinson, assim W_m é o limite máximo do PIB per capita corrigido e então,

$$IR_i = \frac{W_i - 100}{W_m - 100} \quad [2.5]$$

Finalmente, o índice de desenvolvimento humano, IDH_i é dado pela média aritmética simples dos três anteriores:

$$IDH_i = \frac{IL_i + IE_i + IR_i}{3} \quad [2.6]$$

Varia entre 0 e 1 e possui a seguinte classificação estabelecida pelo PNUD:

$0 \leq IDH \leq 0,5$	Baixo desenvolvimento humano
$0,5 \leq IDH \leq 0,8$	Médio desenvolvimento humano
$0,8 \leq IDH \leq 1$	Alto desenvolvimento humano

Em 1999 houve uma mudança no cálculo do índice de Renda na tentativa de minimizar a distorção ocasionada pela renda per capita cujo crescimento não encontrava reflexo direto no aumento do IDH, causando uma redução artificial especialmente nos países de renda mais elevada. Deste modo, foi adotada uma nova metodologia, cuja curva de renda per capita em PPC é mostrada abaixo, figura 2.5, caracterizando uma suavização do crescimento do IDH com a variação da renda. Esta é uma demonstração de uma forma de incorporar melhor os elementos que retrata com maior proximidade a realidade em estudo. Essa metodologia incorpora melhor os elementos e retrata mais fidedignamente a realidade.

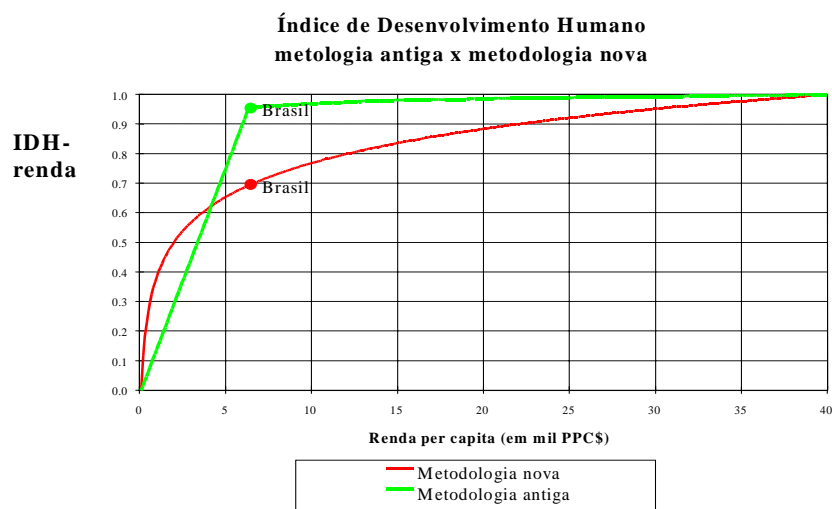


Figura 2.5 – Comparativo das duas metodologias do IDH-renda até 1998 e de 1999.
(Fonte: Programa das Nações Unidas para o Desenvolvimento, 1999)

Neste momento, já é possível antever a utilidade da elaboração de um IDM (Índice de Desenvolvimento Municipal) e de um IDR (Índice de Desenvolvimento Regional) no âmbito desta abordagem, como elementos de apoio à decisão para implemento de políticas públicas.

2.2.2 ÍNDICE DE CARÊNCIA SOCIAL E ÍNDICE DE DESENVOLVIMENTO SOCIAL

O índice de carência social faz parte da metodologia apresentada na publicação “Índice de Desenvolvimento Social (IDS)” de Silveira e Sampaio (1996) para a Fundação de Economia e Estatística, a qual é baseada nos estudos de Maria Cecília Prates Rodrigues, publicados na revista Conjuntura Econômica (1991, 1993 e 1994). Segue uma formulação de cálculo semelhante à do IDH, estabelecendo uma relação entre os valores das variáveis para a melhor e a pior ocorrência em cada município, expressa como segue:

$$I_{iJ} = 1 - \frac{X_{iJ} - X_{iP}}{X_{iM} - X_{iJP}} \quad [2.7]$$

onde

I_{iJ} é o Índice de Carência Social;

x_{iP} é o pior valor da taxa J;

x_{iM} é o melhor valor da taxa J;

i é o número de municípios;

J é o número de variáveis.

As variáveis utilizadas pelos pesquisadores da FEE foram a taxa de alfabetização de adultos e o grau de escolarização no 1º grau para a área da educação; a taxa de mortalidade infantil para a saúde, com a média dos óbitos infantis no triênio 90-92 relativo à população censitária (1991) dos menores de um ano de idade; e para a renda uma ponderação entre o grau de indigência calculado pelo Instituto de Pesquisa Econômica Aplicada (IPEA) e o PIB per capita, com pesos de 0,4 e 0,6 respectivamente. Resulta, então, que o ICS é dado pela média dos três valores calculados para educação, saúde e renda:

$$ICS_i = \frac{ICS_{e_i} + ICS_{s_i} + ICS_{r_i}}{3} \quad [2.8]$$

onde

ICS_i é o Índice de carência social do município i ;

ICS_{e_i} é o Índice de carência social para educação no município i ;

ICS_{s_i} é o Índice de carência social para saúde no município i ;

ICS_{r_i} é o Índice de carência social para renda no município i .

Finalmente, chega-se ao Índice de Desenvolvimento Social (IDS) que seria de 100% na situação ideal de desenvolvimento máximo, portanto o IDS_i para o município i será dado pela relação abaixo:

$$IDS_i = 1 - ICS_i \quad [2.9]$$

A metodologia original para cálculo do IDS apresentada por Rodrigues (1994) segue mais de perto o método do IDH com uma diferença substancial no cálculo do índice de renda que se baseava, em 1991, no grau de distribuição de renda composto pela renda média da população economicamente ativa (PEA) com rendimento ou coeficiente de Gini, o qual para valores elevados revela uma má distribuição de renda. Posteriormente, o índice de renda passa a ser calculado por uma ponderação entre a renda média da PEA remunerada (com peso 0,4) dividida pela razão ricos/pobres (com peso de 0,6) que é a “relação entre a proporção da renda total da PEA apropriada pelos seus 10% mais ricos e a proporção apropriada pelos seus 20% mais pobres” (Rodrigues, op.cit.).

2.2.3 ÍNDICE DE DESENVOLVIMENTO URBANO

O Índice de Desenvolvimento Urbano foi proposto pelo economista Carrion (1997), sob uma ótica preliminar e experimental, como indicador do desenvolvimento urbano dos municípios da Região Metropolitana de Porto Alegre com o fim de subsidiar estudos regionais e municipais criando um panorama “da capacidade diferenciada de atendimento dos principais serviços à população”.

2.3 ASPECTOS ANALÍTICOS

2.3.1 ESTATÍSTICA MULTIVARIADA

Os objetivos principais da investigação experimental através de estatística multivariada são (Johnson, 1998):

- Simplificação: o fenômeno é representado através de todas as variáveis consideradas significativas, o que implica em uma redução que deverá conduzir à interpretação mais direta.
- Agrupamentos: os objetos são agrupados de acordo com suas características similares o que define classes homogêneas.
- Interdependência: as variáveis são analisadas segundo suas interdependências estabelecendo-se aquelas que são independentes e as que são dependentes.
- Previsão: os relacionamentos entre as variáveis são determinados possibilitando prever o comportamento de umas em relação a outras.
- Verificação: as hipóteses são cotejadas através das técnicas de estatística multivariada verificando a validade das assertivas.

Grande parte da informação contida em um conjunto de dados é melhor apropriada pelo cálculo de determinadas medidas designadas como estatísticas descritivas (Johnson, op.cit): média, variância, desvio padrão, covariância e coeficiente de correlação ou momento do produto de Pearson.

O desvio padrão é uma medida do grau de dispersão dos valores em relação à média. A covariância é a média dos produtos dos desvios para cada par de pontos de dados. Usa-se a covariância para determinar a relação entre dois conjuntos de dados. O coeficiente de correlação ou momento do produto Pearson, r , é um índice sem dimensão situado entre -1,0 e 1,0 inclusive, que reflete a extensão de uma relação linear entre dois conjuntos de dados.

2.3.2 ANÁLISE ESPACIAL

Existem dois artigos que serão de importância basilar no desenvolvimento do presente modelo de regionalização, no que tange à construção da metodologia que será adotada, em especial sobre o uso da autocorrelação espacial. O primeiro deles é o artigo intitulado “A Integração de Análise Espacial e SIG” (Ding, 1992), cujo método é a base para o lançamento do segundo, “SIG e Autocorrelação Espacial” (Shen, 1994). Há também um terceiro trabalho, chamado “Modelos de Localização-Alocação como Suporte à Decisão na Definição de Regiões Administrativas” (Lolonis, 1993), não menos importante, cujas definições de coeficiente de diferenciação espacial e distância transformada, são apropriadas no presente estudo. Cabe citar ainda, à título introdutório, o trabalho de Lee e Culhane, “Um índice de agrupamento baseado no perímetro para a medida de segregação espacial: uma abordagem cognitiva SIG”, do qual será aproveitado a metodologia do cálculo do índice de agrupamento. E, por fim, o método de análise de agrupamento apresentado em “Determinação de zonas homogêneas dentro de áreas urbanas: aspectos metodológicos” (Haddad, 1997), cujo conceito de taxonomia regional apropriado de Fischer (1979) será também utilizado, enquanto que a análise de agrupamento será estendida para a realidade regional. Da associação destes trabalhos mencionados, chegar-se-á ao método aqui proposto.

2.3.2.1 MEDIDAS DE ANÁLISE ESPACIAL

À época em que foi publicado o artigo, 1992, os autores pertenciam ao Centro Nacional de Análise e Informação Geográfica de Buffalo (N.Y.). Utilizaram como exemplo de aplicação as províncias da China, com dados do censo de 1990 para o atributo população.

O artigo desdobra-se em três partes: cálculo das medidas de autocorrelação e associação espacial, um módulo de análise espacial e um exemplo de aplicação sobre os dados da população chinesa de 1990.

A primeira é de fundamental importância para este trabalho. Nela, os autores introduzem o tema demonstrando a independência de origem do Sistema de Informação Geográfica (SIG) e da Análise Espacial, entendida como análise estatística espacial. Contudo, justificam a aproximação das duas áreas relatando a deficiência do SIG de

rotinas de análise estatística espacial e as vantagens que um módulo de análise espacial traria às capacidades visuais e de análise geográfica do SIG.

A análise espacial, segundo os autores, trabalha com dois tipos distintos de informação: atributos espaciais dos objetos, como área, população, tipo de solo, precipitação pluviométrica, etc; e informação locacional sobre os objetos, descrita por medidas de posição em mapas ou sistemas de coordenadas geográficas.

A **autocorrelação espacial** tem a ver com o grau de similaridade dos objetos ou atividades localizados próximos, podendo ser positiva, negativa ou nula. Para os objetos que têm similaridade locacional e que tendem a ter atributos similares é dito que apresentam autocorrelação espacial **positiva**. Autocorrelação espacial **negativa** é revelada pelos objetos próximos entre si que apresentam atributos dissimilares. Uma distribuição de atributos independente da localização determina autocorrelação igual a **zero**.

A **medida de proximidade espacial** é dada por uma matriz binária (que assume valores de 1 ou 0) cujos elementos serão 1(um) para polígonos adjacentes ou dentro de uma distância d entre dois pontos, que representam cada polígono, caso contrário 0 (zero). Assim temos:

$$W = [w_{ij}]_{n \times n} \quad [2.10]$$

sendo $w_{ij} = 1$ se o polígono i é adjacente ao polígono j ou;

$w_{ij} = 0$ se não houver adjacência;

$w_{ii} = 0$, a diagonal da matriz possui elementos zerados.

A **medida de autocorrelação espacial** é o coeficiente de **Moran**, I^M , cuja medida de similaridade dos atributos é dado por:

$$C_{ij} = (X_i - X_{\text{médio}}) (X_j - X_{\text{médio}}) \quad [2.11]$$

onde x_i valor do atributo para o objeto i ;

x_j valor do atributo para o objeto j ;

$x_{\text{médio}}$ valor médio do atributo para todos os n objetos.

Então, o índice ou coeficiente de Moran, I^M varia entre +1 e -1 e é dado pela relação:

$$I^M = \frac{\sum_i \sum_j (w_{ij} c_{ij})}{(s^2 \sum_i \sum_j w_{ij})} \quad [2.12]$$

onde o primeiro termo é a medida de covariância entre o objeto i e todos os outros objetos e s^2 é a variância dada por:

$$s^2 = \frac{\sum_i (x_i - x_{\text{médio}})^2}{n} \quad [2.13]$$

A interpretação correspondente para o coeficiente I^M de Moran diz que será positivo para atributos similares, negativo para dissimilares e zero para atributos independentes no espaço.

A **medida de associação espacial** é chamada de estatística G e mede a concentração da distribuição espacial dos valores de um atributo. Está baseada na matriz de distâncias entre os centros de todos os polígonos em questão. Apresenta a vantagem sobre o coeficiente de Moran de desagregação por ponto i , de modo que um conjunto de $G_i(d)$ expressa a medida do grau de associação entre o ponto i e todos os demais dentro de um raio de distância d . Para cada ponto poderia ser calculado por:

$$G_i(d) = \frac{\sum_j w_{ij}(d) x_j}{\sum_j x_j} \quad \text{em que } j \neq i; i = 1, 2, \dots, n \quad [2.14]$$

Para grandes valores de G_i significará que grandes valores do atributo estão agrupados próximos de i . Quando nulo, significará a inexistência de associação espacial entre o valor x_j distante d do ponto i .

Os valores esperados para $G_i(d)$ e variância são dados por:

$$E(G_i(d)) = W_i / (n-1) \quad [2.15]$$

sendo

$$W_i = \sum_j w_{ij}(d) \quad [2.16]$$

e

$$\text{Var}(G_i(d)) = \frac{(W_i (n-1 - W_i) / (n-1)^2 (n-2)) \cdot (Y_2/Y_1^2)}{\quad} \quad [2.17]$$

onde

$$Y_1 = \sum_j x_j / (n-1) \quad [2.18]$$

e

$$Y_2 = \sum_j x_j^2 / (n-1) - Y_1^2 \quad [2.19]$$

Então a curva normal representada pelo valor z para cada ponto i ou polígono será:

$$z_i = [G_i(d) - E(G_i(d))] / \sqrt{(E(\text{Var}(G_i(d))))} \quad [2.20]$$

Um resultado de z positivo e significativo implica em que grandes valores do atributo estão associados espacialmente com o ponto ou polígono i ; se for negativo e significativo representa que pequenos valores do atributo x_j têm associação espacial com o ponto ou polígono i . Os níveis de associação espacial poderão então ser mapeados com facilidade gráfica por um SIG, deixando à vista os níveis de associação nas diversas áreas de estudo. O exemplo a seguir tenta clarificar os tópicos abordados acima, utilizando para tanto uma matriz simples de 7x7.

- Medida de Proximidade Espacial (W_{ij})

Tomando-se como exemplo sete centróides de polígonos, calcula-se a distância d entre os pontos em um determinado raio (lag), e construiu-se uma matriz binária como a que segue abaixo:

$$\begin{array}{c}
 i \rightarrow \\
 \left| \begin{array}{ccccccc}
 0 & 1 & 1 & 0 & 1 & 0 & 0 \\
 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
 1 & 0 & 0 & 1 & 0 & 1 & 0 \\
 0 & 1 & 1 & 0 & 1 & 1 & 0 \\
 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
 0 & 0 & 1 & 1 & 0 & 0 & 1 \\
 0 & 0 & 0 & 0 & 0 & 1 & 0
 \end{array} \right| \begin{array}{c}
 j \\
 \downarrow
 \end{array}
 \end{array}$$

Esta matriz nos dá os atributos de proximidade entre dois pontos. Se o centróide do polígono se encontra dentro da distância d correspondente ao lag, o valor será 1, do contrário o valor será 0. Não é levada em consideração a distância do ponto a si mesmo, adotando-se o valor zero na matriz.

- Medida de Similaridade (c_{ij})

A medida para o atributo similaridade entre dois objetos utilizada no cálculo da autocorrelação com o coeficiente Moran (I^M), é dado pelo produto da diferença entre o atributo no primeiro ponto e a média dos atributos e diferença entre o atributo no segundo ponto e a média dos atributos, ou seja:

$$C_{ij} = (X_i - X_{\text{médio}}) \cdot (X_j - X_{\text{médio}}) \quad [2.11]$$

No exemplo adotado temos os seguintes atributos:

ponto	atributo
1	10
2	5
3	3
4	8
5	7
6	6
7	3

Aplicando-se a fórmula teremos os seguintes valores de c_{11} , c_{12} , c_{13} , e assim em diante até c_{77} , com os quais construiu-se a seguinte matriz:

i →								j ↓
16	-4	-12	8	4	0	-12		
-4	1	3	2	-1	0	3		
-12	3	9	-6	-1	0	9		
8	2	-6	4	2	0	6		
4	-1	-1	2	1	0	-3		
0	0	0	0	0	0	0		
-12	3	9	6	-3	0	9		

- Medida de Autocorrelação Espacial (I^M)

Mede a covariância entre o valor de um atributo em uma localização i e o valor em todas as outras localizações, em um coeficiente constricto entre os valores de +1 e -1. É dada pelo produto dos somatórios da similaridade e proximidade medidas e divididas por um fator de padronização, como segue:

$$I^M = \sum_i \sum_j (w_{ij} c_{ij}) / (s^2 \sum_i \sum_j w_{ij}) \quad [2.12]$$

onde S^2 denota a variância do atributo sendo medido por:

$$s^2 = \sum_i (x_i - x_{\text{médio}})^2 / n \quad [2.13]$$

- Matriz de Autocorrelação ($w_{ij} \cdot c_{ij}$)

Multiplicando-se os valores das matrizes de W_{ij} por C_{ij} teremos a seguintes matriz da similaridade entre os atributos pesados pela proximidade:

$$\begin{array}{c} i \rightarrow \\ \left| \begin{array}{ccccccc} 0 & -4 & -12 & 0 & 4 & 0 & 0 \\ -4 & 0 & 0 & 2 & 0 & 0 & 0 \\ -12 & 0 & 0 & -6 & 0 & 0 & 0 \\ 0 & 2 & -6 & 0 & 2 & 0 & 0 \\ 4 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right| \begin{array}{c} j \\ \downarrow \end{array} \end{array}$$

Desmembrando-se a fórmula do cálculo do índice de Moran para a autocorrelação espacial e aplicando a primeira parte, temos o valor fornecido pela matriz $W_{ij} C_{ij}$:

$$\sum_i \sum_j (W_{ij} C_{ij}) = -28$$

Calculando-se o fator de padronização, temos:

$$s^2 = 5,714$$

$$\sum_i \sum_j W_{ij} = 18$$

E agrupando-se as partes já calculadas, temos:

$$I^M = -28 / 5,714 * 18$$

$$I^M = -0,272$$

O coeficiente é positivo quando áreas ou pontos próximos tendem a ser similares em atributos, negativos quando tendem a ser diferentes e aproximadamente zero quando o valor dos atributos está disposto aleatoriamente e independentemente no espaço. Neste exemplo hipotético o valor é negativo significando que os valores dos atributos em pontos próximos tendem a ser diferentes entre si.

- Medida de Associação Espacial ($G_i(d)$)

A medida formal de associação espacial, através da estatística G , nos dá o valor da concentração de um atributo variável, espacialmente distribuído. Ela está baseada na matriz de pesos $W(d)$ determinada pelo raio de distância adotado (lag). Da mesma forma que na autocorrelação, todos os polígonos são representados pelos seus centróides. Esta medida tem como vantagem o fato de que pode ser desagregada por pontos, dando o grau de associação entre um ponto e os demais, e permite o teste de hipóteses sobre a junção de atributos em determinadas localizações. O cálculo é determinado pela seguinte fórmula:

$$G_i(d) = \sum_j w_{ij}(d) x_j / \sum_j x_j \quad [2.14]$$

No exemplo hipotético adotado, toma-se a matriz de pesos de proximidade espacial W_{ij} ,

$$\begin{array}{c}
 i \rightarrow \\
 \left| \begin{array}{ccccccc}
 0 & 1 & 1 & 0 & 1 & 0 & 0 \\
 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
 1 & 0 & 0 & 1 & 0 & 1 & 0 \\
 0 & 1 & 1 & 0 & 1 & 1 & 0 \\
 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
 0 & 0 & 1 & 1 & 0 & 0 & 1 \\
 0 & 0 & 0 & 0 & 0 & 1 & 0
 \end{array} \right| \begin{array}{c}
 j \\
 \downarrow
 \end{array}
 \end{array}$$

e a relação dos atributos em todos os pontos:

ponto	atributo	
1	10	
2	5	$\sum_j X_j$ (atributos) = 42
3	3	
4	8	
5	7	
6	6	
7	3	

multiplica-se a proximidade entre o ponto inicial (i) e final (j) pelo atributo do ponto final (j), dividindo-se o resultado do produto pela soma de todos os atributos no ponto (j). Como resultado temos o valor da associação espacial de cada ponto em relação aos demais. Como resultados temos:

$$G_1(d) = (5+3+7)/42 = 0,357$$

$$G_2(d) = (10+8)/42 = 0,428$$

$$G_3(d) = (10+8+6)/42 = 0,571$$

$$G_4(d) = (5+3+7+6)/42 = 0,500$$

$$G_5(d) = (10+8)/42 = 0,428$$

$$G_6(d) = (3+8+3)/42 = 0,333$$

$$G_7(d) = 6/42 = 0,143$$

Sendo que:

$$0 \leq G_i(d) < 1$$

Se maiores valores dos atributos estão próximos a i os valores de $G_i(d)$ serão maiores, não significando, necessariamente, uma maior associação espacial deste ponto com os demais. Para testar esta hipótese devemos medir os valores esperados e a variância de $G_i(d)$ para $i=1,2,\dots,n$ definidos por cálculos.

Para os valores esperados:

$$\text{Var}(G_i(d)) = [W_i(n-1 - W_i) / (n-1)^2 (n-1)] [((\sum_j X_j^2 / (n-1) - (\sum_j X_j / (n-1))^2) / (\sum_j X_j / (n-1))^2)]$$

Então o valor Z_i para cada ponto pode ser calculado pela seguinte equação:

$$Z_i = [G_i(d) - E(G_i(d))] / \sqrt{E(\text{Var}(G_i(d)))} \quad [2.20]$$

Um valor de z positivo implica em que valores altos do atributo estão espacialmente associado com o ponto ou polígono i , e do mesmo modo para z negativo valores baixos do atributo de j estão associados ao ponto ou polígono i .

2.3.2.2 GIS e AUTOCORRELAÇÃO ESPACIAL

O artigo de Qing Shen (Shen, 1994) toma por base o anterior de Ding, revê o conceito de autocorrelação espacial e avança, em qualidade e exatidão, o processo de elaboração do modelo tendo por região de aplicação a área da Baía de São Francisco, Califórnia, U.S.A.

Segundo o autor, a vantagem do SIG reside no fato de sua capacidade de integrar as características locacionais e os atributos dos objetos espaciais, permitindo “coletar, armazenar, mostrar e manipular” informações sobre o espaço. Por isso, o SIG torna-se uma ferramenta efetiva que possibilita a análise espacial.

Shen refere a larga utilização da autocorrelação espacial como técnica de apreensão da realidade em muitos campos do conhecimento como estudos ambientais, ecológicos e de geografia urbana, contudo, adverte sobre a limitação de sua aplicação para casos em que um grande número de objetos espaciais estão envolvidos, o que é justamente a situação do presente trabalho. A unidade espacial de estudo é o município que, no estado do Rio Grande do Sul, para efeito de estudo de caso, foi configurado com 427 entidades (1996).

O problema detectado por Shen é que a geração da matriz de proximidade espacial torna-se inviável para um número muito expressivo de objetos, especialmente porque envolve múltiplas formas e tamanhos, o que, dependendo da escala e qualidade do mapa base, geraria um processo de duração impraticável, mesmo com o uso de programas de computador e máquinas de alta performance.

A outra questão criticada por Shen e na qual procura avançar, refere-se ao fato de que, no método de Ding, os polígonos são representados por pontos, isto é, centróides (centro geométrico da figura) inclusive para estabelecer a matriz de distâncias que é a base para a determinação da relação espacial entre os objetos. Entretanto, as áreas são completamente diferentes dos pontos que por ventura as representem, implicando em uma redução da realidade nem sempre oportuna, como nos casos de polígonos de área muito grande em relação aos demais ou quando da determinação da matriz de proximidade que, por vezes, seria mais adequada se fosse uma matriz de adjacência.

Shen salienta que nem sempre é verdadeira a premissa de que pontos próximos entre si são mais semelhantes uns aos outros do que aqueles afastados entre si. Além disso aponta a deficiência que alguns estudos poderiam apresentar, caso considerassem a adjacência direta como definidora da matriz de proximidade, o que acarreta na falta de flexibilidade para o estudioso considerar situações em que a adjacência está interrompida por algum obstáculo, como o leito de um rio, por exemplo, e nem por isso os polígonos não adjacentes deixam de apresentar uma relação de proximidade e semelhança.

Há uma diferença fundamental na definição de proximidade locacional em Shen com relação a Ding. O primeiro trabalha com o objeto área diretamente, enquanto o segundo lida com pontos que representam áreas. Shen trabalha com um polígono de abrangência predeterminada que estipula a zona de entorno da área de estudo para fins de estabelecer a matriz de proximidade locacional. Desta forma, o enfoque de Ding pode ser revisto como um caso particular da abordagem de Shen, se for considerado um polígono mínimo como zona de entorno, então a matriz de proximidade se confundiria com a de adjacência resultando em valores finais iguais. Observe-se que o polígono de controle utilizado por Shen pode ser aumentado ou diminuído de acordo com o objetivo do estudo, o que resulta em maior flexibilidade de utilização e melhor capacidade de representação da realidade em estudo. Para exemplificar o modo como este conceito de proximidade é utilizado, Shen cria alguns polígonos fictícios e demonstra o que significa e como funciona seu método⁹.

⁹ No artigo "SIG e Autocorrelação espacial" o exemplo é apresentado com todos os detalhes dos comandos computacionais empregados na geração da Matriz de proximidade pelo uso do programa AML (ARC Macro Language) do ARC/INFO associado a um programa específico em FORTRAN desenvolvido por Shen e apresentados ao final como apêndices.

2.3.2.3 AUTOCORRELAÇÃO ESPACIAL, INTEGRAÇÃO ESPACIAL E MEDIDA DE SEGREGAÇÃO: ONDE A SOCIOLOGIA ENCONTRA A ANÁLISE ESTATÍSTICA ESPACIAL E A GEOGRAFIA.

A questão da proximidade espacial é retomada e mais profundamente explorada em um trabalho bastante recente de Lee (1998) em que é proposto novo índice de medida da segregação espacial. O autor baseia-se no conceito de contigüidade com origem na sociologia, Geary (1954) e na geografia, Dacey (1968).

Contigüidade é, assim, definida como a medida de similaridade com relação à concentração de um dado conteúdo por unidade de área em comparação com as unidades de áreas adjacentes. Esta característica de contigüidade da distribuição espacial no âmbito da análise estatística espacial é conhecida como autocorrelação espacial.

Medidas de contigüidade são interpretadas como índices de agrupamento modificados. Mesmo que, eventualmente, um alto grau de agrupamento possa não ter correspondência direta com um alto grau de contigüidade. Assim, um novo índice de medida do grau de agrupamento, diferente dos índices de segregação, é necessário (Lee, op. cit.).

Lee define três atributos a partir dos quais gera seu índice de agrupamento: tamanho total do agrupamento, suas formas e o entorno imediato. O tamanho da área agrupada não é tão importante quanto o número de pessoas que habitam a área, em se tratando de aglomerados urbanos.

Estando determinada a população em cada extrato ou setor censitário, há que determinar a concentração do grupo objeto de estudo. Utiliza-se para tanto o **quociente de localização, Q^l** . Se Q^l for maior que 1 ($Q^l > 1$) significa que a localidade tem uma concentração alta de indivíduos em relação às outras localidades da região.

Para quantificar o tamanho, a forma e o entorno do agrupamento, a medida indicada é o **perímetro total, P** . O índice de agrupamento de Lee é determinado como:

$$I^c = 1 - \frac{\sum_i \sum_j |I_i - I_j| b_{ij}}{\sum_i \sum_j b_{ij}} \quad [2.21]$$

onde I_i é o valor binário do setor censitário (unidade de área censitária) e será 1 (um) se $Q^i \geq 1$ e 0 (zero) se $Q^i < 1$, portanto alta e baixa concentração respectivamente;

b_{ij} é o comprimento do limite comum entre o setor censitário i e j , logo se i não for adjacente a j , então $b_{ij} = 0$.

Observando a fórmula 2.21 pode-se verificar que quando os setores censitários tiverem o mesmo valor binário, isto é, ambos de alta ou baixa concentração o termo superior fica zerado, o que significa dizer que o limite entre eles desaparece.

Lee faz um estudo comparativo utilizando os cinco municípios da cidade de Nova York para três atributos: população de cor negra, taxas de pobreza e população sem-teto. Para realizar esta comparação ele calcula cinco índices de segregação a saber: o índice de dissimilaridade (I^d), o índice de proximidade espacial de White (I^{sp}), o índice de Moran (I^M), o índice de dissimilaridade modificado de Wong (I^{md}) e o índice de agrupamento (I^c).

A partir dos resultados obtidos para estes cinco índices aplicados sobre os cinco municípios de N.Y. para os três atributos, é possível apresentar os resultados tabulados por Lee em forma de gráficos, modo de apresentação dos resultados que propicia um outro tipo de leitura diferenciada em relação ao dados simplesmente tabulares.

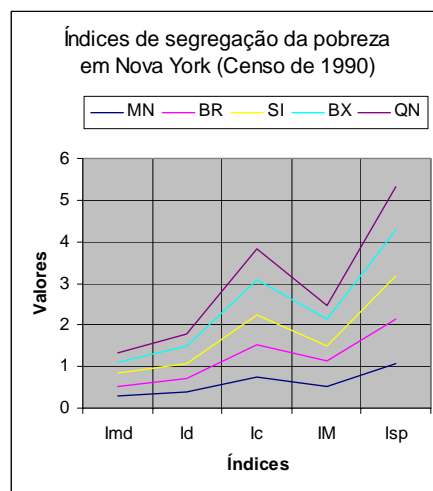


Figura 2.6 – Índices de segregação da pobreza em Nova York (MN, Manhattan; BR, Brooklyn; SI, Staten Island; BX, Bronx; QN, Queens).

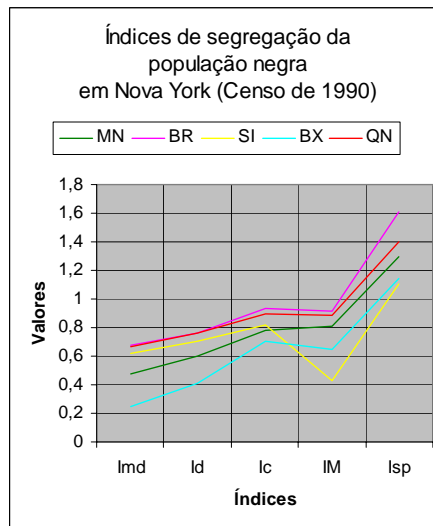


Figura 2.7 - Índices de segregação da população negra em Nova York (MN, Manhattan; BR, Brooklyn; SI, Staten Island; BX, Bronx; QN, Queens)

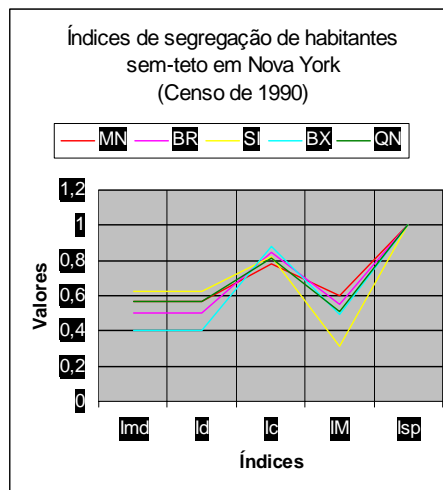


Figura 2.8 - Índices de segregação de habitantes sem-teto em Nova York (MN, Manhattan; BR, Brooklyn; SI, Staten Island; BX, Bronx; QN, Queens).

Observa-se nos gráficos acima que o comportamento dos cinco índices quando representam o fenômeno de segregação da pobreza (figura 2.6) têm um comportamento bastante similar entre si, delineando uma gradação bem definida em cada um dos cinco distritos, o que conduziria à interpretação de que o fenômeno já está consolidado em N.Y., uma vez que todos os índices acusam igual tendência.

Já a segregação negra, representado na figura 2.7, teria o mesmo comportamento não fosse o distrito de Staten Island que para os índices de Moran e White apresenta uma variação diferenciada em relação à tendência esperada.

Com relação ao último fenômeno estudado por Lee, segregação dos sem teto, cabe observar que a partir do gráfico (figura 2.8) pode-se dizer que os índice de White perde a capacidade de diferenciação enquanto que o índice de agrupamento apresenta valores muito próximos para todos os distritos.

Em geral é possível observar que estes três fenômenos estudados são altamente correlacionados, especialmente segregação negra e sem-teto, e quando da observação dos mapas de agrupamento gerados por Lee, verifica-se a vasta sobreposição de áreas, indicando com isso que a pobreza, os sem-teto e os negros são, em grande medida características de uma mesma população marginalizada.

Os índices de Moran [2.12] e de Lee [2.21] já foram apresentados, resta a formulação dos demais:

O índice de dissimilaridade (I^d),

$$I^d = \sum_{i=1}^n \left[\frac{t_i |r_i - R|}{2PR(1-R)} \right] \quad [2.22]$$

onde

t_i é a população da unidade de área i ;

r_i é a proporção do grupo na unidade de área i ;

P é a população total;

R é a proporção do grupo na região completa;

n é o número de unidades de área.

O índice de proximidade espacial de White (I^{sp}),

$$I^{sp} = \frac{1}{TB_{tt}} (XB_{xx} + YB_{yy}) \quad [2.23]$$

e

$$B^X = \sum_{i=1}^n \frac{1}{\chi^2} x_i x_j c_{ij} \quad [2.24]$$

onde

B_{xx} , B_{yy} e B_{tt} são as proximidades médias entre os membros X, entre os Y e entre todos os membros T da população;

X, Y e T são o total dos membros de X, Y e T;

x_i é o número de membros de X nas unidades de área i;

x_j é o número de membros de X nas unidades de área j;

c_{ij} é a proximidade entre as unidades de área i e j.

Por fim, o índice de dissimilaridade modificado de Wong (I^{md})

$$I^{md} = I^d - \frac{1}{2} \sum_i \sum_j w_{ij} |z_i - z_j| \quad [2.25]$$

e

$$w_{ij} = \frac{d_{ij}}{\sum_i \sum_j d_{ij}} \quad [2.26]$$

onde

z_i é a proporção de um grupo na unidade de área i;

d_{ij} é o comprimento do limite comum entre a unidade de área i e j.

2.3.2.4 TAXONOMIA REGIONAL: MÉTODO DE ANÁLISE DE AGRUPAMENTO

Conforme já enunciou Fischer (1979) em seu estudo comparativo sobre metodologias de classificação regional, sob o enfoque da teoria dos sistemas, uma região pode ser vista como um sistema em que um agrupamento de unidades espaciais básicas é caracterizado por um conjunto de k atributos sendo $k \geq 1$. Quando este agrupamento apresenta contigüidade espacial e um alto grau de correspondência entre os atributos selecionados diz-se tratar de uma região homogênea.

Em última instância, pode-se reduzir o problema de regionalização a um procedimento de classificação e desta forma afeito à ciência que estuda os métodos classificatórios, que é a taxonomia (Haddad, 1997).

A taxonomia numérica seria a ciência da classificação a partir da aplicação de técnicas de análise matemática. Neste contexto, coloca-se como importante o método da análise de agrupamento (*cluster analysis*).

Supondo a existência do conjunto de objetos $O = \{O_1, O_2, \dots, O_n\}$, que são unidades espaciais caracterizadas pelo conjunto de atributos $A = \{A_1, A_2, \dots, A_p\}$, em que cada elemento X_{ij} é o valor assumido por cada atributo A_j de um objeto O_i , pode-se escrever a matriz abaixo (Fischer, op.cit.):

$$X_{ij} = \begin{array}{c} \begin{array}{c} X_{11} \quad X_{21} \dots X_{ij} \dots X_{1p} \\ X_{21} \quad X_{22} \dots X_{2j} \dots X_{2p} \\ \vdots \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \\ X_{k1} \quad X_{k2} \dots X_{kj} \dots X_{kp} \\ \vdots \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \\ X_{n1} \quad X_{n2} \dots X_{nj} \dots X_{np} \end{array} \\ \downarrow \\ \text{n objetos} \end{array} \begin{array}{c} \longrightarrow \\ \text{p atributos} \end{array}$$

Uma partição desta matriz X_{ij} de objetos ou unidades básicas, que satisfaça o princípio da homogeneidade interna e/ou o princípio da separação externa será considerada uma região ou taxon.

- Princípio da homogeneidade interna: os objetos componentes da região devem ser tão homogêneos quanto possível espacialmente em relação a atributos específicos.
- Princípio da separação externa: os diferentes objetos regionais devem ser espacialmente o mais distinto possível com relação aos atributos específicos.

2.3.2.5 ANÁLISE REGIONAL

Desde o final da década de 60 e especialmente durante os anos 70, quando prevaleceu o pensamento dito positivista de abordagem do planejamento urbano e regional, são feitas investidas na direção de obter-se maior rigor matemático na solução de problemas que envolvam critérios múltiplos. De forma a procurar revestir de maior objetividade o papel dos planejadores, metodologias de enfoque quantitativo foram criadas.

É a partir deste tipo de abordagem e utilizando estes estudos como base que se procura recriar o presente modelo.

São exemplos desta linha de estudos o artigo de Roy publicado em 1974 e o de Eastin de 1977. No primeiro, é introduzido o conceito de **raio de termo médio** (r_i) que é uma medida de performance de cada critério particular em alguma unidade apropriada, varia entre 0 e 1, sendo 1 correspondente a melhor performance, portanto, seria o valor máximo de r_i e 0 seria o valor mínimo ou a pior performance do critério (Roy, 1974). Como meio de introduzir as preferências gerenciais dos planejadores e pesquisadores, Roy criou um conjunto de **ponderadores relativos** que expressam a performance relativa de cada critério. No segundo artigo, Eastin utiliza estes dois conceitos de Roy e faz uma aproximação mais genérica da questão buscando com isso que, ao invés de resultados de performance absoluta, sejam obtidas metas de satisfação relativa, forçando os planejadores a ordenar, por prioridades, as diversas alternativas (Eastin, 1977).

Roy dicotomiza o conjunto de soluções para problemas de planejamento em eficientes e ineficientes. Define eficiente como sendo o conjunto de raios de termo médio R , cujo elemento típico r_i não encontra além dele nenhuma outra solução viável, o que significa dizer que não há nenhum outro elemento maior ou igual a r_i . R será ineficiente se existir um conjunto R^* cujos elementos $r_i^* \geq r_i$ ($i = 1, \dots, n$) pertencerem a R (Roy, op. cit.).

O autor estabelece uma condição ideal de otimização que cria uma equivalência entre as relações de raios com pesos (w_i) como segue:

$$\frac{r_1}{w_1} = \frac{r_2}{w_2} = \dots = \frac{r_i}{w_i} = \dots = \frac{r_n}{w_n} \quad [2.27]$$

Então, Roy define os conjuntos R' localizados sobre o plano P definido pelas retas $(0,0) - (0,1)$ e $(0,0) - (1,1)$. Chegando a demonstrar que o conjunto $(R')^o$ representa o conjunto de soluções ideais e que quanto menor for a distância entre o elemento de R' e de $(R')^o$, mais próximos do plano ótimo estaremos, idealmente $D = 0$ quando $R' = (R')^o$ onde

$$(R')^o = \left\{ r'_j / r'_j = \frac{w_j}{\sum_{i=1}^n w_i} \right\} \quad [2.28]$$

de modo que $r'_i \in R'$ e $(r'_i)^o \in (R')^o$, portanto resulta:

$$D = \left[\sum_{i=1}^n \left(\frac{r_i}{\sum_{j=1}^n r_j} - \frac{w_j}{\sum_{j=1}^n w_j} \right)^2 \right]^{1/2}. \quad [2.29]$$

Já Eastin utilizando os conceitos de raio de termo médio e pesos relativos descreve um processo de maximização da distância dando prioridade a uma possível função de utilidade U^T onde D^{\max} é a maior distância entre um ponto e $(R')^0$ rescrevendo a função de Roy em termos de maximização $[D^{\max} - D(w_i, r_i)]$, na qual $D(w_i, r_i)$ assume a forma da expressão 2.14 e o modelo a fórmula geral que segue:

$$U = (1-\alpha) \sum_{i=1}^n u_i r_i + \alpha [D^{\max} - D(w_i, r_i)] \quad [2.30]$$

onde o parâmetro α varia entre 0 e 1 e representa a importância relativa entre o critério de Roy e a função utilidade total (Easting, 1977).

Estes dois exemplos poderiam responder a questão relacionada ao conflito existente entre o objetivo de eficiência e o de uniformidade, ambos desejáveis em sistemas de regionalização administrativa.

2.3.2.6 MODELOS DE LOCAÇÃO-ALOCAÇÃO

Como as variáveis de decisão tanto nos modelos de localização e alocação como nos de regionalização administrativa são as mesmas, convém aplicar a mesma sistemática dos primeiros nos segundos visando a solucionar o problema de divisão regional administrativa (Lolonis op. cit.), a partir das seguintes premissas:

- a área de estudo é uma partição contínua de unidades e sem sobreposição, que possui um centro administrativo e limites fixos (no RS seriam os municípios e suas sedes municipais);
- cada unidade administrativa é representada como um ponto de demanda;
- a demanda de cada ponto é representada pela população da unidade administrativa;
- o custo de oferta de serviços cresce com a distância que separa um ponto de demanda de seu centro;
- a alocação de um ponto a um centro depende de sua similaridade a este centro.

A dissimilaridade espacial é um tipo de fricção ou efeito de atrito do espaço que aumenta o custo de alocação da demanda aos centros regionais, quanto mais dissimilar é um ponto de demanda em relação ao seu centro com respeito a características

socioeconômicas e físicas, maior a dificuldade do ponto ser alocado ao centro. A medida deste acréscimo de custo é determinada pelo coeficiente de diferenciação espacial (CDE) que define o grau de dissimilaridade entre dois pontos. Multiplicando-se o CDE pela distância atual entre dois pontos i e j obtém-se a distância transformada (DT) que é uma medida do custo total de alocação de uma unidade de demanda do ponto i para o ponto j (Lolonis op. cit.).

Pode-se, então, definir uma função Z que representa o custo total de alocação de todos os pontos de demanda aos centros regionais. Em termos de regionalização administrativa o problema se coloca como determinação do menor custo:

$$\min Z = \sum_{i=1}^n \left(\sum_{j=1}^n w_i c_{ij} d_{ij} X_{ij} \right) \quad [2.31]$$

onde

Z é o valor da função;

n é o número de pontos de demanda;

w_i peso de demanda do ponto i ;

c_{ij} o coeficiente de diferenciação espacial entre os pontos i e j ;

d_{ij} a menor distância entre i e j ;

X_{ij} variável binária (0,1), se i é ponto alocado para o centro j , $X_{ij} = 1$, senão $X_{ij} = 0$.

A adjacência regional S_{ij} determina se as unidades espaciais i e j estão na mesma região contígua, será 1 se as unidades i e j estiverem na mesma região contígua e 0 (zero) caso contrário (Lolonis, op. cit.). A matriz de adjacência é bidimensional e binária, se o ponto de demanda i for adjacente ao ponto j então o elemento a_{ij} é 1. Como todo ponto de demanda é adjacente a si próprio significa dizer que a diagonal principal da matriz tem elementos iguais a 1. A expressão para S_{ij} é:

$$S_{ij} = \max \left\{ \begin{array}{l} a_{ij} X_{ik} X_{jk}, k = 1, \dots, n \\ a_{ik} S_{kj} X_{il} X_{kl} X_{jl}, k, l = 1, \dots, n \end{array} \right\} \quad [2.32]$$

onde

S_{ij} adjacência regional da unidade espacial i e j ;

X_{ij} variável binária (0,1), se i é ponto alocado para o centro j $X_{ij} = 1$, senão $X_{ij} = 0$;

n é o número de pontos de demanda;

a_{ij} elemento da matriz de adjacência.

2.3 ESTRUTURA DE DADOS CENSITÁRIOS

O Censo Demográfico Brasileiro de 1991 apresentou várias novidades, em termos de métodos aplicados, em relação ao anterior de 1980 (Oliveira, et al., 1996). O IBGE realizou uma parte da apuração dos dados descentralizada em 20 unidades da Federação, utilizou a codificação assistida por computador, aplicou o sistema DIA de Detecção e Imputação Automática de erros para dados qualitativos, fez a redução da amostra com a utilização de duas frações amostrais distintas: de 1/5 para municípios de até 15.000 habitantes e de 1/10 para municípios acima de 15.000 habitantes (Silva, 1990) e fez a expansão dos dados da amostra através de uma metodologia de estimativa denominada GLSEP, *Generalized Least Square Estimation Procedure*, também chamada Mínimos Quadrados Generalizados (MQG2), utilizada no Censo Canadense de 1991.

Analisando o questionário da amostra verifica-se que, após a identificação geral do local e do informante, vem o primeiro bloco de informações que caracterizam o domicílio, a seguir, informações pessoais sobre o chefe da família e uma página para cada uma das demais pessoas do domicílio. Para cada uma destas pessoas existe uma bateria de questões que visa caracterizar os níveis de escolaridade, fluxos migratórios, fecundidade e natalidade, mercado de trabalho, mão de obra e níveis de renda.

Com base na tabela 2.2 abaixo, que é extraída de uma tabela maior onde aparece a distribuição dos bons e maus registros para o Brasil e Regiões (Oliveira op.cit., p.62), depreende-se quais os nichos de dados mais confiáveis com vistas a uma possível aplicação no presente trabalho.

Supondo que até 5% de registros maus seja um nível correspondente à alta confiabilidade dos dados, de 5 a 15% média confiabilidade e acima de 15% baixa confiabilidade, então os dados censitários que dizem respeito à Educação e Migração apresentam o maior grau de confiabilidade. As informações relativas à Mão de obra apresenta dois subconjuntos, sendo o primeiro com alta incidência de bons registros e o segundo com alta incidência de maus registros. O Domicílio e as pessoas que não o Chefe de família têm níveis altos de maus registros e para o Chefe de família os dados tem média incidência de maus registros.

Tabela 2.2 – Distribuição dos registros bons e maus, para Brasil e Região Sul, segundo as aplicações

Aplicações		Brasil		Sul	
		abs	rel	abs	rel
Dom102	Bons	3.058.766	75,99	574.078	84,06
Aplic.1	Maus	966.327	24,01	108.828	15,94
Pchef2	Bons	3.896.910	89,48	667.406	91,38
Aplic.1	Maus	458.154	10,52	62.958	8,62
Nchef2	Bons	7.387.774	85,25	1.195.189	87,41
Aplic.2	Maus	1.278.027	14,75	172.077	12,59
Educ.09	Bons	4.041.805	97,75	569.881	98,38
Aplic.3	Maus	92.956	2,25	9.412	1,62
Migra1	Bons	15.716.631	92,63	2.473.687	92,41
Aplic.4	Maus	1.249.811	7,37	203.236	7,59
Migra2	Bons	16.713.682	98,05	2.629.117	98,21
Aplic.5	Maus	332.030	1,95	47.806	1,79
Mobra1	Bons	12.460.331	96,13	2.000.951	96,63
Aplic.6	Maus	501.520	3,87	69.679	3,37
Mobra2	Bons	10.607.819	81,67	1.744.777	83,18
Aplic.7	Maus	2.381.032	18,33	352.853	16,82

Fonte: IBGE, Textos para discussão, n.86, p.62.

Censo 91 – Resultados da Correção Automática do CD 1.02.

Contudo, o volume de dados fornecidos pelo IBGE, com origem em outras fontes que não a censitária, é imenso e também poderão ser consideradas como elementos de alimentação do presente modelo na busca de divisões regionais criteriosas.

São exemplos de fonte de dados todos aqueles oferecidos pelo Sistema IBGE de Recuperação Automática (SIDRA) disponíveis no site do IBGE, cujo endereço é <http://www.sidra.ibge.gov.br>. Cita-se a Pesquisa de Orçamentos Familiares, o Censo Agropecuário 1995-1996, a Contagem da População de 1996, a Pesquisa Mensal de Emprego e a Pesquisa Industrial Mensal, entre outros, todos disponíveis via Internet.

2.5 GEOPROCESSAMENTO APLICADO À ANÁLISE ESPACIAL

2.5.1 NOÇÕES GERAIS SOBRE GEOPROCESSAMENTO

Geoprocessamento pode ser definido como o conjunto de técnicas e metodologias que implicam no arquivamento, processamento, representação e visualização de dados georreferenciados. Dados georreferenciados são aqueles que se referem a localizações na superfície da terra determinadas por um sistema padrão de coordenadas. O sistema de coordenadas pode ser de uso local, nacional ou internacional como o sistema de projeção UTM (*Universal Transverse Mercator Coordinate System*) ou o sistema tradicional de Latitude / longitude relativo a Greenwich.

O Geoprocessamento, como pode ser previsto pela própria palavra composta, geo + processamento, engloba o uso de técnicas computacionais e matemáticas para o

tratamento da informação geográfica. Naturalmente que a informação geográfica só poderá ser tratada adequadamente se reduzida e conformada de modo a ser processada como dado computacional. Para tanto, seguindo a praxe da topologia, a informação geográfica é representada por três conceitos básicos: o ponto, a linha e a área (Burrough, 1986).

A informação geográfica será, então, um conjunto de coordenadas que definem a posição do evento geográfico, um conjunto de atributos que caracterizam o evento em termos de qualidade e valores, e um conjunto de inter-relações espaciais que definem suas conexões topológicas (Leão, 1997). Desta forma, um mapa ou carta, poderá ser descrito como um conjunto de pontos, linhas e áreas definidos por suas características locacionais, referidas pelo sistema de coordenadas, e por seus atributos não espaciais (Burrough, op.cit.).

2.5.2 O PROCESSO DE MODELAGEM

A transformação da realidade concreta tal como se apresenta para um outro universo em que seja possível a sua compreensão científica chama-se processo de modelagem, cujo objetivo é a criação de um modelo de dados.

Modelo de dados é uma representação da realidade através de conceitos matemáticos, formas geométricas e algoritmos de tratamento dos dados. Particularmente, em relação à realidade geográfica tem-se: o universo do mundo real onde se processam os fenômenos espaciais (tipo de vegetação, topografia, hidrografia, clima, etc); o universo conceitual com duas classes de dados geográficos (contínuos e discretos); o universo de representação onde as entidades formais são mapeadas por diferentes representações geométricas que variam em função da projeção cartográfica, escala, período de aquisição do dado, etc; e o universo de implementação onde se executa o modelo de dados via linguagens de programação (Câmara, 1999).

2.5.3 O UNIVERSO DO MUNDO REAL

2.5.3.1 MAPAS TEMÁTICOS

Mapas temáticos apresentam a distribuição espacial de um dado geográfico de modo qualitativo. Um mapa temático pode ser obtido de dados de levantamento de campo ou a partir de imagens de satélite. Os primeiros são digitalizados para serem incorporados

ao sistema e os segundos são submetidos a um processo de classificação de imagens a partir de níveis de cinza que revelam a assinatura espectral de cada objeto detectado.

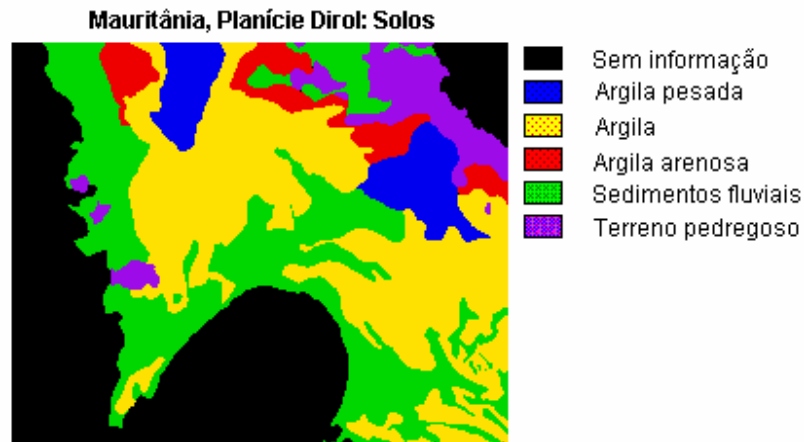


Figura 2.9 – Exemplo de mapa temático: classificação de solos da Planície Dirol (Mauritânia)
(Fonte: Eastman, 1998)

2.5.3.2 MAPAS CADASTRAIS

Mapa cadastral é um conjunto de objetos geográficos que possuem atributos armazenados por um sistema gerenciador de banco de dados (SGBD).

A diferença fundamental entre mapas temáticos e cadastrais é que nos primeiros os limites não são precisamente definidos enquanto nos segundos as informações para os diversos atributos são exatas (Câmara, op.cit.).

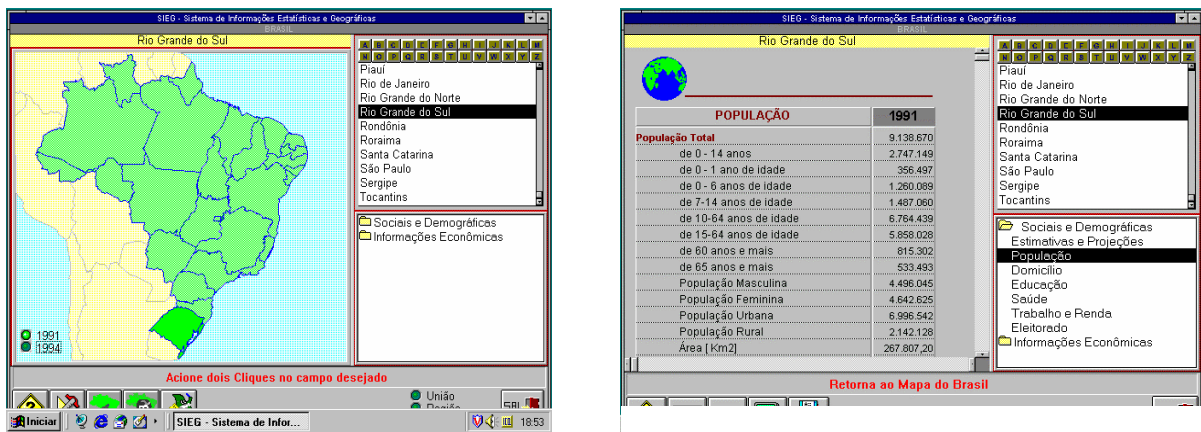


Figura 2.10 – Exemplo de mapa cadastral extraído do Sistema de Informações Estatísticas e Geográficas (SIEG) do IBGE (versão beta II de demonstração).
(Fonte: IBGE, 1998)

2.5.3.3 MODELOS NUMÉRICOS DE TERRENO

Modelo numérico de terreno ou MNT é um termo genérico que representa um modelo matemático definido por um conjunto de amostras (x,y,z) onde x e y são as coordenadas de localização espacial e z o valor do atributo em cada ponto do sistema, determinando uma superfície contínua no espaço.

2.5.3.4 REDES E IMAGENS

Complementando os dados encontrados no universo do mundo real apresenta-se os conceitos de redes e de imagens.

Redes são sistemas gráficos de topologia arco-nó onde as informações são armazenadas em coordenadas vetoriais que fornecem o sentido de fluxo, a localização geográfica e os atributos associados a cada arco e nó, constituindo um grafo (Câmara, op.cit.).

Imagens são informações espaciais obtidas por satélites, fotos aéreas ou sensores remotos multiespectrais armazenadas como matrizes de pixels (*picture element*) que tem um valor correspondente à energia eletromagnética refletida por uma dada área da superfície terrestre em cada ponto.

2.5.4 O UNIVERSO CONCEITUAL

O espaço geográfico pode ser representado por campos e objetos. Como campo, considera-se o espaço geográfico um plano contínuo e os fenômenos variáveis e distribuídos, de modo que a cada ponto corresponda uma característica ou atributo. Como objeto, o espaço geográfico é considerado um conjunto de entidades com atributos que as caracterizam e diferenciam (Câmara, op.cit.).

Relacionando estes conceitos com o presente modelo de regionalização pode-se rescrever o conceito de região como um conjunto de objetos homogêneos cujos atributos estão teoricamente concentrados na sede municipal, para efeito de tratamento via análise estatística.

Outras definições encontráveis neste universo e, algumas, utilizadas neste trabalho são (Câmara, op.cit.):

- Geo-campo é a representação da distribuição espacial de uma variável que possui valores em cada ponto, dentro de uma região, atrelados a um tempo 't'.
- Geo-objeto é um elemento único que possui atributos não-espaciais e está associado a múltiplas localizações bem determinadas.
- Objeto não-espacial é aquele que não possui localizações espaciais associadas.
- Mapa cadastral é um conjunto de representações geométricas de geo-objetos amarrados a uma dada projeção cartográfica.
- Banco de dados geográfico é uma forma de armazenamento de informações que associa, de modo recuperável, um conjunto de geo-campos ou objetos cadastrais com conjuntos de geo-objetos e objetos não-cadastrais.

A maneira como estas informações geográficas, que compõe o universo conceitual, são representadas difere em duas técnicas básicas de representação: a vetorial e a matricial ou raster. Estas duas formas de representação fazem parte do próximo universo a ser focado que será apresentado a seguir: o universo de representação.

2.5.5 O UNIVERSO DE REPRESENTAÇÃO

Na forma de representação matricial, o espaço é uma matriz $M(i, j)$ de i linhas por j colunas, onde cada célula é endereçada pelas suas coordenadas e contém um valor do atributo na célula (Câmara, op.cit.). A resolução do mapa é uma função do tamanho de cada célula em relação à cobertura que ela abrange no terreno.

A representação vetorial é uma composição de três elementos topológicos básicos: o ponto, a linha e a área. O ponto refere-se à entidade geográfica que é definida por um par de coordenadas (x, y) . A linha é um conjunto de pontos conectados composta por todos os seus pares de coordenadas. A área é formada por conjuntos de linhas e os atributos que as relacionam com o entorno.

Segundo Burrough (1986), embora as duas técnicas de representação sejam contrastantes entre si, porque a raster seria a representação explícita do dado espacial enquanto que a vetorial seria a representação implícita, ambas são complementares.

A forma matricial ou raster de representação das entidades espaciais integra em um único arquivo a solução gráfica e os seus atributos. Como pode ser observado abaixo, a realidade é particionada em um conjunto de células que compõem uma grade. Pelo

armazenamento diferenciado de informação ou atributos em cada célula, seja por um valor numérico ou escala de cinzas, é possível recompor a realidade (Figura 2.11). A forma vetorial utiliza um conjunto de linhas definidas por um ponto de início e um ponto de fim, submetidos a um sistema de coordenadas, criando vetores que representam a forma da área considerada. Também é necessário algum tipo de indicador que determine como se dá a conexão entre os diferentes vetores nos vários vértices. É isto o que significa a conectividade.

Os mapas temáticos podem ser representados tanto de modo matricial como vetorial. Quando a exatidão é um elemento significativo do modelo a representação vetorial é mais indicada, mas se o que se pretende é um cruzamento entre mapas diretamente através da álgebra de mapas a forma matricial é mais adequada. Contudo a área de armazenamento em termos de tamanho de arquivos é muito maior na matricial do que na vetorial.

Um exemplo mais próximo da realidade geográfica é tomado do Manual do Usuário do IDRISI¹⁰ para Windows (versão 2) (Eastman, 1998):

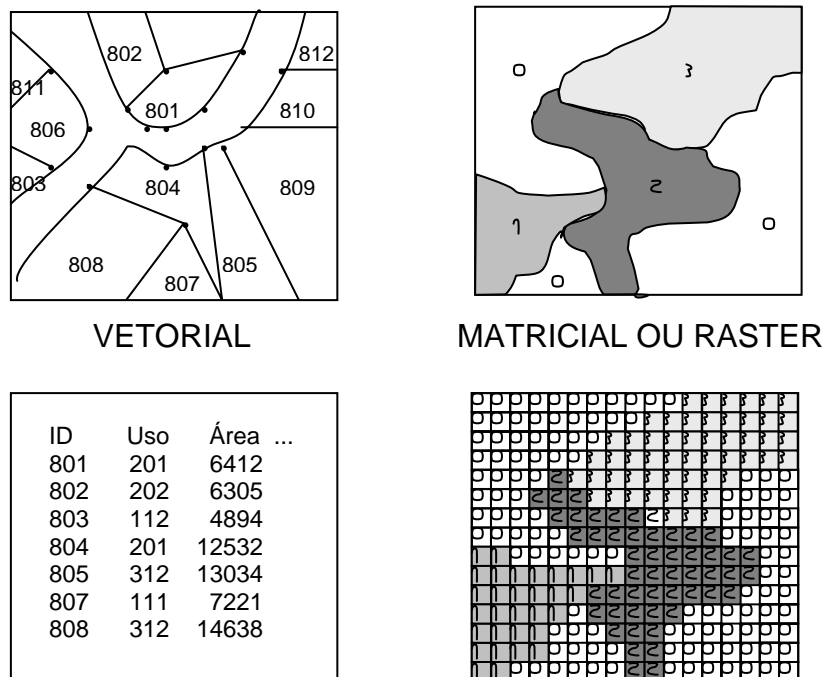


Figura 2.11 – Representação de dados nos formatos vetorial e matricial ou raster
(Fonte: Adaptado de Eastman, 1988)

¹⁰ IDRISI é um Sistema de Informação Geográfica desenvolvido pela Clark University.

Uma mesma região pode ser representada pelos dois tipos de representação. A escolha da forma de representação mais adequada deve levar em consideração quais aspectos são mais importantes em cada caso.

Basicamente, este trabalho faz uso da representação vetorial em que cada município é representado por um polígono fechado, a sede municipal é um ponto de coordenadas (x, y) e as regiões são conjuntos de polígonos. Os atributos de cada município estão teoricamente concentrados em suas sedes para efeito dos cálculos de distância, das relações de proximidade e coeficientes de correlação.

Quanto aos modelos numéricos de terreno há duas formas de representação que são: grade regular e grade triangular.

A grade regular é uma representação matricial que associa a cada par de coordenadas uma cota ou valor de elevação $(x, y \rightarrow z)$, sendo que para os pontos em que não há valor medido, são calculados valores através de processos de interpolação, tendo por base os valores dos vizinhos conhecidos.

A grade triangular é uma representação vetorial que gera uma superfície formada pela justaposição de triângulos cujos vértices possuem três coordenadas (x, y, z) sendo x e y de localização e z de atributo (Câmara, op. cit.).

Ainda caberia citar o universo de implementação, para completar o paradigma dos quatro universos (Câmara, op. cit.), que diz respeito ao conjunto de opções feitas para viabilizar um dado estudo na área de geoprocessamento e que é basicamente a escolha dos algoritmos de tratamento de dados geográficos, estrutura dos dados e desempenho do hardware.

2.5.6 SISTEMAS DE INFORMAÇÃO GEOGRÁFICA

Um Sistema de Informação Geográfica (SIG) é um *software* composto por um conjunto de rotinas, procedimentos e funções automatizadas que permitem, o armazenamento, o processamento, a análise, a representação e a visualização de dados geográficos. Para cada um dos processos que compõe um SIG, existe um módulo ou subsistema responsável pela execução das tarefas, integrados entre si. Então, os componentes de um SIG são: Banco de Dados espaciais e de atributos, Sistema de gerenciamento de banco de dados, Sistema de digitalização de mapas, Sistema de

visualização cartográfica, Sistema de processamento de imagens, Sistema de análise geográfica, Sistema de análise estatística e Sistema de apoio à decisão.

O Banco de Dados é uma área de armazenamento do sistema que é composta por uma coleção de mapas e informações associadas gravadas na forma digital. O Banco de Dados de um SIG apresenta duas gamas de informações: um banco de dados espaciais sobre a superfície terrestre, que descreve forma e posição geográficas e outro com dados de atributos que caracterizam e qualificam esta superfície. O Sistema de Gerenciamento de Banco de Dados (SGBD) de um SIG, além das tarefas usuais de entrada, gerenciamento e análise de dados de atributos, também gerencia os dados espaciais através de utilitários específicos para este fim. A característica especial de um SGBD em um SIG é sua capacidade de análise de dados de atributos e a partir daí gerar uma saída com informação espacial. Por exemplo, é possível, via SGBD, utilizando um banco de dados de atributos que contenha a renda municipal e a população economicamente ativa (PEA) de cada município de um estado, gerar um mapa que identifique apenas aqueles municípios com renda municipal maior do que um valor médio determinado e ainda, criar classes de municípios que associem a PEA com níveis de renda municipal, gerando um novo mapa que, a partir da análise de dados de atributo, gera informações espacializadas, mas não exatamente geográficas.

Os sistemas de digitalização de mapas, em geral, são programas compatíveis com o SIG ou são módulos oferecidos opcionalmente com o SIG. É uma forma de entrada de dados gráficos no banco de dados espaciais. Trata-se do sistema que permite a transferência de mapas impressos para o formato digital, através do uso de um cursor que percorre o mapa sobre uma mesa digitalizadora e registra sucessivos pontos componentes das linhas do mapa. A mesa é dotada de uma malha eletromagnética que a cada toque do cursor transfere as coordenadas do ponto, inserido em um sistema de coordenadas, para uma tela e daí para um arquivo, naturalmente que sob comando de um programa próprio de digitalização.

O Sistema de visualização cartográfica possibilita a saída de mapas na tela, na impressora, no *plotter* ou em arquivos para posterior impressão ou visualização. Os SIG mais sofisticados permitem uma editoração dos mapas bastante qualificada resultando em soluções gráficas bem elaboradas ou permitindo a gravação em formatos compatíveis com *software* de editoração gráfica.

O Sistema de processamento de imagens permite a análise de imagens obtidas por sensoriamento remoto através de rotinas de análises estatísticas específicas. Em sensoriamento remoto existem duas categorias de dados: os espectrais e os de formação de imagem (Swain e Davis, 1978). É sobre este segundo tipo de sistema de formação de imagens que o SIG atua e por processamento é capaz de convertê-las em dados através de um processo de classificação.

O Sistema de análise geográfica parece ser o cerne de um SIG e o seu potencial de aplicação é capaz de distinguir um SIG verdadeiramente eficiente de um outro ineficiente. Porque é através do sistema de análise geográfica que o SIG é capaz de realizar operações e relações de dados baseados em sua posição geográfica, por um processo de sobreposição de mapas pode-se efetuar operações lógicas e até mesmo desenvolver-se uma Álgebra com mapas, ou seja, o uso de imagens como variáveis em operações aritméticas normais.

O Sistema de análise estatística de um SIG em geral oferece apenas os procedimentos estatísticos mais usuais e algumas rotinas específicas para a descrição de dados espaciais. Esta deficiência tem sido uma lacuna bastante importante nos SIG que, no presente modelo, procura ser compensada pelo desenvolvimento e aplicação de uma metodologia de análise baseada em processos estatísticos mais aprofundados.

O Sistema de apoio à decisão em um SIG diz respeito a módulos que determinam o erro no processo, constrói mapas de aptidão aplicando critérios múltiplos e definem melhor localização quando múltiplos objetivos estão em jogo.



CAPÍTULO 3

METODOLOGIA

3. METODOLOGIA

O objetivo deste capítulo é demonstrar de que forma será possível representar a realidade regional com vistas a estabelecer similaridades e disparidades quantificáveis, passíveis de uma classificação e definidas em grupos de entidades homogêneas no espaço ou, independentes e não homogêneas. Para tanto, é de fundamental importância, nesta abordagem, o vínculo que se possa estabelecer entre um dado atributo e sua exata localização no espaço.

Procura-se analisar as relações entre as características socioeconômicas dos municípios do estado do Rio Grande do Sul e sua estrutura geográfica. Isto significa dizer que as estratégias aqui apresentadas investigam as relações entre os padrões espaciais e os padrões dos dados.

Para tornar claro o caminho que será percorrido apresenta-se a seguir um quadro metodológico, figura 3.1, que expõe quatro blocos de desenvolvimento: obtenção dos dados, análise estatística multivariada, validação dos resultados e apresentação dos resultados.

O primeiro bloco comporta o tratamento dos dados originais, obtidos da fonte diretamente e os dados trabalhados e adequados para o presente método, o que é explicado em dois itens: pré processamento e análise exploratória.

O segundo bloco diz respeito à análise estatística multivariada e compreende as análises das similaridades espacial e de atributos distribuindo-se em itens que computam as coordenadas das sedes e os cálculos da distância euclidiana e de outros tipos de distâncias entre os dados.

O terceiro bloco trata-se da validação dos resultados e inclui o teste de Mantel para cada matriz de distâncias de agrupamentos com correlação espacial.

O quarto bloco refere-se à apresentação dos resultados e subdivide-se em itens que explicam a análise de agrupamento, o que são os dendogramas, o método de aglomeração hierárquica e sua interpretação, todos visando o agrupamento de regiões similares com continuidade espacial. Finalmente, este quadro inclui ainda a análise comparativa com outras divisões regionais existentes.

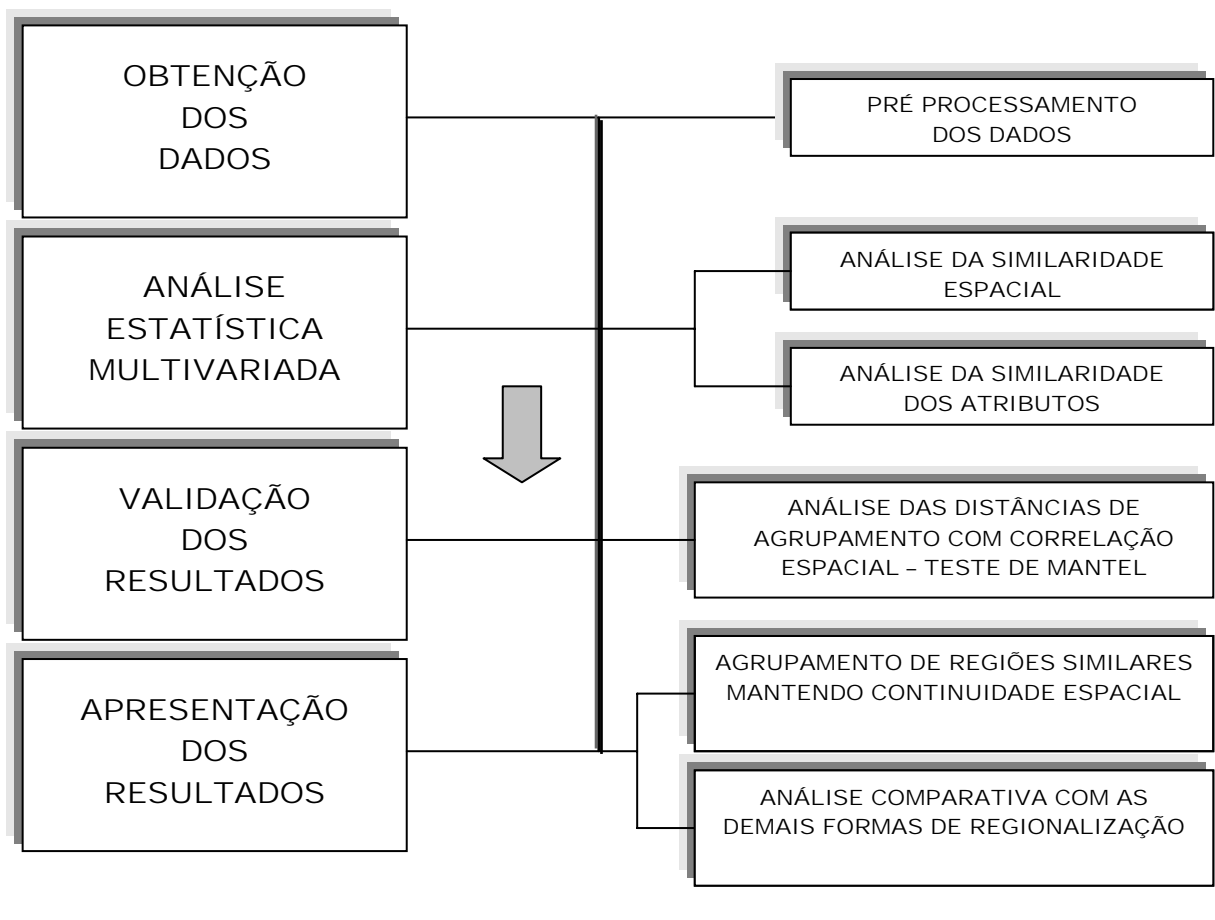


Figura 3.1 – Quadro metodológico

Para cada um dos quadros à direita será apresentada uma figura correspondente à parte do método que será descrita. Da conjugação de todas estas etapas tem-se a visão completa da rotina metodológica e que será apresentada integralmente no capítulo 4, quando da aplicação ao estudo de caso sobre os dados para o estado do Rio Grande do Sul.

3.1 TRATAMENTO DOS DADOS

3.1.1 PRÉ PROCESSAMENTO DOS DADOS

Três elementos definem um objeto espacial qualquer: sua posição espacial, seu valor de atributo e uma unidade temporal. Representado sob a forma de um modelo tridimensional seria a figura 3.2 abaixo:

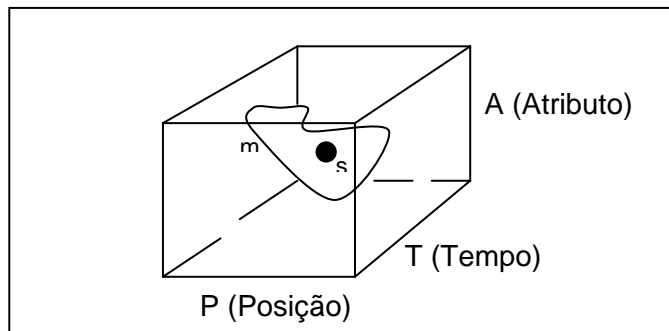


Figura 3.2 – Representação esquemática de um objeto espacial.

A determinação da posição é dada pelo sistema de coordenadas, no caso em estudo, a latitude e a longitude da sede do município, a unidade de tempo é determinada pelo período em que foram levantados os dados para cada município referente aos anos de 1995 e 96, e os atributos estão representados pelas 22 variáveis socioeconômicas selecionadas para cada município.



Figura 3.3 – Representação dos polígonos municipais e seu código identificador.

A figura 3.3 acima representa a forma como os polígonos que representam os municípios são associados a um código chave que funcionará no sistema como o meio de identificar cada município e fazer a associação de informações municipais por esta chave de acesso, originando por exemplo, uma tabela como esta a seguir.

Tabela 3.1 – Extrato da matriz de indicadores para análise espacial (10 municípios).

MATRIZ DE INDICADORES PARA ANÁLISE ESPACIAL (427 MUN.)						
IDENTIFICADORES		COORDENADAS		POPULAÇÃO	SAÚDE	
MUNICÍPIO	ID_IBGE	LATITUDE	LONGITUDE	POP_TOTAL	MORT.INF. 1000 NASC.	TAXA DE MORBIDADE
Agua Santa	430005	-28,17294	-52,03367	4 196	2,445	0,000
Agudo	430010	-29,64062	-53,23740	16 253	1,468	2,111
Ajuricaba	430020	-28,23706	-53,76921	10 759	1,215	2,758
Alecrim	430030	-27,65305	-54,75847	9 369	2,621	1,580
Alegrete	430040	-29,77666	-55,78859	82 527	4,529	2,934
Alegria	430045	-27,82649	-54,05695	5 800	2,004	1,000
Alpestre	430050	-27,24282	-53,03435	11 116	1,038	3,704
Alto Alegre	430055	-28,77107	-52,98774	2 073	6,329	0,000
Alto Feliz	430057	-29,38859	-51,30746	2 538	0,000	0,000
Alvorada	430060	-29,98740	-51,08350	162 005	6,512	2,867
IDENTIFICAÇÃO DO OBJETO		POSIÇÃO		ATRIBUTOS		

(Fonte: Base de Informações Municipais, IBGE ,1996)

Na tabela acima, retirada como exemplo de uma matriz completa com 427 linhas de municípios e 22 colunas de variáveis tem-se a posição definida pelas colunas de latitude e longitude das sedes, representada na figura 3.2 pelo ponto 's', dois atributos: população e saúde, com três variáveis, sendo uma de população total e duas de saúde, mortalidade infantil e taxa de morbidade, todas ocorridas no ano de 1996 que define a unidade de tempo. Além disso é necessário para identificar o polígono 'm' da figura 3.2 que representa um objeto, no caso um município, uma chave de acesso ou código identificador que é a coluna Id_IBGE com um número (código oficial) que está associado a cada município descrito na primeira coluna da tabela 3.1.

3.1.2 ANÁLISE EXPLORATÓRIA

A grande maioria dos dados efetivamente utilizados aqui foram obtidos da Base de Informações Municipais (BIM) do Instituto Brasileiro de Geografia e Estatística, IBGE (1998) relativos a 427 municípios do Rio Grande do Sul para os anos de 1995 e 1996. Os dados relativos ao PIB municipal foram tomados da Fundação de Economia e Estatística (FEE) para o ano de 1996. Existem dados que foram apropriados diretamente, ainda que

não sejam dados primários. A formulação, tanto para os dados diretos como para os dados cuja metodologia de cálculo foi proposta neste trabalho é apresentada a seguir:

ANO DE 1996

POPULAÇÃO

- População total
- População residente urbana
- População residente rural
- Percentual de população urbana
- Percentual de população rural
- Pessoas residentes de 0 a 4 anos de idade
- Taxa de migrantes:

$$M_i = \frac{P_i^m}{P_i^r} \quad [3.1]$$

onde

M_i é a taxa de migrantes (pessoas que saem) do município i ;

P_i^m é a população migrante de 4 anos ou mais;

P_i^r é a população residente de 4 anos ou mais.

EMPREGO

- População em Idade Ativa (PIA) é igual a população com 10 anos ou mais, obtida por somatório da diversas faixas de população fornecidas pela BIM;
- População economicamente ocupada (PEO) é igual a parcela da PIA que está ocupada, isto é, possuem trabalho remunerado exercido com regularidade;
- Taxa de ocupação absoluta (não considera a população desempregada nem os inativos):

$$O_i^a = \frac{PEO_i}{PIA_i} \quad [3.2]$$

onde

O_i^a é a taxa de ocupação absoluta do município i ;

PEO_i é a população economicamente ocupada do município i ;

PIA_i é a população em idade ativa do município i .

- População ocupada no setor secundário (POSS) é composta pelo somatório de seis variáveis da BIM (1998):
 - Pessoal ocupado em alojamento e alimentação;
 - Pessoal ocupado em comércio, reparação de veículos automotores, objetos pessoais e domésticos;

- Pessoal ocupado em construção;
- Pessoal ocupado em indústria de transformação;
- Pessoal ocupado em indústrias extrativas;
- Pessoal ocupado em distribuição de energia, gás e água.
- População ocupada no setor terciário (POST) é composta pelo somatório de oito variáveis da BIM (1998):
 - Pessoal ocupado em administração pública, defesa e seguridade social;
 - Pessoal ocupado em educação;
 - Pessoal ocupado em imobiliárias, aluguéis e serviços prestados às empresas;
 - Pessoal ocupado em intermediação financeira;
 - Pessoal ocupado em organismos internacionais e em outras instituições extraterritoriais;
 - Pessoal ocupado em outros serviços coletivos, sociais e pessoais;
 - Pessoal ocupado em transporte, armazenagem e comunicações.
- Taxa de ocupação no setor secundário (indústria e comércio):

$$O_i^s = \frac{POSS_i}{PIA_i} \quad [3.3]$$

onde

O_i^s é a taxa de ocupação no setor secundário do município i ;

$POSS_i$ é a população ocupada no setor secundário do município i ;

PIA_i é a população em idade ativa do município i .

- Taxa de ocupação no setor terciário (serviços):

$$O_i^t = \frac{POST_i}{PIA_i} \quad [3.4]$$

onde

O_i^t é taxa de ocupação no setor terciário do município i ;

$POST_i$ é população ocupada no setor terciário do município i ;

PIA_i é a população em idade ativa do município i .

O conceito mais usual relacionado com emprego e desemprego, e utilizado na Pesquisa Emprego Desemprego (PED) da Fundação de Economia e Estatística é o de População economicamente ativa (PEA) que é a parcela da PIA que está ocupada ou desempregada. Entretanto, optou-se por utilizar o conceito de PEO, População economicamente ocupada, em função de não haver dados na BIM que contemple o número de desempregados para todos os municípios do Rio Grande do Sul, uma vez que a PED tem como área de estudo a Região Metropolitana de Porto Alegre.

SAÚDE

- Óbitos de pessoas com menos de 1 ano de idade.
- Óbitos de pessoas de 0 a 4 anos de idade.
- Percentual de mortalidade infantil.
- Número de internações hospitalares.
- Número de leitos hospitalares.
- Pessoal ocupado na área da saúde e social.

$$MI_i = \frac{O_i^{<1} + O_i^{1,4}}{P_i^{0,4}} \quad [3.5]$$

onde

MI_i é o percentual de mortalidade infantil do município i ;

$O_i^{<1}$ é o número de óbitos de pessoas com menos de 1 ano de idade no município i ;

$O_i^{1,4}$ é o número de óbitos de pessoas de 1 a 4 anos de idade no município i ;

$P_i^{0,4}$ é a população de 0 a 4 anos de idade no município i .

- Taxa de mortalidade infantil:

$$TMI_i = MI_i \times 10 \quad [3.6]$$

onde

TMI_i é a taxa de mortalidade infantil por mil nascidos vivos no município i .

- Taxa de morbidade:

$$Mb_i = \frac{I_i}{L_i} \quad [3.7]$$

onde

Mb_i é a taxa de morbidade do município i ;

I_i é o número de internações do município i ;

L_i é o número de leitos hospitalares no município i .

- Capacidade de atendimento hospitalar:

$$Cah_i = \frac{Ps_i}{I_i} \quad [3.8]$$

onde

Cah_i é a capacidade de atendimento hospitalar do município i ;

Ps_i é o total de pessoal ocupado na área da saúde e social do município i ;

I_i é o número de internações do município i .

- Disponibilidade hospitalar:

$$Dh_i = \frac{L_i}{I_i} \quad [3.9]$$

onde

Dh_i é a disponibilidade hospitalar do município i ;

L_i é o número de leitos hospitalares no município i ;

I_i é o número de internações do município i .

Obs.: Para os indicadores Mb_i , Cah_i e Dh_i foi colocada uma condição de cálculo que zera o indicador para o caso de o divisor ser zero, a fim de evitar o surgimento de células com erro na planilha.

EDUCAÇÃO

- Taxa de alfabetização
- Taxa de escolarização de 1º grau
- Taxa de escolarização de 2º grau
- Taxa de escolarização de 3º grau

$$A_i = \frac{P_i^{1,3}}{Pr_i} \quad [3.10]$$

onde

A_i é a taxa de alfabetização do município i ;

$P_i^{1,3}$ é o número de pessoas residentes com 1 a 3 anos de estudo no município i ;

Pr_i é a população residente de 4 anos ou mais no município i .

$$E_i^1 = \frac{P_i^{4,7}}{Pr_i} \quad [3.11]$$

onde

E_i^1 é a taxa de escolarização de 1º grau para o município i ;

$P_i^{4,7}$ é o número de pessoas residentes com 4 a 7 anos de estudo no município i ;

Pr_i é a população residente de 4 anos ou mais no município i .

$$E_i^2 = \frac{P_i^{8,10}}{Pr_i} \quad [3.12]$$

onde

E_i^2 é a taxa de escolarização de 2º grau para o município i ;

$P_i^{8,10}$ é o número de pessoas residentes com 8 a 10 anos de estudo no município i ;

Pr_i é a população residente de 4 anos ou mais no município i .

$$E_i^3 = \frac{P_i^{11,14}}{Pr_i} \quad [3.13]$$

onde

E_i^3 é a taxa de escolarização de 3º grau para o município i ;

$P_i^{11,14}$ é o número de pessoas residentes com 11 a 14 anos de estudo no município i ;

Pr_i é a população residente de 4 anos ou mais no município i .

ECONOMIA

- PIB total per capita de cada município.

Esta variável foi obtida diretamente na FEE para o ano de 1996.

ANO DE 1995

AGROPECUÁRIA

- Área de estabelecimentos agropecuários em 31/12/1995 (hectares).
- Pessoal ocupado em estabelecimentos agropecuários.
- Área total / pessoal ocupado em estabelecimentos agropecuários (arredondado).
- Índice de evasão produtiva rural.

$$Epr_i = \frac{A_i}{Poa_i} \quad [3.14]$$

onde

Epr_i é a evasão produtiva rural do município i ;

A_i é a área de estabelecimentos agropecuários no município i ;

Poa_i é o pessoal ocupado em estabelecimentos agropecuários no município i .

AMBIENTAL

- Área antropizada pela agricultura;
- Número de indústrias poluentes;
- Área total do município i ;
- Área de mata nativa do município i ;
- Índice de antropia agrícola:

$$I_i^A = \frac{A_{ai}}{A_{ti}} \quad [3.15]$$

onde

I_i^A é o índice de antropia agrícola do município i ;

A_{ai} é a área antropizada pela agricultura do município i ;

A_{ti} é a área total do município i .

GEOFÍSICAS

- Área do município em Km^2 .
- Altitude da sede municipal (invariável).

Estes dados geofísicos foram extraídos da BIM, o primeiro considera a área municipal para os 427 municípios, malha que vigorou entre 1993 e 1996, o segundo é invariável, a menos que a sede municipal seja deslocada, o que raramente acontece.

De forma esquemática e ordenada conforme a fonte e ano do dado tem-se a figura 3.4 abaixo:

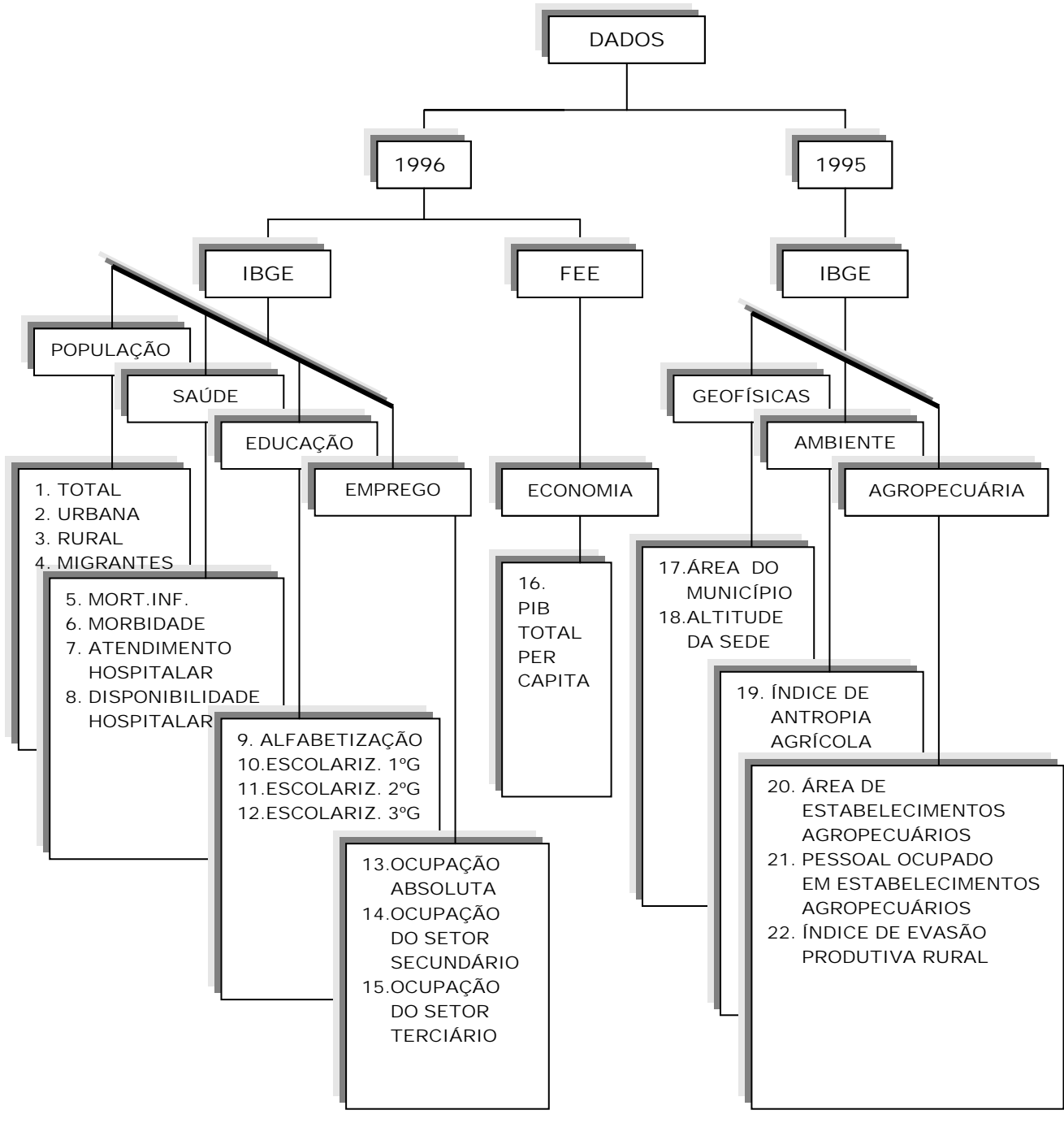


Figura 3.4 – Origem e seleção de variáveis (22).

Todos os dados gerados a partir das formulações acima são organizados em uma matriz e a seguir são normalizados para que os resultados se tornem comparáveis. A fórmula de normalização considera os valores de máximo e mínimo para cada variável entre todas as ocorrências para os municípios (427), como segue:

$$V_normal(i, j)_{0-1} = \frac{V_{ij} - V_{ij\ min}}{V_{ij\ max} - V_{ij\ min}} \quad [3.16]$$

onde

i varia de 1 a 22 variáveis;

j varia de 1 a 427 municípios;

V_{ij} é o valor de uma variável genérica no município j;

V_{ij min} é o menor valor no conjunto de todas as ocorrências da variável i para os 427 municípios;

V_{ij max} é o maior valor no conjunto de todas as ocorrências da variável i para os 427 municípios.

TABELA 3.2 Matriz Normalizada (valores entre 0 e 1)

MUNICÍPIO	POPULAÇÃO			EMPREGO		SAÚDE				EDUCAÇÃO			ETC...
	V ₁	V ₂	V ₃	V ₄	V ₅	V ₆	V ₇	V ₈	V ₉	V ₁₀	V ₁₁	V ₁₂ V _n
Mun ₁	V ₁₁	V ₂₁	V ₃₁	V ₄₁	V ₅₁	V ₆₁	V ₇₁	V ₈₁	V ₉₁	V ₁₀₁	V ₁₁₁	V ₁₂₁	V _{n1}
Mun ₂	V ₁₂	V ₂₂	V ₃₂	V _{n2}
Mun ₃	V ₁₃	V ₂₃	V ₃₃	V _{n3}
....
....
Mun _i	V ₁₄₂₇

Tendo como entrada a matriz normalizada gera-se uma matriz de correlação que utiliza como equação de regressão o produto de Pearson, o qual é dado pela fórmula:

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2] [n\sum y^2 - (\sum y)^2]}} \quad [3.17]$$

Fazendo o quadrado de r (r²) obtém-se o coeficiente de correlação que varia entre 0 e 1 e pode ser interpretado como a proporção da variância em y que pode ser atribuída à variância em x. Onde Y é a matriz de dados dependente e X é a matriz de dados independente. Interpreta-se que valores próximos de 1(um) representam alta correlação entre as variáveis podendo uma delas ser descartada porque a outra a representa e valores próximos a 0 (zero) representam conjuntos de dados independentes.

Exemplificando com a matriz desenhada acima, o conjunto da matriz X seria inicialmente composto de todos os valores de V_1 e Y seria composto pelos valores de V_2 , então tomando todas as variáveis combinadas duas a duas (V_1 com V_3 , V_1 com V_4 , etc.) chega-se a matriz de correlação de $n \times n$ onde n é o número de variáveis consideradas.

TABLEA 3.3 Matriz de correlação (valores entre 0 e 1)

MUNICÍPIO	POPULAÇÃO			EMPREGO		SAÚDE				EDUCAÇÃO			ETC...
	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	V_{10}	V_{11}	V_{12} V_n
V1	1	V_{21N}	V_{31N}	V_{41N}	V_{51N}	V_{61N}	V_{71N}	V_{81N}	V_{91N}	V_{101N}	V_{111N}	V_{121N}	V_{n1N}
V2	V_{12N}	1	V_{32N}	V_{n2N}
V3	V_{13N}	V_{23N}	1	V_{n3N}
V4	1
V5	1
V6	1
V7	1
V8	1
V9	1
V10	1
V11	1
V12	1
....
....
Vn	1

Note-se que os valores V_{21N} e V_{12N} ; V_{31N} e V_{13N} ; V_{32N} e V_{23N} ; etc. são iguais o que caracteriza uma matriz simétrica em relação à diagonal que tem valor 1 (um).

De um modo esquemático a figura 3.5, abaixo, representa esta primeira etapa:

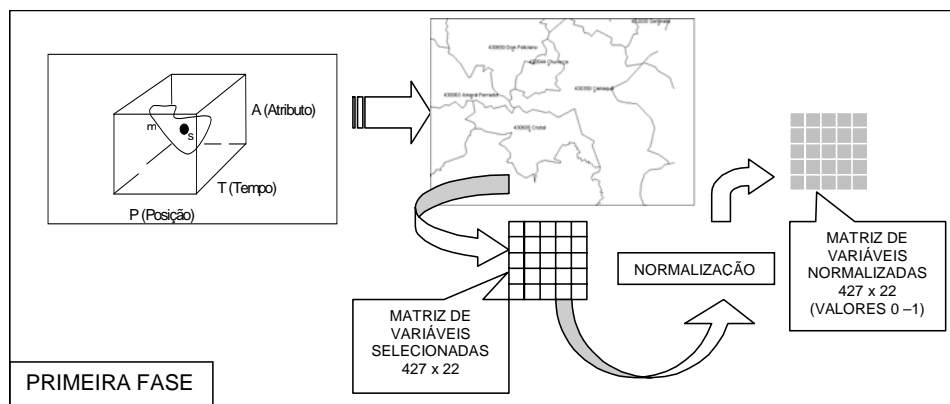


Figura 3.5 – Primeira fase da metodologia, pré processamento dos dados.

3.2 ANÁLISE DA SIMILARIDADE: ESPACIAL E DE ATRIBUTOS

3.2.1 COORDENADAS DAS SEDES MUNICIPAIS E DISTÂNCIAS EUCLIDIANAS

Utilizando-se como ponto de partida as latitudes e longitudes (as coordenadas UTM seriam mais indicadas por resultarem em distâncias efetivas) das sedes dos municípios gaúchos pode-se determinar uma matriz de distância “euclidiana” entre todos os municípios pela aplicação da fórmula abaixo. Observe-se que no exemplo numérico adiante, calcula-se uma distância em unidades de medida não métricas, porque utiliza-se como partida latitudes e longitudes que são dadas em graus e décimos de graus. Entretanto são adequadas para o estudo uma vez que resultam em medidas que permitem uma comparação ponderada em termos de distância.

$$D_{(i,j)} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad [3.18]$$

sendo

$D_{i,j}$ a distância euclidiana entre os municípios i e j ;

x_i e x_j são as abcissas dos municípios i e j ;

y_i e y_j são as ordenadas dos municípios i e j .

De modo genérico a fórmula da distância euclidiana (linha i e linha j) é a seguinte:

$$d(i, j) = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad [3.19]$$

Desta forma se origina uma matriz que terá a diagonal igual a zero e será simétrica em relação à diagonal:

Tabela 3.4 Matriz de distância euclidiana

MUNICÍPIOS	Mun ₁	Mun ₂	Mun ₃	Mun _n
Mun ₁	0	D ₂₁	D ₃₁	D _{n1}
Mun ₂	D ₁₂	0	D ₃₂	D _{n2}
Mun ₃	D ₁₃	D ₂₃	0	D _{n3}
....	0
....	0
Mun _n	D _{1n}	D _{2n}	D _{3n}	0

Esta segunda etapa metodológica pode ser sintetizada pela figura 3.6 a seguir:

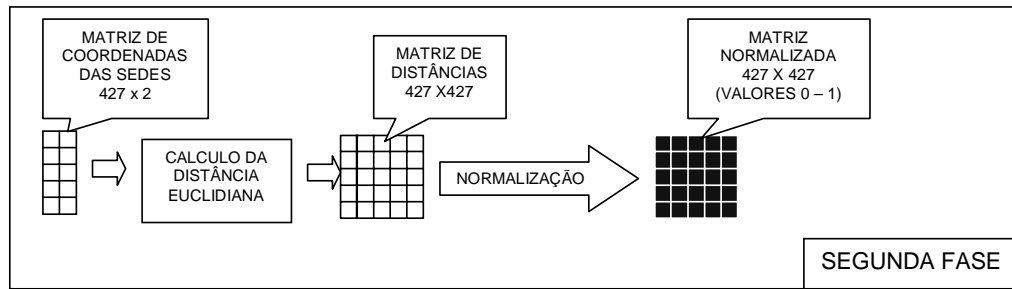


Figura 3.6 – Segunda fase da metodologia, análise da similaridade espacial.

3.2.2 CÁLCULO DE OUTROS TIPOS DE DISTÂNCIAS ENTRE OS DADOS

Considere-se a seguir a matriz da tabela 3.1 dos dados normalizados para todos os municípios do estado (427). Neste caso, se considerados como coordenadas, haveria uma variação aleatória das grandezas nos diferentes eixos coordenados. Portanto, uma distância euclidiana que atribui igual peso para todas as coordenadas não seria aceitável, do mesmo modo que é insatisfatória para a maioria das finalidades estatísticas (Johnson, op. cit.). Assim chega-se à definição de distância estatística que atribui maior peso para as coordenadas em eixos de maior variação e vice-versa.

3.2.2.1 DISTÂNCIA ESTATÍSTICA

A distância estatística, por considerar as diferentes variações ao longo de cada eixo, precisa de formas de medição destas variações, o que pode ser dado pela variância e covariância. Uma forma de ponderar o peso relativo de cada eixo, supondo que fossem dois eixos perpendiculares, seria padronizar as coordenadas dividindo pelo desvio padrão (Johnson, op. cit.). Tomando-se duas variáveis quaisquer P (variável populacional) e S (variável de saúde) representadas pelas coordenadas $P = (x_1, x_2)$ que são os valores que a variável P assume no município 1 e no município 2 respectivamente, e $S = (y_1, y_2)$ cujos valores y_1 e y_2 são os valores para a variável S nos municípios 1 e 2, e ainda supondo-se que P e S variem de forma independente, pode-se determinar a distância estatística entre P e S pela expressão a seguir:

$$d(P,S) = \sqrt{\frac{(x_1 - y_1)^2}{\sigma_1^2} + \frac{(x_2 - y_2)^2}{\sigma_2^2}} \quad [3.20]$$

$$S_{11} \quad S_{22}$$

onde

$d(P,S)$ é a distância estatística entre os “pontos” P e S;

x_1 é o valor de P no município 1;

x_2 é o valor de P no município 2;

y_1 é o valor de S no município 1;

y_2 é o valor de S no município 2;

s_{11} é a variância de P e S na direção 1;

s_{22} é a variância de P e S na direção 2.

Para um caso genérico de n dimensões a fórmula [3.20] assume o seguinte aspecto:

$$d(P,S) = \sqrt{\frac{(x_1 - y_1)^2}{s_{11}} + \frac{(x_2 - y_2)^2}{s_{22}} + \dots + \frac{(x_n - y_n)^2}{s_{nn}}} \quad [3.21]$$

Onde os pontos P e S possuem n coordenadas como segue: $P = (x_1, x_2, \dots, x_n)$ e $S = (y_1, y_2, \dots, y_n)$, enquanto que $s_{11}, s_{22}, \dots, s_{nn}$ são as variâncias para as n medidas de x e de y. Equiparando à realidade do presente estudo tem-se $n = (1,2,\dots, 427)$ municípios com seus respectivos valores de uma variável (x_1, x_2,\dots, x_{10}) de população (por exemplo: Taxa de Migrantes) e outra (y_1, y_2,\dots, y_{10}) de saúde (por exemplo: Capacidade de atendimento hospitalar) para os dez primeiros municípios do estado (ordem alfabética) e conseqüentes variâncias ($s_{11}, s_{22},\dots, s_{1010}$) de cada variável em todos os municípios (base municipal de 1994).

Tabela 3.5 Latitude e longitude das sedes dos primeiros dez (10) municípios gaúchos (ordem alfabética):

MUNNOME	LATITUDE	LONGITUDE
	Y (°,décimos)	X (°,décimos)
AGUA SANTA	-28,17294	-52,03367
AGUDO	-29,64062	-53,2374
AJURICABA	-28,23706	-53,76921
ALECRIM	-27,65305	-54,75847
ALEGRETE	-29,77666	-55,78859
ALEGRIA	-27,82649	-54,05695
ALPESTRE	-27,24282	-53,03435
ALTO ALEGRE	-28,77107	-52,98774
ALTO FELIZ	-29,38859	-51,30746
ALVORADA	-29,9874	-51,0835

Fonte: Ministério da Saúde (<http://www.datasus.com.br>)

A partir dos dados acima (Tabela 3.5) e pelo emprego da fórmula [3.19] obtém-se os resultados expressos na matriz abaixo.

Tabela 3.6 Cálculo de distância “euclidiana” entre os dez primeiros municípios gaúchos

MUNICÍPIOS	Água Santa	Agudo	Ajuricaba	Alecrim	Alegrete	Alegria	Alpestre	Alto Alegre	Alto Feliz	Alvorada
Água Santa	0,00000	1,89817	1,73672	2,77395	4,08306	2,05273	1,36619	1,12606	1,41605	2,04819
Agudo		0,00000	1,50093	2,50282	2,55481	1,99066	2,40638	0,90468	1,94633	2,18164
Ajuricaba			0,00000	1,14878	2,53934	0,50136	1,23634	0,94650	2,71776	3,20573
Alecrim				0,00000	2,36027	0,72264	1,77225	2,09415	3,86284	4,35369
Alegrete					0,00000	2,60801	3,74248	2,97590	4,49790	4,70981
Alegria						0,00000	1,17745	1,42669	3,16225	3,67572
Alpestre							0,00000	1,52896	2,75436	3,36727
Alto Alegre								0,00000	1,79016	2,25955
Alto Feliz									0,00000	0,63932
Alvorada										0,00000

A seguir, utilizando-se os dados da matriz normalizada para a variável de população, Taxa de Migrantes, e da variável de saúde, Capacidade de atendimento hospitalar, e empregando-se a fórmula [3.20] obtém-se a seguinte matriz:

Tabela 3.7 Cálculo de distância estatística para duas variáveis entre os dez (10) primeiros municípios gaúchos (Taxa de Migrantes X Capacidade de atendimento hospitalar) - 1996

Municípios	Água Santa	Agudo	Ajuricaba	Alecrim	Alegrete	Alegria	Alpestre	Alto Alegre	Alto Feliz	Alvorada
Água Santa	0,00000	1,80891	1,87973	1,76336	1,68063	2,46539	1,30442	3,03962	2,17671	5,16621
Agudo		0,00000	1,51747	1,51747	1,42049	2,29596	0,94594	2,90389	1,98276	5,08753
Ajuricaba			0,00000	1,53403	1,43817	2,30694	0,97228	2,91258	1,99546	5,09249
Alecrim				0,00000	1,40937	2,28910	0,92915	2,89847	1,97481	5,08444
Alegrete					0,00000	2,27703	0,89900	2,88894	1,96080	5,07901
Alegria						0,00000	1,19901	2,99590	2,11522	5,14060
Alpestre							0,00000	2,85110	1,90460	5,05758
Alto Alegre								0,00000	2,25515	5,19974
Alto Feliz									0,00000	5,11532
Alvorada										0,00000

As tabelas 3.6 e 3.7 serão as entradas para o teste de Mantel que define o grau de correlação entre as distâncias entre as sedes e as “distâncias” entre o valores dos atributos para cada município em relação a todos os demais, como será explicado nas seções subsequentes.

3.2.2.2 OUTROS TIPOS DE DISTÂNCIAS ENTRE OS DADOS

Existe uma variedade bastante extensa de medidas de distância que podem ser utilizadas como dimensão de similaridade entre os dados, em consequência, já foram desenvolvidos algoritmos de cálculo como o ADE-4 (*Analyses des Données Écologiques*)¹⁰.

¹⁰ O ADE-4 é um *software* de domínio público que apresenta uma metodologia exploratória e euclidiana para as Ciências Ambientais, encontrado para *download* em <http://pbil.univ-lyon1.fr/ADE-4>.

Estas medidas podem ser métricas ou euclidianas, ou não euclidianas. Apresenta-se a seguir a formulação de alguns destes coeficientes de similaridade:

- Medida de Minkowski

$$d(x, y) = \left[\sum_{i=1}^p |x_i - y_i|^m \right]^{1/m} \quad (\text{Euclidiana}) \quad [3.22]$$

onde

x é o conjunto de p observações da variável x ;

y é o conjunto de p observações da variável y ;

p é o número de dimensões observadas (neste caso duas, x e y);

m é o peso dado às maiores e menores diferenças.

Observe-se que para $m = 1$ tem-se a distância entre dois pontos, também chamada de distância *city block* ou Manhattan (Johnson, op. cit.) e quando $m = 2$ é a própria distância euclidiana.

- Distância de Canberra

$$d(x, y) = \sum_{i=1}^p \frac{|x_i - y_i|}{(x_i + y_i)} \quad (\text{não Euclidiana}) \quad [3.23]$$

- Coeficiente de Czekanowski

$$d(x, y) = 1 - \frac{2 \sum_{i=1}^p \min(x_i, y_i)}{\sum_{i=1}^p (x_i + y_i)} \quad (\text{não Euclidiana}) \quad [3.24]$$

Demais tipos de distâncias são calculadas através do formulário abaixo, sendo X uma matriz de n linhas por p colunas representada por $X = [x_{ij}]$.

1. *City Block = Manhattan – Range* ou D3 de Gower & Legendre (1986) – distância não euclidiana:

$$d_1(i, j) = 1/p \sum_{k=1}^p \frac{|x_{ik} - x_{jk}|}{r_k} \quad [3.25]$$

onde

$$r_k = \text{Max}_{i=1}^n (x_{ik}) - \text{Min}_{i=1}^n (x_{ik}) \quad [3.26]$$

2. *City Block = Manhattan – Standard deviation* de Cain & Harrison – distância não euclidiana:

$$d_2(i, j) = 1/p \sum_{k=1}^p \frac{|x_{ik} - x_{jk}|}{r_k} \quad [3.27]$$

onde

$$r_k = \sqrt{1/n \sum_{i=1}^n (x_{ik} - m_k)^2} \quad [3.28]$$

e

$$m_k = 1/n \sum_{i=1}^n x_{ik} \quad [3.29]$$

3. *Canberra = Lance & Williams ou D7* de Gower & Legendre – distância não euclidiana

$$d_3(i, j) = 1/p \sum_{k=1}^p \frac{|x_{ik} - x_{jk}|}{|x_{ik}| + |x_{jk}|} \quad [3.30]$$

4. *Bray-Curtis ou D8* de Gower & Legendre – distância não euclidiana

$$d_4(i, j) = \frac{1/p \sum_{k=1}^p |x_{ik} - x_{jk}|}{\sum_{k=1}^p (x_{ik} + x_{jk})} \quad [3.31]$$

5. *Gower & Legendre D5* – distância euclidiana

$$d_5(i, j) = 1/p \sum_{k=1}^p \frac{(x_{ik} - x_{jk})^2}{(x_{ik} + x_{jk})^2} \quad [3.32]$$

6. *Gower & Legendre D9* - distância não euclidiana

$$d_6(i, j) = \frac{\sum_{k=1}^p |x_{ik} - x_{jk}|}{\sum_{k=1}^p \text{Max}_{i, j} (x_{ik}, x_{jk})} \quad [3.33]$$

7. Gower & Legendre D10 - distância não euclidiana

$$d_7(i, j) = 1/p \sum_{k=1}^p \left(1 - \frac{\text{Min}_{i,j}(x_{ik}, x_{jk})}{\text{Max}_{i,j}(x_{ik}, x_{jk})} \right) \quad [3.34]$$

Todas as últimas sete distâncias foram efetivamente utilizadas como determinantes das matrizes de similaridade dos atributos e fazem parte do módulo de “Distâncias: variáveis quantitativas” do ADE-4, sendo geradas automaticamente pelo uso do aplicativo, bastando para isso alimentar com a matriz de entrada, definir um nome para a matriz de saída e escolher um tipo de distância entre as sete anteriores.

Contudo, um oitavo tipo de distância foi também considerado, e que não faz parte do módulo de distâncias do ADE-4, é aquele que utiliza a fórmula do Chi-2 (Chi-quadrado), que faz uso da mesma lógica da análise de correspondência, desenvolvida a seguir com maior aproximação.

Trata-se de uma forma gráfica de representação de associações em tabelas de freqüências ou tabelas de contingências. Para uma tabela $T(i, j)$ onde i é o número de linhas e j o número de colunas, a distribuição do conjunto dos pontos I e dos pontos J em relação a suas posições revelam associações: proximidade de pontos de linha entre si significa linhas de perfis similares através das colunas; proximidade de pontos de colunas entre si significa similaridade de perfis de colunas através das linhas; e categorias de linhas e colunas não relacionadas são representadas pela proximidade de pontos de linha dos pontos de coluna. Os resultados da análise de correspondência são: uma representação bidimensional do melhor conjunto de dados levando em conta as coordenadas locadas e uma medida da quantidade de informação contida em cada dimensão, designada por “inertia” ou inércia (Johnson, 1998).

A fórmula do Chi-2 (Chi-quadrado) para medição do grau de associação entre as variáveis de linha e coluna em uma tabela de contingência $T(i, j)$ é:

$$\chi^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad [3.35]$$

onde

$O_{ij} = t_{ij}$ é a freqüência observada para a célula (i, j) ésima;

$E_{ij} = n l_i c_j$ é a freqüência esperada na célula (i, j) ésima, se a variável de linha é independente da variável de coluna.

A matriz de correspondência é o conjunto dos pesos de linhas e colunas relativas ao total das freqüências da tabela de contingência T (i, j) assim expressos:

$$l_i = \sum_{j=1}^J p_{ij} = \sum_{j=1}^J \frac{t_{ij}}{n}, \quad i = 1, 2, \dots, I \quad [3.36]$$

$$c_j = \sum_{i=1}^I p_{ij} = \sum_{i=1}^I \frac{t_{ij}}{n}, \quad j = 1, 2, \dots, J \quad [3.37]$$

onde

n é o total das freqüências;

t_{ij} é a freqüência observada na célula (i, j) ésima;

p_{ij} é o peso relativo da freqüência na célula (i, j) ésima;

l_i é o somatório dos pesos nas linhas;

c_j é o somatório dos pesos nas colunas.

Assim, a expressão do Chi-2 pode ser rescrita como segue:

$$\chi^2 = n \sum_{i,j} \frac{(p_{ij} - l_i c_j)^2}{l_i c_j} \quad [3.38]$$

Chega-se, por fim, à definição de *inertia* (inércia) total que é o somatório do quadrado das distâncias do perfil da linha (ou da coluna) ao centróide, logo uma medida de variação global, ou diferenças, nos pontos representantes dos perfis de linha ou de coluna (Johnson, op. cit.).

$$\text{Inércia} = \sum_i l_i \sum_i (p_{ij} / l_i - c_j)^2 / c_j = \sum_i c_j \sum_i (p_{ij} / c_j - l_i)^2 / l_i \quad [3.39]$$

$$\text{Inércia} = \sum_i \sum_j (p_{ij} - l_i c_j)^2 / l_i c_j \quad [3.40]$$

$$\text{Inércia} = \chi^2 / n \quad [3.41]$$

A figura 3.7 abaixo expressa de forma esquemática esta terceira fase da metodologia:

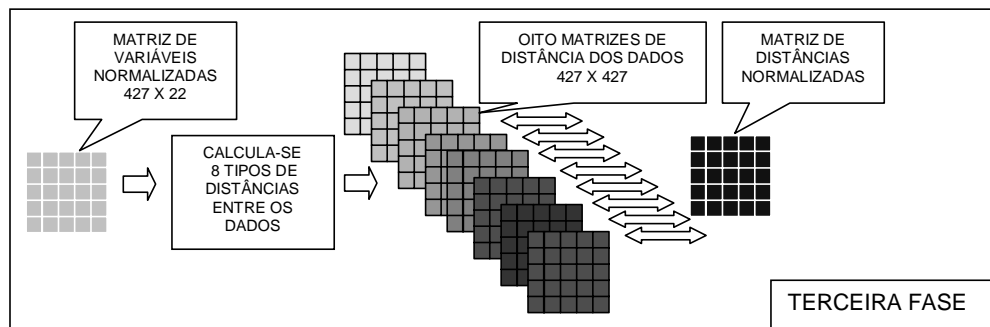


Figura 3.7 – Terceira fase da metodologia, análise da similaridade dos atributos.

3.2.2.3 MONTE CARLO

O método de Monte Carlo é uma técnica que gera, por repetição de múltiplas combinações aleatórias dos valores de entrada, uma amostra de distribuição das saídas cujos parâmetros da função geradora podem ser estimados da própria amostra (Mendes, 1994). Apresenta extensas utilizações em modelos de simulação em pesquisas de Físico-Química (Binder, 1987; Netz, 1992; e Ferrão, 1992), Estatística, Hidrologia e Engenharia (Mortensen, 1987).

3.3 TESTE DE MANTEL

O teste de Mantel relaciona duas matrizes de distâncias, sendo uma de distâncias físicas e outra de distâncias entre atributos. O problema central reside em descobrir a ordenação dos elementos em uma matriz de modo a maximizar ou minimizar a correlação entre a primeira e a segunda matriz (Manly, 1994).

Para o presente estudo toma-se um caso mais particular da abordagem de Mantel, considerando-se duas matrizes simétricas com diagonal zero, uma vez que os termos em posições simétricas em relação à diagonal são iguais e as distâncias de cada elemento em relação a si mesmo é zero.

$$D = \begin{bmatrix} 0 & d_{21} & d_{31} & \dots & \dots & d_{n1} \\ d_{21} & 0 & d_{32} & \dots & \dots & d_{n2} \\ d_{31} & d_{32} & 0 & \dots & \dots & d_{n3} \\ \dots & \dots & \dots & 0 & \dots & \dots \\ \dots & \dots & \dots & \dots & 0 & \dots \\ d_{n1} & d_{n2} & d_{n3} & \dots & \dots & 0 \end{bmatrix} \text{ e } A = \begin{bmatrix} 0 & a_{21} & a_{31} & \dots & \dots & a_{n1} \\ a_{21} & 0 & a_{32} & \dots & \dots & a_{n2} \\ a_{31} & a_{32} & 0 & \dots & \dots & a_{n3} \\ \dots & \dots & \dots & 0 & \dots & \dots \\ \dots & \dots & \dots & \dots & 0 & \dots \\ a_{n1} & a_{n2} & a_{n3} & \dots & \dots & 0 \end{bmatrix}$$

Portanto, pode-se calcular a correlação entre as duas matrizes a partir dos m elementos acima ou abaixo da diagonal onde m é determinado por:

$$m = n(n - 1) / 2 \quad [3.42]$$

onde

n é o número total de linhas (= colunas).

Assim, a correlação entre as matrizes D e A se estabelece a partir da área dos “triângulos” matriciais abaixo da diagonal (Manly, op. cit.) e é dada por:

$$r = \frac{\sum d_{ij} a_{ij} - \sum d_{ij} \sum a_{ij} / m}{\sqrt{[\sum d_{ij}^2 - (\sum d_{ij})^2 / m] [\sum a_{ij}^2 - (\sum a_{ij})^2 / m]}} \quad [3.43]$$

onde

d_{ij} é um elemento genérico da matriz D ;

a_{ij} é um elemento genérico da matriz A .

Desta equação é possível verificar que somente o somatório do produto dos elementos das matrizes D e A , primeiro termo do numerador, seriam alterados pela mudança da ordenação dos elementos em uma delas. Designando por Γ (gama), este produto poderia ser escrito como segue (Jacquez, 1999):

$$\Gamma = D \times A = \sum_{i=1}^N \sum_{j=1}^N d_{ij} a_{ij} \quad [3.44]$$

De modo geral, a expressão [3.44], em estudos para a área de saúde, assume a seguinte forma (Jacquez, op. cit.):

$$\Gamma = \sum_{i=1}^N \sum_{j=1}^N \delta_{ij} c_{ij} \quad [3.45]$$

onde

n é o número de locais de ocorrência;

δ é uma medida de proximidade;

c é uma variável calculada a partir de observações.

Para este modelo de regionalização, n seria o número de sedes municipais, δ seria uma forma de distância (geográfica, estatística, pesos espaciais, adjacência ou vizinho mais próximo), e c uma observação numérica de um atributo em cada município.

O produto gama (Γ), extensamente utilizado em trabalhos de pesquisa que buscam estabelecer padrões de distribuição espacial dos dados, será perfeitamente adequado à esta investigação que, por sua vez, busca determinar padrões de agrupamento dos dados associados à sua distribuição espacial, de modo a indicar regionalizações prováveis.

O quadro a seguir traduzido e adaptado (Jacquez, op. cit.) presta-se a elucidação das diversas aplicações que o produto gama teve nos últimos 50 anos.

Quadro 3.1 – Testes de agrupamentos sob a forma de variações do produto gama (Γ)

Estatística	Dado	δ	c	Técnica	Referência
Mantel	(x,y,t) local de eventos relativos à saúde	Distância entre par de casos	Tempo estimado entre casos	Teste de interação espaço-temporal	Mantel, 1967
Mantel	(x,y, z ₁ ,z ₂ ,..., z _p)	Distância entre localizações da amostra	Distância multivariada calculada de z ₁ ,z ₂ ,..., z _p	Teste de estrutura espacial em dados multivariados	Smouse, 1986
Knox	(x,y,t)	Adjacência espacial	Adjacência temporal	Teste de interação espaço-temporal	Knox, 1964
Jacquez	(x,y,t)	Vizinho espacial mais próximo	Vizinho temporal mais próximo	Teste de interação espaço-temporal	Jacquez, 1996
Cuzick e Edwards	(x,y,c), onde c é caso identificado	Vizinho mais próximo	c _{ij} = 1 se i e j são casos, senão 0	Teste de agrupamento espacial	Cuzick e Edwards, 1990
Moran	(x,y,z), onde z é um atributo	Conexão espacial por pesos entre localizações i e j	(z _i - z _{médio}) x (z _j - z _{médio})	Teste de autocorrelação espacial	Moran, 1950
Contagem agrupada	(x,y,z), onde z = 1 se x, y são rotulados	Adjacência espacial	1 se ambos, i e j são rotulados	Teste de agrupamento espacial de localizações rotuladas	Cliff e Ord, 1981

(Fonte: Adaptado de BioMedware Press, 1999).

O segundo Mantel do quadro acima é o exato modelo de teste que aplicaremos no presente trabalho, em que x e y são as coordenadas das sedes municipais e z₁,z₂,..., z_p são os atributos selecionados como representativos das características municipais e, mais adiante, regionais.

O objetivo do teste de Mantel é, através da permutação aleatória dos elementos de uma das duas matrizes D e A, obter a melhor correlação entre a matriz de distâncias físicas (D) e a matriz de distâncias entre os atributos (A). Em outras palavras, significa fazer variar o produto gama (Γ) entre estas duas matrizes por sucessivas permutações aleatórias de seus elementos, de modo a obter a correlação mais significativa em termos de relação entre os dados e o espaço.

Continuando com o exemplo genérico das matrizes D e A acima, pode-se verificar que a rotulação de uma das matrizes é independente de posição em relação a outra, de modo que a nova ordenação da matriz A poderia assumir a forma abaixo (matriz randômica) pela permutação de alguns rótulos de linha e coluna:

$$A_R = \begin{bmatrix} 0 & a_{23} & a_{53} & \dots & \dots & a_{13} \\ a_{23} & 0 & a_{52} & \dots & \dots & a_{12} \\ a_{53} & a_{52} & 0 & \dots & \dots & a_{15} \\ \dots & \dots & \dots & 0 & \dots & \dots \\ \dots & \dots & \dots & \dots & 0 & \dots \\ a_{13} & a_{12} & a_{15} & \dots & \dots & 0 \end{bmatrix}$$

Caso esta matriz A_R representasse a melhor correlação obtida pela fórmula [3.43] e supondo que estes rótulos representem de fato municípios, pode-se começar a vislumbrar uma indicação de agrupamento que associa de forma mais significativa padrões espaciais com dados municipais. Aqui começa a aparecer a idéia de uma regionalização com base nas informações e vinculada diretamente à sua localização no espaço.

A justificativa para as permutações aleatórias de posição nas matrizes A ou D necessárias para o teste de Mantel está apoiada no fato de que os dados espaciais são não experimentais e, desta forma, a inferência é um tipo de abordagem exploratória dos dados espaciais, cujo objetivo é determinar estruturas e padrões (Jacquez, op. cit.). Além disso, o critério de ordenação dos dez primeiros municípios em ordem alfabética, no exemplo adotado aqui, é por si só casual.

Mais uma vez apresenta-se uma figura que resume esta etapa metodológica, trata-se da figura 3.8 abaixo:

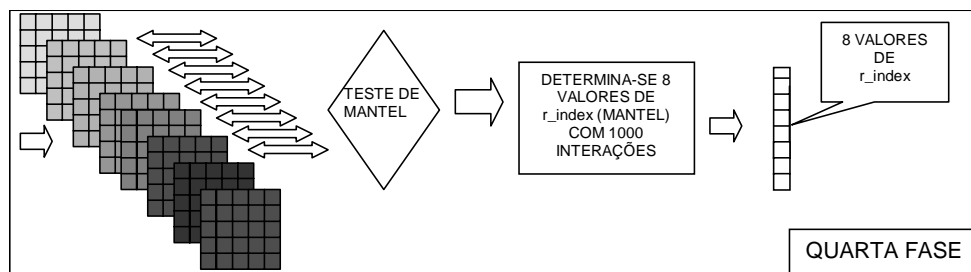


Figura 3.8 – Quarta fase da metodologia, análise da similaridade dos atributos.

3.4 ANÁLISE DE AGRUPAMENTOS

3.4.1 ESTATÍSTICA MULTIVARIADA: ANÁLISE DE AGRUPAMENTOS

Análise de agrupamento é a designação de um grupo de técnicas multivariadas que agrupa objetos a partir e através de suas características (Hair, 1998).

No caso em estudo, os objetos são os municípios do estado do Rio Grande do Sul representados por um conjunto de atributos. O objetivo da análise de agrupamento, neste caso, é estabelecer a similaridade com relação às variáveis que quantificam estes atributos pré-selecionados. Portanto, o conjunto das variáveis utilizadas para comparação define o caráter dos objetos. Deste modo, a análise de agrupamento é a única técnica multivariada que não estima a variação empiricamente, mas utiliza a variação especificada pelo pesquisador.

Análise de agrupamento é uma técnica exploratória que não serve para inferências estatísticas, a partir de uma amostra, extensiva a toda a população. As soluções obtidas não são únicas, o que vale dizer que uma dada regionalização conseguida por uma interação do modelo não é exclusiva. Os objetos formadores do agrupamento, isto é, os municípios que constituem uma dada região obtida por análise de agrupamento depende de vários elementos do processo, isto significa que múltiplas soluções podem ser alcançadas pela variação de um ou mais elementos (Hair, op. cit.). Portanto, a solução de agrupamento é dependente do conjunto de variáveis usadas como base para a medida de similaridade. Por exemplo, se uma região é gerada a partir de variáveis para os atributos população, saúde, educação, agropecuária e, no momento, seguinte inclui-se variáveis relativas a meio ambiente, a regionalização resultante será outra que não a primeira quando o conjunto de variáveis era menor. Desta forma, deve-se ter em mente que a decisão que envolve a escolha do caráter da região, enquanto seleção de atributos e variáveis pertinentes, é determinante sobre as soluções regionais resultantes.

3.4.2 DENDOGRAMA

Um dendograma é uma representação gráfica em forma de árvore que expressa de modo hierárquico a relação entre os objetos, desde o nível mais baixo, em que cada qual pode ser considerado um agrupamento individual, até o nível em que todos os objetos formam um único agrupamento. Esta representação se dá ao longo de um eixo em que os

níveis de agrupamento são determinados por medidas de similaridade. A seguir, no item 3.4.3 serão apresentados um série de exemplos de dendogramas.

3.4.3 CLUSTER – AGRUPAMENTOS HIERÁRQUICOS

A técnica de análise de agrupamento não estabelece a priori o número de grupos ou mesmo a estrutura de agrupamento. Tem por base medidas de similaridade ou distâncias que são informações inerentes aos próprios dados, a partir dos quais e através da elucidação de suas formas de associação, estabelecem critérios de aglomeração (Johnson, 1998).

Os métodos de agrupamentos hierárquicos são de duas naturezas que se desenvolvem em sentidos opostos: um divide, o outro unifica. O método de divisão hierárquica em que um só grupo é dividido em dois, de modo que os objetos de um subgrupo sejam o mais dissimilares dos objetos do outro, e assim a subdivisão continua até que, no limite, cada objeto é um grupo. O método de agrupamento hierárquico que parte dos objetos unitários que são agrupados em conjuntos mais similares que, por sua vez são reagrupados, por determinação de suas similaridades (Johnson op. cit.). Tanto uma técnica como a outra podem ser expressas sob a forma de dendogramas, os quais são formas gráficas que através de linhas retas que unem os objetos demonstram suas vinculações em vários níveis hierárquicos.

Como é objetivo deste trabalho desenvolver um modelo de regionalização cujo propósito é agrupar objetos por suas similaridades, supõe-se que o método de agrupamento hierárquico se coaduna melhor com este fim.

Existem pelo menos cinco métodos de agrupamento hierárquico mais usuais (Hair, op. cit.) que são: conexão única, conexão completa, conexão média, método de Ward e método dos centróides. A diferença entre eles reside na forma como a distância ou similaridade é medida.

Todos estes métodos seguem um algoritmo básico que pode ser descrito em quatro passos (Johnson, op. cit.):

1. Elabora-se uma matriz simétrica de distâncias, $D = \{d_{ij}\}$ de $n \times n$ objetos onde n é o número total de objetos a serem agrupados.

2. Na matriz de distâncias encontra-se a menor, $\min (d_{ij})$, e cria-se um par de objetos 'u' e 'v', formando um agrupamento mais similar (uv).
3. Atualiza-se a matriz D, eliminando-se as linhas e colunas correspondentes a 'u' e 'v' e adiciona-se uma linha e coluna que estabelece as distâncias entre o agrupamento (uv) e os demais objetos.
4. Repete-se os passos 2 e 3 por n-1 vezes, ao final, o algoritmo conclui quando todos os objetos formarem um só agrupamento. É preciso gravar o rótulo dos agrupamentos unidos e os níveis em que acontece a união.

A partir da Tabela 3.5 será exemplificado cada um dos cinco métodos acima. Toma-se para exemplificar a conexão única, os primeiros cinco municípios gaúchos em ordem alfabética. Será feita uma simplificação na tabela abaixo de modo que o exemplo seja mais claro.

Tabela 3.8 – Tabela de distâncias para cinco municípios gaúchos (1ª parte da tabela 3.6).

MUNICÍPIOS	Água Santa	Agudo	Ajuricaba	Alecrim	Alegrete
Água Santa	0,00000	1,89817	1,73672	2,77395	4,08306
Agudo		0,00000	1,50093	2,50282	2,55481
Ajuricaba			0,00000	1,14878	2,53934
Alecrim				0,00000	2,36027
Alegrete					0,00000

Os valores circulados indicam valores mínimos que estabelecem, segundo o método, quais as linha e coluna que serão eliminadas para continuar afinando os agrupamentos.

$$\begin{array}{c}
 a \\
 b \\
 c \\
 d \\
 e
 \end{array}
 \left[
 \begin{array}{ccccc}
 a & b & c & d & e \\
 0 & 1,90 & 1,74 & 2,77 & 4,08 \\
 & 0 & 1,50 & 2,50 & 2,55 \\
 & & 0 & \textcircled{1,15} & 2,54 \\
 & & & 0 & 2,36 \\
 & & & & 0
 \end{array}
 \right]$$

$$\min (d_{ij}) = d_{cd} = 1,15$$

$$d_{(cd)a} = \min \{d_{ca}, d_{da}\} = \min \{1,74; 2,77\} = 1,74$$

$$d_{(cd)b} = \min \{d_{cb}, d_{db}\} = \min \{1,50; 2,50\} = 1,50$$

$$d_{(cd)e} = \min \{d_{ce}, d_{de}\} = \min \{2,54; 2,36\} = 2,36$$

	(cd)	a	b	e
(cd)	0	1,74	1,50	2,36
a		0	1,90	4,08
b			0	2,55
e				0

$$d_{(bcd)a} = \min \{d_{(cd)a}, d_{ba}\} = \min \{1,74; 1,90\} = 1,74$$

$$d_{(bcd)e} = \min \{d_{(cd)e}, d_{be}\} = \min \{2,36; 2,55\} = 2,36$$

	(bcd)	a	e
(bcd)	0	1,74	2,36
a		0	4,08
e			0

$$d_{(abcd)e} = \min \{d_{(bcd)e}, d_{ae}\} = \min \{2,36; 4,08\} = 2,36$$

	(abcd)	e
(abcd)	0	2,36
e		0

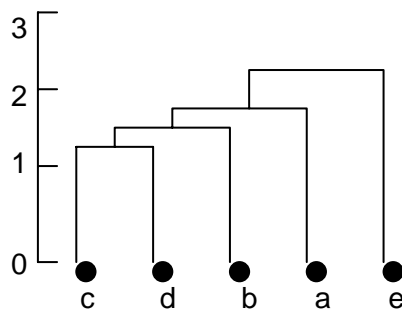


Figura 3.9 – Dendrograma de conexão única para cinco municípios gaúchos.

Para exemplificar a conexão completa retoma-se a matriz simplificada do exemplo anterior:

	a	b	c	d	e
a	0	1,90	1,74	2,77	4,08
b		0	1,50	2,50	2,55
c			0	1,15	2,54
d				0	2,36
e					0

$$d_{(cd)a} = \max \{d_{ca}, d_{da}\} = \max \{1,74; 2,77\} = 2,77$$

$$d_{(cd)b} = \max \{d_{cb}, d_{db}\} = \max \{1,50; 2,50\} = 2,50$$

$$d_{(cd)e} = \max \{d_{ce}, d_{de}\} = \max \{2,54; 2,36\} = 2,54$$

$$\begin{array}{c}
 \begin{matrix} & (cd) & a & b & e \\
 (cd) & \left[\begin{array}{cccc}
 0 & 2,77 & 2,50 & 2,54 \\
 a & & 0 & \textcircled{1,90} & 4,08 \\
 b & & & 0 & 2,55 \\
 e & & & & 0 \end{array} \right]
 \end{matrix} \\
 d_{(cd)(ab)} = \max \{d_{(cd)a}, d_{b(cd)}\} = \max \{2,77; 2,50\} = 2,77 \\
 d_{(ab)e} = \max \{d_{ae}, d_{be}\} = \max \{4,08; 2,55\} = 4,08 \\
 \begin{matrix} & (cd) & (ab) & e \\
 (cd) & \left[\begin{array}{ccc}
 0 & 2,77 & \textcircled{2,54} \\
 (ab) & & 0 & 4,08 \\
 e & & & 0 \end{array} \right]
 \end{matrix} \\
 d_{(cde)(ab)} = \max \{d_{(cd)(ab)}, d_{(ab)e}\} = \max \{2,77; 4,08\} = 4,08
 \end{array}$$

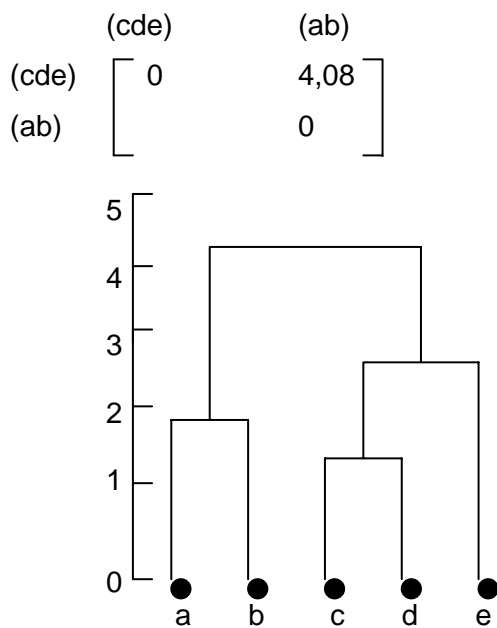


Figura 3.10 – Dendrograma de conexão completa para cinco municípios gaúchos.

Pela comparação das figuras 3.9 e 3.10, verifica-se que no primeiro dendrograma para conexão única, após a formação do primeiro agrupamento, todos os demais elementos são adicionados um a um, formando a cada vez um novo agrupamento com o primeiro já existente. Já no segundo dendrograma, figura 3.10, para conexão completa, há a formação de dois grupos independentes e posterior inclusão do último elemento a um dos dois grupos. Obedecem, portanto, leis de formação diversas.

A conexão média utiliza como critério de agrupamento a média das distâncias de todos os objetos de um grupo até todos os demais, assim este método tende a agrupar conjuntos com pequena variação, isto é, com variância semelhante (Hair, op.cit.).

Seguindo com o exemplo dos cinco municípios gaúchos tem-se:

	a	b	c	d	e
a	0	1,90	1,74	2,77	4,08
b		0	1,50	2,50	2,55
c			0	1,15	2,54
d				0	2,36
e					0

$$d_{(cd)a} = \text{med} \{d_{ca}, d_{da}\} = \text{med} \{1,74; 2,77\} = 2,26$$

$$d_{(cd)b} = \text{med} \{d_{cb}, d_{db}\} = \text{med} \{1,50; 2,50\} = 2,00$$

$$d_{(cd)e} = \text{med} \{d_{ce}, d_{de}\} = \text{med} \{2,54; 2,36\} = 2,45$$

	(cd)	a	b	e
(cd)	0	2,26	2,00	2,45
a		0	1,90	4,08
b			0	2,55
e				0

$$d_{(cd)(ab)} = \text{med} \{d_{ca}, d_{cb}, d_{da}, d_{db}\} = \text{med} \{1,74; 1,50; 2,77; 2,5\} = 2,13$$

$$d_{(ab)e} = \text{med} \{d_{ae}, d_{be}\} = \text{med} \{4,08; 2,55\} = 3,32$$

	(cd)	(ab)	e
(cd)	0	2,13	2,45
(ab)		0	3,32
e			0

$$d_{(cdab)e} = \text{med} \{d_{ce}, d_{de}, d_{ae}, d_{be}\} = \text{med} \{2,54; 2,36; 4,08; 2,55\} = 2,88$$

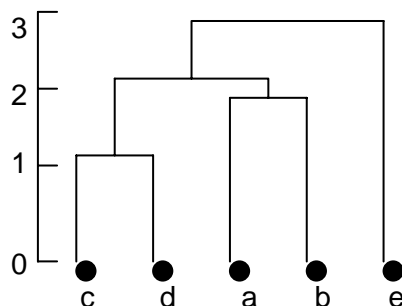


Figura 3.11 – Dendrograma de conexão média para cinco municípios gaúchos.

Neste terceiro dendrograma houve inicialmente a formação de dois agrupamentos que, por sua vez agrupam-se entre eles para somente na última etapa incluir o município

'e' (Alegrete) que, em todas as técnicas exemplificadas até aqui, sempre é o objeto mais dissimilar.

Observe-se que em nenhuma situação acima os dendogramas resultam iguais. Este fato aponta para a necessidade de critérios adicionais para efeito de eleger o melhor resultado de agrupamento hierárquico quando da utilização de técnicas de agrupamento hierárquico.

O método de Ward baseia-se em um processo chamado de menor perda de informação (Johnson, op.cit.). Utiliza o critério do erro da soma dos quadrados, ESS (*error sum of squares*), cujo cálculo é dado pela soma dos quadrados dos desvios em relação à média, como definido na expressão abaixo:

$$ESS = \sum_{i=1}^n (x_i - \mu) (x_i - \mu) \quad [3.46]$$

onde

ESS é o erro da soma dos quadrados;

n é o número total de objetos;

x_i é uma medida multivariada associada ao i-ésimo objeto;

μ é a média de todas as medidas x_i .

Segue-se com o exemplo dos cinco municípios gaúchos e sua matriz simplificada de distâncias.

a	0	1,90	1,74	2,77	4,08
b		0	1,50	2,50	2,55
c			0	1,15	2,54
d				0	2,36
e					0

$$\text{média } (d_{ij}) = \mu (d_{ij}) = 2,31$$

a	0	0,17	0,32	0,21	3,13
b		0	0,66	0,04	0,06
c			0	1,34	0,05
d				0	0,002
e					0

$$ESS_5 = 5,982$$

$$ESS_{(de)a} = \sum [d_{da} - \mu (d_{ij})]^2 + [d_{ea} - \mu (d_{ij})]^2$$

$$ESS_{(de)a} = (2,77 - 2,31)^2 + (4,08 - 2,31)^2 = 2,57$$

$$ESS_{(de)b} = \sum [d_{db} - \mu (d_{ij})]^2 + [d_{eb} - \mu (d_{ij})]^2$$

$$ESS_{(de)b} = (2,50 - 2,31)^2 + (2,55 - 2,31)^2 = 0,09$$

$$ESS_{(de)c} = \sum [d_{dc} - \mu (d_{ij})]^2 + [d_{ec} - \mu (d_{ij})]^2$$

$$ESS_{(de)c} = (1,74 - 2,31)^2 + (1,15 - 2,31)^2 = 1,67$$

	a	b	c	(de)
a	0	0,17	0,32	2,57
b		0	0,66	0,09
c			0	1,67
(de)				0

$$ESS_4 = 5,48$$

$$ESS_{(bde)a} = \sum [d_{ba} - \mu (d_{ij})]^2 + [d_{da} - \mu (d_{ij})]^2 + [d_{ea} - \mu (d_{ij})]^2$$

$$ESS_{(bde)a} = (1,90 - 2,31)^2 + (2,77 - 2,31)^2 + (4,08 - 2,31)^2 = 2,74$$

$$ESS_{(bde)c} = \sum [d_{bc} - \mu (d_{ij})]^2 + [d_{dc} - \mu (d_{ij})]^2 + [d_{ec} - \mu (d_{ij})]^2$$

$$ESS_{(bde)c} = (1,50 - 2,31)^2 + (1,74 - 2,31)^2 + (1,15 - 2,31)^2 = 2,22$$

	a	c	(bde)
a	0	0,32	2,74
c		0	2,22
(bde)			0

$$ESS_3 = 5,28$$

$$ESS_{(bde)(ac)} = \sum [d_{ba} - \mu (d_{ij})]^2 + [d_{bc} - \mu (d_{ij})]^2 + [d_{da} - \mu (d_{ij})]^2 + [d_{dc} - \mu (d_{ij})]^2 + [d_{ea} - \mu (d_{ij})]^2 + [d_{ec} - \mu (d_{ij})]^2$$

$$ESS_{(bde)(ac)} = (1,90 - 2,31)^2 + (1,50 - 2,31)^2 + (2,77 - 2,31)^2 + (1,15 - 2,31)^2 + (4,08 - 2,31)^2 + (2,54 - 2,31)^2 = 5,35$$

	(ac)	(bde)
(ac)	0	5,35
(bde)		0

$$ESS_2 = 5,35$$

5.35	
------	--

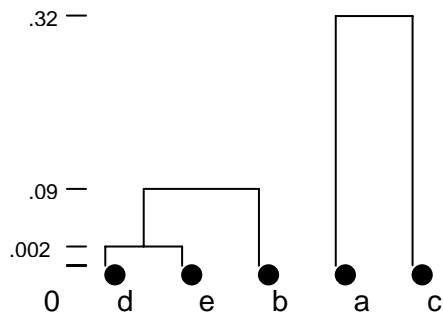


Figura 3.12 – Dendrograma do método de Ward para cinco municípios gaúchos.

A representação do dendrograma acima seria mais exata caso a escala fosse logarítmica porque as ordens de grandeza dos diversos níveis são muito afastadas, especialmente nos níveis superiores, isto é, neste ponto, os agrupamentos são antes um cumprimento do método do que uma correlação significativa.

O quinto e último método de agrupamento hierárquico a ser abordado nesta seção é o método do centróide, onde a distância entre os agrupamentos é dado pela distância euclidiana entre os seus centros geométricos, que são a média das coordenadas de seus pontos componentes. Cada vez que se forma um novo agrupamento, ele gera consigo uma nova posição do centróide (Hair, op. cit.). Esse método é muito utilizado na áreas de Biologia, contudo, pode gerar resultados confusos e desorganizados, chamados de inversões que ocorrem sempre que a distância entre um par agrupado for menor do que a distância entre um outro par consolidado anteriormente. No dendrograma este fenômeno poderia ser expresso de duas formas: ou pelo aparecimento de um cruzamento de linhas (*crossover*) ou pela mudança da escala do eixo com a inclusão de uma marcação não monotônica (Johnson, op. cit.). Apesar das inversões, o método dos centróides é o menos afetado pelo surgimento de valores fora de série (*outliers*).

3.4.4 INTERPRETAÇÃO DO MÉTODO DE AGRUPAMENTO HIERÁRQUICO

A interpretação do método de agrupamento hierárquico passa por três etapas bem definidas: a primeira procura identificar características marcantes que possam rotular com clareza os agrupamentos resultantes do método; a segunda busca confrontar os dados originais através do método do centróide (Hair, op.cit.); a terceira testa a estabilidade dos agrupamentos obtidos pela inclusão de pequenas perturbações nos dados, de modo que, após a inserção destes erros, os agrupamentos resultantes do método de agrupamento hierárquico deveriam corresponder a aqueles anteriores às perturbações (Johnson, op. cit.).

Para o caso do presente modelo de regionalização e de forma complementar, poderia ser adicionado ao processo de interpretação acima, um conjunto de quatro itens de verificação:

1. Criar perfis médios de variação para cada grupo de variáveis (população, saúde, emprego, etc.) usando um modelo normal de distribuição dos dados.
2. Desenhar as relações mais estáveis, através de expressões matemáticas, entre os conjuntos de dados para cada grupo original de variáveis.
3. Verificar se, nos agrupamentos finais, estas relações são mantidas ou não.
4. Utilizar os índices de correlação espacial, calculados sobre os agrupamentos obtidos pelo método, como forma de validação dos resultados.

A interpretação dos resultados não deverá descuidar do fato de que, na maioria dos métodos de agrupamento hierárquico, as fontes de erro e variações não são rigorosamente consideradas, o que significa dizer que é um método sensível a valores fora de série (*outliers*). Além disso, não há formas previstas de redistribuição de objetos impropriamente agrupados, portanto, é razoável aplicar vários métodos de agrupamento aglomerativo e utilizar, para cada um deles, dois ou três modos de cálculo das distâncias ou similaridades entre os objetos (Johnson, op. cit.). Em qualquer dos casos, ajustes deverão ser feitos em direção a uma formação de agrupamentos mais homogêneos sempre que houver incongruência, inseqüência ou inconsistência nos resultados oferecidos pelos métodos em relação aos dados originais.

Finalmente, a solução de agrupamento obtida deverá passar por um processo de validação e submeter-se a um recorte que trace os perfis de cada agrupamento.

A questão da validação passa por algumas características que os agrupamentos finais devem apresentar, tais como representatividade em relação ao universo total, possibilidade de generalização para outros objetos e relativa estabilidade ao longo do

tempo (Hair, op. cit.). A característica de estabilidade, contudo, precisa ser levada em consideração com algum cuidado neste trabalho, por causa de duas tendências opostas que precisam estar presentes: os planos dos administradores públicos só se tornam viáveis se houver um tempo compatível de execução, portanto, é necessária uma certa estabilidade regional, em oposição à proposta de que para cada atualização do modelo, que pode ser feita a cada ano para o qual se possuem dados, uma nova fisionomia ou configuração seria resultado final do método de agrupamento, o que garante a característica de dinamismo objetivada inicialmente.

Entre as formas de abordar a validação da solução de agrupamento existe uma que, provavelmente, é melhor adequada ao modelo de regionalização aqui proposto. Trata-se de utilizar variáveis de controle que não fazem parte daquelas selecionadas inicialmente para gerar os agrupamentos, e cuja variação é bem conhecida através dos agrupamentos. Outra forma seria dividir os dados em dois grupos, submetê-los à divisão hierárquica aglomerativa e comparar os resultados (Hair, op.cit.).

Em relação ao traço dos perfis dos agrupamentos significa caracterizar cada agrupamento de modo a estabelecer suas diferenças mais significativas. Esta é uma questão que pode ser bem resolvida por análise discriminante. Em geral, para isso, são utilizados outros dados que não aqueles que fizeram parte do processo de aglomeração dos agrupamentos, tais como, dados demográficos e padrões de consumo. Estes dados, na análise discriminante, tornam-se as variáveis independentes, enquanto que os agrupamentos finais constituem a variável dependente (Hair, op.cit.).

Resumindo esta quinta fase, da mesma forma que as anteriores, tem-se a figura 3.13 que procura representar em forma de esquema a etapa de agrupamento das regiões.

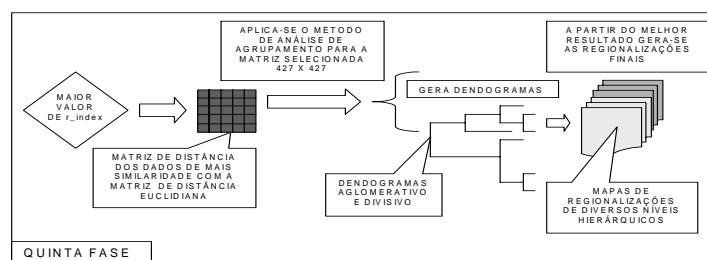


Figura 3.13 – Quinta fase metodológica, agrupamento de regiões similares.
3.5 ANÁLISE COMPARATIVA COM OUTRAS DIVISÕES REGIONAIS

A última etapa do método visa buscar relações com as demais divisões regionais já estabelecidas. Entre elas pode-se citar as mesorregiões e microrregiões do IBGE, as Delegacias Regionais de Saúde, as Delegacias de Educação e os Conselhos Regionais de Desenvolvimento.

Do mesmo modo, a figura que resume esta fase final é apresentada abaixo:

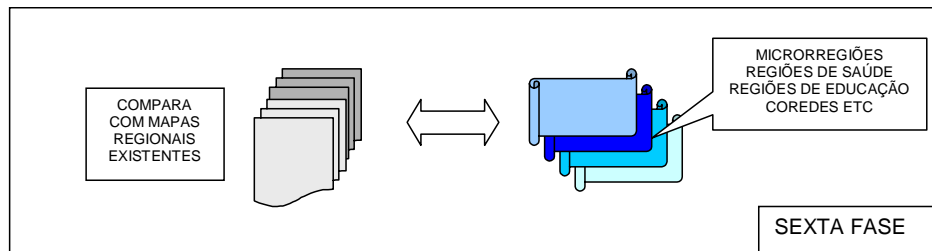


Figura 3.14 – Sexta fase metodológica, análise comparativa com outras regionalizações.



CAPÍTULO 4

ESTUDO DE CASO:
RIO GRANDE DO SUL

4. ESTUDO DE CASO: ESTADO DO RIO GRANDE DO SUL

Neste capítulo aplica-se a metodologia, explicada no capítulo anterior, aos dados do estado do Rio Grande do Sul. Inicialmente é feita a delimitação espacial da área de aplicação do estudo de caso, definindo-se a divisão política adotada. Gera-se, então, a matriz de distâncias “euclidianas” para todos os municípios. A seguir, utiliza-se oito coeficientes de similaridade para os dados que se expressam como matrizes de distâncias. Logo após, realiza-se o teste de Mantel para determinar o mais similar em relação à matriz de distâncias. Uma vez escolhido o melhor resultado aplica-se as técnicas de análise de agrupamento que dão origem aos dendogramas e por fim visualiza-se os resultados através dos diversos mapas que representam cada nível de agrupamento e, portanto, diferentes níveis de regionalização. Como complemento, pode-se traçar uma análise paralela entre as regionalizações dos diversos níveis hierárquicos resultantes do método, aqui aplicado, com algumas divisões regionais já existentes.

Vale ressaltar o caráter dinâmico dos resultados, tendo em vista que outro conjunto de informações, ou outro conjunto de coeficientes de similaridade, ou ainda outra base cartográfica (467 municípios) gerariam resultados, provavelmente, muito diferentes. Este fato, dentro das metas desta dissertação, ao invés de ser uma falha, é uma característica de reforço do tratamento da questão da regionalização de uma forma não cristalizada e estanque.

4.1 FLUXOGRAMA DE APLICAÇÃO AO ESTUDO DE CASO

A seguir apresenta-se esquematicamente a seqüência de aplicação ao estudo de caso dividido em seis fases, passo a passo. Após o que descreve-se de modo subsequente cada um dos passos apresentados no esquema. Trata-se da figura 4.1 abaixo, composta de seis quadros apresentados isolados no capítulo 3, que procura elucidar de que modo chega-se às regionalizações aqui apresentadas como resultado do método desenvolvido.

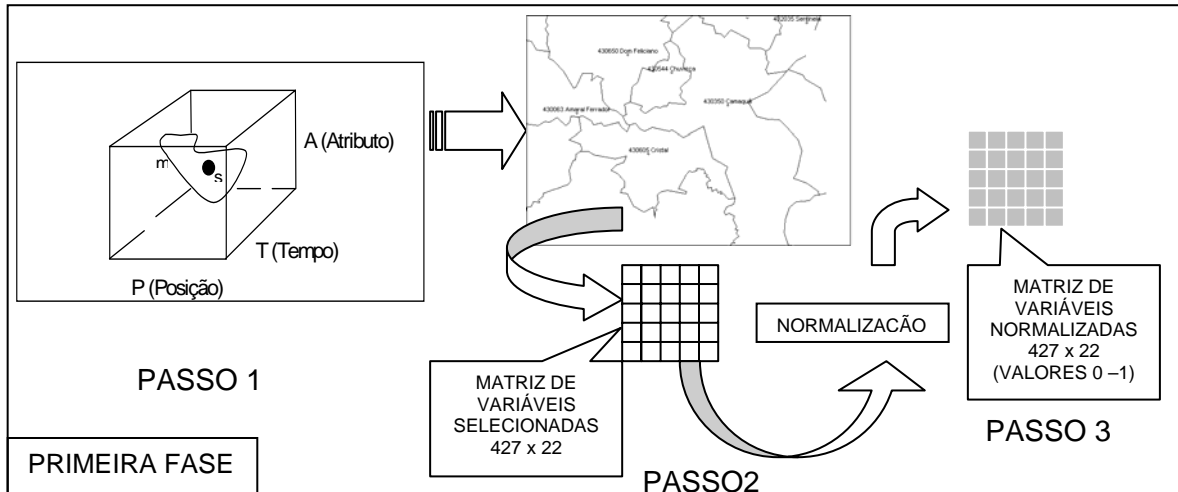


Figura 4.1.1 – Pré processamento.

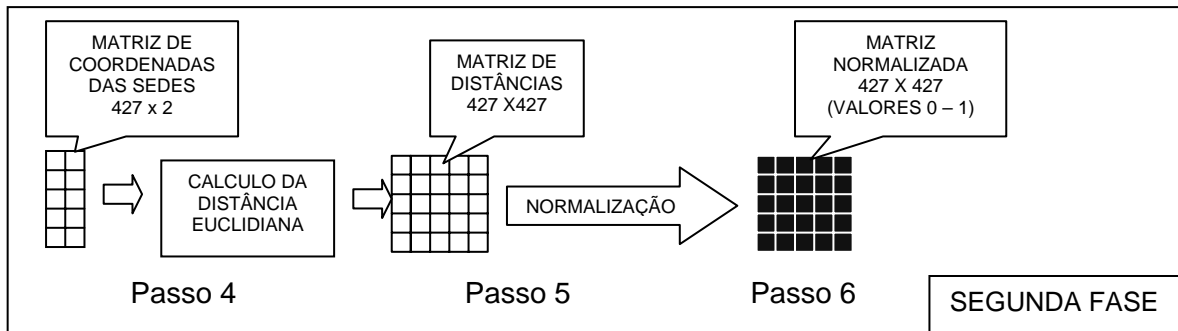


Figura 4.1.2 – Similaridade espacial.

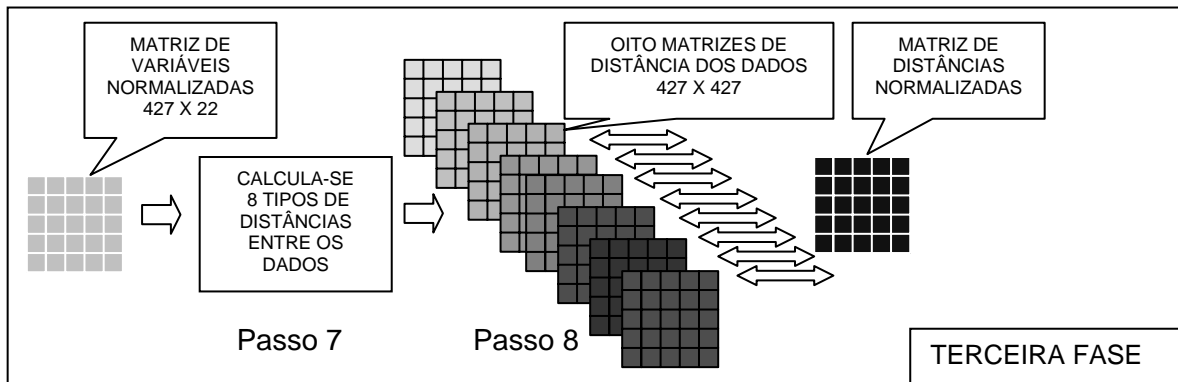


Figura 4.1.3 – Análise da similaridade das características medidas (atributos).

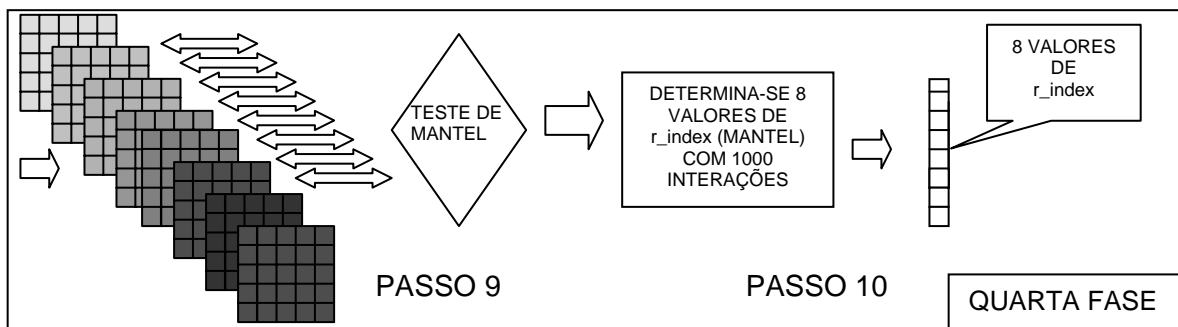


Figura 4.1.4 – Análise das distâncias de agrupamento com correlação espacial – Mantel.

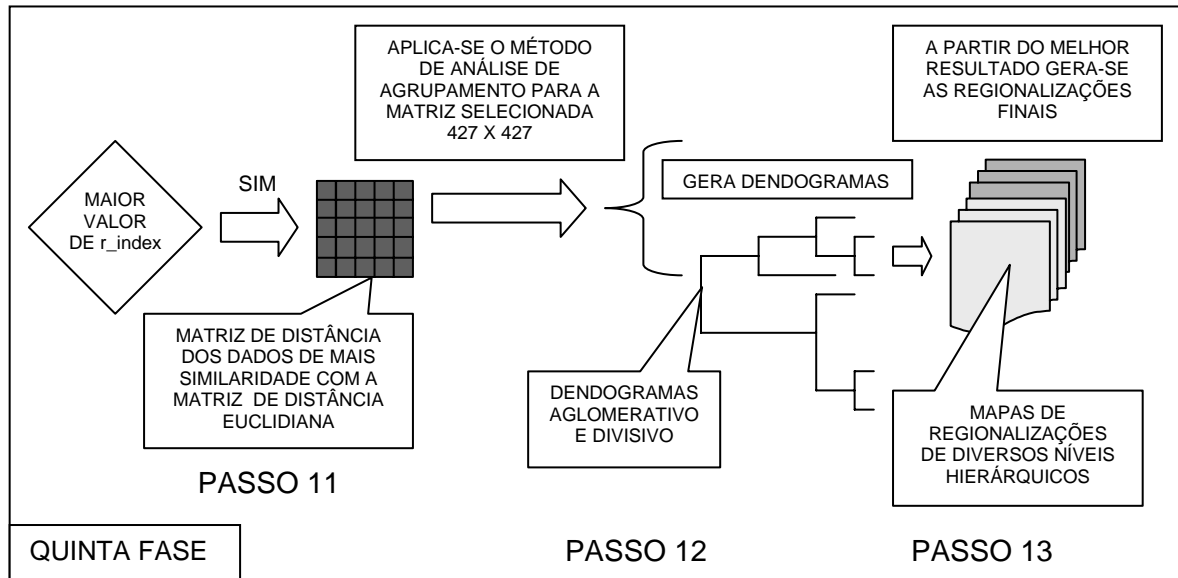


Figura 4.1.5 – Agrupamento de regiões similares mantendo a estrutura espacial.

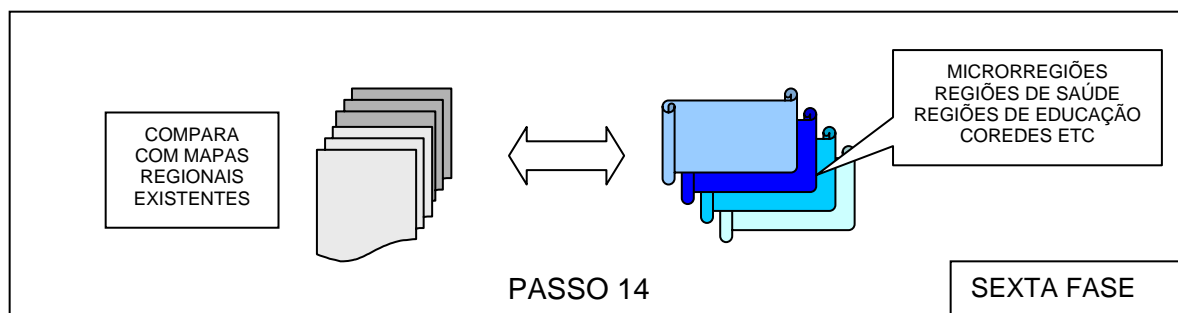


Figura 4.1.6 – Análise comparativa com as demais formas de regionalização.

Figura 4.1 – Fluxograma esquemático de aplicação ao estudo de caso.

Os passos são os seguintes, como pode ser acompanhado na figura acima:

- Primeira fase: pré processamento.
 1. Assume-se que os municípios serão representados pelas coordenadas de suas sedes e que os atributos de cada município serão dados sobre população, emprego, saúde, educação e economia para o ano de 1996, e também dados sobre agropecuária e meio ambiente para o ano de 1995, além de características geofísicas dos municípios, perfazendo um total de 22 variáveis. Sobre a origem e forma de cálculo desta variáveis consulte-se o capítulo 3 na parte sobre análise exploratória.
 2. Monta-se uma matriz composta por 427 linhas (municípios) e 22 colunas (variáveis).
 3. Executa-se a normalização desta matriz confinando os valores entre 0 e 1, o que guarda a relação entre eles e permite comparações.

- Segunda fase: similaridade espacial.
 4. Aplica-se a fórmula da distância sobre a matriz de coordenadas das sedes.
 5. Gera-se uma matriz de distâncias físicas de 427 x 427.
 6. Normaliza-se os resultados.
- Terceira fase: análise da similaridade das características medidas (atributos).
 7. Entra-se com a matriz normalizada de 427 municípios x 22 variáveis socioeconômicas.
 8. Calcula-se vários (8) tipos de distâncias entre os dados gerando-se matrizes de distância dos dados, com a mesma fórmula de distância euclidiana.
- Quarta fase: análise das distâncias de agrupamento com correlação espacial (Teste de Mantel).
 9. Compara-se cada uma delas em termos de similaridade com a matriz de distâncias físicas, através do teste de Mantel, com 1000 interações.
 10. Entre o valores de Mantel (r_{index}) obtidos seleciona-se o maior que indica a matriz que melhor mantém a estrutura espacial.
- Quinta fase: agrupamento de regiões similares mantendo a estrutura espacial.
 11. Aplica-se a análise de agrupamento para esta matriz pelos métodos aglomerativo e divisivo.
 12. Gera-se dois dendogramas: um aglomerativo e outro divisivo.
 13. Tomando-se por base o melhor dendograma, que neste caso é o divisivo gera-se os mapas de regionalizações para diversos níveis hierárquicos (15).
- Sexta fase: análise comparativa com as demais formas de regionalização.
 14. Compara-se as regionalizações resultantes com algumas divisões regionais existentes.

4.2 CONFIGURAÇÃO DA BASE DE APLICAÇÃO

O estado do Rio Grande do Sul, base de aplicação do estudo de caso, foi tomado com 427 municípios, configuração criada em 1992 e que não inclui as emancipações geradas em 1995, a partir de um arquivo digitalizado e georreferenciado no sistema de coordenadas de Greenwich, latitude e longitude, fornecido pelo sistema Datasus, do Ministério da Saúde (download gratuito em <ftp.datasus.gov.br>). Os dados, já mencionados anteriormente, são relativos às variáveis selecionadas e informadas para os anos de 1995 e 1996 pelo IBGE, através da BIM (Base de Informações Municipais) e FEE (Fundação de Economia e Estatística - RS) para estes 427 municípios. É importante ressaltar que disponível no *síte* DATASUS existem três malhas municipais fornecidas pelo IBGE: malha

de 91 que vigorou em 1991 e 1992; malha de 94 válida de 1993 a 1996 e a de 97, mais atual. Utiliza-se como base a malha de 1994 e como referência a de 1997.

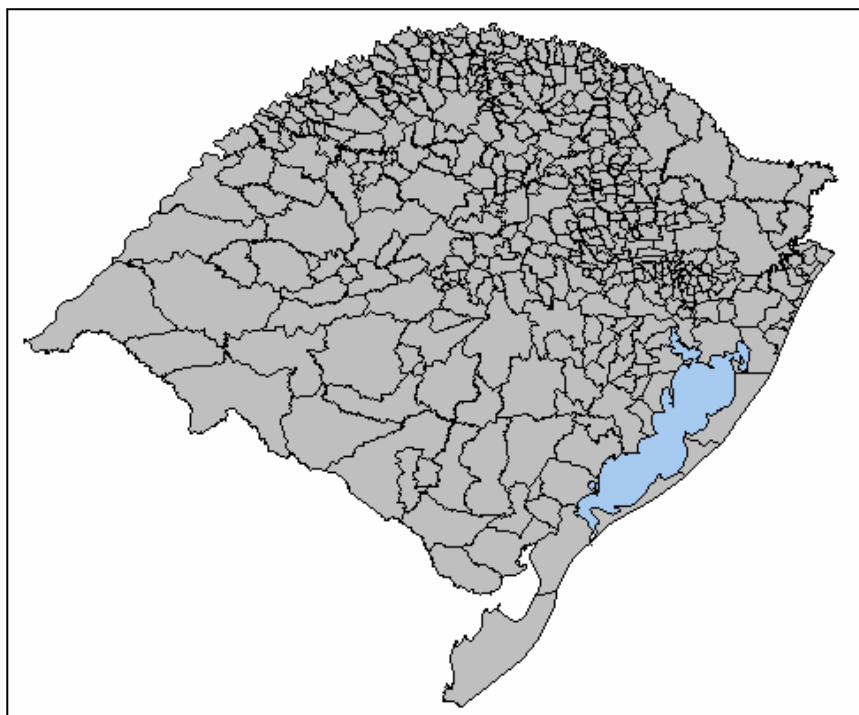


Figura 4.2 – Mapa político do Rio Grande do Sul – malha de 1994 (427 municípios).
(Fonte: IBGE, via DATASUS – Ministério da Saúde).

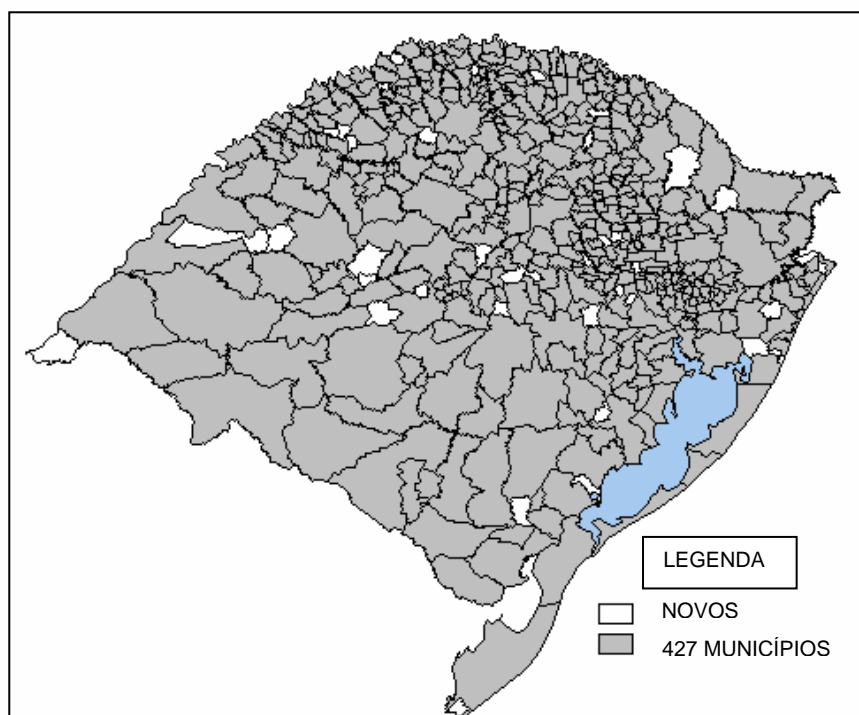


Figura 4.3 – Mapa político do Rio Grande do Sul – malha de 1997 (467 municípios).
(Fonte: IBGE, via DATASUS – Ministério da Saúde).

4.3 MATRIZ DE “DISTÂNCIAS EUCLIDIANAS”

Sobre a matriz de distâncias adotada com base nas latitudes e longitudes das sedes municipais, é importante lembrar que, de fato, não se trata de uma matriz de distâncias físicas, contudo, o emprego da fórmula de distância euclidiana sobre os valores de latitude e longitude conferem aos resultados uma qualidade que guarda a relação entre os municípios como se fossem utilizadas as coordenadas UTM, verdadeiras determinantes das distâncias físicas. Além disso, foi feita a normalização da matriz resultante, o que para efeito desta análise gera resultados equivalentes.

Aplicando a fórmula da distância euclidiana (linha i e linha j) [3.18] sobre a matriz abaixo, de 427 linhas ($n=427$) por 2 colunas ($p=2$), gera-se uma nova matriz de $n=427$ e $p=427$ já normalizada (valores entre o intervalo 0 e 1) pela fórmula [3.16]:

Tabela 4.1 – Matriz de 427 X 2 de latitude e longitude para os municípios do RS.

(Obs.: As duas primeiras colunas, município e id são apenas referenciais).

MUNICÍPIO	ID	LATITUDE	LONGITUDE
AGUA SANTA	1	-28,17294	-52,03367
AGUDO	2	-29,64062	-53,23740
AJURICABA	3	-28,23706	-53,76921
ALECRIM	4	-27,65305	-54,75847
ALEGRETE	5	-29,77666	-55,78859
ALEGRIA	6	-27,82649	-54,05695
ALPESTRE	7	-27,24282	-53,03435
ALTO ALEGRE	8	-28,77107	-52,98774
ALTO FELIZ	9	-29,38859	-51,30746
ALVORADA	10	-29,98740	-51,08350
AMARAL FERRADOR	11	-30,87378	-52,25458
AMETISTA DO SUL	12	-27,35644	-53,17582
ANDRE DA ROCHA	13	-28,62514	-51,56972
ANTA GORDA	14	-28,96887	-52,00288
ANTONIO PRADO	15	-28,85508	-51,27649
ARAMBARE	16	-30,90915	-51,49214
ARATIBA	17	-27,38994	-52,30017
ARROIO DO MEIO	18	-29,40068	-51,94045
ARROIO DO SAL	19	-29,55085	-49,88672
ARROIO DOS RATOS	20	-30,07311	-51,72429
ETC...	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
XANGRI-LA	427	-29,80051	-50,03960

A matriz acima apresenta as duas colunas à direita sombreadas querendo com isto evidenciar que a matriz efetiva de entrada para geração da matriz de distâncias é aquela composta por 427 linhas (uma linha para cada município) e 2 colunas (duas coordenadas de posição geográfica para cada sede municipal).

A matriz normalizada de 427X427 será simétrica em relação à diagonal principal zerada, porque $d_{12} = d_{21}$, $d_{13} = d_{31}$ e assim por diante. Uma janela para os primeiros vinte municípios gaúchos segue abaixo:

Tabela 4.2 – Matriz de distâncias normalizadas para os primeiros vinte municípios RS.

0.000	0.321	0.294	0.470	0.691	0.348	0.231	0.191	0.240	0.347	0.459	0.238	0.110	0.135	0.173	0.472	0.140	0.208	0.432	0.326
0.321	0.000	0.254	0.424	0.433	0.337	0.407	0.153	0.330	0.369	0.267	0.387	0.331	0.238	0.358	0.365	0.413	0.223	0.568	0.266
0.294	0.254	0.000	0.195	0.430	0.085	0.209	0.160	0.460	0.543	0.515	0.180	0.378	0.324	0.435	0.594	0.287	0.367	0.694	0.465
0.470	0.424	0.195	0.000	0.400	0.122	0.300	0.355	0.654	0.737	0.691	0.273	0.564	0.517	0.624	0.781	0.419	0.561	0.885	0.657
0.691	0.433	0.430	0.400	0.000	0.442	0.634	0.504	0.762	0.797	0.627	0.603	0.740	0.655	0.780	0.752	0.716	0.655	1.000	0.690
0.348	0.337	0.085	0.122	0.442	0.000	0.199	0.242	0.535	0.622	0.599	0.169	0.442	0.398	0.502	0.679	0.306	0.447	0.764	0.548
0.231	0.407	0.209	0.300	0.634	0.199	0.000	0.259	0.466	0.570	0.629	0.031	0.341	0.340	0.404	0.673	0.127	0.410	0.661	0.528
0.191	0.153	0.160	0.355	0.504	0.242	0.259	0.000	0.303	0.383	0.377	0.242	0.241	0.170	0.290	0.442	0.261	0.207	0.541	0.307
0.240	0.330	0.460	0.654	0.762	0.535	0.466	0.303	0.000	0.108	0.298	0.467	0.137	0.138	0.090	0.259	0.378	0.107	0.242	0.136
0.347	0.369	0.543	0.737	0.797	0.622	0.570	0.383	0.108	0.000	0.249	0.569	0.245	0.232	0.194	0.171	0.486	0.176	0.216	0.109
0.459	0.267	0.515	0.691	0.627	0.599	0.629	0.377	0.298	0.249	0.000	0.616	0.398	0.325	0.380	0.129	0.590	0.255	0.459	0.163
0.238	0.387	0.180	0.273	0.603	0.169	0.031	0.242	0.467	0.569	0.616	0.000	0.347	0.338	0.410	0.666	0.148	0.404	0.669	0.522
0.110	0.331	0.378	0.564	0.740	0.442	0.341	0.241	0.137	0.245	0.398	0.347	0.000	0.094	0.063	0.387	0.243	0.146	0.325	0.247
0.135	0.238	0.324	0.517	0.655	0.398	0.340	0.170	0.138	0.232	0.325	0.338	0.094	0.000	0.124	0.340	0.272	0.074	0.372	0.193
0.173	0.358	0.435	0.624	0.780	0.502	0.404	0.290	0.090	0.194	0.380	0.410	0.063	0.124	0.000	0.350	0.303	0.146	0.263	0.220
0.472	0.365	0.594	0.781	0.752	0.679	0.673	0.442	0.259	0.171	0.129	0.666	0.387	0.340	0.350	0.000	0.611	0.266	0.356	0.147
0.140	0.413	0.287	0.419	0.716	0.306	0.127	0.261	0.378	0.486	0.590	0.148	0.243	0.272	0.303	0.611	0.000	0.346	0.548	0.465
0.208	0.223	0.367	0.561	0.655	0.447	0.410	0.207	0.107	0.176	0.255	0.404	0.146	0.074	0.146	0.266	0.346	0.000	0.349	0.120
0.432	0.568	0.694	0.885	1.000	0.764	0.661	0.541	0.242	0.216	0.459	0.669	0.325	0.372	0.263	0.356	0.548	0.349	0.000	0.323
0.326	0.266	0.465	0.657	0.690	0.548	0.528	0.307	0.136	0.109	0.163	0.522	0.247	0.193	0.220	0.147	0.465	0.120	0.323	0.000

Evidentemente que a matriz completa tem 427 X 427, uma vez que o cruzamento de todos os municípios tomados dois a dois gera uma matriz quadrada da qual apresenta-se acima, tão somente, uma seção de 20 X 20 dos primeiros municípios em ordem alfabética.

4.4 DETERMINAÇÃO DAS MATRIZES DE DISTÂNCIAS DOS DADOS COM OITO COEFICIENTES DE SIMILARIDADE

Distância é uma função matemática d em $M \times M$ cujos elementos pertencem aos R (conjunto dos números reais) e para quaisquer x , y e z pertencentes a M tem-se: $d(x, y) \geq 0$; $d(x, y) = 0 \Leftrightarrow x = y$; $d(x, y) = d(y, x)$ e $d(x, y) \leq d(x, z) + d(z, y)$.

Dissimilaridade é uma função estatística δ de $N \times N$ em R cujos elementos quaisquer i e j pertencentes a N satisfazem as seguintes condições: $\delta_{ij} \geq 0$; $\delta_{ii} = 0$; $\delta_{ij} = \delta_{ji}$ (ADE-4, 1998).

Combinando as duas definições das funções acima é possível dizer que uma matriz de distâncias é aquela que contém dissimilaridades observadas e que toda matriz de distância é quadrada ($n \times n$), formada por valores positivos, simétricos em relação à diagonal principal que necessariamente será composta por zeros.

Sete distâncias foram calculadas através do ADE-4, cuja formulação foi apresentada no capítulo 3 e que estão indicadas abaixo pela numeração:

1. *City Block = Manhattan – Range* ou D3 de Gower & Legendre (1986) – distância não euclidiana [3.25];
2. *City Block = Manhattan – Standard deviation* de Cain & Harrison – distância não euclidiana [3.27];
3. *Canberra = Lance & Williams* ou D7 de Gower & Legendre – distância não euclidiana [3.30];
4. *Bray-Curtis* ou D8 de Gower & Legendre – distância não euclidiana [3.31];
5. *Gower & Legendre D5* – distância euclidiana [3.32];
6. *Gower & Legendre D9* - distância não euclidiana [3.33];
7. *Gower & Legendre D10* - distância não euclidiana [3.34].

O oitavo coeficiente de similaridade ou fórmula [3.35] de distância não faz parte do módulo de “Distâncias: variáveis quantitativas” do ADE-4, contudo, foi o melhor resultado, aquele de maior similaridade com a matriz de distâncias físicas: Chi-2 (Chi-quadrado).

4.5 TESTE DE MANTEL PARA OITO DISTÂNCIAS

São utilizadas como entradas para o teste de Mantel as matrizes de distâncias dos dados resultantes do módulo de Distâncias do ADE-4 calculadas a partir do item variáveis quantitativas e empregando a formulação indicada acima, que compõe a totalidade de opções do módulo (7 tipos de distâncias).

Utilizando como entrada a matriz dos dados de 427 X 20 (427 municípios por 20 colunas de atributos com valores para cada município) e empregando de cada vez uma opção de distância de tipo 1 ao tipo 7, gera-se sete matrizes de 427 X 427 e procede-se, logo após, a normalização. O Teste de Mantel permite determinar a correlação entre duas matrizes de distâncias ou de similaridades, executando um número determinado de interações aleatórias nos dados, o que foi de 1000 permutações, neste trabalho. Como resultado, obtém-se um “r index” para cada distância, as quais montamos em uma tabela abaixo, juntamente com os demais dados do processos realizados com 1000 permutações para cada distância:

Tabela 4.3 – Resultados do teste de Mantel para 8 distâncias.

ORDEM	MATRIZES	DISTÂNCIA / SIMILARIDADE	INTERAÇÕES	r_index
1	“Distâncias Euclidianas” X Distâncias dos “dados” (427X427)	City Block = Manhattan - Range	1000	0,2634
2		City Block = Manhattan - Standard dev		0,2440
3		Camberra = Lance & Williams		0,1873
4		Bray-Curtis		0,2617
5		Gower & Legendre D5		0,1764
6		Gower & Legendre D9		-0,01506
7		Gower & Legendre D10		0,2053
8		Chi-2		0,4885

Concluindo, verifica-se através do teste de Mantel que a matriz de distâncias entre os dados que é a melhor, porque mantém parcialmente a estrutura espacial, é aquela obtida pela distância Chi-2 com índice $r = 0,4885$. Sabendo-se que r varia de -1 a $+1$ pode-se afirmar que os agrupamentos resultantes guardam uma relação com a estrutura espacial de caráter moderado apenas.

4.6 ANÁLISE DE AGRUPAMENTO PARA A MATRIZ COM MAIS SIMILARIDADE

Tabulando as saídas para os quinze (15) níveis mais representativos de agrupamento pode-se gerar a tabela 4.4 abaixo. A partir do nível 16 em diante os agrupamentos tornam-se desordenados em função das distâncias tornarem-se muito próximas o que ocasiona a formação de dezenas de pequenos grupos esparsos, por esta razão o método considera somente até 14 agrupamentos, uma vez que o primeiro nível inclui todos os municípios (427).

Tabela 4.4 – Descrição de 14 categorias para 15 níveis de agrupamento
(gerada a partir do ADE-4).

CAT	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1 Q	427													
F	1													
2 Q	297	130												
F	,696	,304												
3 Q	96	201	130											
F	,225	,471	,304											
4 Q	96	201	25	105										
F	,225	,471	,058	,246										
5 Q	96	201	24	105	1									
F	,225	,471	,056	,246	,002									
6 Q	96	201	24	105	1									
F	,225	,471	,056	,246	,002									
7 Q	96	201	24	105	1									
F	,225	,471	,056	,246	,002									
8 Q	96	65	101	35	24	38	67	1						
F	,225	,152	,237	,082	,056	,089	,157	,002						
9 Q	72	65	101	35	24	38	67	24	1					
F	,169	,152	,237	,082	,056	,089	,157	,056	,002					
10Q	72	65	101	35	8	38	67	16	24	1				
F	,169	,152	,237	,082	,019	,089	,157	,037	,056	,002				
11Q	72	65	101	35	8	38	67	16	24	1				
F	,169	,152	,237	,082	,019	,089	,157	,037	,056	,002				
12Q	72	65	101	35	8	38	67	16	24	1				
F	,169	,152	,237	,082	,019	,089	,157	,037	,056	,002				
13Q	72	65	101	35	8	38	67	16	24	1				
F	,169	,152	,237	,082	,019	,089	,157	,037	,056	,002				
14Q	72	65	101	35	8	18	44	23	16	40	20	24	1	
F	,169	,152	,237	,082	,019	,042	,103	,054	,037	,094	,047	,056	,002	
15Q	72	25	101	35	8	18	24	23	16	40	20	20	24	1
F	,169	,058	,237	,082	,019	,042	,056	,054	,037	,094	,047	,047	,056	,002

Obs.: Q indica o número de municípios componentes de cada categoria ou região e F a frequência relativa.

4.7 DENDOGRAMAS DAS SOLUÇÕES

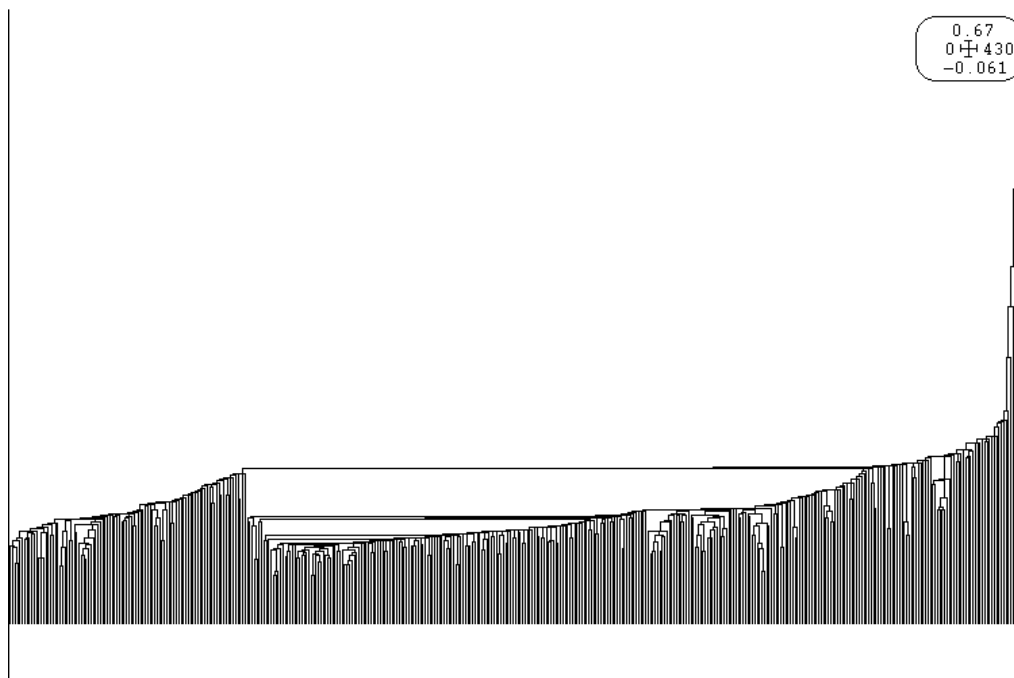


Figura 4.4 - Dendrograma da Análise de agrupamento: método aglomerativo.

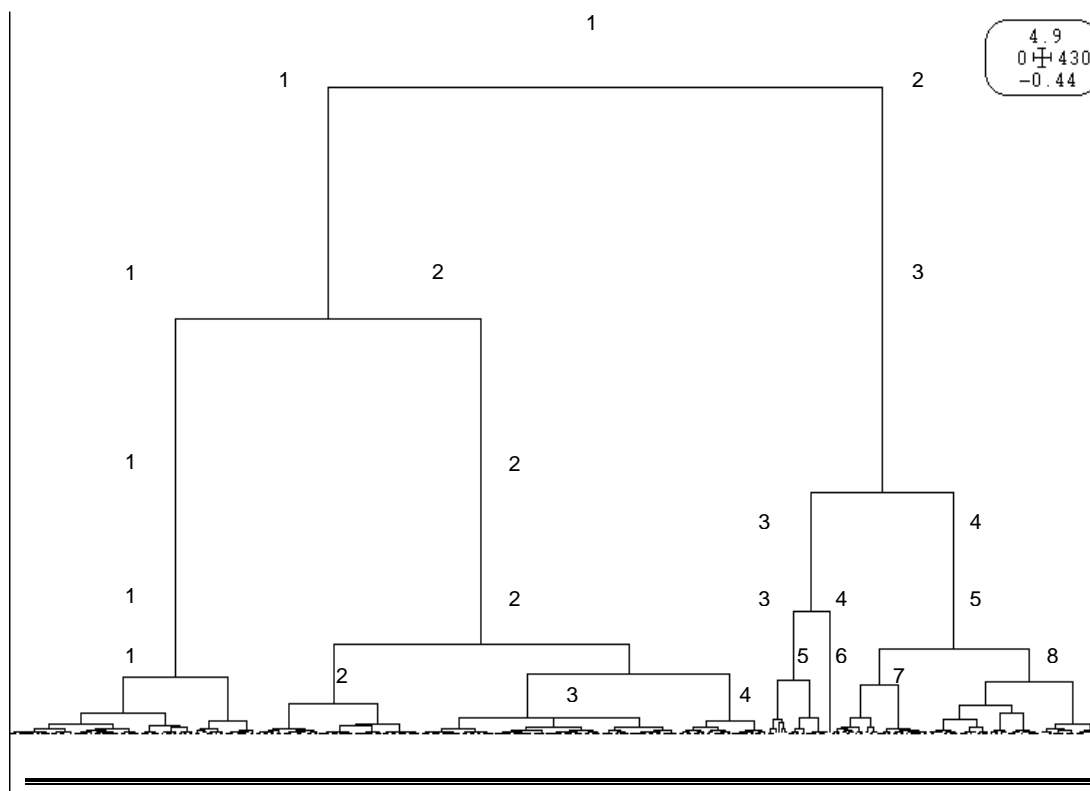


Figura 4.5 - Dendrograma da Análise de agrupamento: método divisivo.

A interpretação das figuras 4.4 e 4.5 acima é indicada pela figura em forma de cruz no canto superior direito dos dendogramas. Ali verifica-se que o eixo horizontal é definido pelo intervalo de 0 a 430 contendo os 427 municípios considerados. O eixo vertical na figura 4.4 varia de $-0,061$ a $+0,67$, enquanto que, na figura 4.5 varia de $-0,44$ até $+4,9$ indicando com isto os níveis hierárquicos de distância entre os dados de maior correlação com a matriz de distâncias físicas, portanto preservando nos agrupamentos alguma estrutura espacial.

Sobre o melhor resultado de similaridade com as distâncias físicas que é representado na figura 4.5, assinalando-se por números justapostos às linhas até o oitavo nível hierárquico (em função da limitação física do desenho), compõe-se os diversos mapas de agrupamentos ou regionalizações em até 14 categorias. Abaixo deste nível não é possível continuar porque tem lugar uma desagregação em função dos municípios apresentarem distâncias muito próximas uns dos outros.

4.7 VISUALIZAÇÃO DOS RESULTADOS: MAPAS REGIONAIS

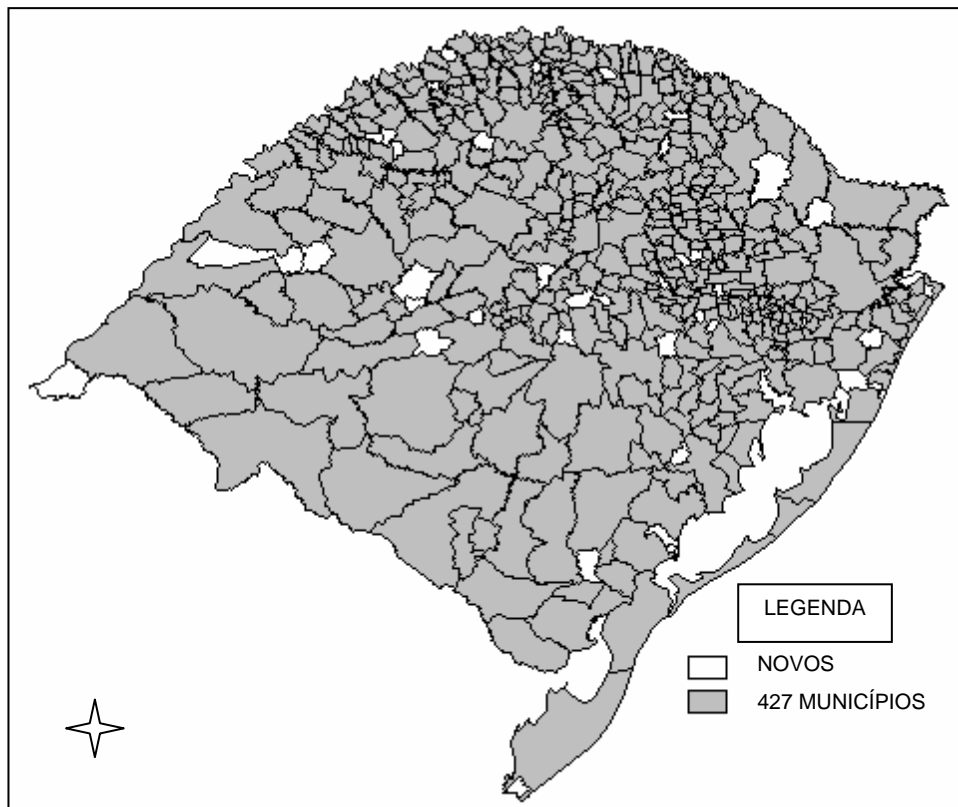


Figura 4.6 – Mapa do Rio Grande do Sul – Nível 1 – (427 + 40 municípios novos) – 1997
Escala 1:250.000

4.7.1 ANÁLISE DO NÍVEL HIERÁRQUICO 2

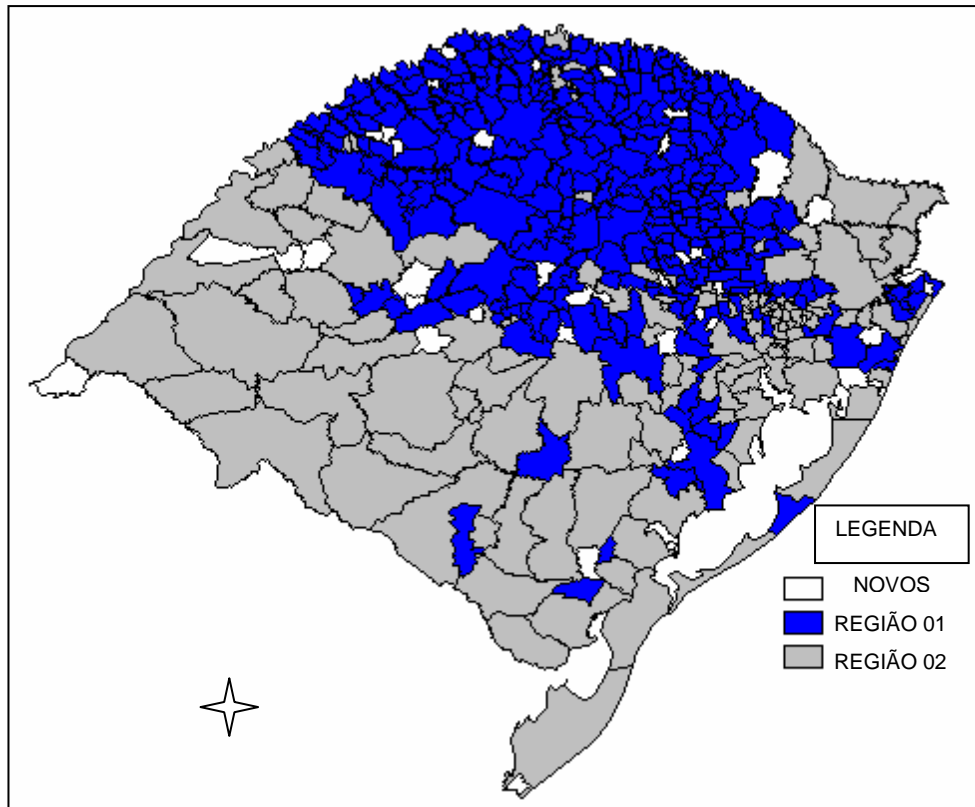


Figura 4.7 – Regionalizações do RS: Nível hierárquico 2
Método dos agrupamentos divisivos – Escala 1:250.000.

Observa-se, em linhas gerais, dois grandes blocos regionais diferenciados e com características próprias.

O primeiro bloco formado por municípios de dimensões pequenas e médias, ao norte, noroeste e centro do estado, além de alguns municípios no entorno da RMPA, designado por região 01 com 297 municípios. O segundo bloco formado pela área sul, sudoeste, leste e nordeste do estado, constituído predominantemente por municípios de grandes e médias áreas chamado de região 02 com 130 municípios. Deve-se fazer o destaque de que a RMPA se alinha, pelas características aferidas pelo modelo, com o bloco sul (região 02) e não com o norte (região 01) como poderia ser esperado.

Cabe destacar ainda a existência de “ilhas” (municípios soltos) de uma região localizadas no interior da outra, o que significa dizer que em termos de similaridade estas ilhas pouco se afinam com o seu entorno, sendo mais dissimilares em relação aos demais municípios que compõem a região onde elas estão encravadas.

4.7.2 ANÁLISE DO NÍVEL HIERÁRQUICO 3

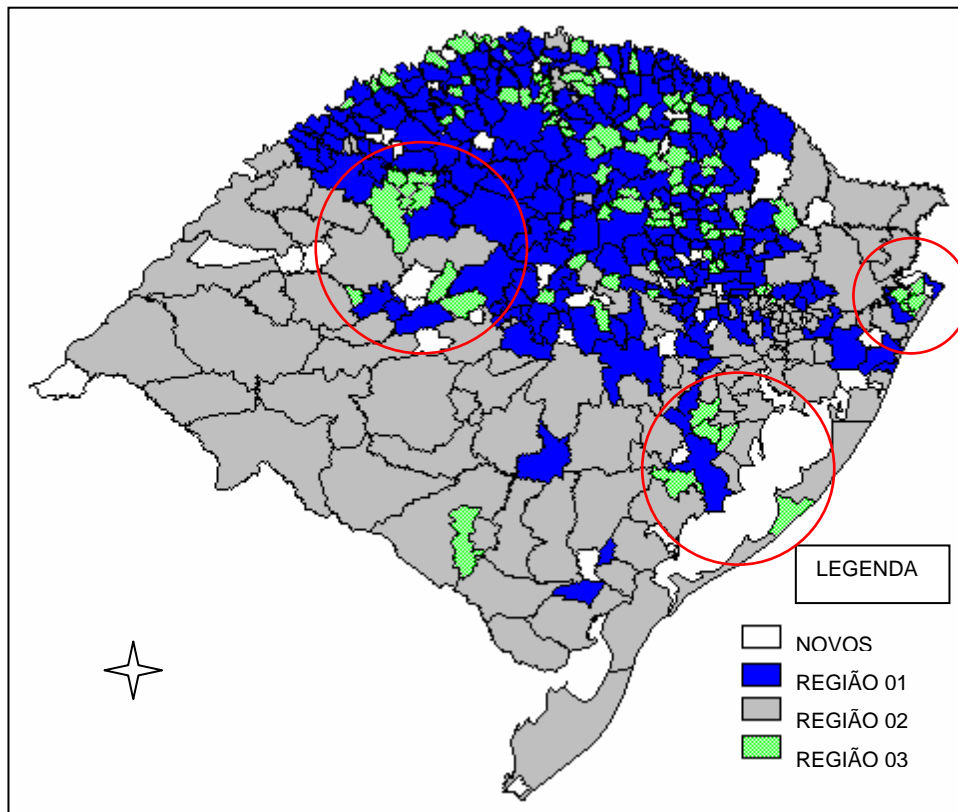


Figura 4.8 – Regionalizações do RS: Nível hierárquico 3
Método dos agrupamentos divisivos – Escala 1:250.000.

Neste nível inicia-se a diferenciação da região 01, com a formação de subgrupos fragmentados de municípios espalhados principalmente pelo norte do estado. Refere-se a esta área como região 03 com 96 municípios de dimensões pequenas. Em relação ao nível hierárquico anterior observa-se a quebra de continuidade espacial desta região derivada da região 01. Note-se que a região 02 permanece inalterada com 130 municípios e a região 01 passa a contar com 201 municípios.

Assinalados por três círculos vermelhos observa-se subgrupos desta região 03 com características peculiares, porque localizam-se em áreas periféricas em relação a maior concentração de municípios ao norte e em zonas de transição entre a região 01 e 02, além de apresentarem alguma continuidade em cada subgrupo. Em relação às duas regiões anteriores esta apresenta bastante descontinuidade.

4.7.3 ANÁLISE DO NÍVEL HIERÁRQUICO 4

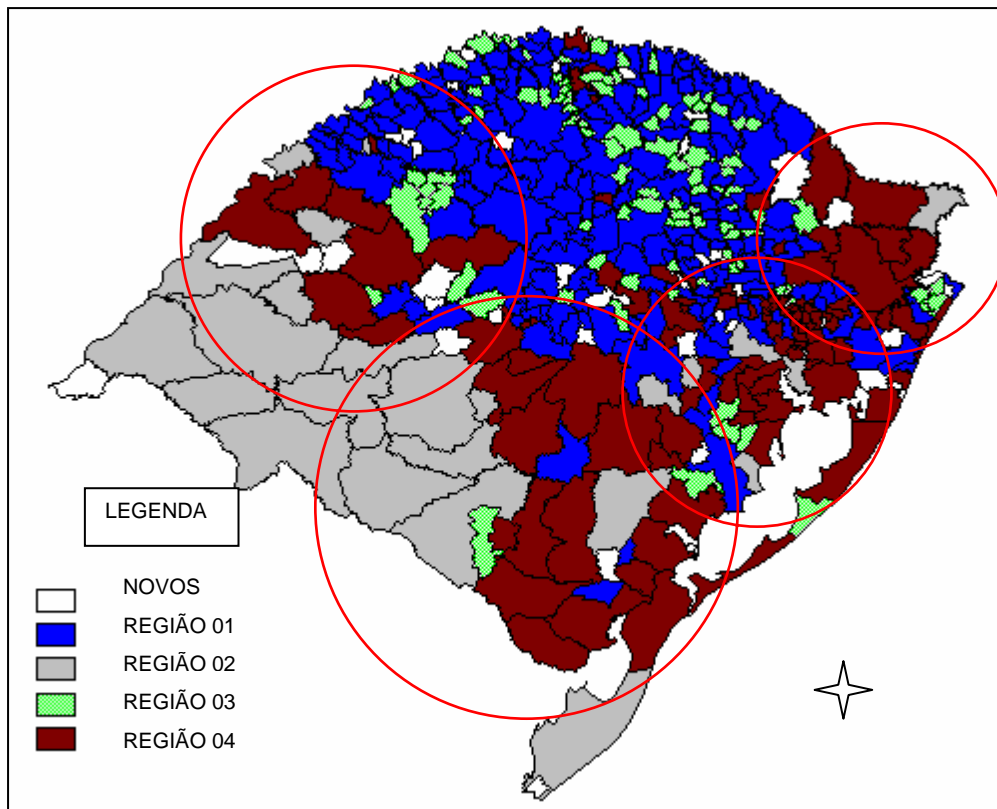


Figura 4.9 – Regionalizações do RS: Nível hierárquico 4
Método dos agrupamentos divisivos – Escala 1:250.000.

No nível 4 tem lugar uma acentuada diferenciação da região 02, especialmente a leste e centro-sul do estado envolvendo 105 municípios que passam a compor a região 04 formada por municípios que constituem quatro agrupamentos contínuos: a nordeste, a sudeste, a centro-oeste e grande parte da RMPA, exclusive Porto Alegre. Estes quatro agrupamentos estão assinalados pelos círculos vermelhos no mapa. É notória a característica de blocos contínuos que eles constituem, formados em sua maioria por municípios de grandes áreas com exceção da área que inclui parte da RMPA.

Um outro aspecto a ser considerado é que, ignorando-se alguma descontinuidade que de fato ocorre, pode-se imaginar a forma desta região como um grande “V” que se estende de norte ao sul na parte leste do estado e de oeste a leste na parte central.

A nova contagem fica então assim expressa: região 01 permanece com 201; região 02 com apenas 25; região 03 com os mesmos 96 e região 04 com 105 municípios. Observe-se que Porto Alegre continua fazendo parte da região 02.

4.7.4 ANÁLISE DOS NÍVEIS HIERÁRQUICOS 5, 6 e 7

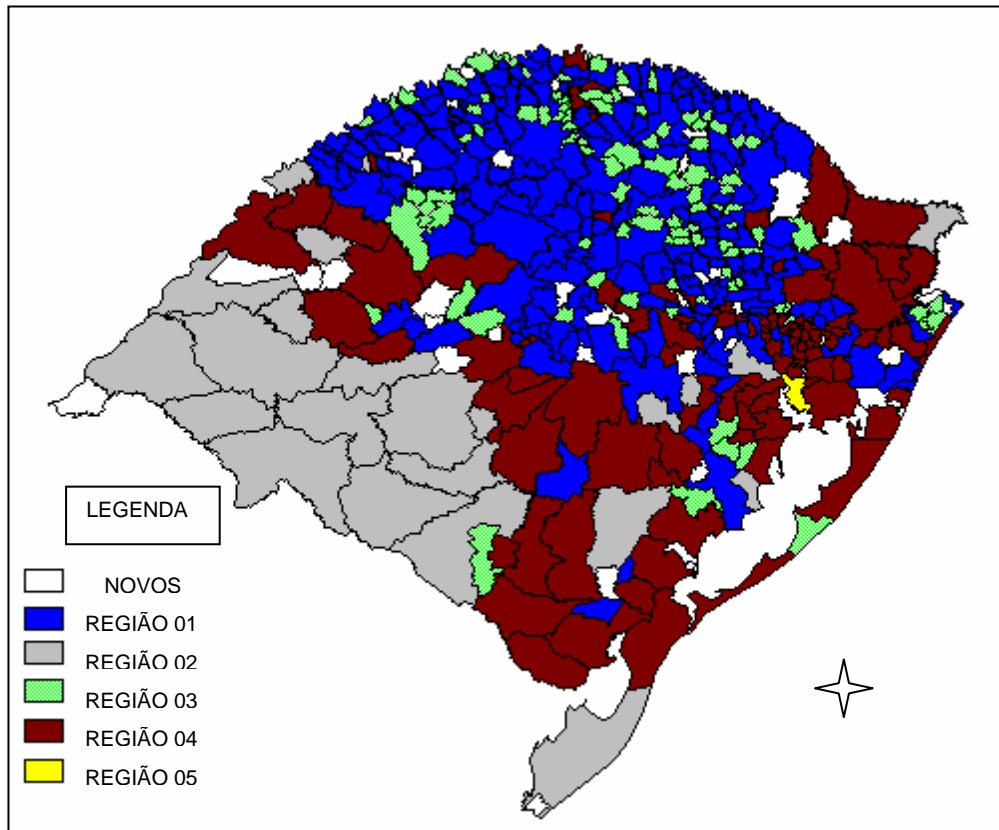


Figura 4.10 – Regionalizações do RS: Níveis hierárquicos 5, 6 e 7
Método dos agrupamentos divisivos – Escala 1:250.000.

No nível hierárquico 5, o município de Porto Alegre se apresenta diferenciado e é o único caso de um agrupamento unitário. Nos próximos dois níveis hierárquicos: 6, 7 a configuração regional permanece constante com 201 municípios na região 01, 24 municípios na região 02, 96 municípios na região 03, 105 municípios na região 04 e 1 município na região 05.

Até aqui verifica-se a estabilidade da região 01 que permanece com a mesma configuração desde o nível hierárquico 3. Conclui-se, então, que as demais regiões vem surgindo de uma diferenciação da antiga região 02 (130) e que agora possui 96 municípios.

Outra constatação que pode ser adiantada é de que a região 05, Porto Alegre constitui-se na única região unitária e permanecerá assim ao longo de todos os níveis que se sucedem.

4.7.5 ANÁLISE DO NÍVEL HIERÁRQUICO 8

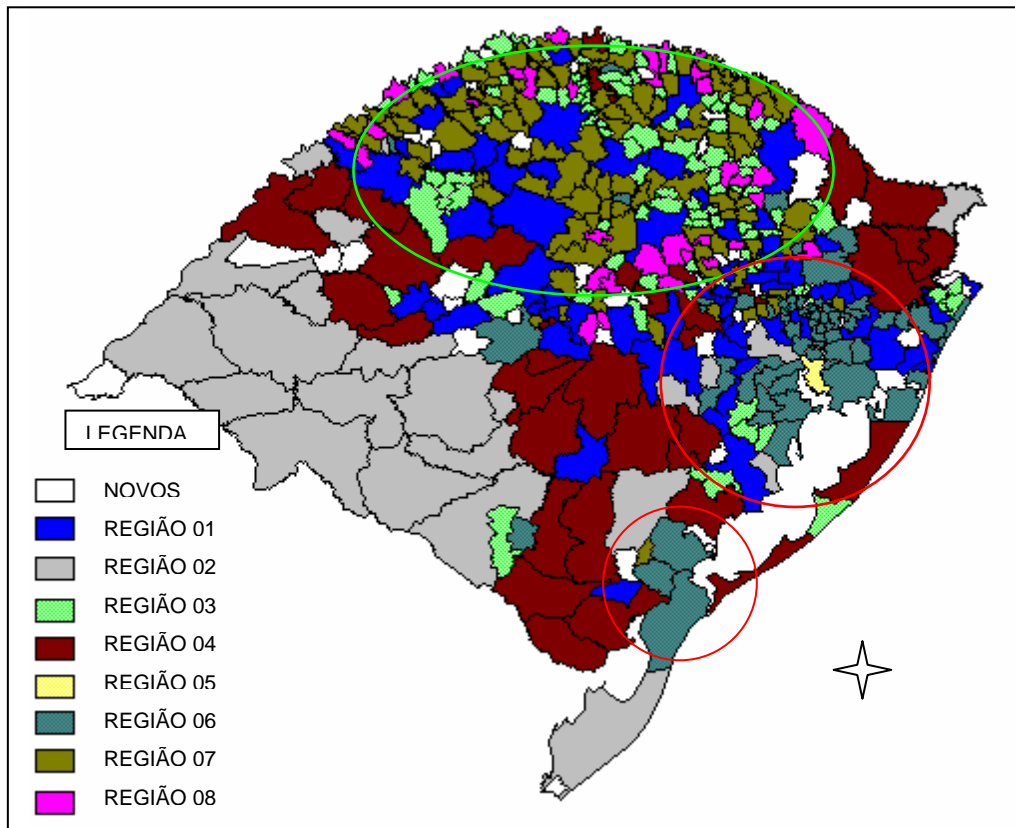


Figura 4.11 – Regionalizações do RS: Nível hierárquico 8
Método dos agrupamentos divisivos – Escala 1:250.000.

Aqui surgem três novos agrupamentos: a região 06 com 67 municípios composta por dois conjuntos, um que contém a RMPA (sem Porto Alegre) e seu entorno com ramificações ao norte e ao sul na base da Lagoa dos Patos, derivada da região 04 que fica com 38 municípios, assinalada no mapa acima pelos dois círculos vermelhos; a região 07 com 101 municípios, que é derivada da região 01 e guarda alguma similaridade formal com a região 03 pelo seu tipo de espalhamento ainda que de modo mais concentrado, indicada pela forma elíptica envolvente, na cor verde; e a região 08 com 35 municípios, esparsos ao longo da divisa com Santa Catarina e com a fronteira da Argentina, distribuindo-se ao longo de três quartas partes do contorno da elipse verde referida acima, além de um pequeno grupo descontínuo no centro-norte, também originários da partição da região 01 a qual passa a ter 65 municípios neste nível.

Observa-se que as duas regiões “mães” que deram origem às demais tem um comportamento em contraste, enquanto a região 02, ao ser partida, se mantém com

grandes blocos relativamente contínuos, a região 01 tem seu tecido esgarçado formando região que apresentam maior descontinuidade. É oportuno lembrar que a região 01 é conformada por municípios de área menor e a região 02, em maior parte, por grandes municípios.

4.7.6 ANÁLISE DO NÍVEL HIERÁRQUICO 9

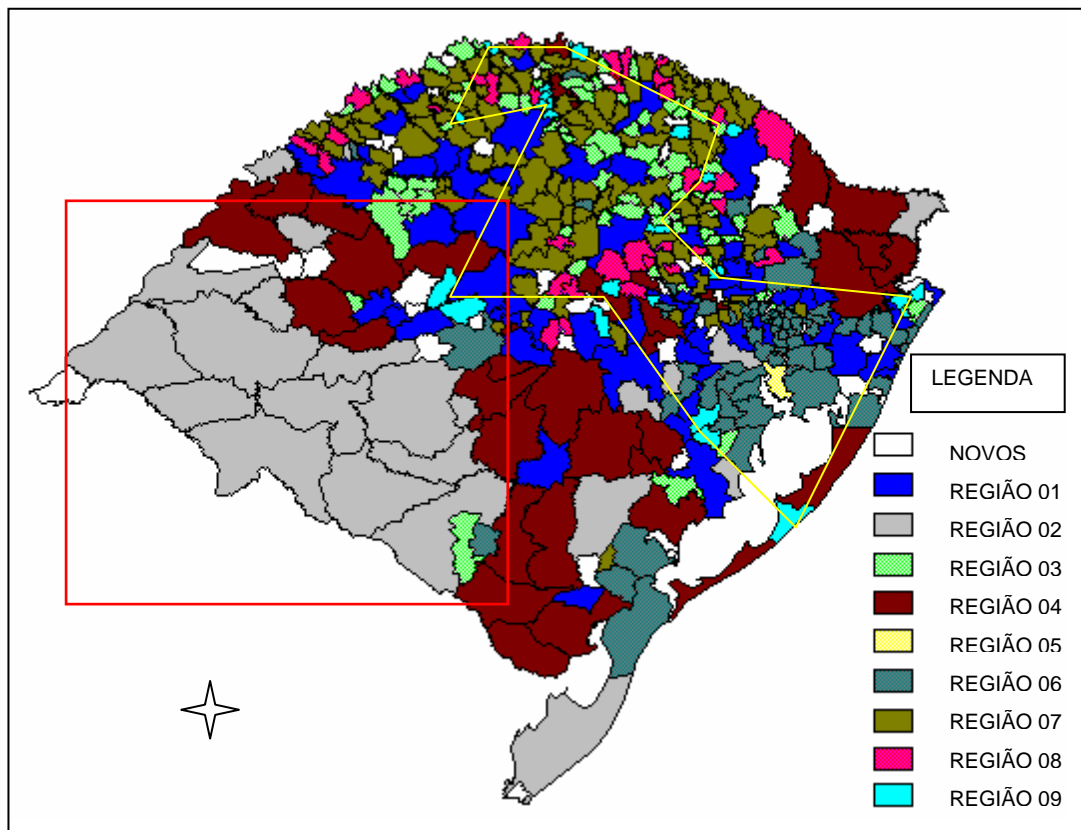


Figura 4.12 – Regionalizações do RS: Nível hierárquico 9
Método dos agrupamentos divisivos – Escala 1:250.000.

Aqui surge a região 09 com 24 municípios dispersos e diferenciados a partir da região 03 que fica com 72 municípios.

Note-se que neste ponto duas regiões apresentam a menor frequência, isto é, tem menos municípios componentes. São elas a região 09 que acabou de agrupar-se e a região 02 que existe estável desde o nível 5, ambas com 24 municípios. Lembrando que a primeira origina-se da região “mãe” 01, pode-se verificar a grande diferença que existe em suas configurações. A região 09 apresenta-se completamente desagregada e composta por pequenos municípios, indicada no mapa pela poligonal irregular amarela e a região 02 agrupada em um bloco contínuo de mais da metade de seus municípios envolvido pelo

quadrado vermelho no mapa, os maiores do estado, além de mais alguns outros espalhados, e de área um pouco menor.

4.7.7 ANÁLISE DOS NÍVEIS HIERÁRQUICOS 10,11,12 e 13

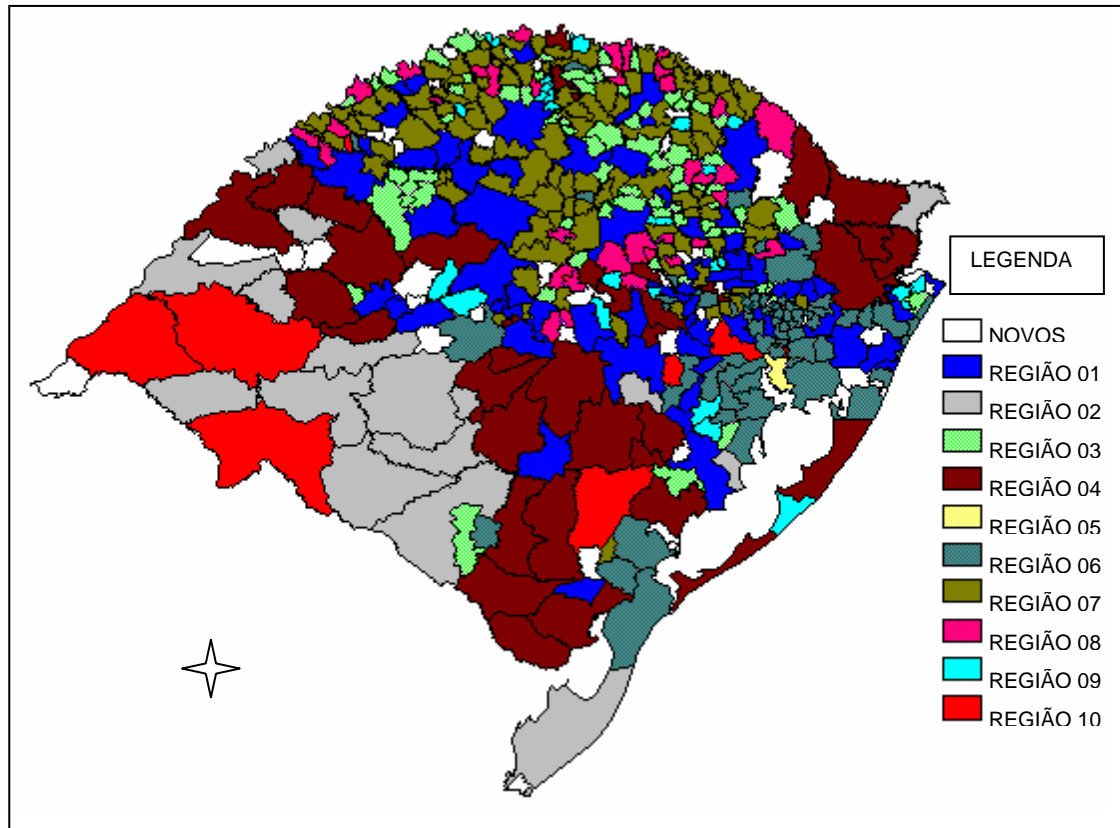


Figura 4.13 – Regionalizações do RS: Níveis hierárquicos 10, 11, 12 e 13
Método dos agrupamentos divisivos – Escala 1:250.000.

Novamente aqui ocorre o fenômeno de estabilidade ao longo de quatro níveis hierárquicos, que ocorreu anteriormente nos níveis 5, 6 e 7. A única formação nova é a região 10 com apenas 8 municípios originários da região 02 que permanece com 16 municípios.

Assim, a atualização da frequência dos municípios fica região 01, 65 municípios; região 02, 165 municípios; região 01, 65 municípios; região 01, 65 municípios; região 01, 65 municípios; região 02, 16 municípios; região 03, 72 municípios; região 04, 38 municípios; região 05, 1 município; região 06, 67 municípios; região 08, 16 municípios; região 09, 24 municípios e região 10 com 8 municípios.

4.7.8 ANÁLISE DO NÍVEL HIERÁRQUICO 14

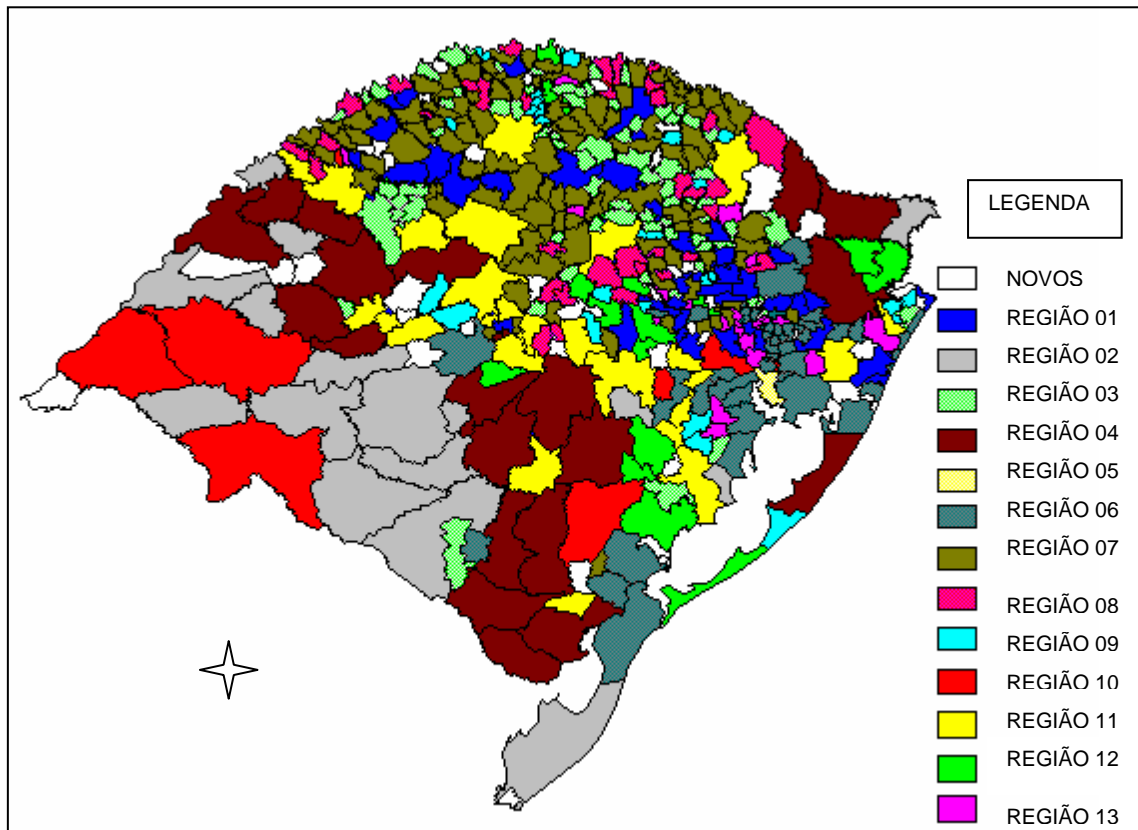


Figura 4.14 – Regionalizações do RS: Nível hierárquico 14
Método dos agrupamentos divisivos – Escala 1:250.000.

Neste nível hierárquico surgem três novos agrupamentos: a região 11 derivada da região 01, com 25 municípios; a região 12 a partir da região 04, com 18 municípios e a região 13 partição da região 06, com 23 municípios.

A região 11 é formada por municípios de área média ao longo do centro e norte do estado. A região 12 é constituída por municípios um pouco menores na parte central e sudeste do estado. A região 13 com municípios pequenos espalhados no entorno e dentro da RMPA em direção ao norte.

Por este ser o penúltimo nível passível de análise neste trabalho, cabe ressaltar a homogeneidade da região 7, a mais numerosa, com 101 municípios derivados originalmente da região 01 e que desde o nível 8 até o último mantém-se estável. Enquanto que a região de menor frequência é a região 10 com apenas 8 municípios de origem na região 02.

4.7.9 ANÁLISE DO NÍVEL HIERÁRQUICO 15

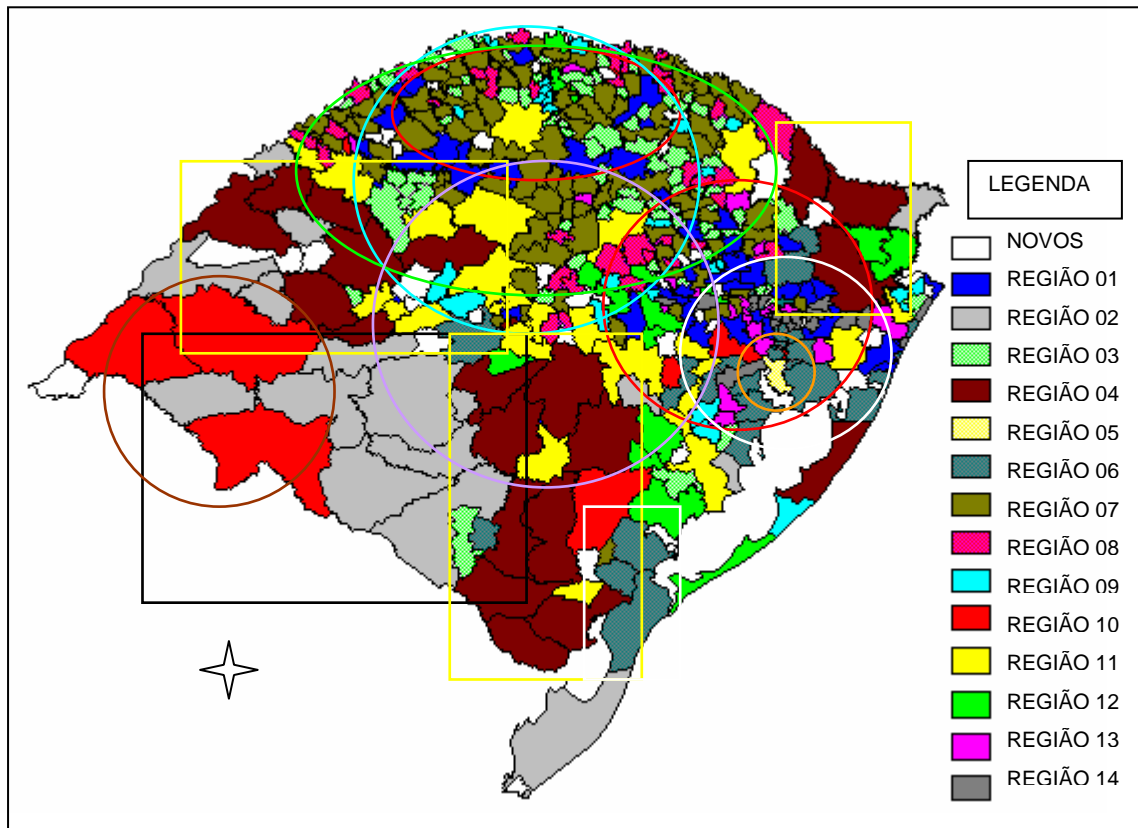


Figura 4.15 – Regionalizações do RS: Nível hierárquico 15
Método dos agrupamentos divisivos – Escala 1:250.000.

A região 14 que aparece neste último nível deriva da região 06 e é formada por 20 municípios pequenos localizados, em maioria, ao norte da RMPA de forma espalhada.

Por fim, indica-se na figura 4.15 as áreas de maior concentração de cada região, especialmente naquelas em que devido ao tipo de espalhamento isto seja possível de ser indicado. Para a região 01 corresponde a duas áreas definidas na figura pela elipse vermelha ao norte e pelo círculo vermelho a nordeste. Para a região 02 o retângulo preto ao sul. A região 03 com a maior parte de seus municípios dispersos dentro do círculo azul claro. A região 04 possui três áreas de concentração bem definidas por retângulos amarelos, o primeiro retângulo vertical indica a maior, o segundo retângulo horizontal indica a média e o terceiro retângulo vertical indica a pequena. A região 05 é indicada pelo pequeno círculo laranja e trata-se do município de Porto Alegre. Para a região 06 há dois tipos de indicações: um círculo branco marca a maior concentração na RMPA e no entorno, e o outro indicado pelo retângulo vertical branco a sudeste. A região 07 encontra-se praticamente inscrita no interior da elipse verde claro ao norte e noroeste, enquanto que a

região 08 distribui-se ao longo do perímetro desta elipse verde. A região 09 não permite este tipo de enquadramento geométrico porque está dispersa e é composta de poucos municípios. Para a região 10 um círculo marrom mostra uma pequena concentração de três municípios grandes a sudoeste. Para a região 11 o círculo lilás indica a principal concentração. A região 12 não está indicada em razão de seu tipo de espalhamento irregular. E finalmente, tanto a região 13 como a 14 encontram-se distribuídas dentro do círculo branco que já indicou parte da região 06.

O critério para a adoção da nomenclatura da região “mãe” em relação às regiões “filhas” foi aquele que considera a maior área remanescente como mãe, e a menor como filha, independente da frequência de municípios de cada região.

4.8 ANÁLISE COMPARATIVA COM AS DIVISÕES REGIONAIS EXISTENTES

A primeira comparação que pode ser feita é entre o nível hierárquico 15, figura 4.13 e os COREDES com 22 agrupamentos, figura 4.14. Para facilitar a análise pode-se iniciar pela metade sul do estado onde os municípios são grandes e relativamente poucos. As regiões dos COREDES da Fronteira Oeste e Campanha tem relação com municípios que compõem as regiões 02 e 10, contudo alguns municípios que pertencem à Fronteira Oeste têm maior similaridade com a Campanha. A região Sul (COREDES) tem alguma afinidade com a região 04, embora na área correspondente à região Sul haja, pelo menos sete tipos de municípios em termos de similaridade. A região Metropolitano – Delta do Jacuí é formada por cinco tipos de agrupamentos com mais representantes da região 06 do nível 15. Quanto as demais regiões da metade norte, pouco se pode afirmar em termos de comparação visual, contudo é possível dizer que os COREDES desta área do estado congregam muitas classes diferenciadas de municípios e que não encontrariam maneira de serem agrupados de forma tão contínua se levados em consideração critérios de similaridade em função de dados sócioeconômicos.

Com relação às microrregiões (35) do IBGE para o estado, figura 4.15 mais adiante, observa-se uma melhor correspondência em termos de comparação genérica com os resultados do modelo, uma vez que elas apresentam-se mais subdivididas do que os COREDES e acompanham a tendência de partição da metade norte do estado. Portanto, grosso modo, é possível dizer que a diferenciação no interior da microrregião se dá de modo bem menos intenso do que dentro dos COREDES, ou seja, há mais encontros de

municípios similares que fazem parte de uma mesma microrregião. Entretanto, caberia questionar sua homogeneidade que relativa à similaridade está longe de existir.

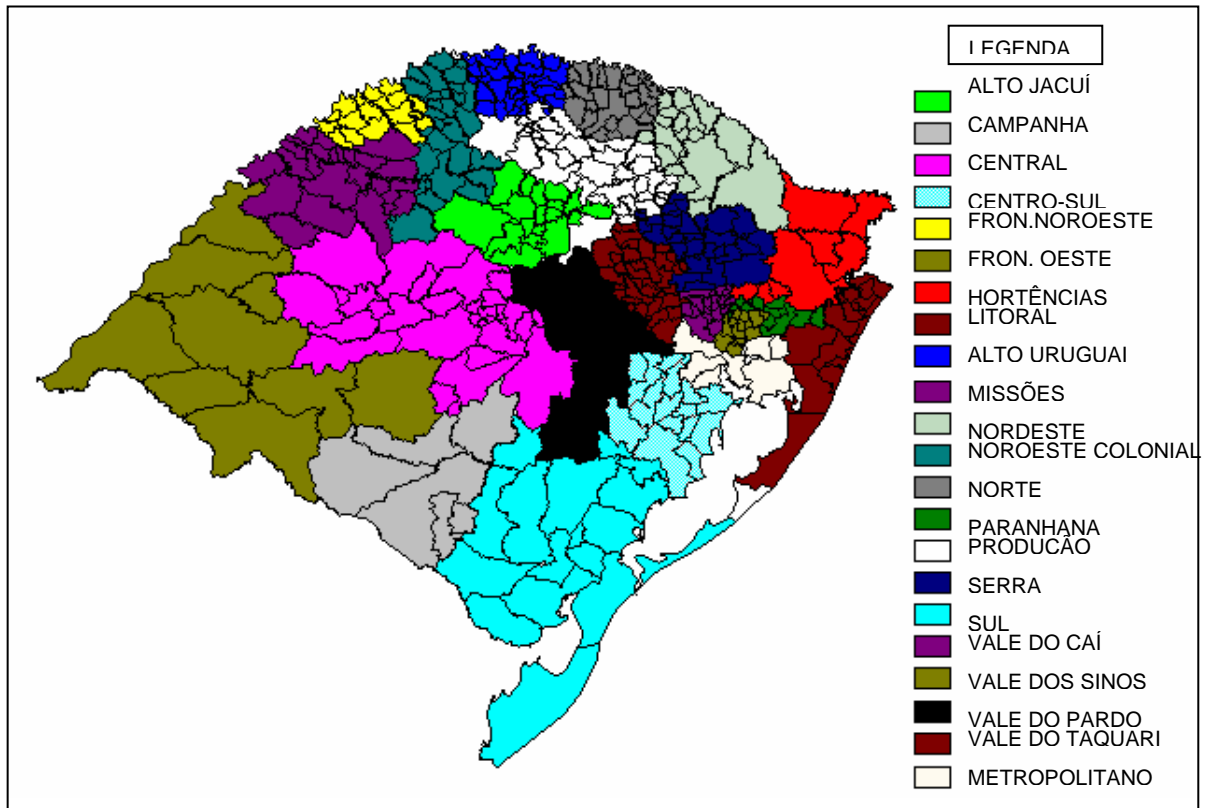


Figura 4.16 – Conselhos Regionais de Desenvolvimento do RS (COREDES)
Escala 1:250.000.

As Mesorregiões, figura 4.18 e as Regionais de Saúde, figura 4.19, por abrangerem áreas muito grandes apresentam o mesmo problema analisado para as microrregiões, muita dissimilaridade entre os municípios componentes.

Mas, se for feita a comparação com níveis hierárquicos menores nos quais o número de agrupamentos seja numericamente próximo às mesorregiões, como por exemplo nos níveis hierárquicos 8 e 9 verifica-se que há algum grau de correspondência entre as mesorregiões Sudoeste e Sudeste em relação às regiões 05 e 06 do modelo e em menor grau à região Nordeste. Já as mesorregiões Noroeste, Centro Ocidental, Centro Oriental e Porto Alegre registram uma alta diversidade na composição de seus municípios, não sendo possível fazer comparações consistentes.

Desta análise, conclui-se que as regionalizações em grande escala são sempre menos reveladoras da realidade do que aquelas mais fragmentadas, justamente porque a similaridade entre os municípios se dá de forma descontínua em grande medida.

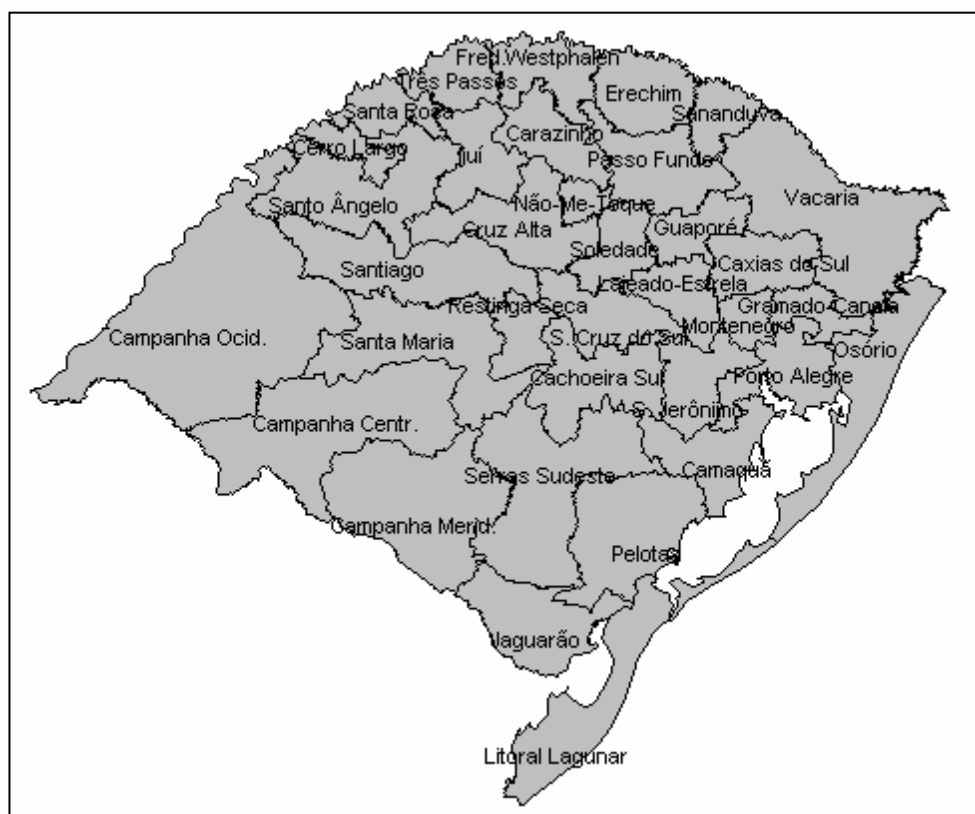


Figura 4.17 – Microrregiões Homogêneas do RS (IBGE)
Escala 1:250.000.

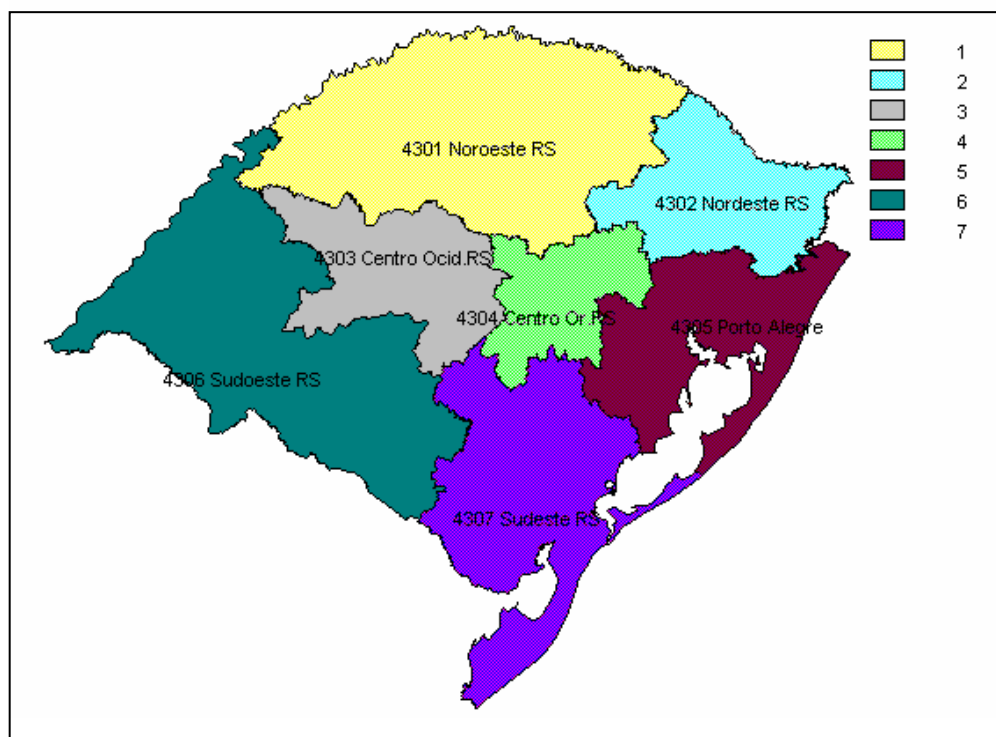


Figura 4.18 – Mesorregiões do RS (IBGE)



Figura 4.19 – Regionais de Saúde do RS (IBGE) - Escala 1:250.000.



CAPÍTULO 5

CONCLUSÕES

5. CONCLUSÕES

O capítulo final objetiva abordar quatro aspectos principais: tecer comentários analíticos sobre os resultados finais, indicar possíveis deficiências da análise e onde seria necessário maior aprofundamento, propor novos enfoques ao tema a partir da utilização do instrumental apresentado e suscitar algumas questões que poderiam ser respondidas por trabalhos futuros na área de regionalizações com base no presente modelo.

5.1 ANÁLISE FINAL

5.1.1 DOS OBJETIVOS

Em relação ao objetivo geral supõe-se que o modelo tenha atingido suficiente sensibilidade a ponto determinar com exatidão as regiões através da determinação da similaridade entre os municípios. Para que possa auxiliar mais diretamente na distribuição de recursos públicos se faz necessário criar uma rotina de avaliação que poderia classificar as diversas regiões segundo a média ponderada do PIB per capita municipal, ou outra forma de medida como um índice de desenvolvimento social. Contudo, as relações de similaridades estão bem definidas e são ponto de partida para qualquer distribuição de recursos públicos.

Sobre os objetivos específicos pode-se argumentar que apesar de na fase de seleção terem sido listadas mais variáveis do que aquelas que efetivamente participaram das matrizes de dados deve-se ao fato de serem necessárias duas filtragens: primeiro a adequação à existência da informação na fonte, e em segundo lugar o fato de que das 26 variáveis que iniciaram o processo apenas 22 foram consideradas porque pela verificação dos valores de autocorrelação retira-se todos aqueles que são mais altos para que o

atributo não seja considerado de modo repetido. Relativamente aos demais objetivos específicos supõe-se que tenham sido atingidos.

5.1.2 DA QUESTÃO ESPACIAL

Ao longo do presente trabalho foi sempre tratada como uma questão de fundamento e conceitualmente importante a característica dinâmica do espaço. Uma forma versátil de apreender este aspecto foi constantemente uma preocupação subjacente e por vezes, explícita ao longo do texto. Haja visto todo o esforço de clareza de definição realizado no capítulo 2, item 2.1.1, com relação à diferença entre divisão regional e regionalização como modo de incorporar ao modelo esta permanente transformação a que o espaço está submetido, o que é razoavelmente alcançado através da forma de equacionar os atributos e do método de correlacionar as variáveis.

Um segundo aspecto de grande relevância é o objetivo de preservar a estrutura espacial, o que está estreitamente relacionado com as distâncias euclidianas e não-euclidianas utilizadas como critério de similaridade para garantir o princípio de homogeneidade regional. As dissimilaridades levantadas representam problemas e demandas que devem ser tratados de forma diferenciada o que acusa uma descontinuidade espacial e que é refletida no método de agrupamento hierárquico divisivo. Ainda que o espaço como um todo seja de natureza contínua, as similaridades e por conseguinte, em grande escala, as dissimilaridades são de caráter descontínuo, logo a homogeneidade de áreas circunscritas ao domínio municipal pode ou não ser descontínua, ainda que do ponto de vista administrativo alguma continuidade seja de todo desejável e até necessária para a aplicação de políticas públicas. Contudo, a informação que refere demandas diferenciadas serve como referencial para uma estruturação orçamentária que poderia, se quantificada ou dimensionada, ser direcionada aos municípios e regiões de maior carência.

5.1.3 DO ENFOQUE METODOLÓGICO

Em geral o processo decisório está ancorado em dois pilares: a quantidade de informação e a qualidade da informação. De pouco adianta existirem técnicos da mais alta capacitação, com acentuada experiência em planejamento se os dados que alimentam suas decisões forem de natureza imprecisa ou insuficientes. Nesta medida, o modelo

heurístico responde com vigor à questão de transformar o dado simples, que é o reflexo do fenômeno, na informação interpretável ou decodificada, que é a expressão comparável do fenômeno. Aí reside a capacidade interpretativa do método em razão de que a sua aplicação resulta em elementos que contribuem claramente para o apoio à tomada de decisão, porque é a partir das saídas originárias do processo e da medição do grau de relacionamento entre os dados, e mesmo em função das informações geradas, que as soluções regionais são determinadas, como um efeito do dado evoluído à informação e de seu tratamento estatístico.

O ponto central do método desenvolvido visa estabelecer de forma comparativa o grau de relacionamento entre os padrões espaciais e os padrões dos dados. Reduzida a informação espacial a uma matriz de distância euclidiana e o comportamento dos dados a várias (8) matrizes de distância entre eles, define-se uma forma de detectar a melhor relação em termos de similaridade entre a primeira e cada uma das demais, o que é assumido pelo valor mais alto do índice gerado pelo teste de Mantel. Aplica-se, então, um método de análise de agrupamento que visa criar grupos hierárquicos, no caso, de municípios que configuram as diversas regionalizações finais. Sabe-se que estas sucessivas reduções impostas pelo método e que são aplicadas durante o processo terminando por corporificar o modelo, não carregam consigo a amplitude total dos fenômenos socioeconômicos e geográficos em estudo, por esta razão uma criteriosa análise de erro deve ter lugar. Além do que aponta para o cuidado que se deve ter em fazer interpretações demasiado largas ou abrangentes que o modelo não contempla. É preciso fazer a leitura até onde o instrumento é capaz de aferir e não além.

5.1.4 DOS RESULTADOS APRESENTADOS

Os níveis de agregação dos agrupamentos localizados na metade norte do estado são menores do que aqueles localizados na metade sul. Isto explica-se em parte porque as áreas dos municípios ao norte são bem menores do que as áreas dos municípios ao sul, o que amplia a possibilidade de combinações e conseqüentemente do surgimento de lacunas intermunicipais. E porque também o grau de homogeneidade ao sul é aparentemente maior, o que origina agrupamentos mais monolíticos e formados por poucos municípios que cobrem vastas áreas mais interligadas.

Quanto maior o nível hierárquico de agrupamento maior a definição de entrelaçamento entre os municípios, isto é, mais se aproximam as diferenças, de modo que

os patamares de níveis hierárquicos começam a quase se sobrepor o que implica em menores diferenças ou maior proximidade entre as dissimilaridades.

Verifica-se que Porto Alegre é um município altamente diferenciado, porque configura uma delimitação regional unitária e única nesta condição.

No outro extremo a região 07 com 101 municípios similares identifica uma área regional ao norte e noroeste de característica bem definida.

A existência de “ilhas” (município isolado ou subgrupo de dois ou três municípios) em meio a regiões pode significar nichos de excelência ou de deficiência em relação ao entorno.

Alguns resultados como as regiões formadas por poucos municípios grandes ou aquelas constituídas por muitos municípios pequenos são previsíveis porque confirmam a teoria de similaridade entre os próximos (juntos) isto é, característica de continuidade espacial. Já as regiões de poucos municípios espalhados apontam para a questão de ser possível haver similaridade na contra mão do princípio citado acima, o que significaria desenvolver estudos específicos sobre estes municípios em especial, para estabelecer políticas públicas direcionadas.

É oportuno lembrar em concordância que os mapas não se constituem em respostas definitivas, porém são úteis para expor relações e auxiliar a investigação na demonstração de novas correlações (Talen, op. cit.).

5.2 ASPECTOS PROBLEMÁTICOS

Um primeiro problema que pode ser mencionado é a distorção das relações de distância ao considerar, para o cálculo da matriz, as coordenadas de municípios cujas sedes se localizam em posição excêntrica, especialmente naqueles municípios muito extensos ou de geometria muito irregular. A escolha dos centróides como pontos de referência para calcular a matriz de distâncias seria mais indicada assim como o uso de coordenadas UTM. Entretanto, preferiu-se utilizar as coordenadas das sedes porque supõe-se que a maior concentração de população ocorre nas proximidades das sedes e nas próprias sedes que, portanto, polarizam os aspectos de demanda, assim as relações de distância entre as sedes representam melhor porque expressam a concentração de demanda que se localiza nas sedes e não necessariamente nos centróides.

Outro problema ocorre em função dos quarenta (40) novos municípios que são desconsiderados na análise espacial para determinação da matriz de distância euclidiana, mas os dados socioeconômicos utilizados incluem estas áreas, porque se referem a um período em que estas áreas novas ainda participavam dos municípios considerados aqui. Este é mais um elemento de propagação de erro.

5. 3 ASPECTOS DE RELEVÂNCIA

Uma expansão do presente trabalho que incluísse uma série histórica de dados permitiria uma comparação evolutiva da regionalização, podendo tratar tanto os dados escolhidos aqui, bem como dados específicos de uma ou mais áreas de estudo. Por exemplo, dados de saúde considerados ao longo de um período (série de 5 ou mais anos) geraria perfis regionais que poderiam revelar o comportamento dinâmico da questão saúde no estado, naturalmente que dentro de uma mesma configuração geográfica. Um outro exemplo com todo o conjunto de variáveis já tratados aqui geraria um quadro evolutivo que poderia registrar com clareza o movimento regional dinâmico no tempo.

Uma possível automatização do processo poderia ser pensado em termos de sistema computacional que permitisse certo grau de interatividade, de modo que, escolhidas as entradas: arquivos de dados submetidos a um padrão de entrada e fórmulas de distâncias escolhidos de uma lista, seria feito o teste de Mantel e a geração de uma tabela que representasse a saída do dendograma automaticamente. A partir daqui esta tabela seria escolhida como nova entrada para a geração de mapas regionais, além disso poderia ser pensado um módulo que oferecesse rotinas de cálculo de erro e determinação do grau de certeza nas saídas do sistema.

A utilização de software livre para desenvolvimento desta aplicação garantiria a abertura do código e um possível aprimoramento do sistema pelos usuários potenciais ou um redirecionamento para a área de interesse particular de estudo, uma vez que o presente modelo já foi desenvolvido utilizando-se recursos de *software* de domínio público como o ADE-4 (Universidade de Lyon 1), "Análise de Dados Ecológicos", aplicativo de análise multivariada e métodos de análise espacial de dados, o TabWin do Sistema Datasus, do Ministério da Saúde, tabulador e visualizador de gráficos e mapas, além de alimentar-se de dados oficiais do IBGE e FEE/RS disponíveis para *download* via Internet.

5.4 ABORDAGENS ESPECÍFICAS

A metodologia apresentada se presta perfeitamente para aplicações em áreas específicas, como aquelas das quais advém a base de dados que lhe deu origem. Pode-se pensar em desenvolver estudos de natureza ambiental, ou na área da saúde, ou educação, ou sobre a questão do desemprego, entre outros. Estes estudos poderiam utilizar a mesma estratégia apresentada neste trabalho de variadas maneiras.

5.5 INDICAÇÕES DE DESAFIOS POSSÍVEIS

É fácil imaginar que outros universos de dados poderiam facilmente ser utilizados como entrada do modelo, assim poderiam ser realizados estudos específicos voltados para a educação ou para a saúde ou mesmo para a política.

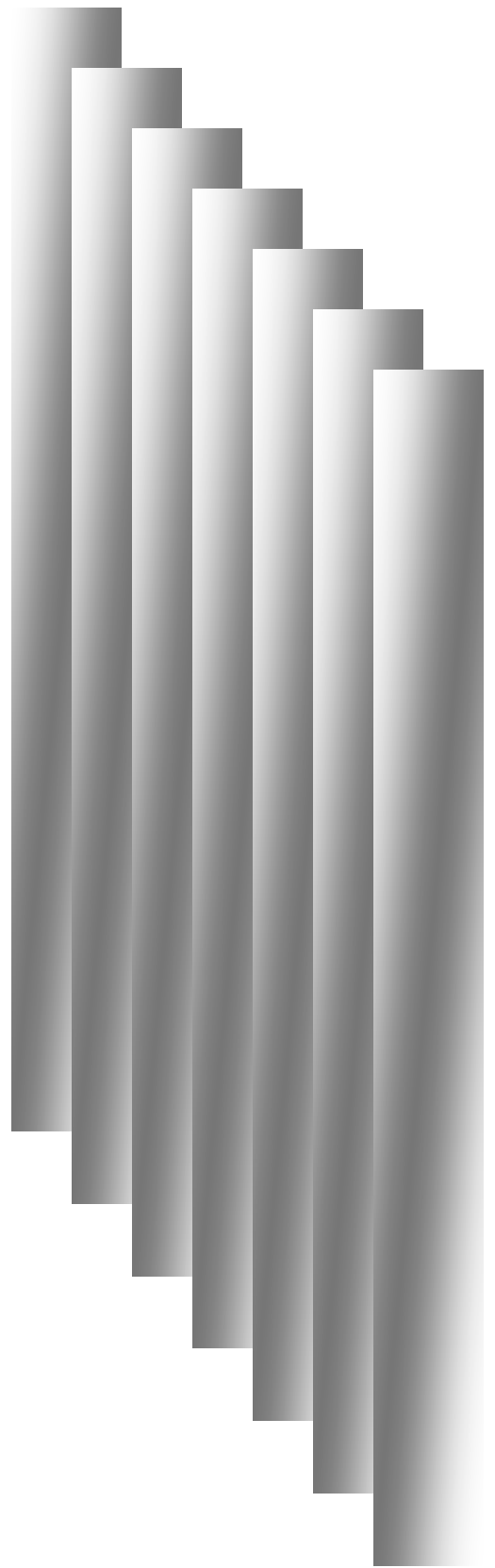
Novos operadores estatísticos poderiam ser considerados assim como outras formas de medição de distância além daquelas aqui consideradas.

Pode-se, também, pensar em encaminhar a pesquisa para uma linha de atributos sócio-políticos e não apenas socioeconômicos o que resultaria em outros perfis regionais e em novos cenários de análise.

Por si só o método, da forma como foi desenvolvido, capacita o analista para outros tipos de análise que não pressupõe um conhecimento prévio da área de estudo, senão pelo acesso aos dados, mesmo assim outras informações adicionais que já se possua, somente enriquecerão o processo.

5.6 FECHAMENTO

A título de encerramento vale dizer que a melhor visão da realidade é a que consegue superar os clássicos impedimentos com os quais nos deparamos ao proceder um processo de conhecimento. São impedimentos para uma visão apurada da realidade nossas próprias convicções, nossa linha de pensamento, nossa inserção social e política. Por isso, um método que busca certa isenção tem a qualidade de fornecer com generosidade um conglomerado de informações não encontráveis por abordagens mais descritivas da realidade. Muitas vezes, tais expedientes metodológicos recorrem a dados numéricos e à aplicação de conceitos estatísticos, em que pese sua origem positivista. No presente estudo, que considerou a estrutura espacial através de correlações de distâncias, o recurso a esses métodos mostrou ser de relevante valia para subsidiar estudos qualitativos ou auxiliar no planejamento e implementação de programas governamentais ou políticas públicas em geral.



REFERÊNCIAS
BIBLIOGRÁFICAS



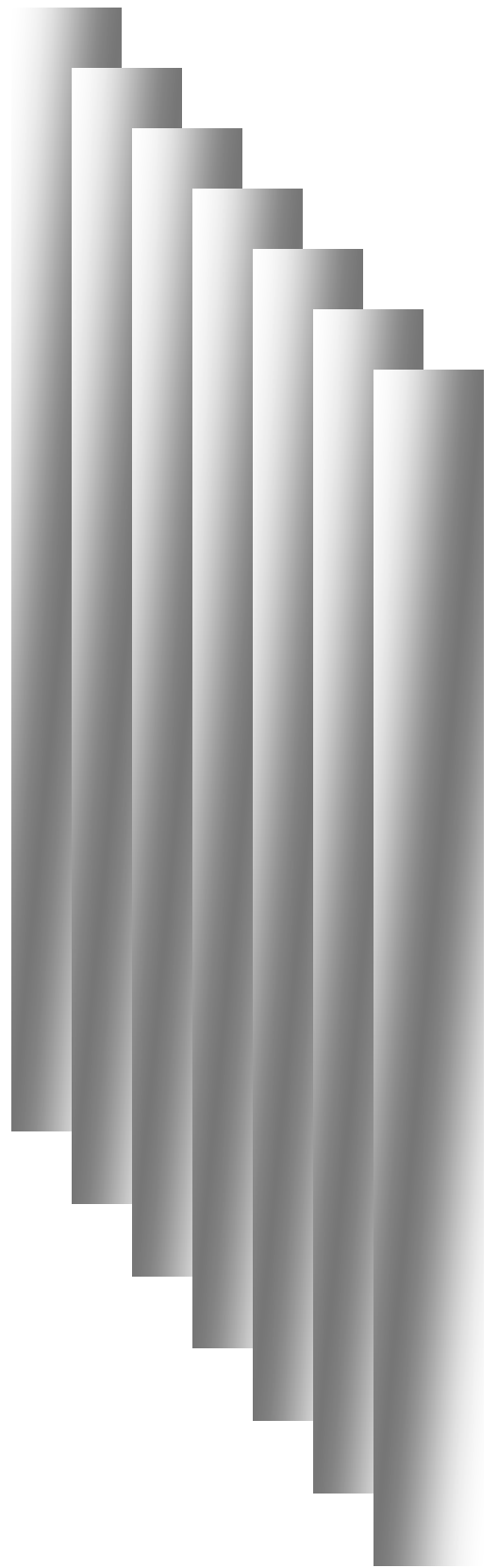
REFERÊNCIAS BIBLIOGRÁFICAS

- ADE-4, 1998. Analyses des Données Écologiques, Université Lyon 1. France. (Software de análise estatística espacial em <http://pbil.univ-lyon1.fr/ADE-4/ADE-4F.html>).
- ALONSO, J.A.F., et alli, 1984. Áreas estatisticamente comparáveis no Rio Grande do Sul: 1940-75, Fundação de Economia e Estatística S.E.H. Porto Alegre.
- ALONSO, J.A.F., Benetti, M.D. & Bandeira, P.S. 1994. *Crescimento Econômico da Região Sul do Rio Grande do Sul – Causas e Perspectivas*, Fundação de Economia e Estatística S.E.H. Porto Alegre.
- ANSON, J., 1991. Demographic indices as social indicators , *Environment and Planning A*, 23(433-446).
- BALLARD, K.P. & Wendling, R.M. 1980. *The National-Regional impact evaluation system: a spatial model of U.S. economic and demographic activity*, *Journal of Regional Science*, 20(143-158).
- BANDEIRA, P.S. & Gründling, N.A. 1988. *Distribuição geográfica do crescimento industrial no Rio Grande do Sul – década de 70*, FEE, Porto Alegre.
- BARKLEY, D.L., Henry, M.S., e Bao, S. 1996. Identifying “Spread” versus “Backwash” Effects in Regional Economic Areas: A Density Functions Approach, *Land Economics*, 72(3): 336-357.
- BERNARDES, N. 1963. *Bases Geográficas do povoamento do Rio Grande do Sul*. Boletim Geográfico, Rio de Janeiro, IBGE, 21(171):587-620.
- BERRY, B. 1964. *Approaches to Regional Analysis: A Synthesis*, *Annals, Association of American Geographers*, 54 (2-11).
- BINDER, K. 1987. *Applications of the Monte Carlo Method in Statistical Physics*. Springer-Verlag, New York.
- CABELLO, O. 1960. The use of statistics in the formulation and evaluation of social programmes, *Journal of American Statistical Association*, 55(454-468).
- CÂMARA, G. e Monteiro, A.M.V. 1999. *Análise Espacial de Dados Geográficos*, Instituto Nacional de Pesquisas Espaciais, V Congresso e Feira para Usuários de Geoprocessamento da América Latina, Salvador.

- CARRION, E.S. 1997. Região Metropolitana de Porto Alegre – Índice de Desenvolvimento Urbano – Subsídios para Análise Regional e Municipal, METROPLAN, Porto Alegre.
- CHRISTALLER, W. 1966. *Central Places in Germany*, Prentice-Hall.
- COOMBES, M. e Wong, C. 1994. Methodological steps in the development of multivariate indexes for urban and regional policy analysis, *Environment and Planning A*, 26 (1297-1316).
- CORRÊA, R.L., 1987. *Algumas considerações sobre análise regional*. *Revista Brasileira de Geografia*, Rio de Janeiro, 49(4):47-52.
- DING, Y. e Fotheringham A.S. 1992. *The integration of spatial analysis and GIS*, *Comput. Environment and Urban Systems*, 16 (3-19).
- DUARTE, A.C., 1980. Regionalização – considerações metodológicas. *Boletim de Geografia Teórica*, Rio Claro, s.ed., 10(20):5-32.
- DUARTE, A.C., 1988. *O conceito de totalidade aplicado à identificação de uma região*. *Revista Brasileira de Geografia*, Rio de Janeiro, 50(2):99-106.
- EASTIN, R.V. 1977. *Multiple-criteria optimization: a more general approach*, *Environment and Planning A*, 9 (345-350).
- EASTMAN, J.R. 1998. *IDRISI for Windows: Introdução e Exercícios Tutorais*, Centro de Recursos IDRISI, UFRGS, Porto Alegre.
- FERRÃO, M.F. 1992. *Aplicação do Método de Simulação Monte Carlo ao estudo dos mecanismos de cura co-reativa de resinas Epoxi com anidridos dicarboxílicos em presença de amina terciária*. Dissertação de Mestrado, UFRGS, Curso de Pós-graduação em Química, Porto Alegre.
- FERREIRA, A.B. de H. 1986. *Novo Dicionário da Língua Portuguesa*, Ed. Nova Fronteira, Rio de Janeiro.
- FISCHER, M.M. 1980. Regional Taxonomy: a comparison of some hierarchic and non-hierarchic strategies, *Regional Science and Urban Economics*, 10(503-537).
- GRIFFITH, D. 1987. *Spatial autocorrelation a primer*, Resource Publications Advisory Board, Boston.
- GUIMARÃES, F.de M.S., 1963. Observações sobre o problema da divisão regional. *Revista Brasileira de Geografia*, Rio de Janeiro, IBGE, 25(3):289-312.
- HADDAD, E. 1997. Determinação de zonas homogêneas dentro de áreas urbanas: aspectos metodológicos – parte II – Da solução do problema taxonômico: a análise de agrupamento, Congresso ANPUR (1997).
- HAIR, J.F.; Anderson, R.E.; Tatham, R.L. e Black, W.C. 1998 *Multivariate Data Analysis*, 5 ed., Prentice-Hall, Inc.
- HARVEY, D. 1973. *Social Justice, Postmodernism, and the City* (p. 96– 97).

- HUGHES, D.W. e Holland, D.W. 1994. Core-Periphery Economic Linkage: A Measure of Spread and Backwash Effects for the Washington Economy, *Land Economics*, 70(3): 364-377.
- IBGE, 1980. *Divisão territorial do Brasil*, Fundação Instituto Brasileiro de Geografia e Estatística, 9 ed. - Rio de Janeiro.]
- JACQUEZ, G.M.; e Jacquez, J.A. 1999 *Disease Clustering for Uncertain Locations*. In: *Disease Mapping and Risk Assessment for Public Health*. Lawson et al., 1999. John Wiley & Sons Ltd. Chichester (151-168).
- JOHNSON, R.A. e Wichern, D.W. 1998 *Applied Multivariate Statistical Analysis*, 4 ed., Prentice-Hall, Inc.
- LEÃO, S.Z., 1997. O Planejamento Locacional Urbano em um contexto de disputa pelo uso do solo: seleção de áreas para aterro sanitário de resíduos sólidos urbanos em Porto Alegre / RS. Porto Alegre. Dissertação de Mestrado. Faculdade de Arquitetura, PROPUR, UFRGS.
- LÖSCH, A., 1954. *The economics of location*. New Haven, Yale University Press.
- LOLONIS, P. e Armstrong, M.P. 1993. *Location-allocation models as decision aids in delineating administrative regions*, *Environment and Urban Systems*, 17 (153-174).
- MANLY, B.F.J. 1994 *Multivariate Statistical Methods: A Primer*. Chapman & Hall. London.
- MARQUES, A., Erbert A. e Silva J. 1997. *Análises Estatísticas Espaciais Associadas ao SIG*, disciplina de Geoprocessamento, não publicado.
- MARTIN, D., Senior, M.L. e Williams H.C.W.L. 1994. *On measures of deprivation and spatial allocation of resources for primary health care*, *Environment and Planning A*, 26(1911-1929).
- MESQUITA, Z., 1984. *Divisões Regionais do Rio Grande do Sul: uma revisão*. Ensaio FEE, Fundação de Economia e Estatística S.E.H., Porto Alegre, 5(2):95-146.
- MORTENSEN, R. 1987. *Random Signals and Systems*. John Wiley & Sons, New York.
- NETZ, P.A. 1992. Simulação computacional de processos de reticulação. Dissertação de Mestrado, UFRGS, Curso de Pós-graduação em Química, Porto Alegre.
- OLIVEIRA, L.C.S.; et al., 1996. Apuração dos dados investigados pelo questionário da Amostra – CD 1.02 do Censo Demográfico de 1991, *Textos para Discussão da Fundação Instituto Brasileiro de Geografia e Estatística*, Rio de Janeiro, n.86.
- PALACIOS, J.J., 1983. *El concepto de region: la dimension espacial de los procesos sociales*. *Revista Interamericana de Planificación*, México, 17(66):56-68.
- PFEIFFER, D. 1980. Disparidades de desenvolvimento no Brasil – um exemplo de análise de cluster. *Revista Brasileira de Estatística*, Rio de Janeiro, 41(164): 559-576.
- ROCHE, J. 1969. A colonização alemã e o Rio Grande do Sul I e II, Porto Alegre, Globo.
- RODRIGUES, M.C.P. 1991. O índice do desenvolvimento social. *Revista Conjuntura Econômica*, jan/91: 73-77.

- RODRIGUES, M.C.P. 1993. Os pobres e os ricos do Brasil. *Revista Conjuntura Econômica*, mar/93: 46-49.
- RODRIGUES, M.C.P. 1994. O desenvolvimento social nos estados brasileiros. *Revista Conjuntura Econômica*, mar/94: 52-56.
- ROY, G.G. 1974. *A multiple criteria approach to regional planning problems*, *Environment and Planning A*, 6 (313-320).
- RS, Secretaria da Agricultura e Abastecimento, SAA, 1994. *Macrozoneamento agroecológico e econômico do Rio Grande do Sul*, Porto Alegre.
- RS, Secretaria do Desenvolvimento Regional e Obras Públicas, SDO, 1974. *Proposição de Organização Territorial do Estado para Fins de Programação Territorial e Urbana*, cap. VII, Porto Alegre.
- SHEFER, D., Amir, S., Frenkel, A. e Law-Yone, H. 1997. Generating and evaluating alternative regional development plans, *Environment and Planning B: Planning and Design*, 15(7-22).
- SHEN, Q. 1994. An application of GIS to measurement of Spatial Autocorrelation, *Environment and Urban Systems*, 18(167-191).
- SILVA, L.F.; Bianchini, Z.M., 1990. A redução da amostra e a utilização de duas frações amostrais no Censo Demográfico de 1990. Textos para Discussão da Fundação Instituto Brasileiro de Geografia e Estatística, Rio de Janeiro, n.33.
- SILVEIRA, F.G.; Sampaio, M.H. 1996. Índice de Desenvolvimento Social: uma estimativa para os municípios do Rio Grande do Sul, Fundação de Economia e Estatística Siegfried Emanuel Heuser, Porto Alegre.
- TALEN, E. 1998. Visualizing Fairness: Equity Maps for Planners, *Journal of American Planning*, 64/1(22-38).



ANEXOS



ANEXOS

O Anexo I apresenta o TabWin, tabulador oferecido gratuitamente pelo DATASUS do Ministério da Saúde. O Anexo II é constituído pelas duas páginas iniciais do ADE-4 da Universidade de Lyon 1. O Anexo III é uma indicação da origem dos dados utilizados no presente trabalho.

ADE-4: um software de análise multivariada e visualização gráfica

Jean Thioulouse (1), Daniel Chessel (2) & Jean-Michel Olivier (2)

(1) *Laboratoire de Biométrie, Génétique et Biologie des Populations, UMR CNRS 5558, Université Lyon 1, 69622 Villeurbanne Cedex, France.*

(2) *Laboratoire d'Ecologie des Eaux Douces et des Grands Fleuves, URA CNRS 1974, Université Lyon 1, 69622 Villeurbanne Cedex, France.*

Referência:

Thioulouse J., Chessel D., Dolédec S., & Olivier J.M. (1997)

ADE-4: a multivariate analysis and graphical display software. *Statistics and Computing*, 7, 1, 75-83

Resumo

1. Introdução
2. A interface do usuário
 - 2.1 Módulos computacionais
 - 2.2 Módulos Gráficos
 - 2.3 Interface WinPlus e Hypercard
3. Métodos de análise de dados
 - 3.1 Métodos de uma tabela
 - 3.2 Uma tabela com estruturas espaciais
 - 3.3 Uma tabela com grupos de linhas
 - 3.4 Regressão Linear
 - 3.5 Método de duas tabelas acopladas
 - 3.6 Método de análise de Coinércia
 - 3.7 Método de análise de K-tabelas

4 Representações Gráficas

4.1 Gráficos de uma dimensão

4.2 Curvas

4.3 Scatters

4.4 Módulos cartográficos

5. Conclusão

Disponibilidade

Agradecimentos

Referências

Resumo

Apresentamos o ADE-4, um *software* de análise multivariada e visualização gráfica. Os métodos de análise multivariada disponíveis no ADE-4 incluem os métodos usuais de uma tabela como análise de componente principal e análise de correspondência, métodos de análise espacial de dados (usando uma decomposição de variância total em componentes locais e globais, análogos aos índices de Moran e de Geary), análise discriminante e análise intra/entre grupos, muitos métodos de regressão linear incluindo regressão polinomial, regressão múltipla e PLS (*partial least squares* – mínimos quadrados parciais) regressão ortogonal (regressão de componente principal), métodos de projeção como análise de componente principal em variáveis instrumentais, análise de correspondência canônica e muitas outras variantes, análise de coinércia e método RQL, e vários métodos de análise de tabelas (k-tabelas). As técnicas de visualização gráfica incluem uma coleção automática de gráficos elementares correspondente a grupos de linhas e colunas na tabela de dados, então proporciona um modo muito eficiente de gráficos automáticos para k-tabelas e opções de mapeamento geográfico. Um módulo automático gráfico permite operações interativas como busca, ampliação, seleção de pontos, e visualização de valores de dados sobre mapas. A *interface* do usuário é simples e homogênea por todos os programas; isto contribui para fazer o uso de ADE-4 muito fácil para não especialistas em estatística, análise de dados ou ciência da computação.

Palavras-chave:

Análise multivariada, análise de componente principal, análise de correspondência, variáveis instrumentais, análise de correspondência canônica, regressão dos mínimos

quadrados parciais, análise de coinerência, gráficos, gráficos multivariados, gráficos interativos, Macintosh, Hypercard, Windows 95.

Autor correspondente:

Jean Thioulouse

Laboratoire de Biométrie – Université Lyon 1

69622 Villeurbanne Cedex – France

1. Introdução

ADE-4 é um *software* de análise multivariada e visualização gráfica para microcomputadores Macintosh Apple e Windows 95. É composto por várias aplicações independentes, chamados módulos, que abrangem um vasto conjunto de métodos de análise multivariada. Também fornece muitas possibilidades de auxílio de visualização gráfica no processo de análise multivariada de conjuntos de dados. Foi desenvolvido no contexto de análise de dados ambientais, mas pode ser usado em outras disciplinas científicas (por exemplo, sociologia, quimiometria, geociências, etc.), onde a análise de dados é freqüentemente usada. É obtida gratuitamente na Internet. Aqui, queremos apresentar as principais características do ADE-4, de três pontos de vista: (1) interface do usuário, (2) métodos de análise de dados, e (3) capacidades de visualização gráfica.

As primeiras páginas do ADE-4 na Internet foram traduzidas acima como modo de apresentação do *software* e indicação de suas abrangentes possibilidades. Maior aprofundamento pode ser obtido no <http://pbil.univ-lyon1.fr/ADE-4/ref/ADE-4-Web.htm>.