

**ANÁLISE DE DADOS DE ÁREA APLICADA A DOIS
INDICADORES ECONÔMICOS DE MESORREGIÕES
DO ESTADO DE MINAS GERAIS**

NAJE CLÉCIO NUNES DA SILVA

2010

NAJE CLÉCIO NUNES DA SILVA

**ANÁLISE DE DADOS DE ÁREA APLICADA A DOIS INDICADORES
ECONÔMICOS DE MESORREGIÕES DO ESTADO DE MINAS GERAIS**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para obtenção do título de “Mestre”.

Orientador

Prof. Dr. Marcelo Silva de Oliveira

LAVRAS
MINAS GERAIS-BRASIL
2010

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da
Biblioteca Central da UFLA**

Silva, Naje Clécio Nunes da.

Análise de dados de área aplicada a dois indicadores econômicos de mesorregiões do Estado de Minas Gerais / Naje Clécio Nunes da Silva. – Lavras: UFLA, 2010.

85 p. : il.

Dissertação (Mestrado) - Universidade Federal de Lavras, 2010.

Orientador: Marcelo Silva de Oliveira

Bibliografia.

1. Dados de área. 2. Distribuição espacial 3. Autocorrelação. 4. Dependência. I. Universidade Federal de Lavras. II. Título.

CDD - 519.537

NAJE CLÉCIO NUNES DA SILVA

**ANÁLISE DE DADOS DE ÁREA APLICADA A DOIS INDICADORES
ECONÔMICOS DE MESORREGIÕES DO ESTADO DE MINAS GERAIS**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para obtenção do título de “Mestre”.

APROVADA em 3 de março de 2010

Prof. Dr. Eric Batista Ferreira

UNIFAL-MG

Prof. Dr. Renato Ribeiro de Lima

UFLA

Prof. Dr. Marcelo Ângelo Cirillo

UFLA

Prof. Dr. Marcelo Silva de Oliveira
UFLA
(Orientador)

LAVRAS
MINAS GERAIS - BRASIL

Dedico esta vitória:

A DEUS

*Pela sua graça que alcançamos por meio de seu filho Jesus Cristo;
Aos meus pais, Raimundo Ribeiro e Graça Nunes;
À minha esposa, Mailcia, pelo apoio e amor;
À Meu filho querido, Estêvão.*

*“ Porque Deus amou o mundo de tal maneira que deu o seu Filho unigênito,
para que todo aquele que nele crê não pereça, mas
tenha a vida eterna.”
João 3:16*

*“ A saber: Se com a tua boca confessares ao Senhor Jesus, e em teu coração
creres que Deus o ressuscitou dentre os mortos,
serás salvo.”
Romanos 10:9*

AGRADECIMENTOS

A Deus, pela sua graça que alcançamos por meio de seu filho nosso Senhor Jesus Cristo.

À minha esposa querida Mailcia, por ser a minha princesa que Deus me deu.

Ao meu filho querido, Estêvão, príncipe do Senhor Jesus.

Ao professor Marcelo Silva de Oliveira, pela responsabilidade que tem com seu trabalho, pela confiança que me concedeu, pelos ensinamentos e orientações, muitas vezes tirando dúvidas, outras vezes dando sugestões, criticando e aconselhando.

Aos meus pais queridos, Raimundo Ribeiro e Graça Nunes, pela confiança, fé, compreensão e amor.

Às minhas irmãs, Núbia, Iris e Maressa, por serem bênçãos de Deus em minha vida.

Aos meus sobrinhos Pedro Werner, Isis Richelle, Mateus, Lucas, Jenyfer, Jhonatan e Jhames, pelo respeito e amor.

Aos meus cunhados Emerson, Antônio Werner e Moisés, e cunhada Maira, pela amizade, respeito e carinho.

Ao meu sogro Moisés e à minha sogra Maria, pelo amor e carinho.

Aos meus pastores Disney e Rosane, por serem bênçãos de Deus em minha vida.

À todos do Ministério Ultra-Jovem, pelas orações e companheirismo nessa jornada, por amor a Jesus Cristo.

Aos irmãos Donizete, Nilson e Kennedy, pelos conselhos e palavras de confiança.

Aos meus irmãos Jucenildo, Aleksandres, Dean, Nilberto, Jonas, Nenê, Nelson e Cidiclei, por serem pessoas que me ajudaram muito no início da minha caminhada com Jesus Cristo.

Aos meus amigos Cosmo, Damião, Everton, Arlex, José Nunes, Francisco Antônio, Assis, Heberton, Júlio (paroara), Arnaldo Paixão, Ronildo, Weverton, Edcarlos, Paulo e Ricardo, pela amizade, respeito e carinho.

Ao grande irmão Altemir da Silva Braga, pelo incentivo, companheirismo e

ajuda desde o início do mestrado até a conclusão desse trabalho, e a sua esposa Carmem e filhos, por fazerem parte da minha família.

Ao casal Professor Eustáquio e Francisca, por estarem sempre me apoiando nessa jornada.

Ao amigo Arionaldo, pelos ensinamentos.

À todos os colegas de mestrado e doutorado em Estatística, em especial a minha turma Paulo, Jair, Deyse, Gilson, Manoel, Walmes, Rosiana, Simone, Púpio, Flávia e Franciella pela amizade.

Aos professores Antônio Carlos e José Marques, pelas horas de estudo e amizade.

Aos Professores Tadeu Alexandre, Aroldo Alexandre, Manoel Domingos, Simone, Salete, José Ivan e Geirto (o abençoado), pelos ensinamentos.

À meu co-orientador e doutorando, Professor Gérson Rodrigues dos Santos (brother), pelos ensinamentos e companheirismo.

Ao grande irmão João Marcos, pelo incentivo e companheirismo.

À todos os meus professores da graduação e da especialização.

Aos professores do Departamento de Ciências Exatas, pelos ensinamentos.

À Universidade Federal de Lavras e ao Departamento de Ciências Exatas, e todos os funcionários, em especial, a Josi e Selminha.

Ao CNPQ, pela bolsa de estudo, essencial para a realização deste trabalho.

Aos demais que, direta ou indiretamente, contribuíram para a elaboração deste trabalho.

Najé Elcio Nunes da Silva

SUMÁRIO

LISTA DE TABELAS	i
LISTA DE FIGURAS	ii
LISTA DE SÍMBOLOS	vii
RESUMO	viii
ABSTRACT	ix
1 INTRODUÇÃO	1
2 REFERENCIAL TEÓRICO	3
2.1 Análise de dados espaciais	3
2.1.1 Dados de processos pontuais	6
2.1.2 Dados de superfície aleatória	7
2.1.3 Dados de interação espacial	7
2.1.4 Dados de área	8
2.1.5 Sistema de informação geográfica (SIG)	9
2.2 Análise de dados de área	12
2.2.1 Problema típico que demanda análise de dados de área	13
2.2.2 Visualização de dados	14
2.2.3 Modelos de distribuição de dados de área	15
2.2.4 Análise de autocorrelação espacial	16
2.2.5 Matriz de proximidade espacial	17
2.2.6 Média móvel local	20
2.2.7 Índice de Moran	24
2.2.8 Teste de autocorrelação espacial	29
2.2.8.1 Distribuição aproximada de I	29
2.2.8.2 Teste de permutação aleatória	30
2.2.9 Índice de Moran local	32
2.2.10 Gráfico de espalhamento de Moran	34
2.2.11 <i>Box Map, LISA Map e Moran Map</i>	37
2.2.11.1 <i>Box Map</i>	37
2.2.11.2 <i>LISA Map</i>	38
2.2.11.3 <i>Moran Map</i>	39

2.2.12	Resumo das técnicas de análise de dados de área	40
3	MATERIAL E MÉTODOS	41
3.1	Origem dos dados	41
3.2	Produto interno bruto (PIB)	41
3.3	Valor adicionado agropecuário (VAA)	42
3.4	Análise exploratória	43
3.5	Média móvel local	43
3.6	Índice de Moran global	44
3.7	Simulação do valor-p para o índice I de Moran utilizando teste de permutação aleatória	45
3.8	Gráfico de espalhamento de Moran e <i>Box Map</i>	46
3.9	Índice de Moran local e <i>Moran Map</i>	46
3.10	Softwares computacionais utilizados	47
4	RESULTADOS E DISCUSSÃO	48
4.1	Estudo da variável IPIB na mesorregião Metropolitana de Belo Horizonte e na mesorregião Sul/Sudeste de Minas	48
4.1.1	Análise exploratória	48
4.1.2	Média móvel local	49
4.1.3	Índice de Moran global e o teste de permutação aleatória	51
4.1.4	Gráfico de espalhamento de Moran e <i>Box Map</i>	54
4.1.5	Índice de Moran local e <i>Moran Map</i>	56
4.2	Estudo da variável IVAA na mesorregião Metropolitana de Belo Horizonte e na mesorregião Sul/Sudeste de Minas	58
4.2.1	Análise exploratória	58
4.2.2	Média móvel local	59
4.2.3	Índice de Moran global e o teste de permutação aleatória	61
4.2.4	Gráfico de espalhamento de Moran e <i>Box Map</i>	63
4.2.5	Índice de Moran local e <i>Moran Map</i>	65
5	CONCLUSÕES	69
5.1	Para a variável IPIB	69
5.1.1	Quanto ao padrão espacial	69
5.1.2	Quanto à autocorrelação espacial	69

5.2 Para a variável IVAA	69
5.2.1 Quanto ao padrão espacial	69
5.2.2 Quanto à autocorrelação espacial	69
5.3 Trabalhos futuros	70
REFERÊNCIAS BIBLIOGRÁFICAS	71
ANEXOS	74

LISTA DE TABELAS

1	Matriz de proximidade espacial por vizinhança de primeira ordem.	19
2	Matriz de proximidade espacial por vizinhança de segunda ordem.	19
3	Matriz de proximidade espacial por vizinhança de primeira ordem, padronizada pelas linhas da Tabela 1.	20
4	Matriz de proximidade espacial por vizinhança de segunda ordem, padronizada pelas linhas da Tabela 2.	21
5	Índice do produto interno bruto dos municípios localizados na me- sorregião Campo das Vertentes - MG, no ano de 2004.	27
6	Cálculos Auxiliares para a determinação do índice global de Moran.	28
7	Sumarização das técnicas de análise de dados de área e suas funções.	40

LISTA DE FIGURAS

1	Localização de poços de água identificados por cruces e óbitos por cólera representados por pontos. Londres, 1854. Fonte: Câmara et al. (2001).	5
2	Mapa com pontos de coleta do pH do solo em uma superfície aleatória hipotética. Fonte: Assunção (2001).	8
3	Mapa da interação espacial entre os municípios de uma determinada região hipotética. Fonte: Assunção (2001).	9
4	Mapa da distribuição espacial do IDH dos municípios do Vale do Paraíba paulista, 2007. Fonte: Oliveira & Nascimento (2007).	10
5	Estrutura básica de sistema de informação geográfica. Fonte: Câmara et al. (2001).	12
6	Mapa do percentual de idosos (com mais de 70 anos) dos 96 bairros da cidade de São Paulo (censo 1991). Fonte: Câmara et al. (2001).	14
7	Mapa construído por elaboração própria de sete municípios localizados na mesorregião Campo das Vertentes - MG. Lavras, 2009. Fonte: Base Cartográfica do Instituto Brasileiro de Geografia e Estatística - IBGE (2002).	18
8	Distribuição dos idosos na cidade de São Paulo (censo de 1991). À esquerda, mapa de desvios-padrão. À direita, mapa de desvios-padrão com a média móvel local. Fonte: Câmara et al. (2001).	24
9	Gráfico de espalhamento de Moran, onde a é inclinação da reta. Fonte: Santos & Souza (2007).	34
10	<i>Box Map</i> da variável percentagem de viagens de ônibus na cidade de Bauru - SP. Fonte: Krempi (2004).	38

11	<i>LISA Map</i> da variável percentagem de viagens de ônibus na cidade de Bauru - SP. Fonte: Krempi (2004).	39
12	<i>Moran Map</i> da variável percentagem de viagens de ônibus na cidade de Bauru - SP. Fonte: Krempi (2004).	40
13	Mapa construído por elaboração própria da mesorregião Metropolitana de Belo Horizonte - MG. Lavras, 2010. Fonte: Base cartográfica do IBGE (2002).	42
14	Mapa construído por elaboração própria da mesorregião Sul/Sudeste de Minas - MG. Lavras, 2010. Fonte: Base cartográfica do IBGE (2002).	43
15	Mapa de intervalos iguais com gráfico de barras para a variável IPIB na mesorregião Metropolitana de Belo Horizonte - MG, 2004. Fonte: Construído por elaboração própria com base no software TerraView (INPE, 2007).	49
16	Mapa de intervalos iguais com gráfico de barras para a variável IPIB na mesorregião Sul/Sudeste de Minas - MG, 2004. Fonte: Construído por elaboração própria com base no software TerraView (INPE, 2007).	50
17	Mapa da distribuição da variável IPIB na mesorregião Metropolitana de Belo Horizonte - MG, 2004. À esquerda, apresentação da variável por quintil. À direita, média móvel local. Lavras, 2010. Fonte: Construído por elaboração própria com base no software TerraView (INPE, 2007).	51
18	Mapa da distribuição da variável IPIB na mesorregião Sul/Sudeste de Minas - MG, 2004. À esquerda, apresentação da variável por quintil. À direita, média móvel local. Fonte: Construído por elaboração própria com base no software TerraView (INPE, 2007).	52

19	Teste de significância para o índice de Moran global da variável IPIB sob a hipótese nula de ausência de autocorrelação espacial. Fonte: Construído por elaboração própria com base no software GeoDA (Anselin, 2004).	53
20	Teste de significância para o índice de Moran global da variável IPIB sob a hipótese nula de ausência de autocorrelação espacial. Fonte: Construído por elaboração própria com base no software GeoDA (Anselin, 2004).	53
21	À esquerda, diagrama de espalhamento de Moran da variável IPIB. À direita, <i>Box Map</i> da variável IPIB. Fonte: Construído por elaboração própria com base nos softwares GeoDA (Anselin, 2004) e TerraView (INPE, 2007).	54
22	À esquerda, diagrama de espalhamento de Moran da variável IPIB. À direita, <i>Box Map</i> da variável IPIB. Fonte: Construído por elaboração própria com base nos softwares GeoDA (Anselin, 2004) e TerraView (INPE, 2007).	55
23	Mapa do índice de Moran local da variável IPIB na mesorregião Metropolitana de Belo Horizonte-MG, 2004. À esquerda, valores positivos e negativos do índice de Moran local. À direita, <i>Moran Map</i> . Fonte: Construído por elaboração própria com base nos softwares TerraView (INPE, 2007) e GeoDA (Anselin, 2004).	56
24	Mapa do índice de Moran local da variável IPIB na mesorregião Sul/Sudeste de Minas-MG, 2004. À esquerda, valores positivos e negativos do índice de Moran local. À direita, <i>Moran Map</i> . Fonte: Construído por elaboração própria com base nos softwares TerraView (INPE, 2007) e GeoDA (Anselin, 2004).	57
25	Mapa de intervalos iguais com gráfico de barras para a variável IVAA na mesorregião Metropolitana de Belo Horizonte - MG, 2004. Fonte: Construído por elaboração própria com base no software TerraView (INPE, 2007).	59

26	Mapa de intervalos iguais com gráfico de barras para a variável IVAA na mesorregião Sul/Sudeste de Minas - MG, 2004. Fonte: Construído por elaboração própria com base no software TerraView (INPE, 2007).	60
27	Mapa da distribuição da variável IVAA da mesorregião Metropolitana de Belo Horizonte - MG, 2004. À esquerda, apresentação da variável por quintil. À direita, média móvel local. Fonte: Construído por elaboração própria com base no software TerraView (INPE, 2007).	61
28	Mapa da distribuição da variável IVAA na mesorregião Sul/Sudeste de Minas - MG, 2004. À esquerda, apresentação da variável por quintil. À direita, média móvel local. Fonte: Construído por elaboração própria com base no software TerraView (INPE, 2007).	62
29	Teste de significância para o índice de Moran global da variável IVAA sob a hipótese nula de ausência de autocorrelação espacial. Fonte: Construído por elaboração própria com base no software GeoDA (Anselin, 2004).	62
30	Teste de significância para o índice de Moran global da variável IVAA sob a hipótese nula de ausência de autocorrelação espacial. Fonte: Construído por elaboração própria com base no software GeoDA (Anselin, 2004).	63
31	À esquerda, diagrama de espalhamento de Moran da variável IVAA. À direita, <i>Box Map</i> da variável IVAA. Fonte: Construído por elaboração própria com base nos softwares GeoDA (Anselin, 2004) e TerraView (INPE, 2007).	64
32	À esquerda, diagrama de espalhamento de Moran da variável IVAA. À direita, <i>Box Map</i> da variável IVAA. Fonte: Construído por elaboração própria com base nos softwares GeoDA (Anselin, 2004) e TerraView (INPE, 2007).	65

33	<p>Mapa do índice de Moran local da variável IVAA na mesorregião Metropolitana de Belo Horizonte - MG, 2004. À esquerda, valores positivos e negativos do índice de Moran local. À direita, <i>Moran Map</i>.</p> <p>Fonte: Construído por elaboração própria com base nos softwares TerraView (INPE, 2007) e GeoDA (Anselin, 2004).</p>	66
34	<p>Mapa do índice de Moran local da variável IVAA na mesorregião Sul/Sudeste de Minas - MG, 2004. À esquerda, valores positivos e negativos do índice de Moran local. À direita, <i>Moran Map</i>.</p> <p>Fonte: Construído por elaboração própria com base nos softwares TerraView (INPE, 2007) e GeoDA (Anselin, 2004).</p>	67

LISTA DE SÍMBOLOS

AA	Alto-Alto
AB	Alto-Baixo
ADA	Análise de D ados de Á rea
BA	Baixo-Alto
BB	Baixo-Baixo
BM	Box-Map
ESRI	Environmental Systems R esearch I nstitute - Instituto de Pesquisa de Sistemas Ambientais
GEM	Gráfico de E spalhamento M oran
GeoDA	G eographical D ata A nalisis - Análise de Dados Geográficos
GIS	G eographical I nformation S ystem - Sistema de Informação Geográfica
GPS	G lobal P osition S ystem - Sistema de Posicionamento Global
IBGE	Instituto B rasileiro de G eografia e E statística
IDH	Índice de D esenvolvimento H umano
IMG	Índice de M oran G lobal
IML	Índice de M oran L ocal
INPE	I nstituto de P esquisas E spaciais
IPIB	Índice do P roduto I nterno B ruto
IVAA	Índice do V alor A dicionado A gropecuário
LM	Lisa-Map
MG	Minas G erais
MM	M oran- M ap
MMBH	M esorregião M etropolitana de B elo H orizonte
MML	M édia M óvel L ocal
MSSM	M esorregião S ul/ S udeste de M inas
pH	P otencial H idrogeniônico
SUFRAMA	S uperintendência da Zona F ranca de M anus
TPA	T este de P ermutação A leatória
ZEE	Z oneamento E cológico E conômico

RESUMO

SILVA, Naje Clécio Nunes da. **Análise de dados de área aplicada a dois indicadores econômicos de mesorregiões do Estado de Minas Gerais**. 2010. 85 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras. *

No presente trabalho buscou-se analisar a distribuição espacial do índice do produto interno bruto (IPIB) e do índice do valor adicionado agropecuário (IVAA) nas mesorregiões Metropolitana de Belo Horizonte (MMBH) e Sul/Sudeste de Minas (MSSM), no Estado de Minas Gerais, no ano de 2004, visando a entender a dependência ou semelhança entre os municípios dessas mesorregiões por meio desses índices. Esse entendimento poderá servir de base para tomada de decisões na implantação de políticas públicas para o desenvolvimento regional. Para tanto, utilizou-se das técnicas de dados de área, destacando-se a média móvel local, o índice de Moran global, o índice de Moran local, gráfico de espalhamento de Moran e Mapas (*Box Map* e *Moran Map*); como também, da base de dados dos IPIB e IVAA, produzidas pelo Zoneamento Ecológico Econômico do Estado de Minas Gerais, referente ao ano de 2004, e dos softwares GeoDA e TerraView. Os resultados alcançados quanto ao padrão espacial pela média móvel local, no estudo do IPIB e no estudo do IVAA, permitiram observar *clusters* espaciais de municípios tanto na MMBH quanto na MSSM. E, também trouxeram benefícios na redução da variância amostral para as duas variáveis em ambas as mesorregiões. Os resultados alcançados quanto à autocorrelação espacial nas mesorregiões foram: na MMBH, no estudo da variável IPIB e IVAA, apresentou autocorrelação espacial positiva em termos globais, pelo índice de Moran global e o teste de permutação aleatória, indicando que existe similaridade entre os municípios dessa mesorregião. E constatou-se também autocorrelação espacial em termos locais, pelo índice de Moran local e o *Moran Map*, indicando também similaridade entre os municípios dessa mesorregião. Na MSSM, no estudo da variável IPIB e IVAA, também apresentou autocorrelação espacial positiva em termos globais, pelo índice de Moran global e o teste de permutação aleatória, indicando que existe similaridade entre os municípios dessa mesorregião. E constatou-se autocorrelação espacial em termos locais, pelo índice de Moran local e o *Moran Map*, indicando também similaridade entre os municípios dessa mesorregião.

Palavras-chave: dados de área, distribuição espacial, autocorrelação, dependência.

*Comitê Orientador: Prof. Dr. Marcelo Silva de Oliveira - UFLA (Orientador) e Prof. Msc. Gérson Rodrigues dos Santos - FIAMG.

ABSTRACT

SILVA, Naje Clécio Nunes da. **Analysis of area data applied to two economic indicators of mesoregions in Minas Gerais State.** 2010. 85 p. Dissertation (Master Science in Statistics and Agricultural Experimentation) - Federal University of Lavras, Lavras. *

The current work aimed to analyze the spatial distribution of gross domestic product index (GDPI) and the agricultural added value index (ADVI) on the mesoregion metropolitan area of Belo Horizonte (MABH) and South/Southeast Minas areas (SSMA), in Minas Gerais State, in 2004, looking for comprehending the dependence or similarity amongst the municipalities within these regions through those indexes. This understanding may become an important tool for decision making in implanting public policies for regional development. Therefore, the area data techniques was used, detaching the local mobile mean, the global Moran index, the local Moran index, Moran scatterplot and maps (Box Map and Moran Map); as well from the GDPI and ADVI database, produced through Ecological and Economical Zoning of Minas Gerais State, referring to 2004 year, and the softwares GeoDa and TerraView. The results regarding to spatial pattern by the local mobile mean, on both GDPI and ADVI studies, permitted to heed spatial clusters of municipalities for both studied regions, i.e., MABH and SSMA. Also, they brought benefits in reducing the sample variance for the two variables in both regions. The results concerning to the study of GDPI and ADVI variables for the MABH region presented a positive autocorrelation overall, according to Moran global and random permutation test, pointing there are likeness among the municipalities composing this region. It was evidenced also, positive spatial autocorrelation by Moran local index and Moran Map, again showing similarities among the municipalities. Concerning the study of GDPI and ADVI variables for the SSMA region, it has shown positive spatial autocorrelation overall, by both Moran global and random permutation test, suggesting there are resemblance among the municipalities within this region. It was verified spatial autocorrelation in local extension, by Moran local index and Moran Map, again showing similitude among the municipalities.

Key-words: area data, spacial distribution, autocorrelation, dependence.

*Guidance Committee: Marcelo Silva de Oliveira - UFLA (Advisor) and Gérson Rodrigues dos Santos - FIAMG.

1 INTRODUÇÃO

A análise de dados de área é um dos ramos de pesquisa da Estatística Espacial, em que a aplicação está associada a áreas A_i , com $i = 1, 2, \dots, n$, contidas numa região R , de forma que $\bigcup_{i=1}^n A_i = R$ e $A_i \cap A_j = \emptyset$ se $i \neq j$, qualquer que seja a forma de A_i e de R . Tal análise ocorre com muita frequência quando lida-se com dados agregados por municípios, bairros, setores censitários ou zonas de tráfego, entre outros, em que não se dispõe da localização exata dos dados, mas de um valor por área. Nessas análises, procura-se identificar padrões espaciais de distribuição nos valores observados.

As técnicas de análise de dados de área foram desenvolvidas para tentar identificar regiões onde a distribuição dos valores possa apresentar um padrão específico associado a sua localização espacial. A informação que se busca é o quanto o valor de uma variável em uma determinada área assemelha-se com os valores da mesma variável localizada em sua vizinhança próxima, e o quanto é similar ou dissimilar aos valores de sua vizinhança distante.

Essas técnicas são apresentadas na forma de indicadores, que medem, tanto o padrão, quanto a associação espacial, dentre as quais destacam-se: o índice de Moran global para a associação (ou correlação) espacial e a média móvel local para o padrão espacial. Em especial, o índice de Moran global constitui uma aproximação mais tradicional do efeito da dependência espacial, em que o esquema geral de dependência pode ser resumido em um único valor e pode ser visualizado por meio do gráfico de espalhamento de Moran e de um mapa (*Box Map*). Tem-se também o índice de Moran local. Esse indicador é usado quando se dispõe de grande número de áreas, resultantes, por exemplo, de escalas espaciais detalhadas, e é muito provável a existência de diferentes regimes de correlação espacial em diferentes sub-regiões. Esses diferentes regimes também podem ser visualizados por meio de mapas (*LISA Map* e o *Moran Map*).

A análise de dados oriundos de áreas está bem desenvolvida e descrita em detalhes em diversas referências, tais como Cliff & Ord (1981), Bailey & Gatrell (1995), Anselin (1995,1996), Câmara et al. (2001) e Assunção (2001). Apesar dessa já substancial produção bibliográfica sobre o tema, esta dissertação pro-

cura contribuir com o desenvolvimento dessa área da Estatística Espacial, fazendo sua apresentação, formulação, e aplicação, por meio de duas questões centrais de pesquisa:

(i) existe um padrão espacial para uma dada variável de interesse?

(ii) existe uma estrutura de dependência espacial (autocorrelação espacial) da variável de interesse?

Visando a responder a essas questões, neste trabalho, objetivou-se estudar técnicas de análise de dados de área aplicadas à distribuição espacial de duas variáveis de interesse, a saber, o índice do produto interno bruto (IPIB) e o índice do valor adicionado agropecuário (IVAA) das mesorregiões Metropolitana de Belo Horizonte e Sul/Sudeste de Minas - MG, no ano de 2004. Essas variáveis foram estudadas no Zoneamento Ecológico Econômico do Estado de Minas Gerais.

2 REFERENCIAL TEÓRICO

Inicialmente, esse capítulo estuda a análise de dados espaciais no geral, abordando os principais conceitos de Estatística Espacial, os processos e ferramentas de análise espacial, a classificação dos dados espaciais e algumas considerações sobre sistema de informação geográfica. Em seguida, apresenta-se a análise de dados de área, um problema típico que demanda análise de dados de área e um conjunto de técnicas de análise de dados agregados por área que serão utilizados para o desenvolvimento dessa dissertação.

2.1 Análise de dados espaciais

Sabe-se que, para realizar qualquer estudo na área de Estatística, precisa-se da “matéria-prima”, que são os dados. Os dados são registros de observações oriundas da realização de variáveis em experimentos ou em levantamentos amostrais.

Quando as variáveis são indexadas espacialmente, isto é, quando a um valor realizado de uma variável se pode associar uma coordenada espacial, então tais dados são chamados dados espaciais (ver definição estatística na subseção 2.2.3).

Estatística Espacial é o conjunto de conceitos e métodos estatísticos que podem ser utilizados para a análise de dados espaciais. Frequentemente, mesmo que não necessariamente, o espaço que indexa as variáveis é o espaço geográfico.

A análise de dados espaciais pode ser definida como a aplicação de um conjunto de técnicas estatísticas que buscam descrever os padrões existentes nos dados espaciais e estabelecer, preferencialmente, de forma quantitativa, os relacionamentos entre as diversas variáveis geográficas, ou de modo mais geral, as diversas variáveis espaciais.

Considerando as etapas do processo de um tratamento estatístico de dados espaciais, de acordo com Bailey & Gatrell (1995), os processos de análise espacial incluem métodos de visualização, métodos que investiguem o padrão espacial dos dados e sua correlação espacial, métodos que auxiliem na escolha de um modelo estatístico e a estimação dos parâmetros desse modelo.

Dessa forma, as ferramentas de Estatística Espacial podem ser destinadas a seleção de variáveis, manipulação, análise exploratória e análise confirmatória. A

seleção é feita por meio da navegação em bancos de dados, identificando os dados de interesse, realizando consultas e apresentando mapas. O processo de manipulação envolve todas as funções que operam sobre dados espaciais. As técnicas de análise exploratória descrevem e visualizam dados espaciais, verificando a existência de padrões de comportamento e de associação espacial. As técnicas de análise confirmatória envolvem o conjunto de modelos de estimação e teste, e procedimentos para sua validação.

Segundo Assunção (2001), a visualização é um fator importante da Estatística Espacial, mas não tão útil se não for possível descrever, comparar e interpretar os dados. A Estatística Descritiva, assim como a Estatística Inferencial são úteis ou, algumas vezes, essenciais para tornarem explícito o que está apenas implícito em mapas, ou para analisá-los, chamando a atenção para características que não são perceptíveis apenas pela visualização.

Tomando agora outro ponto de vista, pode-se considerar os tipos de modelos utilizados no tratamento estatístico de dados espaciais. Para Landim (1998), quando se utilizam modelos da Estatística Clássica para representar as propriedades dos valores amostrais distribuídos espacialmente, suspeita-se que esses sejam realizações de uma variável aleatória, que as posições relativas da amostra são ignoradas e presume-se que todos os valores amostrais tenham a mesma probabilidade de serem escolhidos. Assim, muitos modelos da Estatística Clássica exigem a independência das variáveis.

De acordo com Landim (1998), Estatística Espacial é um outro conjunto de modelos que traz resultados diferentes daqueles obtidos pela Estatística Clássica, sendo os primeiros geralmente mais informativos por incorporarem a dimensão espacial. Para sua análise, são necessárias, pelo menos, as informações sobre a localização e os atributos, que são os valores associados aos dados. Parte-se do pressuposto de que os dados podem ser espacialmente dependentes.

Para mostrar a importância da localização amostral, cita-se Câmara et al. (2001), que apresentam o caso estudado pelo médico inglês John Snow, na cidade de Londres, em 1854, onde consta o caso de uma grave epidemia de cólera, doença desconhecida para a época quanto à forma de contaminação. Como uma das estratégias para lidar com o problema, foi identificada no mapa da cidade a localização exata

das residências com óbitos por cólera e dos poços de água que abasteciam a cidade. Com a representação espacial dos dados, percebeu-se que a maioria dos casos estava concentrada em torno do poço de Broad Street, conforme mostra a Figura 1. Como consequência, este foi lacrado, contribuindo para acabar com a epidemia.

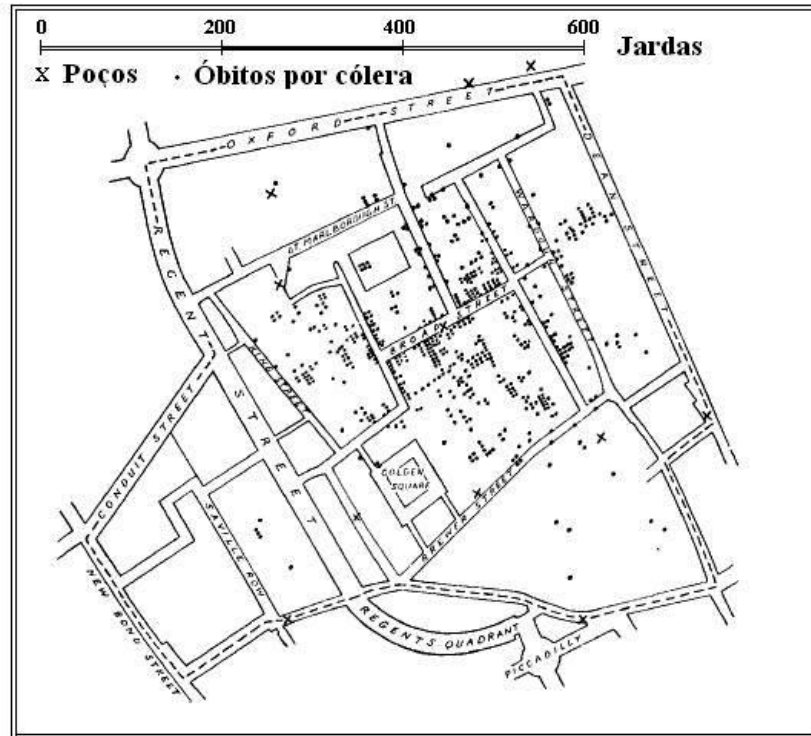


FIGURA 1 Localização de poços de água identificados por cruzes e óbitos por cólera representados por pontos. Londres, 1854.
Fonte: Câmara et al. (2001).

Para Bailey & Gatrell (1995) e Assunção (2001), considerando que a Estatística Espacial é um ramo da Estatística que estuda métodos científicos para a coleta, descrição, visualização e análise de dados que possuem coordenadas espaciais, os dados espaciais podem ser classificados em quatro categorias:

- (i) Dados de processos pontuais;
- (ii) Dados de superfície aleatória;

(iii) Dados de interação espacial;

(iv) Dados de área.

Como o que diferencia essas categorias é o tipo de dado, é natural que em cada um desses quatro tipos existam métodos estatísticos diferentes para descrever ou analisar-los.

Das três definições vistas de estatística espacial, será adotada, nesta dissertação, a proposta por Bailey & Gatrell (1995) e Assunção (2001). Como também, será estudado a quarta categoria, dados de área. Entretanto, serão descritos brevemente, os quatro tipos clássicos de dados que são tratados na Estatística Espacial, e o contexto em que surgem (subseções de 2.1.1 a 2.1.4) e, adicionalmente, será descrito sobre o sistema de informação geográfica (subseção 2.1.5).

2.1.1 Dados de processos pontuais

Segundo Câmara et al. (2001), um padrão pontual representa um conjunto de realizações consistindo de uma série de localizações pontuais ($\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_i, \dots$) em que \mathbf{x}_i é um vetor de coordenadas espaciais, usualmente de duas dimensões, e as localizações indicam a ocorrência de eventos de interesse dentro da área de estudo. Esses eventos ocorreriam de um modo aleatório, fazendo as localizações ($\mathbf{x}_1, \mathbf{x}_2, \dots$) serem estocásticas. A palavra “evento” é utilizada para expressar qualquer tipo de fenômeno localizado no espaço. Esses fenômenos podem representar, por exemplo, crimes, ocorrências de doenças e localizações de espécies vegetais.

Nesse tipo de análise, as observações estão associadas a valores, mas apenas a ocorrência dos eventos é considerada. A área das ocorrências dos eventos não é uma medida usada. A princípio, seriam apenas as localizações, porém, os dados não são compostos exclusivamente pela localização dos eventos, podendo também ter outros atributos associados que podem ser incorporados na análise.

O objetivo da análise espacial de pontos é estudar a distribuição espacial desses pontos, isto é, localizações, testando hipóteses sobre o padrão observado: se é aleatório, se apresenta-se em aglomerados, ou se os pontos estão regularmente distribuídos, buscando entender em qual escala esse padrão ocorre.

O exemplo mais tradicional desse tipo de abordagem é o estudo feito por John

Snow do mapeamento da epidemia de cólera em Londres, visto na Figura 1. A análise de padrão de pontos é a área da Estatística Espacial que usa estes tipos de dados.

2.1.2 Dados de superfície aleatória

Para Assunção (2001), o que distingue esse tipo de dado é o que é considerado aleatório: aleatória é a superfície $Z(\mathbf{x})$ do fenômeno estudado (temperatura, acidez do solo, pH da água de um lago, etc.), e não a localização do evento. Geralmente, existem n pontos de coleta de dados em localizações fixas ($\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$). Nestes pontos, a superfície $Z(\mathbf{x})$ é observada exatamente, gerando as variáveis aleatórias ($Z(\mathbf{x}_1), Z(\mathbf{x}_2), \dots, Z(\mathbf{x}_n)$) cuja distribuição conjunta é induzida pela distribuição da superfície aleatória $Z(\mathbf{x})$. A Geoestatística é a parte da Estatística Espacial utilizada para analisar tais dados.

De acordo com Andriotti (2004), a Geoestatística é uma abordagem probabilística de modelagem que engloba um conjunto de métodos estatísticos para a análise e mapeamento de dados distribuídos no espaço e/ou no tempo, que leva em consideração um modelo explícito de dependência espacial. Sua aplicação faz-se em vários campos, desde as ciências da terra e atmosfera, na agricultura, nas ciências dos solos e hidrologia, estudos ambientais e, mais recentemente, na epidemiologia, entre outras.

As técnicas mais usuais na análise desse tipo de dados são voltadas para a predição espacial, ou seja, para estimar valores da variável em locais não amostrados. Uma das técnicas mais utilizadas para a predição é denominada krigagem. Na Figura 2, mostra-se um exemplo esquemático da krigagem.

Na Figura 2, observa-se uma superfície aleatória com pontos de coleta do pH do solo indicados por setas, e uma localização onde deseja-se predizer o valor não observado da superfície, indicada por um raio, por meio da krigagem.

2.1.3 Dados de interação espacial

Conforme Bailey & Gatrell (1995) e Assunção (2001), em dados de interação espacial, tem-se posições fixas, tal como em dados de superfícies aleatórias. Entretanto, essas posições fixas são tratadas em pares ordenados, sendo que um dos

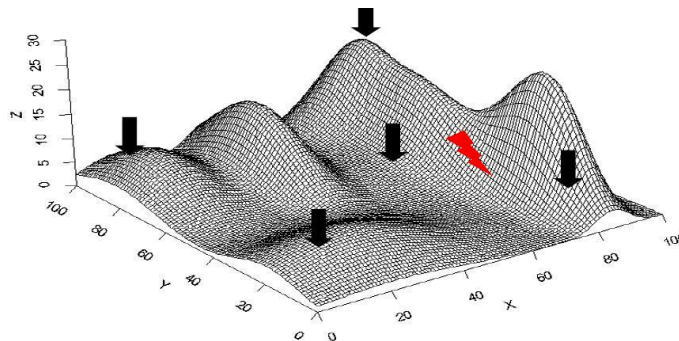


FIGURA 2 Mapa com pontos de coleta do pH do solo em uma superfície aleatória hipotética.
Fonte: Assunção (2001).

elementos do par é considerado uma origem i enquanto o outro é um destino j . O que trafega entre esses pares é um fluxo, considerado como aleatório.

Em geral, o objetivo da análise de interação espacial é encontrar um modelo que se ajusta ao fluxo, ou seja, encontrar um Y_{ij} que representa o fluxo da posição i para a posição j . Exemplos muito comuns são os dados de fluxo migratório, o fluxo de pacientes entre centros de atendimento, o fluxo de passageiros entre cidades, entre outros. Na Figura 3, mostra-se um exemplo esquemático de dados de interação espacial.

Na Figura 3, observa-se que os círculos equivalem aos municípios, sendo o raio proporcional ao tamanho da população. E o segmento de reta entre os círculos é proporcional ao fluxo de passageiros, e esse fluxo é a variável aleatória em estudo. A análise de dados de interação espacial é a área da Estatística Espacial que cuida deste tipo de dado.

2.1.4 Dados de área

Dados de área referem-se a um mapa de uma região R , particionado em áreas A_i , $i = 1, 2, \dots, n$, onde, em cada uma delas, mede-se uma ou mais variáveis aleatórias Y_i , e possivelmente, covariáveis de interesse, que, supostamente, afetam a distribuição de probabilidade de Y_i (Assunção, 2001).

O objetivo da análise é verificar a existência de um padrão espacial e de uma

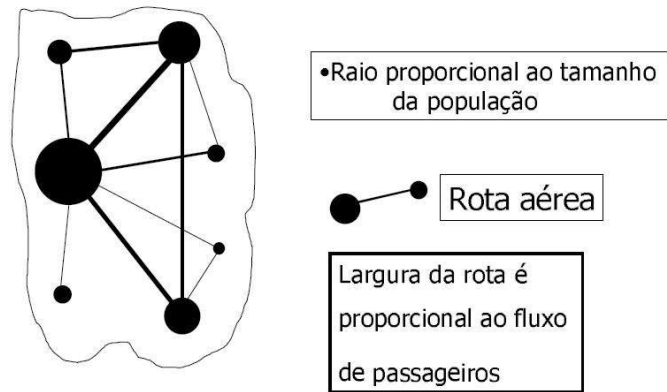


FIGURA 3 Mapa da interação espacial entre os municípios de uma determinada região hipotética.
Fonte: Assunção (2001).

estrutura de correlação espacial nos valores observados. A forma usual de apresentação dos dados agregados por área é por meio de mapas coloridos com o padrão espacial do fenômeno. Quando esse padrão espacial é observado, pode-se verificar se é aleatório ou se apresenta uma agregação definida, ou ainda se esta distribuição pode ser associada a causas mensuráveis, porém, nesta dissertação não se abordará esse problema. Como exemplo de dados de área, a Figura 4 representa o índice de desenvolvimento humano dos municípios do Vale do Paraíba paulista.

Segundo Oliveira & Nascimento (2007), na Figura 4, os municípios com melhores valores de IDH (entre 0,801 e 0,849), concentram-se ao longo da Rodovia Presidente Dutra. Sendo assim, pode-se considerar que o eixo econômico formado pela rodovia favorece o desenvolvimento dos municípios mais próximos, melhorando as condições de vida de suas populações. A análise de dados de área é a parte da Estatística Espacial que estuda tais dados.

2.1.5 Sistema de informação geográfica (SIG)

Compreender a distribuição espacial de dados oriundos de fenômenos ocorridos no espaço é um grande desafio para a elucidação de questões centrais em diversas áreas do conhecimento, seja em agronomia, em saúde, geologia, ecologia, entre tantas outras. Tais estudos tornam-se cada vez mais comuns, muito em

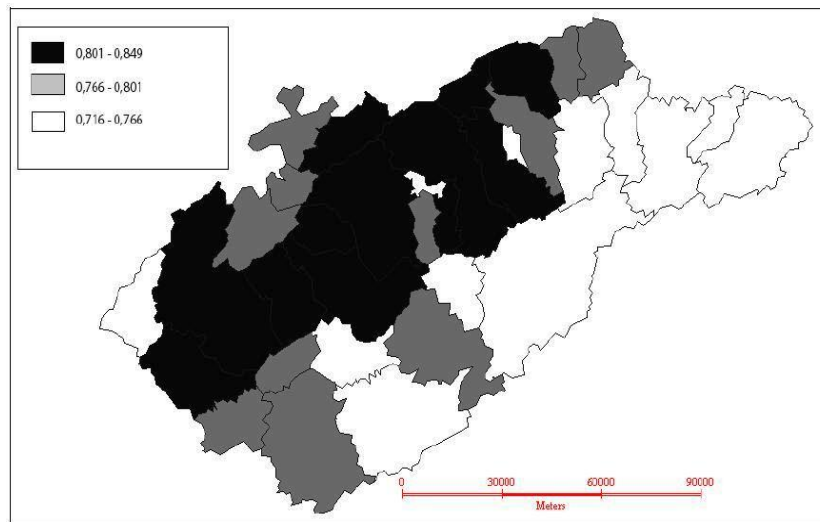


FIGURA 4 Mapa da distribuição espacial do IDH dos municípios do Vale do Paraíba paulista, 2007.
 Fonte: Oliveira & Nascimento (2007).

função das facilidades decorrentes da disponibilidade de sistemas de informações geográficas (SIGs).

Um SIG é um conjunto de equipamentos e programas de computação que integram mapas e gráficos com uma base de dados sobre um espaço geográfico definido, ou seja, um conjunto de ferramentas capazes de coletar, armazenar, manejar, analisar e visualizar informações georreferenciadas (Assunção, 2001).

Para Santos & Barcellos (2006), um SIG possui quatro funções básicas:

- (i) Aquisição de dados: envolve as etapas necessárias à alimentação do sistema, tais como: a captura, importação, validação e edição dos dados. Os dados podem ser confeccionados para introdução no sistema, por meio de técnicas como sensoriamento remoto, restituição aerofotogramétrica, digitalização de levantamentos topográficos, digitação de dados em tabela, etc. Entretanto, existe uma série de condições às quais esses dados devem obedecer no que diz respeito à sua estrutura, para que possam ser utilizados. Por isso, eles precisam ser analisados, e eventuais incoerências e imperfeições devem ser corrigidas;

- (ii) Gerenciamento de banco de dados: envolve o armazenamento dos dados espaciais de forma estruturada, de modo a possibilitar e facilitar a realização de análises. A forma como os dados são estruturados é crucial para o sistema, pois dela dependem os tipos de análises que poderão ser realizados. A um mapa armazenado no SIG, sempre podem ser associadas novas informações, provenientes de diversas fontes, permitindo que se incorpore o trabalho de diferentes órgãos e instituições. Ele permite combinar vários tipos de informações, como, por exemplo, dados obtidos em campo por GPS (Global Positioning System) ou por topografia convencional, tabelas, mapas, imagens, entre outros;
- (iii) Visualização e apresentação cartográfica: um SIG necessita ter agilidade para utilizar as diversas camadas de dados e exibir esse resultado por meio de mapas de síntese com boa qualidade gráfica. Os mapas, anteriormente feitos à mão, transformam-se agora em um produto de todas as operações desenvolvidas dentro do SIG, com inúmeras possibilidades de atualização mais constante;
- (iv) Consulta e análise: uma função que pode ser considerada como a principal de um SIG é a de análise, pois possibilita operações de extração e geração de novas informações sobre o espaço geográfico, a partir de critérios especificados pelo próprio usuário. As operações mais comuns são as pesquisas de dados, a busca de informações de acordo com algum critério de seleção (por exemplo, pela localização, proximidade, tamanho, valor) e a análise espacial que envolve modelagem e análise de padrões espaciais.

Na Figura 5 está representado a estrutura básica de um SIG.

Basicamente, conforme a Figura 5 mostra, um SIG é composto de uma gerência de base de dados espaciais, ferramental de entrada e manipulação de dados, consulta e análises destes, bem como permite geração de resultados em meio digital ou impresso. O acesso a um SIG ocorre via interface tipo console (janela de comandos) ou de forma gráfica.

Os SIGs podem ser comerciais ou livres. Um bom exemplo comercial seria o ArcGIS da ESRI (Economic and Social Research Institute - ESRI, 2009) que é

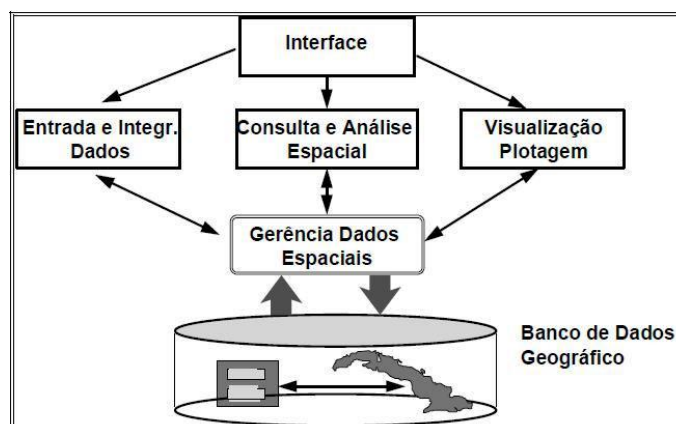


FIGURA 5 Estrutura básica de sistema de informação geográfica.
 Fonte: Câmara et al. (2001).

composto de vários módulos. Esses vão desde um programa SIG básico, como o ArcView, até módulos mais específicos de análise de dados espaciais, tais como: Spatial Analyst, Spatial Statistics, 3D Analyst, Geostatistical Analyst, etc. Entre os livres, destacam-se o TerraView (Instituto Nacional de Pesquisas Espaciais - INPE, 2007) e o GeoDa (Anselin, 2004), utilizados neste trabalho.

2.2 Análise de dados de área

Esta seção aprofundará a análise de dados de área, já apresentada inicialmente na subseção 2.1.4. Como já visto, de acordo com Assunção (2001), dados de área envolvem elementos que associam o mapa geográfico de uma região R a uma base de dados. Este mapa geográfico é dividido por áreas $A_i, i = 1, 2, \dots, n$, de forma que $\bigcup_{i=1}^n A_i = R$ e $A_i \cap A_j = \emptyset$ se $i \neq j$, qualquer que seja a forma de A_i e de R . Tem-se, também, que cada uma dessas áreas possui uma ou mais variáveis aleatórias (Y_1, Y_2, \dots, Y_n) que representam um valor para toda área. Portanto, não se conhece a localização exata (pontual) do evento, mas sim um valor agregado válido para a área toda.

2.2.1 Problema típico que demanda análise de dados de área

Para Câmara et al. (2001), os indicadores mais utilizados em dados de área são contagens, como é o caso da maior parte das variáveis coletadas no censo, por meio do IBGE. Por exemplo, para cada setor censitário, tem-se o número de chefes de família em cada uma das faixas de renda consideradas. Na saúde, pelo Ministério e Secretarias de Saúde, tem-se, por exemplo, para cada município dados de óbitos, partos e doenças transmissíveis.

Além desses, outros indicadores são bastante úteis, tais como: (a) proporções, como proporção de chefes de domicílio com renda superior a 3 salários mínimos, ou seja, números de chefes de domicílios com renda superior a 3 salários mínimos em relação ao total de chefes de domicílios; (b) médias, como renda média do chefe da família por setor censitário; e (c) taxas, como taxa de mortalidade, ou seja, o número de óbitos que ocorre em um determinado período de tempo, dividido pela população no mesmo período.

A forma usual de apresentação de dados agregados por áreas é o uso de mapas coloridos com o padrão espacial do fenômeno. Na Figura 6, mostra-se a distribuição espacial do percentual de idosos dos 96 bairros da cidade de São Paulo, no ano de 1991.

Na Figura 6, nota-se que os dados são indicadores da variabilidade social da cidade, com uma variação entre o centro com a periferia.

Para Câmara et al. (2001), grande parte dos usuários limita seu conhecimento no uso de SIG a essas operações de visualização, tirando conclusões intuitivas. Mas é possível ir muito além. Quando visualiza-se um padrão espacial, é muito útil traduzi-lo em questões objetivas:

- (i) O padrão que se observa, é aleatório ou apresenta uma agregação definida?
- (ii) Esta distribuição pode ser associada a causas mensuráveis?
- (iii) Os valores observados são suficientes para analisar o fenômeno espacial a ser estudado?
- (iv) Existem agrupamentos de áreas com padrões diferenciados dentro da região de estudo?

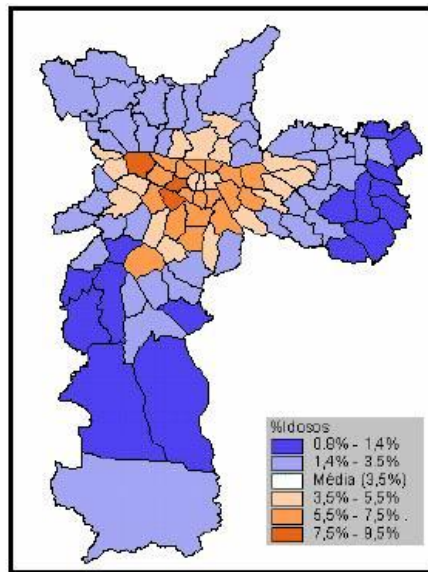


FIGURA 6 Mapa do percentual de idosos (com mais de 70 anos) dos 96 bairros da cidade de São Paulo (censo 1991).
Fonte: Câmara et al. (2001).

As questões acima mostram problemas típicos demandando análise de dados de área (ADA). Para responder a essas perguntas, um conjunto de técnicas de análise espacial de dados agregados por área será apresentado nas subseções de 2.2.2 a 2.2.11.

2.2.2 Visualização de dados

Conforme Câmara et al. (2001), os SIGs dispõem usualmente de três métodos descritivos de visualização de variável: (i) intervalos iguais, (ii) quantis e (iii) número de desvios-padrão.

(i) *Intervalos iguais*: a diferença entre o valor máximo e o valor mínimo da variável é dividida pelo número de classes, obtendo-se o valor do intervalo de cada classe. Se a variável tem uma distribuição muito concentrada de um lado, esta visualização deixa apenas um número pequeno de áreas nas classes da “perna” mais longa da distribuição. Como resultado, a maior parte

das áreas será alocada a uma ou duas cores;

- (ii) *Quantis*: o intervalo de valores de cada classe é calculado por meio do ordenamento dos valores da variável (do menor para o maior valor da variável). Após o ordenamento, cada classe vai conter o mesmo número de valores da variável. É uma forma de separar as classes em percentual de valores da variável;
- (iii) *Número de desvios-padrão*: são geradas classes separadas pelo valor do desvio-padrão σ . Nessa visualização o principal divisor é a média μ . A partir dela, são definidos os valores da variável acima e abaixo da média, em intervalos de valores que podem ser múltiplos de um desvio-padrão (σ). O número de classes resultantes acima e abaixo dependerá do valor da quantidade de desvios-padrão escolhidos, e também de valores da variável.

Um exemplo para o caso (iii) pode ser verificado pela Figura 6 citada na subseção 2.2.1.

2.2.3 Modelos de distribuição de dados de área

Segundo Câmara et al. (2001), o modelo de distribuição mais utilizado para dados de área é o modelo de variação espacial discreta. Considere a existência de um processo estocástico $\{Y_i, i = 1, 2, \dots, n\}$, em que y_i é a realização do processo espacial na área A_i , e n é o número total de áreas. O objetivo último da análise é construir uma aproximação para a distribuição conjunta de variáveis aleatórias (Y_1, Y_2, \dots, Y_n) , estimando sua distribuição, porém, nessa dissertação os métodos contemplarão apenas estimação de momentos da distribuição (média, covariáveis, etc.).

Tem-se que a variável aleatória Y_i poderá ser de qualquer tipo quantitativo (discreta ou contínua), tais como contagens, índices, proporções, medições ou taxas associadas à área A_i . Apesar da possibilidade de Y_i também ser contínua, uma situação comum é tratar com a variável aleatória Y_i , sendo discreta, que descreve o número de ocorrências de um dado evento de interesse em cada área, a qual pode ser associada uma distribuição de probabilidade de Poisson, entre outras. Tal hipótese, justifica-se por ser esta uma distribuição estatística adequada a fenômenos

que envolvem contagens de eventos, como é o caso em grande parte dos dados agregados por áreas. Evidentemente, outras distribuições podem ser adequadas, dependendo da variável a ser analisada.

Quanto às variáveis contínuas, diferentes distribuições podem também ser adequadas para modelagem. Por exemplo, taxas podem ser modeladas utilizando a distribuição normal, entre outras.

2.2.4 Análise de autocorrelação espacial

Para Cliff & Ord (1981), a autocorrelação espacial pode ser entendida como a tendência a que o valor de uma variável associada a uma determinada localização assemelha-se mais aos valores de suas observações vizinhas do que ao restante das localizações do conjunto amostral.

Dessa forma, autocorrelação espacial é a associação dos valores de uma variável em um mapa e visa a identificar a estrutura de correlação espacial que melhor descreva os dados. A preposição “auto” indica que a medida de correlação é realizada com a mesma variável aleatória, porém, medida em locais distintos do espaço.

De acordo com Santos & Souza (2007), a estrutura de dependência espacial entre os valores observados nas várias áreas do fenômeno em estudo, a partir da comparação entre os valores de uma amostra e de seus vizinhos, é analisada pelos indicadores de autocorrelação espacial.

Os indicadores mais utilizados pela ADA são: o índice de Moran e o índice de Geary (também conhecidos como índices globais). Esses indicadores constituem uma aproximação mais tradicional do efeito da dependência espacial, em que o esquema geral de dependência pode ser resumido em um único valor. Tem-se também, os indicadores locais de autocorrelação espacial (Moran e Geary). Esses indicadores são usados quando se dispõe de grande número de áreas, resultantes, por exemplo, de escalas espaciais detalhadas, e é muito provável a existência de diferentes regimes de correlação espacial em diferentes sub-regiões. Desses indicadores, será descrito em subseções posteriores, apenas, os índices de Moran (global e local). Além desses indicadores, serão estudados nessa dissertação a média móvel local, o gráfico de espalhamento de Moran, o *Box Map*, o *LISA Map*

e o *Moran Map*. Todas essas ferramentas dependem da definição de matriz de proximidade espacial adotada, discutida a seguir.

2.2.5 Matriz de proximidade espacial

De acordo com Câmara et al. (2001), quando se trata do estudo da dependência espacial de dados de área, o elemento chave desta análise é o conceito da matriz de proximidade espacial. A matriz de proximidade espacial (W) modela a estrutura de variabilidade espacial de dados de área, sendo uma ferramenta muito útil para descrever o arranjo espacial dos objetos. É uma matriz quadrada, não estocástica, cujos elementos w_{ij} refletem a intensidade da interdependência existente entre as áreas A_i e A_j . Em outras palavras, na matriz $W_{n \times n}$, cada elemento w_{ij} representa uma medida de proximidade espacial entre as áreas A_i e A_j .

Para Bailey & Gatrell (1995), a matriz $W_{n \times n}$ pode ser construída pressupondo diferentes medidas de adjacências, tais como:

- (i) $w_{ij} = 1$, se o ponto de referência de A_j é um dos k pontos mais próximos ao ponto de referência de A_i , e $w_{ij} = 0$, caso contrário;
- (ii) $w_{ij} = 1$, se o ponto de referência de A_j está dentro de uma distância especificada ao ponto de referência de A_i , e $w_{ij} = 0$, caso contrário;
- (iii) $w_{ij} = 1$, se A_i tem fronteira comum com A_j , e $w_{ij} = 0$, caso contrário;
- (iv) $w_{ij} = c_{ij}/c_i$, onde c_{ij} é o comprimento da fronteira comum entre A_i e A_j e c_i é o perímetro de A_i ;
- (v) $w_{ij} = 1/(d_{ij})^p$, onde $p \in \mathbb{N}^*$ e d_{ij} é a distância entre os centróides A_i e A_j .

Para melhor compreensão de como calcular a matriz de proximidade $W_{n \times n}$, usa-se um exemplo ilustrativo utilizando o caso (iii), que será o pressuposto adotado nesta dissertação. O mapa da Figura 7 refere-se aos municípios: Alfredo Vasconcelos (A_1), Barbacena (A_2), Barroso (A_3), Carandaí (A_4), Dolores de Campos (A_5), Prados (A_6) e Ressaquinha (A_7), localizados na mesorregião Campo das Vertentes - MG.

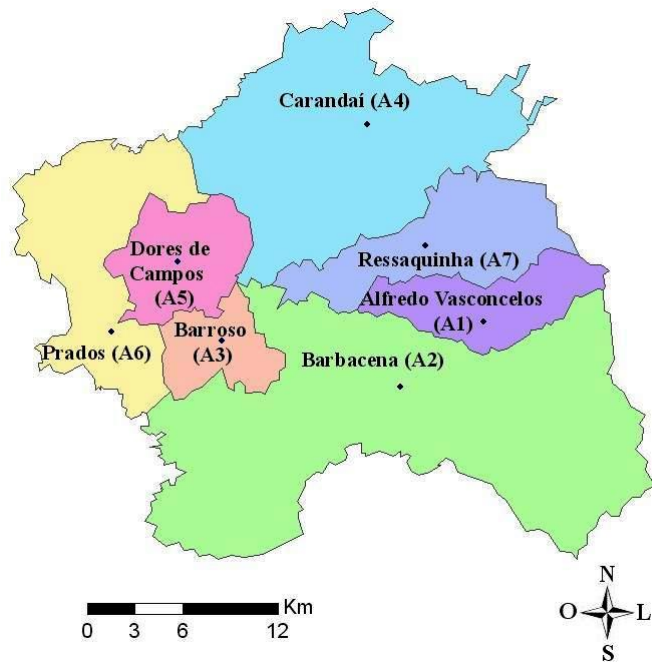


FIGURA 7 Mapa construído por elaboração própria de sete municípios localizados na mesorregião Campo das Vertentes - MG. Lavras, 2009.
 Fonte: Base Cartográfica do Instituto Brasileiro de Geografia e Estatística - IBGE (2002).

Observa-se, pela Figura 7 que: A_1 tem fronteira comum (vizinho de primeira ordem) com A_2 e A_7 ; A_2 tem fronteira comum com A_1 , A_3 , A_4 , A_5 , A_6 e A_7 ; A_3 tem fronteira comum com A_2 , A_4 , A_5 e A_6 ; A_4 tem fronteira comum com A_2 , A_3 , A_5 , A_6 e A_7 ; A_5 tem fronteira comum com A_2 , A_3 , A_4 e A_6 ; A_6 tem fronteira comum com A_2 , A_3 , A_4 e A_5 ; e A_7 tem fronteira comum com A_1 , A_2 e A_4 . Dessa forma, pode-se construir a matriz de proximidade de primeira ordem $W_{7 \times 7}^{(1)}$, dada na Tabela 1.

A idéia da matriz de proximidade espacial pode ser generalizada para vizinhos de maior ordem (vizinhos dos vizinhos). Com critério análogo ao adotado para a matriz de proximidade espacial de primeira ordem $W_{7 \times 7}^{(1)}$ que é obtida pela conexão de duas áreas, conforme visto na Tabela 2.1, pode-se construir a matriz de proximidade espacial de segunda ordem $W_{7 \times 7}^{(2)}$. Por exemplo, na Figura 1, os mu-

TABELA 1 Matriz de proximidade espacial por vizinhança de primeira ordem.

Município	A_1	A_2	A_3	A_4	A_5	A_6	A_7
A_1	0	1	0	0	0	0	1
A_2	1	0	1	1	1	1	1
A_3	0	1	0	1	1	1	0
A_4	0	1	1	0	1	1	1
A_5	0	1	1	1	0	1	0
A_6	0	1	1	1	1	0	0
A_7	1	1	0	1	0	0	0

nicípios Barroso (A_3) e Ressaquinha (A_7) são vizinhos de segunda ordem, já que, Barroso (A_3) é vizinho de primeira ordem de Barbacena (A_2), e Barbacena (A_2) é vizinho de primeira ordem de Ressaquinha (A_7). Neste mesmo raciocínio, tem-se que: A_1 é vizinho de segunda ordem com A_3 , A_4 , A_5 e A_6 ; A_2 não tem vizinho de segunda ordem; A_3 é vizinho de segunda ordem com A_1 e A_7 ; A_4 é vizinho de segunda ordem com A_1 ; A_5 é vizinho de segunda ordem com A_1 e A_7 ; A_6 é vizinho de segunda ordem com A_1 e A_7 ; e A_7 é vizinho de segunda ordem com A_3 , A_5 e A_6 . Assim, tem-se a matriz de proximidade de segunda ordem $W_{7 \times 7}^{(2)}$, mostrada na Tabela 2.

TABELA 2 Matriz de proximidade espacial por vizinhança de segunda ordem.

Município	A_1	A_2	A_3	A_4	A_5	A_6	A_7
A_1	0	0	1	1	1	1	0
A_2	0	0	0	0	0	0	0
A_3	1	0	0	0	0	0	1
A_4	1	0	0	0	0	0	0
A_5	1	0	0	0	0	0	1
A_6	1	0	0	0	0	0	1
A_7	0	0	1	0	1	1	0

De acordo com a Figura 7, para o exemplo dado, pode-se determinar no máximo a matriz de proximidade espacial de segunda ordem, já que não se tem vizinhos de ordem maior. Entretanto, se o número de áreas em uma determinada região em estudo for grande, têm-se matrizes de proximidades de ordem superiores, tais

como: $W_{n \times n}^{(3)}, W_{n \times n}^{(4)}, \dots, W_{n \times n}^{(k)}$.

Conforme Câmara et al. (2001), como a matriz de proximidade espacial $W_{n \times n}$ é utilizada em cálculos de indicadores na fase de análise exploratória, é muito útil padronizar suas linhas, dividindo cada elemento w_{ij} pela soma total dos elementos de cada linha a que pertence, gerando uma nova matriz $W_{n \times n}^*$, que é a matriz de proximidade espacial padronizada de $W_{n \times n}$, de tal modo que a soma de cada linha dessa nova matriz seja igual a um.

Esta dissertação está baseada nos cálculos de indicadores em que utiliza-se a matriz de proximidade espacial padronizada de primeira ordem, e, para simplificação de notação, a matriz $W_{n \times n}^*$, que é a matriz de proximidade espacial padronizada de primeira ordem de $W_{n \times n}$, será denotada pelo mesmo símbolo $W_{n \times n}$, e seus elementos por w_{ij} .

As Tabelas 3 e 4 são exemplos de matrizes de proximidade espacial por vizinhança de primeira e segunda ordem, respectivamente, padronizadas pelas linhas das Tabelas 1 e 2.

TABELA 3 Matriz de proximidade espacial por vizinhança de primeira ordem, padronizada pelas linhas da Tabela 1.

Município	A_1	A_2	A_3	A_4	A_5	A_6	A_7
A_1	0,000	0,500	0,000	0,000	0,000	0,000	0,500
A_2	0,167	0,000	0,167	0,167	0,167	0,167	0,167
A_3	0,000	0,250	0,000	0,250	0,250	0,250	0,000
A_4	0,000	0,200	0,200	0,000	0,200	0,200	0,200
A_5	0,000	0,250	0,250	0,250	0,000	0,250	0,000
A_6	0,000	0,250	0,250	0,250	0,250	0,000	0,000
A_7	0,333	0,333	0,000	0,333	0,000	0,000	0,000

2.2.6 Média móvel local

Conforme Bailey & Gatrell (1995), uma maneira simples de avaliar as variações das tendências espaciais da variável, em termos globais, é estimar a média μ_i associada a variável Y_i relativo à i -ésima área A_i , a partir dos elementos w_{ij} da matriz de proximidade espacial $W_{n \times n}$ padronizada pelas linhas, tomando-se a média dos vizinhos. Assim, a variabilidade espacial é reduzida, pois a operação

TABELA 4 Matriz de proximidade espacial por vizinhança de segunda ordem, padronizada pelas linhas da Tabela 2.

Município	A_1	A_2	A_3	A_4	A_5	A_6	A_7
A_1	0,000	0,000	0,250	0,250	0,250	0,250	0,000
A_2	0,000	0,000	0,000	0,000	0,000	0,000	0,000
A_3	0,500	0,000	0,000	0,000	0,000	0,000	0,500
A_4	1,000	0,000	0,000	0,000	0,000	0,000	0,000
A_5	0,500	0,000	0,000	0,000	0,000	0,000	0,500
A_6	0,500	0,000	0,000	0,000	0,000	0,000	0,500
A_7	0,000	0,000	0,333	0,000	0,333	0,333	0,000

tende a produzir uma superfície com menor flutuação que os dados originais. Tal média denomina-se de média móvel local. Sua fórmula é dada por:

$$\hat{\mu}_i = \sum_{j=1}^n w_{ij} Y_j, \quad (2.1)$$

e o seu modelo estatístico pode ser expresso por:

$$Y_i = \mu_i + \varepsilon_i + \varepsilon, \quad (2.2)$$

em que:

Y_i é o valor da variável na área A_i ;

μ_i é o valor médio de Y inerente a área A_i (ou padrão espacial de Y em A_i);

ε_i é a componente espacial estocástica de Y_i , com $E(\varepsilon_i) = 0$;

ε é o ruído (variável aleatória independente e identicamente distribuída com $E(\varepsilon) = 0$

e $Var(\varepsilon) = \tau^2$).

Demonstração: Aplicando a esperança em ambos os membros do modelo

(2.2), tem-se:

$$\begin{aligned} E [Y_i] &= E [\mu_i + \varepsilon_i' + \varepsilon] \\ &= E [\mu_i] + E [\varepsilon_i'] + E [\varepsilon] \\ &= E [\mu_i] + 0 + 0 \\ &= \mu_i. \end{aligned}$$

Portanto,

$$E [Y_i] = \mu_i.$$

Pelo método dos momentos, conclui-se que:

$$\hat{E} [Y_i] = \hat{\mu}_i = \sum_{j=1}^n w_{ij} Y_j.$$

Uma outra forma utilizada para calcular a média móvel local, é estimar a média α_i associada ao desvio $(Y_i - \mu)$, em que μ é a média geral, sendo sua fórmula dada por:

$$\hat{\alpha}_i = \sum_{j=1}^n w_{ij} (Y_j - \bar{Y}), \quad (2.3)$$

e o seu modelo estatístico pode ser expresso por:

$$Y_i = \mu + \alpha_i + \varepsilon_i' + \varepsilon, \quad (2.4)$$

em que:

Y_i é o valor da variável na área A_i ;

μ é o efeito médio global (efeito que atua igualmente em todas as áreas A_i);

α_i é o efeito espacial próprio ou local da área A_i ;

ε_i' é a componente espacial estocástica de Y_i , com $E (\varepsilon_i') = 0$;

ε é o ruído (variável aleatória independente e identicamente distribuída com $E (\varepsilon) = 0$

e $Var(\varepsilon) = \tau^2$).

Demonstração: Aplicando a esperança em ambos os membros do modelo (2.4), então:

$$\begin{aligned} E[Y_i] &= E[\mu + \alpha_i + \varepsilon_i + \varepsilon] \\ &= E[\mu] + E[\alpha_i] + E[\varepsilon_i] + E[\varepsilon] \\ &= \mu + \alpha_i. \end{aligned}$$

Portanto,

$$E[Y_i] = \mu + \alpha_i \Rightarrow \alpha_i = E[Y_i] - \mu.$$

Pelo método dos momentos, tem-se que:

$$\begin{aligned} \hat{\alpha}_i &= \hat{E}[Y_i] - \hat{\mu} \\ &= \sum_{j=1}^n w_{ij} Y_j - \bar{Y} \cdot 1 \\ &= \sum_{j=1}^n w_{ij} Y_j - \bar{Y} \sum_{j=1}^n w_{ij} \\ &= \sum_{j=1}^n w_{ij} Y_j - \sum_{j=1}^n w_{ij} \bar{Y} \\ &= \sum_{j=1}^n w_{ij} (Y_j - \bar{Y}). \end{aligned}$$

Logo,

$$\hat{\alpha}_i = \sum_{j=1}^n w_{ij} (Y_j - \bar{Y}).$$

Na Figura 8, ilustra-se o uso da média móvel local, utilizando a expressão (2.1), para o percentual de idosos (mais de 70 anos) para os 96 distritos da cidade de São Paulo, comparado com o mapa da Figura 6 visto na subseção 2.1.4.

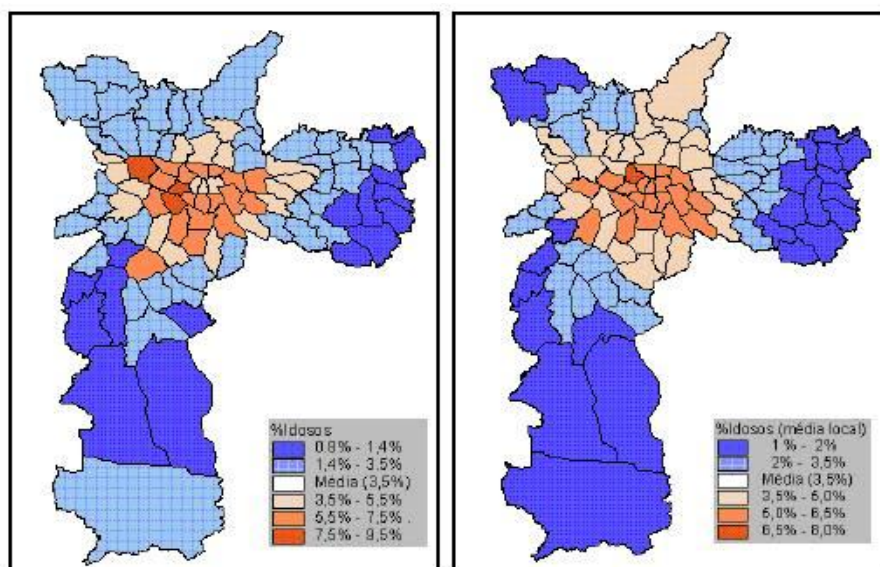


FIGURA 8 Distribuição dos idosos na cidade de São Paulo (censo de 1991). À esquerda, mapa de desvios-padrão. À direita, mapa de desvios-padrão com a média móvel local.
Fonte: Câmara et al. (2001).

Na Figura 8, com o cálculo da média móvel local, há um alisamento: o valor mínimo que era de 0,8% passa a ser de 1% e o máximo que era de 8,5% passa a ser de 8%. Pode-se notar, ao comparar os dois mapas da Figura 8, que a média móvel local fornece uma visão das grandes tendências do fenômeno em estudo e no caso do percentual de idosos, mostra um forte gradiente centro-periferia. Essa forte “tendência” é o que chama-se de “padrão espacial”.

2.2.7 Índice de Moran

De acordo com Magalhães (2006), se X e Y são variáveis aleatórias no mesmo espaço de probabilidade, o coeficiente de correlação entre X e Y é definido por:

$$\rho_{X,Y} = E \left[\left(\frac{X - \mu_X}{\sigma_X} \right) \left(\frac{Y - \mu_Y}{\sigma_Y} \right) \right], \quad (2.5)$$

desde que as esperanças envolvidas na expressão (2.5) existam. μ_X e μ_Y são as médias populacionais e σ_X e σ_Y são os desvios-padrão populacionais de X e Y . Portanto, o coeficiente de correlação entre as variáveis X e Y em n objetos distintos pode ser estimado (pelo método dos momentos) como:

$$r = \hat{\rho}_{X,Y} = \frac{1}{n} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{S_X} \right) \left(\frac{Y_i - \bar{Y}}{S_Y} \right), \quad (2.6)$$

em que \bar{X} e \bar{Y} são as médias aritméticas amostrais e $S_X = \sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 / n}$ e $S_Y = \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2 / n}$ são os desvios-padrão amostrais de X e Y (algumas definições dividem por $n - 1$ ao invés de n). Portanto, o r é a média aritmética dos n pares de produtos dos desvios padronizados, definidos por:

$$\left(\frac{X_i - \bar{X}}{S_X} \right) \quad e \quad \left(\frac{Y_i - \bar{Y}}{S_Y} \right).$$

O índice I de Moran, descrito em Moran (1950), mede a correlação espacial de uma variável Y e é similar em vários aspectos ao coeficiente de correlação r . Considere uma matriz de proximidade espacial binária e simétrica $W_{n \times n}$ com $w_{ij} = 1$ ou $w_{ij} = 0$ e $w_{ij} = w_{ji}$. Como em r , o índice I considera o desvio padronizado $(Y_i - \bar{Y})/S_Y$, mas verifica qual é a sua relação com o desvio padronizado da mesma variável Y numa área A_j vizinha de A_i . Assim, tem-se o seguinte produto cruzado:

$$\left(\frac{Y_i - \bar{Y}}{S_Y} \right) \left(\frac{Y_j - \bar{Y}}{S_Y} \right). \quad (2.7)$$

O interesse é o comportamento típico ou médio desses produtos cruzados. Então, toma-se a média de (2.7) sobre todos os pares de áreas vizinhas de uma determinada região em estudo. Para tirar a média, observa-se antes que a soma sobre todos os pares (A_i, A_j) de vizinhos é equivalente a somar sobre todos os pares

possíveis, vizinhos ou não, mas multiplicando-os por w_{ij} . Então:

$$\sum_{\text{paresdevizinhos}} \left(\frac{Y_i - \bar{Y}}{S_Y} \right) \left(\frac{Y_j - \bar{Y}}{S_Y} \right) \iff \sum_{i \neq j} w_{ij} \left(\frac{Y_i - \bar{Y}}{S_Y} \right) \left(\frac{Y_j - \bar{Y}}{S_Y} \right). \quad (2.8)$$

Tem-se que, na soma da expressão (2.8), aparece tanto o par (A_i, A_j) , quanto o par (A_j, A_i) . O número de pares ordenados (A_i, A_j) de áreas vizinhas é dado por $\sum_{i \neq j} w_{ij}$. O índice procurado é a média, ou a soma dividida pelo número de pares:

$$I = \frac{1}{\sum_{i \neq j} w_{ij}} \sum_{i \neq j} w_{ij} \left(\frac{Y_i - \bar{Y}}{S_Y} \right) \left(\frac{Y_j - \bar{Y}}{S_Y} \right). \quad (2.9)$$

Tem-se que $\sum_{i \neq j} w_{ij}$ é soma todos os elementos fora da diagonal principal da matriz $W_{n \times n}$. Para incluir a diagonal principal em que os valores de $w_{ii} = 0$, basta escrever o índice da seguinte forma:

$$I = \frac{1}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (Y_i - \bar{Y}) (Y_j - \bar{Y})}{S_Y^2}. \quad (2.10)$$

Observa-se que $\sum_{i=1}^n \sum_{j=1}^n w_{ij} = n$, se $W_{n \times n}$ for uma matriz de proximidade espacial padronizada. Então:

$$I = \frac{1}{n} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (Y_i - \bar{Y}) (Y_j - \bar{Y})}{\sum_{i=1}^n (Y_i - \bar{Y})^2 / n}$$

$$I = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (Y_i - \bar{Y}) (Y_j - \bar{Y})}{\sum_{i=1}^n (Y_i - \bar{Y})^2}. \quad (2.11)$$

Quando as variáveis são espacialmente independentes, o valor esperado de I é próximo de zero ($I \approx 0$). Quando há similaridade entre áreas próximas, I tende a ser positivo ($I > 0$). Quando há dissimilaridade entre áreas próximas (fenômeno raro de ser observado), I tende a ser negativo ($I < 0$). Este índice é também chamado de índice de Moran global ou índice de autocorrelação de Moran.

Para compreender o cálculo do índice de Moran, utilizando a expressão (2.11), serão mostrados a seguir, nas Tabelas 5 e 6, os resultados dos cálculos que são necessários para a determinação do I . Esses cálculos, foram realizados para uma massa de dados do Zoneamento Ecológico Econômico, pertencentes a sete municípios da mesoregião Campo das Vertentes - MG, no ano de 2004. Em que $A_1, A_2, A_3, A_4, A_5, A_6$ e A_7 representam, respectivamente, os municípios de Alfredo Vasconcelos, Barbacena, Barroso, Carandaí, Dolores de Campos, Prados e Ressaquinha.

TABELA 5 Índice do produto interno bruto dos municípios localizados na mesoregião Campo das Vertentes - MG, no ano de 2004.

Município	$IPIB_{2004}(y_i)$
A_1	0,1222
A_2	0,0794
A_3	0,0484
A_4	0,0658
A_5	0,0642
A_6	0,0611
A_7	0,1252
Média dos Índices \bar{Y}	0,0809
$\sum_{i=1}^n (y_i - \bar{y})^2$	0,0056

TABELA 6 Cálculos Auxiliares para a determinação do índice global de Moran.

MUNICÍPIOS	w_{ij}	$(y_i - \bar{y})$	$(y_j - \bar{y})$	$w_{ij}(y_i - \bar{y})(y_j - \bar{y})$
$A_1 \times A_2$	0,50000000	0,04130000	-0,0015000	- 0,00003098
$A_1 \times A_3$	0,00000000	0,04130000	-0,0325000	0,00000000
$A_1 \times A_4$	0,00000000	0,04130000	-0,0151000	0,00000000
$A_1 \times A_5$	0,00000000	0,04130000	-0,0167000	0,00000000
$A_1 \times A_6$	0,00000000	0,04130000	-0,0198000	0,00000000
$A_1 \times A_7$	0,50000000	0,04130000	0,0443000	0,00091480
$A_2 \times A_1$	0,16700000	-0,00150000	0,0413000	- 0,00001034
$A_3 \times A_3$	0,16700000	-0,00150000	-0,0325000	0,00000814
$A_4 \times A_4$	0,16700000	-0,00150000	-0,0151000	0,00000378
$A_5 \times A_5$	0,16700000	-0,00150000	-0,0167000	0,00000418
$A_6 \times A_6$	0,16700000	-0,00150000	-0,0198000	0,00000496
$A_7 \times A_7$	0,16700000	-0,00150000	0,0443000	- 0,00001110
$A_3 \times A_1$	0,00000000	-0,03250000	0,0413000	0,00000000
$A_3 \times A_2$	0,25000000	-0,03250000	-0,0015000	0,00001219
$A_3 \times A_4$	0,25000000	-0,03250000	-0,0151000	0,00012269
$A_3 \times A_5$	0,25000000	-0,03250000	-0,0167000	0,00013569
$A_3 \times A_6$	0,25000000	-0,03250000	-0,0198000	0,00016088
$A_3 \times A_7$	0,00000000	-0,03250000	0,0443000	0,00000000
$A_4 \times A_1$	0,00000000	-0,01510000	0,0413000	0,00000000
$A_4 \times A_2$	0,20000000	-0,01510000	-0,0015000	0,00000453
$A_4 \times A_3$	0,20000000	-0,01510000	-0,0325000	0,00009815
$A_4 \times A_5$	0,20000000	-0,01510000	-0,0167000	0,00005043
$A_4 \times A_6$	0,20000000	-0,01510000	-0,0198000	0,00005980
$A_4 \times A_7$	0,20000000	-0,01510000	0,0443000	- 0,00013379
$A_5 \times A_1$	0,00000000	-0,01670000	0,0413000	0,00000000
$A_5 \times A_2$	0,25000000	-0,01670000	-0,0015000	0,00000626
$A_5 \times A_3$	0,25000000	-0,01670000	-0,0325000	0,00013569
$A_5 \times A_4$	0,25000000	-0,01670000	-0,0151000	0,00006304
$A_5 \times A_6$	0,25000000	-0,01670000	-0,0198000	0,00008266
$A_5 \times A_7$	0,00000000	-0,01670000	0,0443000	0,00000000
$A_6 \times A_1$	0,00000000	-0,01980000	0,0413000	0,00000000
$A_6 \times A_2$	0,25000000	-0,01980000	-0,0015000	0,00000742
$A_6 \times A_3$	0,25000000	-0,01980000	-0,0325000	0,00016088
$A_6 \times A_4$	0,25000000	-0,01980000	-0,0151000	0,00007474
$A_6 \times A_5$	0,25000000	-0,01980000	-0,0167000	0,00008266
$A_6 \times A_7$	0,00000000	-0,01980000	0,0443000	0,00000000
$A_7 \times A_1$	0,33300000	0,04430000	0,0413000	0,00060925
$A_7 \times A_2$	0,33300000	0,04430000	-0,0015000	-0,00002213
$A_7 \times A_3$	0,00000000	0,04430000	-0,0325000	0,00000000
$A_7 \times A_4$	0,33300000	0,04430000	-0,0151000	-0,00022275
$A_7 \times A_5$	0,00000000	0,04430000	-0,0167000	0,00000000
$A_7 \times A_6$	0,00000000	0,04430000	-0,0198000	0,00000000
Soma	7	-	-	0,00237173

Dessa forma, o valor de I é dado por:

$$I = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{0,00237173}{0,00560000} = 0,4235.$$

Observa-se que $I > 0$, então, existe um grau de similaridade entre áreas próximas, ou seja, há uma autocorrelação positiva.

2.2.8 Teste de autocorrelação espacial

De acordo com Bailey & Gatrell (1995), há duas principais abordagens para testar valores observados de I , as quais vêm da hipótese de que não há autocorrelação espacial.

2.2.8.1 Distribuição aproximada de I

Se há um suficiente número de áreas observadas (isto é, n suficientemente grande), então um resultado aproximado para a distribuição do índice I sob certas condições pode ser usado para produzir um teste. Um primeiro resultado é que, se Y_1, Y_2, \dots, Y_n são variáveis aleatórias independentes e identicamente distribuídas com distribuição normal, então o índice I de Moran possui uma distribuição aproximadamente normal com média e variância, citadas em Cliff & Ord (1981), dadas por:

$$E[I] = -\frac{1}{(n-1)} \quad (2.12)$$

$$Var[I] = \frac{n^2 S_1 - n S_2 + 3 S_0^2}{(n-1)(n+1) S_0^2} - \left(-\frac{1}{n-1}\right)^2, \quad (2.13)$$

em que S_0, S_1, S_2 são funções de w_{ij} definidas por $S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij}$, $S_1 = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (w_{ij} + w_{ji})^2$ e $S_2 = \sum_{i=1}^n (w_{i.} + w_{.i})^2$. Tem-se que $w_{i.} = \sum_{j=1}^n w_{ij}$ é a

soma da linha i da matriz $W_{n \times n}$ e $w_{.i} = \sum_{j=1}^n w_{ji}$ é a soma da coluna i da matriz $W_{n \times n}$.

Assim, um teste de significância de nível α para a hipótese nula de independência é construído, com *valor - p* dado pela probabilidade:

$$\text{valor} - p = P \left(Z > \frac{I_{obs} - E(I)}{\sqrt{Var(I)}} \right), \quad (2.14)$$

em que:

I_{obs} é o valor realizado (observado) da variável aleatória I ;

Z é a variável aleatória normal padrão.

O fato de supor que I tem uma distribuição aproximadamente normal, depende da existência de muitas áreas no mapa, o que nem sempre ocorre. Porém, tem-se uma outra alternativa, que é o teste de permutação aleatória.

2.2.8.2 Teste de permutação aleatória

Suponha que se tem n valores $y_i, i = 1, 2, \dots, n$, relacionados com as áreas A_i . Então, tem-se $n!$ possíveis permutações desse mapa, cada um correspondendo a um rearranjo diferente dos n valores y_i sobre as áreas. Uma dessas permutações corresponde aos dados que foram observados. O valor de I pode ser calculado para qualquer uma destas $n!$ permutações e se obter uma distribuição empírica para os possíveis valores de I sobre as permutações aleatórias dos n valores. A idéia fundamental por trás deste teste é considerar as variáveis aleatórias Y_1, Y_2, \dots, Y_n como independentes e identicamente distribuídas, estabelecendo esta hipótese como a hipótese nula (H_0) do teste. As $n!$ permutações possíveis são, então, um modo de se realizar na prática esta independência e distribuição idêntica. Se o valor de I correspondente à configuração observada for um valor extremo na distribuição encontrada, então pode-se interpretar isso como evidência de autocorrelação espacial. Tal teste pode ser descrito como um algoritmo, como a seguir:

- Calcula-se o índice I com os dados observados no mapa obtendo $I_{(0)}$.
- Para k variando de 1 a N (N é um número menor ou igual à $n!$), repete-se independentemente: Calcula-se o índice $I_{(k)}$ permutando aleatoriamente

os valores de y_i entre as áreas. A limitação em N deve-se ao aspecto de viabilidade computacional, já que, se n é grande, $n! \gg n$.

- Sob H_0 , as variáveis aleatórias Y_i são independentes e identicamente distribuídas. Então todas as permutações dos valores y_i entre as áreas são igualmente prováveis. Assim, o *valor - p* do teste é:

$$\text{valor} - p = \frac{\text{número de } I_{(j)} > I_{(0)}, j = 1, \dots, N.}{N + 1} \quad (2.15)$$

Rejeita-se a hipótese nula em um teste de nível α se $\text{valor} - p < \alpha$. Este teste está considerando apenas autocorrelação positiva. Para o caso em que a autocorrelação é negativa, o *valor - p* do teste é dado por:

$$\text{valor} - p = \frac{\text{número de } I_{(j)} < I_{(0)}, j = 1, \dots, N.}{N + 1} \quad (2.16)$$

Uma maneira útil de visualizar o resultado é fazer um histograma com todos os valores $I_{(k)}$, $k = 0, 1, 2, \dots, N$, e apontar o local na escala horizontal em que está localizado o valor $I_{(0)}$ observado com os dados originais.

Cliff & Ord (1981) mostram que os momentos de I sob permutação dos índices podem ser obtidos teoricamente e que a distribuição assintótica de I ainda é normal sob determinadas hipóteses bem gerais. Eles mostraram que:

$$E[I] = -\frac{1}{(n-1)} \quad (2.17)$$

e

$$\begin{aligned} \text{Var}[I] = & \frac{n[(n^2 - 3n + 3)S_1 - nS_2 + 3S_0^2] - k[(n^2 - n)S_1 - 2nS_2 + 6S_0^2]}{(n-1)(n-2)(n-3)S_0^2} \\ & - \left(-\frac{1}{n-1}\right)^2, \end{aligned} \quad (2.18)$$

em que, $S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij}$, $S_1 = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (w_{ij} + w_{ji})^2$, $S_2 = \sum_{i=1}^n (w_{i.} + w_{.i})^2$ e $k = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^4 / \left(\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2\right)^2$. Tem-se que $w_{i.} = \sum_{j=1}^n w_{ij}$ é a soma

da linha i da matriz $W_{n \times n}$ e $w_{.i} = \sum_{j=1}^n w_{ji}$ é a soma da coluna j da matriz $W_{n \times n}$.

2.2.9 Índice de Moran local

Para Câmara et al. (2001), quando se tem um grande número de áreas em uma determinada região em estudo, é muito provável que ocorram diferentes regimes de associação espacial, e que apareçam vários locais de autocorrelação espacial, onde a dependência espacial é ainda mais pronunciada. Para tanto, é preciso utilizar indicadores locais de associação espacial que possam ser associados às diferentes localizações de uma variável distribuída espacialmente.

Ainda que o índice I de Moran global seja muito útil para fornecer uma indicação de agrupamento global dos dados, esta medida precisa ser complementada por estatísticas locais. De acordo com Anselin (1995), uma estatística que cumpre este objetivo é o índice de Moran local, que é uma ferramenta estatística que serve para testar a autocorrelação espacial local entre a área A_i e seus vizinhos A_j , permitindo, assim, a identificação de agrupamentos de áreas semelhantes (clusters), áreas fora do padrão da região (outliers) e de regimes espaciais que não são detectados pelo índice de Moran global. A fórmula para o cálculo do índice de Moran local para cada área A_i é dada por:

$$I_i = (Y_i - \bar{Y}) \frac{\sum_{j=1}^n w_{ij} (Y_j - \bar{Y})}{\sum_{i=1}^n (Y_i - \bar{Y})^2 / n}. \quad (2.19)$$

A estatística pode ser interpretada da seguinte forma: valores positivos de I_i significam que existem clusters espaciais com valores similares (alto ou baixo) da variável em estudo; valores negativos significam que existem clusters espaciais com valores dissimilares da variável entre as áreas e seus vizinhos.

Segundo Anselin (1995), a soma dos índices de Moran local, para todas as observações em uma determinada área em estudo, é proporcional ao índice de Moran global.

Demonstração: Aplicando o somatório de i em ambos os membros da equação

(2.19), segue que:

$$\sum_{i=1}^n I_i = \sum_{i=1}^n (Y_i - \bar{Y}) \frac{\sum_{j=1}^n w_{ij} (Y_j - \bar{Y})}{\sum_{i=1}^n (Y_i - \bar{Y})^2 / n},$$

sabe-se que $S_Y^2 = \sum_{i=1}^n (Y_i - \bar{Y})^2 / n$. Então,

$$\sum_{i=1}^n I_i = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (Y_i - \bar{Y}) (Y_j - \bar{Y})}{S_Y^2},$$

comparando com a fórmula do índice de Moran global, expressão (2.10), e chamando $\sum_{i=1}^n \sum_{j=1}^n w_{ij} = S_0$, tem-se que:

$$I = \frac{1}{S_0} \sum_{i=1}^n I_i.$$

Logo, a soma dos índices de Moran locais é proporcional ao índice de Moran global, sendo a constante de proporcionalidade igual a $\frac{1}{S_0}$.

De forma similar aos indicadores globais, a significância do índice de Moran local I_i deve ser avaliada utilizando-se hipótese de normalidade ou simulação de distribuição por permutação aleatória nos valores dos atributos (Anselin, 1995).

Na permutação aleatória do índice de Moran local, a simulação é realizada calculando-se o índice I_i para cada área A_i . Depois, permuta-se, aleatoriamente, o valor das demais áreas A_j até obter uma pseudo-distribuição onde se possam computar os parâmetros de significância. É útil gerar mapas (*LISA Map* e *Moran Map*) indicando as regiões que apresentam correlação local significativamente diferente das demais.

2.2.10 Gráfico de espalhamento de Moran

Conforme Anselin (1996), o gráfico de espalhamento de Moran tem o objetivo de comparar os desvios ($Z_i = Y_i - \bar{Y}$) do atributo numa área A_i , com a média dos desvios de seus vizinhos A_j ponderada pela matriz de proximidade espacial $W_{n \times n}$ padronizada pelas linhas $\left(WZ_j = \sum_{j=1}^n w_{ij} (Y_j - \bar{Y}) \right)$. Constrói-se um gráfico bidimensional de abscissa Z e de ordenada WZ . Esse gráfico é uma forma adicional de visualizar a dependência espacial e indicar os diferentes regimes espaciais presentes nos dados. Dessa forma, dividem-se o diagrama em quatro quadrantes (Q1, Q2, Q3 e Q4), conforme a Figura 9.

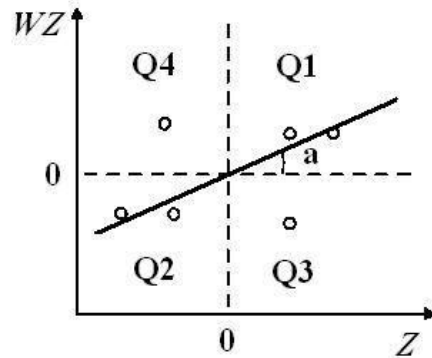


FIGURA 9 Gráfico de espalhamento de Moran, onde a é inclinação da reta.
Fonte: Santos & Souza (2007).

Os quadrantes da Figura 9 podem ser interpretados como:

- O primeiro quadrante (localizado na parte superior direita) mostra as áreas que apresentam altos valores para a variável em análise (valores acima da média) cercada por áreas vizinhas que também apresentam valores acima da média para a variável em análise. Esse quadrante é classificado como alto-alto (AA).
- O segundo quadrante (localizado no canto inferior esquerdo) é constituído pelas áreas com valores baixos para as variáveis em análise cercadas por vizinhos que também apresentam baixos valores. Esse quadrante é classifi-

cado como baixo-baixo (BB).

- O terceiro quadrante (localizado no canto inferior direito) é formado pelas áreas com altos valores para as variáveis em análise cercadas por áreas com baixos valores. Esse quadrante é classificado como alto-baixo (AB).
- O quarto quadrante (localizado na parte superior esquerda) mostra as áreas com valores baixos cercadas por vizinhos que apresentam valores altos. Esse quadrante é geralmente classificado como baixo-alto (BA).

As áreas que estão localizadas nos quadrantes AA e BB apresentam autocorrelação espacial positiva, ou seja, essas áreas formam *clusters* de valores similares. Por outro lado, os quadrantes BA e AB apresentam autocorrelação espacial negativa, ou seja, há dissimilaridade entre as áreas vizinhas com relação à variável em estudo.

Segundo Anselin (1996), o índice de Moran global é equivalente ao coeficiente de regressão linear, que indica a inclinação da reta de regressão de WZ versus Z .

Demonstração: Conforme Draper & Smith (1998), a forma matricial de um modelo de regressão linear WZ versus Z é dada por:

$$\mathbf{WZ} = \mathbf{Z}\beta + \varepsilon, \quad (2.20)$$

em que:

$$\mathbf{WZ} = \begin{bmatrix} \sum_{j=1}^n w_{1j} (y_j - \bar{y}) \\ \sum_{j=1}^n w_{2j} (y_j - \bar{y}) \\ \vdots \\ \sum_{j=1}^n w_{nj} (y_j - \bar{y}) \end{bmatrix},$$

$$\mathbf{Z} = \begin{bmatrix} 1 & y_1 - \bar{y} \\ 1 & y_2 - \bar{y} \\ \vdots & \vdots \\ 1 & y_n - \bar{y} \end{bmatrix},$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \text{ e } \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

O estimador não-tendencioso linear para $\boldsymbol{\beta}$ e de variância mínima é dado por:

$$\hat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{W}\mathbf{Z}, \quad (2.21)$$

em que:

$$\mathbf{Z}' = \begin{bmatrix} 1 & 1 & \dots & 1 \\ y_1 - \bar{y} & y_2 - \bar{y} & \dots & y_n - \bar{y} \end{bmatrix}$$

$$\text{e } \hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}.$$

Resolvendo a expressão (2.21), tem-se:

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \left(\begin{bmatrix} 1 & 1 & \dots & 1 \\ y_1 - \bar{y} & y_2 - \bar{y} & \dots & y_n - \bar{y} \end{bmatrix} \begin{bmatrix} 1 & y_1 - \bar{y} \\ 1 & y_2 - \bar{y} \\ \vdots & \vdots \\ 1 & y_n - \bar{y} \end{bmatrix} \right)^{-1} \begin{bmatrix} \sum_{j=1}^n w_{1j} (y_j - \bar{y}) \\ \sum_{j=1}^n w_{2j} (y_j - \bar{y}) \\ \vdots \\ \sum_{j=1}^n w_{nj} (y_j - \bar{y}) \end{bmatrix}$$

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^n (y_i - \bar{y}) \\ \sum_{i=1}^n (y_i - \bar{y}) & \sum_{i=1}^n (y_i - \bar{y})^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_j - \bar{y}) \\ \sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y}) (y_j - \bar{y}) \end{bmatrix}.$$

Sabendo-se que $\sum_{i=1}^n (y_i - \bar{y}) = 0$ e $nS_Y^2 = \sum_{i=1}^n (y_i - \bar{y})^2$, então:

$$\begin{aligned}
\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} &= \frac{1}{n^2 S_Y^2} \begin{bmatrix} n S_Y^2 & 0 \\ 0 & n \end{bmatrix} \begin{bmatrix} \sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_j - \bar{y}) \\ \sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y}) (y_j - \bar{y}) \end{bmatrix} \\
\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} &= \frac{1}{n^2 S_Y^2} \begin{bmatrix} n S_Y^2 \sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_j - \bar{y}) \\ n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y}) (y_j - \bar{y}) \end{bmatrix} \\
&= \begin{bmatrix} \sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_j - \bar{y}) / n \\ \sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y}) (y_j - \bar{y}) / n S_Y^2 \end{bmatrix}.
\end{aligned}$$

Portanto,

$$\hat{\beta}_1 = I = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y}) (y_j - \bar{y})}{n S_Y^2}.$$

Logo, I é equivalente ao coeficiente de regressão linear de WZ versus Z .

2.2.11 *Box Map, LISA Map e Moran Map*

Para Câmara et al. (2001), esses três dispositivos gráficos (*Box Map*, *LISA Map* e *Moran Map*) são baseados nos resultados obtidos, do gráfico de espalhamento de Moran e dos índices de Moran locais.

2.2.11.1 *Box Map*

Esse mapa é uma extensão do gráfico de espalhamento de Moran, onde os elementos de cada quadrante do gráfico de espalhamento de Moran são representados por uma cor específica com seus respectivos polígonos, como mostra a Figura 10, que indica a percentagem de viagens de ônibus na cidade de Bauru - SP.

Segundo Krempi (2004), na Figura 10, verifica-se que Q1 e Q2 indicam uma associação espacial positiva, no sentido de que uma localização possui vizinhos

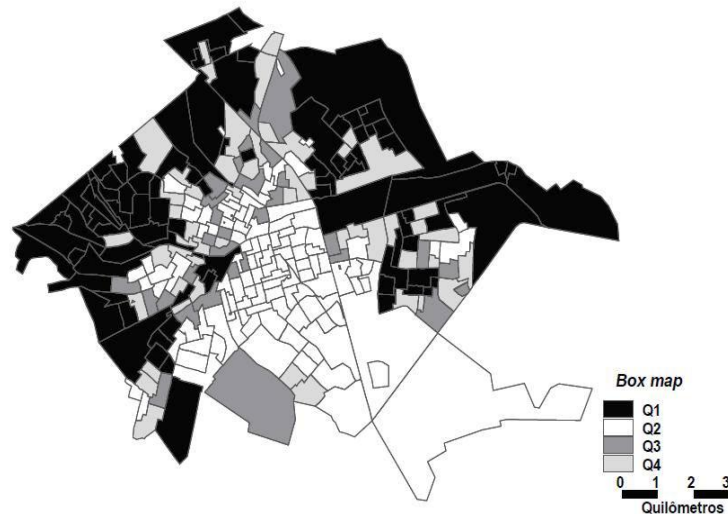


FIGURA 10 *Box Map* da variável percentagem de viagens de ônibus na cidade de Bauru - SP.
 Fonte: Krempi (2004).

com valores semelhantes. Verifica-se que Q1 e Q2 são áreas predominantemente residenciais, localizadas nas zonas Norte, Sudeste e Oeste da cidade de Bauru.

2.2.11.2 *LISA Map*

Para análise dos índices de Moran locais, o mapa denominado *LISA Map* indica as regiões que apresentam correlação local significativamente diferente das demais. A avaliação da significância é feita comparando-se os valores dos índices de Moran locais com uma série de valores obtidos por meio de permutações dos valores dos atributos dos vizinhos (9999 permutações), sob a hipótese nula de não existência de autocorrelação espacial. Na geração desse mapa os índices são classificados nos seguintes grupos, como mostra a Figura 11, não significativos, com significâncias de 5%, 1% e 0,1%, toma-se, como exemplo, a análise da variável percentagem de viagens de ônibus na cidade de Bauru - SP (Krempi, 2004).

De acordo com Krempi (2004), na Figura 11, as zonas que possuem índice de Moran local significativo são representadas no mapa com classe 1 (43 zonas), com 5% de significância, classe 2 (31 zonas), com 1% de significância, e classe

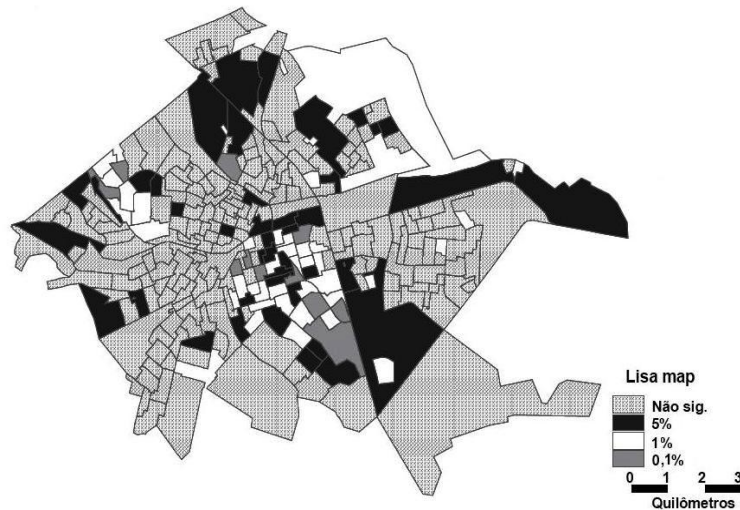


FIGURA 11 *LISA Map* da variável percentagem de viagens de ônibus na cidade de Bauru - SP.
 Fonte: Krempi (2004).

3 (7 zonas), com 0,01% de significância. E essas zonas são de predominância residencial.

2.2.11.3 *Moran Map*

No *Moran Map*, de forma semelhante ao *LISA Map*, somente são apresentadas as áreas para as quais os valores dos índices de Moran locais foram considerados significativos (valor-p < 0,05), porém, com classificação em quatro grupos, conforme os quadrantes aos quais pertençam no gráfico de espalhamento. As demais áreas ficam classificadas como não significativas. Um exemplo é apresentado na Figura 12, para a análise da variável percentagem de viagens de ônibus na cidade de Bauru - SP (Krempi, 2004).

Conforme Krempi (2004), na Figura 12, observa-se que do total de 274 zonas analisadas, 194 não são significativas, 39 pertencem a Q1, 32 pertencem a Q2, 5 pertencem a Q3 e 4 pertencem a Q4. Foram poucas as zonas de transição com significância espacial (Quadrantes 3 e 4). As zonas pertencentes ao quadrante Q1 estão, basicamente, na região sul da cidade, em que predominam as residências.

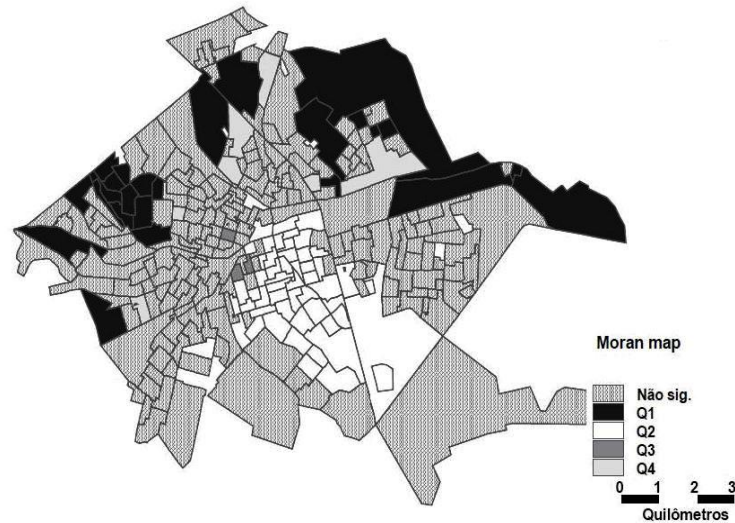


FIGURA 12 *Moran Map* da variável percentagem de viagens de ônibus na cidade de Bauru - SP.
 Fonte: Krempi (2004).

2.2.12 Resumo das técnicas de análise de dados de área

Com o objetivo de compreender as técnicas de análise de dados de área quanto ao padrão espacial e a autocorrelação espacial estudadas neste trabalho, mostra-se, por meio da Tabela 7 a sumarização de cada técnica e suas respectivas funções.

TABELA 7 Sumarização das técnicas de análise de dados de área e suas funções.

Técnica	Padrão espacial	Autocorrelação espacial
MML	Estima as variações e tendências globais	-
IMG	-	Estima a correlação global da região R
IML	-	Estima a correlação local da região R
GEM	-	Gráfico de correlação local da região R
BM	-	Mapa de correlação local da região R
LM	-	Mapa de correlação local usando o TPA aos níveis de 5%, 1% e 0,1% de probabilidade
MM	-	Mapa de correlação local usando o TPA ao nível de 5% de probabilidade

3 MATERIAL E MÉTODOS

Inicialmente, neste capítulo, será mencionada a origem dos dados e definidas as variáveis de estudo. Em seguida, faz-se com detalhes os procedimentos utilizados na aplicação das técnicas de dados de área em variáveis do ZEE - MG, para os cálculos da média móvel local, dos índices de Moran global e local. Descreve-se o teste de permutação aleatória para validação do índice de Moran global. Depois, comenta-se sobre o gráfico de espalhamento de Moran, *Box Map* e *Moran Map*. E, por fim, mencionam-se os softwares que serão utilizados nas análises estatísticas.

3.1 Origem dos dados

Os dados utilizados neste trabalho são oriundos do Zoneamento Ecológico Econômico das mesorregiões Metropolitana de Belo Horizonte e Sul/Sudeste de Minas, do Estado de Minas Gerais e referem-se ao ano de 2004. Tais dados fazem parte do produto maior denominado ZEE - MG, Zoneamento Ecológico Econômico do Estado de Minas Gerais (Scolforo et al., 2008a, 2008b, 2008c). A mesorregião Metropolitana de Belo Horizonte é composta por 105 municípios, como mostra a Figura 13, dos quais, para cada município serão estudados as variáveis: índice do produto interno bruto (IPIB) e o índice do valor adicionado agropecuário (IVAA). A mesorregião Sul/Sudeste de Minas é composta por 146 municípios, como mostra a Figura 14, e também serão estudados as variáveis IPIB e IVAA para cada município.

3.2 Produto interno bruto (PIB)

O produto interno bruto (PIB) representa a soma do valor monetário de todos os bens e serviços finais produzidos em uma determinada região. Esse é um dos indicadores mais utilizados na economia, com o objetivo de mensurar a pujança da atividade econômica de uma região. Seu cálculo é coordenado pelo Instituto Brasileiro de Geografia e Estatística (IBGE), desde 1996, em parceria com órgãos estaduais de pesquisa e estatística, as secretarias estaduais de planejamento e a superintendência da zona franca de Manaus - SUFRAMA (Carvalho & Silva, 2008).

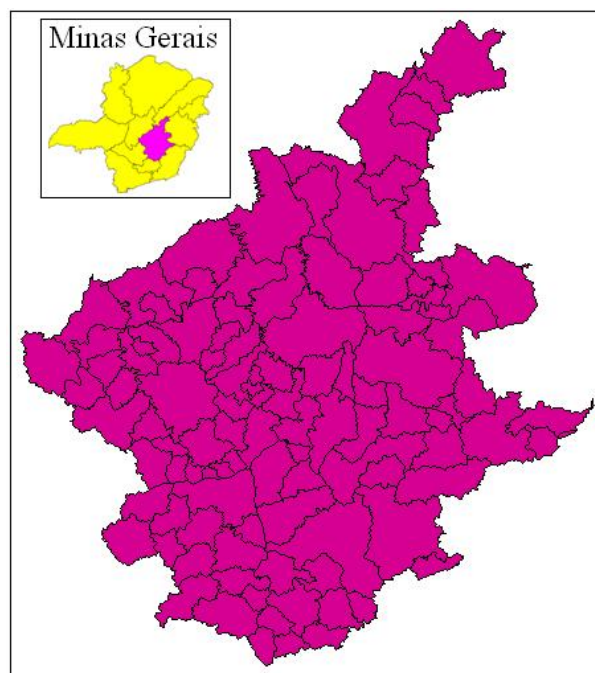


FIGURA 13 Mapa construído por elaboração própria da mesorregião Metropolitana de Belo Horizonte - MG. Lavras, 2010.
Fonte: Base cartográfica do IBGE (2002).

3.3 Valor adicionado agropecuário (VAA)

O valor adicionado do setor agropecuário corresponde, conforme conceituação do departamento de contas nacionais do IBGE, ao valor da produção da atividade agropecuária (lavoura permanente, lavoura temporária, pecuária, horticultura, extrativa vegetal, silvicultura, investimentos em formação de matas plantadas e lavouras permanentes, pesca, indústria rural, produção particular da pessoa residente no estabelecimento rural e serviços auxiliares da agropecuária), menos o consumo intermediário (não computados os impostos). Considera-se que, quanto maior o índice do setor agropecuário, melhor a situação do município em termos de potencialidade produtiva (Scolforo et al., 2008b, 2008c).

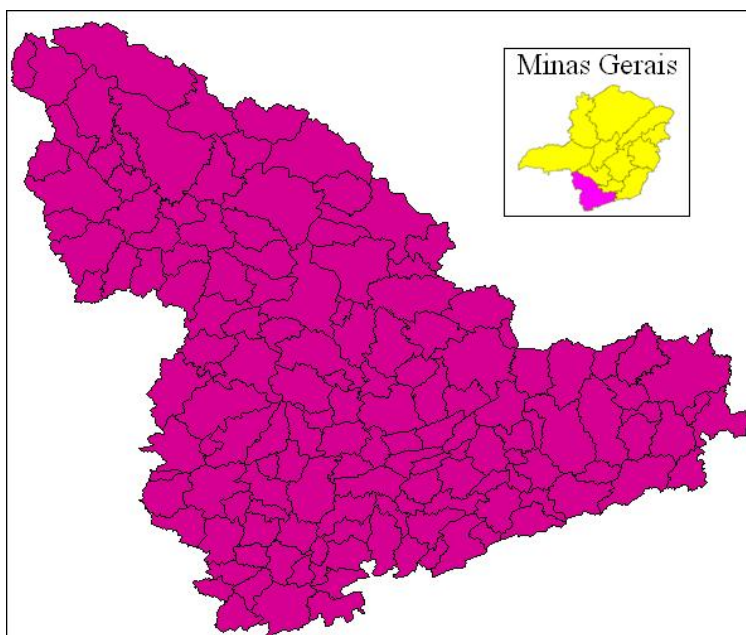


FIGURA 14 Mapa construído por elaboração própria da mesorregião Sul/Sudeste de Minas - MG. Lavras, 2010.
Fonte: Base cartográfica do IBGE (2002).

3.4 Análise exploratória

Com o objetivo de observar o comportamento da variável IPIB (ou IVAA), na mesorregião Metropolitana de Belo Horizonte, e na mesorregião Sul/Sudeste de Minas, será feita uma análise visual, por meio do mapa de intervalos iguais com gráfico de barras.

Como forma de analisar as variações das tendências espaciais da variável IPIB (ou IVAA), utiliza-se a média móvel local.

3.5 Média móvel local

Como já visto no capítulo 2, a fórmula para cálculo da média móvel local para cada município das mesorregiões em estudo, é descrita a seguir:

$$\hat{\mu}_i = \sum_{j=1}^n w_{ij} y_j, \quad (3.1)$$

com $i = 1, \dots, n = 105$ e $j = 1, \dots, n = 105$, para mesorregião Metropolitana de Belo Horizonte, e com $i = 1, \dots, n = 146$ e $j = 1, \dots, n = 146$, para mesorregião Sul/Sudeste de Minas. w_{ij} e y_j , representam, respectivamente, os elementos da matriz de proximidade espacial padronizada de primeira ordem $W_{n \times n}^{(1)}$ (definida pelo método da fronteira comum) e o valor observado da variável IPIB (ou IVAA).

Para mensurar os efeitos de dependências espaciais entre os municípios, utilizam-se os índices de Moran global e local.

3.6 Índice de Moran global

Conforme o capítulo 2, o índice I de Moran global é definido por:

$$I = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y}) (y_j - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (3.2)$$

com $i = 1, \dots, n = 105$ e $j = 1, \dots, n = 105$, para mesorregião Metropolitana de Belo Horizonte, e com $i = 1, \dots, n = 146$ e $j = 1, \dots, n = 146$, para mesorregião Sul/Sudeste de Minas, em que:

w_{ij} são os elementos da matriz de proximidade espacial padronizada de primeira ordem $W_{n \times n}^{(1)}$ (definida pelo método da fronteira comum);

y_i e y_j são os valores observados das variáveis IPIB (ou IVAA), respectivamente, dos municípios A_i e A_j ;

\bar{y} é a média do valor observado da variável IPIB (ou IVAA) de todos os municípios da mesorregião.

Para a validação do índice de Moran global, será utilizado o teste de permutação aleatória, definido a seguir.

3.7 Simulação do valor-p para o índice I de Moran utilizando teste de permutação aleatória

Para simular o valor-p utilizando permutação aleatória para o índice I de Moran serão gerados $N = 8000$ valores de I , conforme o algoritmo dado no capítulo 2, descrito a seguir:

- 1) Calcula-se o índice $I_{(0)}$ para os dados observados do IPIB (ou IVAA) dos n municípios;
- 2) Calcula-se o índice $I_{(1)}$ permutando aleatoriamente os valores do IPIB (ou IVAA) entre os n municípios;
- 3) Calcula-se o índice $I_{(2)}$ permutando novamente de forma aleatória os valores do IPIB (ou IVAA) entre os n municípios;
- 4) Repete-se o item anterior 7998 vezes até calcular o índice $I_{(8000)}$.

Em que: $n = 105$ municípios para mesorregião Metropolitana de Belo Horizonte, e $n = 146$ municípios para mesorregião Sul/Sudeste de Minas.

Para avaliar a significância da estatística I de Moran, as seguintes hipóteses são estabelecidas:

- H_0 : $I = 0$ (não existe autorrelação espacial entre os municípios);
- H_1 : $I > 0$ (existe autorrelação espacial positiva entre os municípios).

Sob a hipótese H_0 , a variável aleatória IPIB (ou IVAA) é independente e identicamente distribuída, então todas as permutações dos valores IPIB (ou IVAA) entre os municípios são igualmente prováveis. Assim, o *valor - p* do teste será:

$$\text{valor} - p = \frac{NIS}{N + 1}, \quad (3.3)$$

em que NIS representa o número de índices simulados com valores maiores do que o valor do índice observado $I_{(0)}$, e N é o número total de índices simulados.

Rejeita-se a hipótese H_0 no teste de nível α se $\text{valor} - p < \alpha$. Para visualizar o resultado será feito o histograma com todos os valores $I_{(k)}$ com $k = 0, 1, \dots, 8000$, e aponta-se o local na escala horizontal por meio de uma reta na vertical no valor $I_{(0)}$.

3.8 Gráfico de espalhamento de Moran e Box Map

O gráfico de espalhamento será utilizado como uma ferramenta para visualizar a autocorrelação existente entre os municípios da mesorregião Metropolitana de Belo Horizonte (ou entre os municípios da mesorregião Sul/Sudeste de Minas) para cada uma das variáveis IPIB e IVAA.

Na geração do *Box Map*, cada município da mesorregião Metropolitana de Belo Horizonte (ou cada município da mesorregião Metropolitana Sul/Sudeste de Minas) será classificado conforme sua posição em relação aos quadrantes do gráfico de espalhamento de Moran, recebendo uma cor correspondente no mapa gerado.

3.9 Índice de Moran local e Moran Map

O índice I_i de Moran local é dado pela seguinte expressão:

$$I_i = (y_i - \bar{y}) \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_j - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2 / n}, \quad (3.4)$$

com $i = 1, \dots, n = 105$ e $j = 1, \dots, n = 105$, para mesorregião Metropolitana de Belo Horizonte, e com $i = 1, \dots, n = 146$ e $j = 1, \dots, n = 146$, para mesorregião Sul/Sudeste de Minas, em que:

w_{ij} são os elementos da matriz de proximidade espacial padronizada de primeira ordem $W_{n \times n}^{(1)}$ (definida pelo método da fronteira comum);

y_i e y_j são os valores observados das variáveis IPIB (ou IVAA), respectivamente, dos municípios A_i e A_j ;

\bar{y} é a média do valor observado da variável IPIB (ou IVAA) de todos os municípios da mesorregião.

No *Moran Map*, será avaliada a significância dos valores do índice de Moran local obtido para cada município, em relação à hipótese de não existência de autocorrelação espacial entre os municípios (hipótese nula). Na avaliação da significância, será utilizada a abordagem de permutação dos atributos dos municípios

vizinhos, conforme descrito em Anselin (1995), sendo feitas 9999 permutações geradas pelo software GeoDA (Anselin, 2004). Os municípios para os quais os valores do índice de Moran local foram considerados significantes ($\text{valor-p} < 0,05$), são destacados, porém, aparecem classificados em quatro grupos, conforme sua localização no quadrante do gráfico de espalhamento. Os demais municípios, ficam classificados como sem significância.

3.10 Softwares computacionais utilizados

Todas as análises estatísticas feitas neste trabalho, foram realizadas com a utilização dos softwares GeoDA (Anselin, 2004) e TerraView (INPE, 2007). As informações para quem deseja trabalhar com o software GeoDa está disponível em Anselin (2005), e com o software TerraView está disponível em INPE (2009).

4 RESULTADOS E DISCUSSÃO

A seguir serão apresentados os principais resultados obtidos neste estudo. Para tal, foram aplicadas as técnicas espaciais de dados de área descritas na metodologia para o cálculo da média móvel local, dos índices de autocorrelação de Moran global e local, e o teste da permutação aleatória para a validação do índice de Moran global. Foram também construídos gráficos de espalhamento e mapas (*Box Map* e *Moran Map*) temáticos com o propósito de melhorar a compreensão da relação existente entre os municípios da mesorregião Metropolitana de Belo Horizonte (ou entre os municípios da mesorregião Sul/Sudeste de Minas), por meio das variáveis, índice do produto interno bruto (IPIB) e o índice do valor adicionado agropecuário (IVAA).

4.1 Estudo da variável IPIB na mesorregião Metropolitana de Belo Horizonte e na mesorregião Sul/Sudeste de Minas

4.1.1 Análise exploratória

Como primeira análise exploratória foi construído o mapa de intervalos iguais com gráfico de barras para a variável IPIB dos municípios da mesorregião Metropolitana de Belo Horizonte, e da mesorregião Sul/Sudeste de Minas, como mostram as Figuras 15 e 16, respectivamente.

Observou-se na Figura 15, os dois maiores índices do PIB da mesorregião Metropolitana de Belo Horizonte que estão localizados nos municípios de Belo Horizonte (0,7604) e Betim (0,4980), seguidos de três índices do PIB com valores moderados localizados em Contagem (0,2869), Sete Lagoas (0,1806) e Ouro Branco (0,1625). Os demais municípios possuem baixos índices do PIB, entre 0,0124 e 0,1621.

Na Figura 16, os quatro municípios que se destacaram com os maiores índices do PIB foram: Poços de Caldas (0,1266), São Bento Abade (0,116), Itanhandu (0,1143) e Extrema (0,1102). E os demais municípios estão com valores dos índices do PIB, distribuídos em diferentes intervalos.

Com a tentativa de reduzir a variabilidade espacial e observar possíveis padrões espaciais, comparou-se o mapa de visualização por quintil com o mapa da média

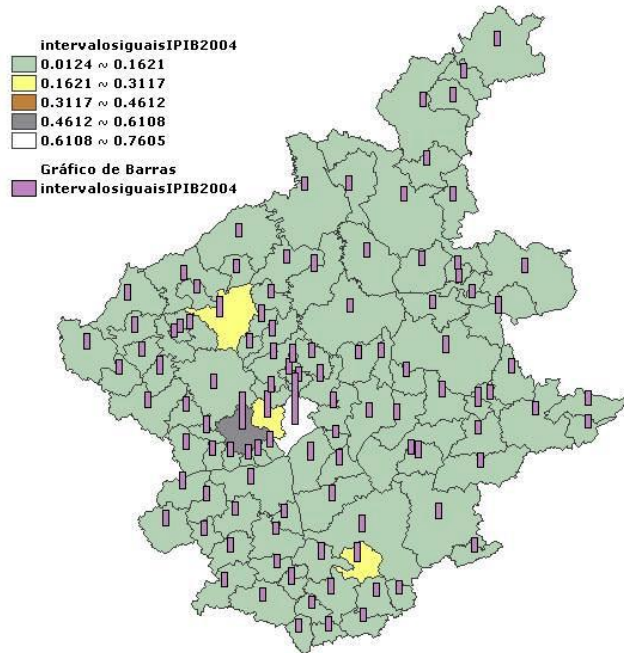


FIGURA 15 Mapa de intervalos iguais com gráfico de barras para a variável IPIB na mesorregião Metropolitana de Belo Horizonte - MG, 2004. Fonte: Construído por elaboração própria com base no software TerraView (INPE, 2007).

móvel local. Apresentado a seguir.

4.1.2 Média móvel local

Inicialmente, criou-se o mapa da mesorregião Metropolitana de Belo Horizonte por meio da visualização por quintil da variável IPIB. Em seguida, comparou-se esse mapa com o mapa dessa mesma mesorregião, utilizando a visualização da média móvel local da variável IPIB como mostra a Figura 17.

Percebeu-se, ao comparar os dois mapas da Figura 17, que a média móvel local forneceu um alisamento na distribuição da variável IPIB, ou seja, o valor mínimo que era de 0,012499 passou a ser de 0,042166 e o máximo que era de 0,760401 passou a ser de 0,347701. Outro fato observado pela média móvel local, foi um leve padrão espacial de valores altos do índice do PIB no centro da mesorregião e dimi-

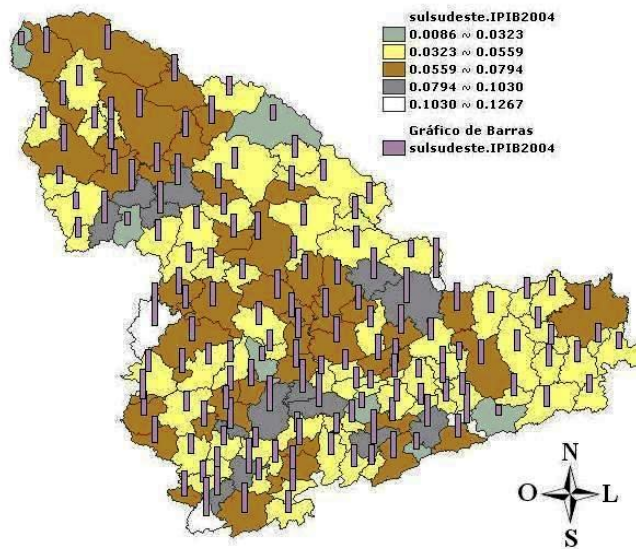


FIGURA 16 Mapa de intervalos iguais com gráfico de barras para a variável IPIB na mesorregião Sul/Sudeste de Minas - MG, 2004.

Fonte: Construído por elaboração própria com base no software TerraView (INPE, 2007).

nuindo gradativamente até chegar nas extremidades da mesorregião. Observou-se, também, que os municípios localizados no centro da mesorregião, foram correlacionados pelos altos índices do PIB dos municípios de Belo Horizonte (0,7604), Betim (0,4980) e Contagem (0,2869), formando um *cluster* espacial de municípios com valores altos do índice do PIB.

De forma análoga, estudou-se a variável IPIB na mesorregião Sul/Sudeste de Minas. Primeiramente, fez-se o mapa da variável IPIB de sua mesorregião com a visualização por quintil, e comparou-se com o mapa dessa mesma mesorregião, utilizando-se a média móvel local, como mostra a Figura 18.

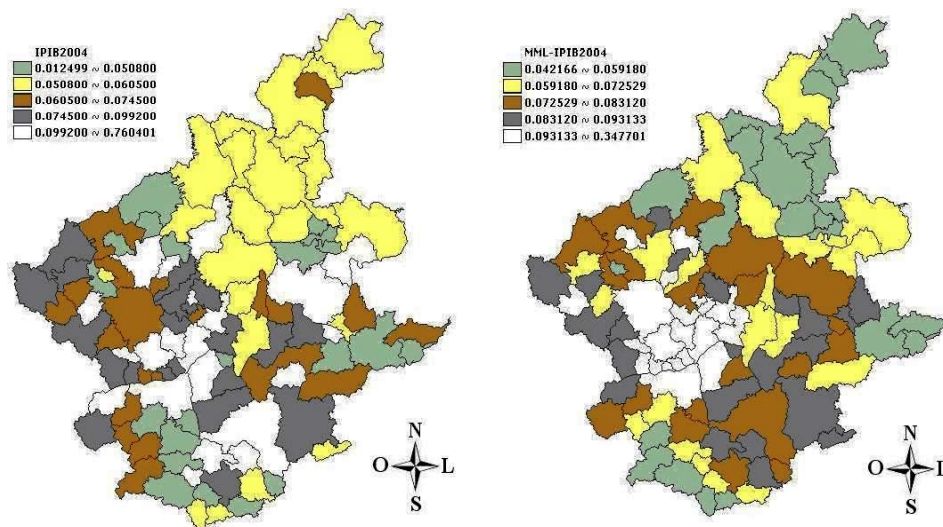


FIGURA 17 Mapa da distribuição da variável IPIB na mesorregião Metropolitana de Belo Horizonte - MG, 2004. À esquerda, apresentação da variável por quintil. À direita, média móvel local. Lavras, 2010. Fonte: Construído por elaboração própria com base no software TerraView (INPE, 2007).

Comparando-se os dois mapas da Figura 18, a média móvel local, assim como na mesorregião Metropolitana de Belo Horizonte, forneceu um alisamento na distribuição da variável IPIB, ou seja, o valor mínimo que era de 0,008699 passou a ser de 0,040349 e o máximo que era de 0,126601 passou a ser de 0,149481. Outro fato registrado por meio da média móvel local foi a formação de um *cluster* espacial de valores baixos do IPIB no leste da mesorregião.

Dessa forma, o uso da média móvel local para a variável IPIB, trouxe benefício na redução da variância amostral para cada mesorregião. E, também, foi eficiente na descrição espacial e identificação de *clusters*.

4.1.3 Índice de Moran global e o teste de permutação aleatória

Os valores do índice I de Moran global para a variável IPIB nas mesorregiões Metropolitana de Belo Horizonte e Sul/Sudeste de Minas foram, respectivamente, de 0,1259 e 0,0818. Tem-se que esses valores são baixos para concluir se houve,

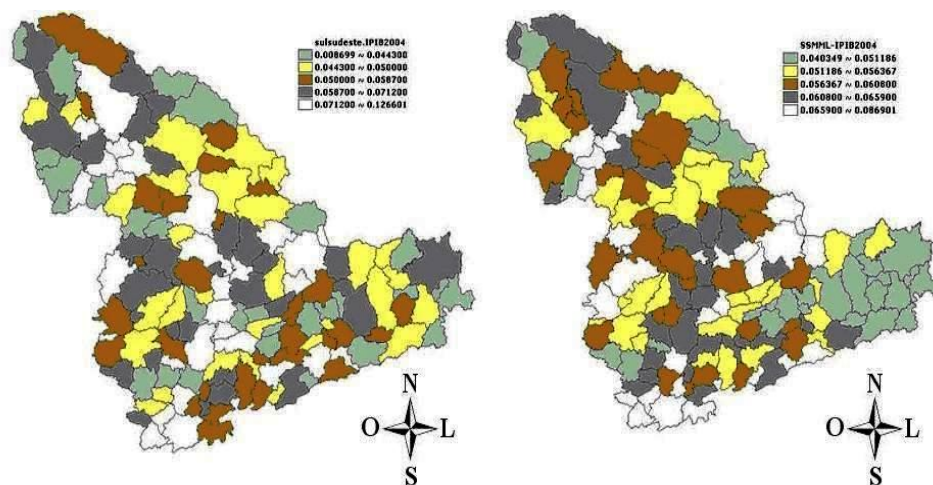


FIGURA 18 Mapa da distribuição da variável IPIB na mesorregião Sul/Sudeste de Minas - MG, 2004. À esquerda, apresentação da variável por quintil. À direita, média móvel local.

Fonte: Construído por elaboração própria com base no software TerraView (INPE, 2007).

ou não, uma autocorrelação espacial positiva. Porém, foi realizado o teste de permutação aleatória do nível de significância de I , para cada uma das mesorregiões, sob a hipótese nula de ausência de autocorrelação espacial entre os municípios de cada uma das mesorregiões em estudo. Os resultados obtidos do teste com a variável IPIB, na mesorregião Metropolitana de Belo Horizonte, estão representados na Figura 19, e na mesorregião Sul/Sudeste de Minas, estão representados na Figura 20.

De acordo com a Figura 19, foram feitas 8000 permutações e obteve-se um valor- $p = 0,0230$, então rejeitou-se a hipótese nula de ausência de autocorrelação espacial entre os municípios da mesorregião Metropolitana de Belo Horizonte com relação a variável IPIB, ao nível de 5% de significância.

Na Figura 20, também foram feitas 8000 permutações e obteve-se um valor- $p = 0,0477$, então rejeitou-se a hipótese nula de ausência de autocorrelação espacial entre os municípios da mesorregião Sul/Sudeste de Minas com relação à variável IPIB também ao nível de 5% de significância.

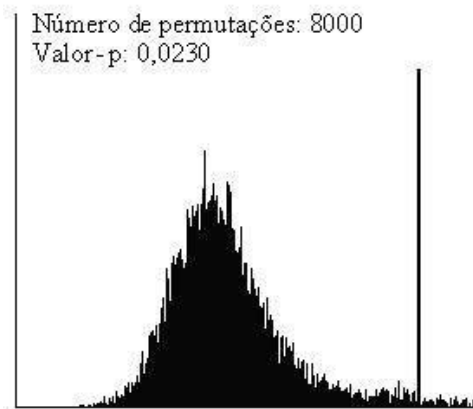


FIGURA 19 Teste de significância para o índice de Moran global da variável IPIB sob a hipótese nula de ausência de autocorrelação espacial. Fonte: Construído por elaboração própria com base no software GeoDA (Anselin, 2004).

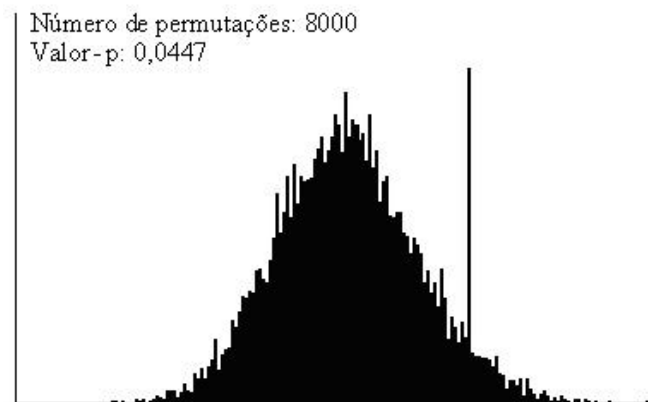


FIGURA 20 Teste de significância para o índice de Moran global da variável IPIB sob a hipótese nula de ausência de autocorrelação espacial. Fonte: Construído por elaboração própria com base no software GeoDA (Anselin, 2004).

Assim, como em ambas as mesorregiões, o valor de I foi positivo e o teste foi significativo a 5% de probabilidade, então existe autocorrelação espacial positiva entre os municípios de cada mesorregião, ou seja, o valor da variável IPIB de um

município dentro de sua mesorregião, tende a ser semelhante aos valores dos seus vizinhos.

A seguir, usou-se o diagrama de espalhamento de Moran e o *Box Map*, como ferramentas para visualizar a autocorrelação existente entre os municípios de cada mesorregião.

4.1.4 Gráfico de espalhamento de Moran e *Box Map*

Para detalhar mais o índice de Moran global foram construídos o gráfico de espalhamento de Moran e o *Box Map* da variável IPIB, para cada mesorregião, mostrados pelas Figuras 21 e 22, respectivamente.

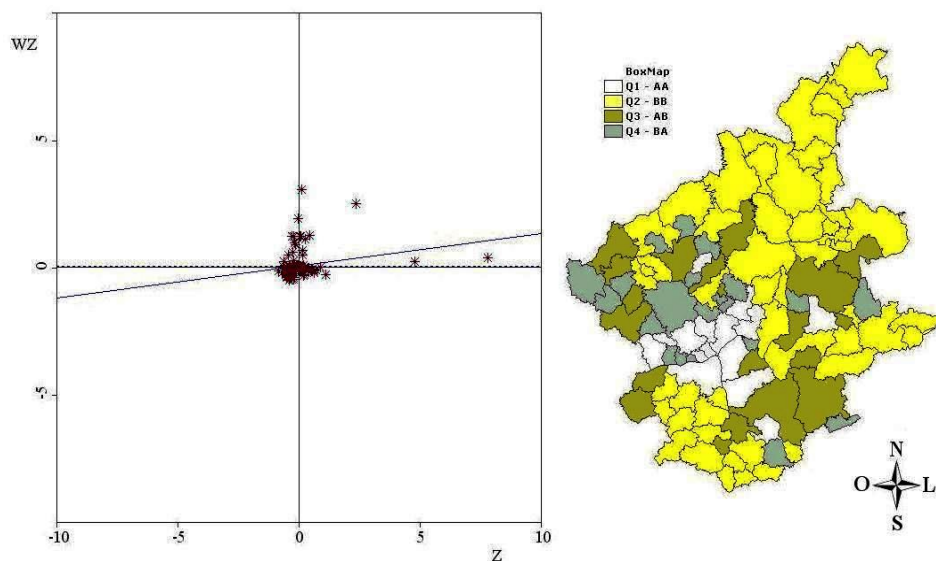


FIGURA 21 À esquerda, diagrama de espalhamento de Moran da variável IPIB. À direita, *Box Map* da variável IPIB.

Fonte: Construído por elaboração própria com base nos softwares GeoDA (Anselin, 2004) e TerraView (INPE, 2007).

Na Figura 21, observou-se que a maior parte dos municípios da mesorregião Metropolitana de Belo Horizonte estão localizados nos quadrantes Alto-Alto (14,3%) e Baixo-Baixo (46,7%). Esses resultados estão de acordo com o *I* de Moran global calculado, já que mostram que a maioria dos municípios observados encontra-se nos quadrantes que representam a existência de autocorrelação espa-

cial positiva: valores do índice PIB dos municípios são similares aos verificados por seus vizinhos. Ademais, a inclinação positiva da reta também comprova a existência de autocorrelação, visto que o I de Moran global é o coeficiente angular da mesma. Observou-se, pelo Box Map, que houve similaridade entre os municípios com valores altos do índice PIB, localizados no centro da mesorregião, e houve similaridade entre os municípios de valores baixos do índice PIB, localizados no norte e sul da mesorregião.

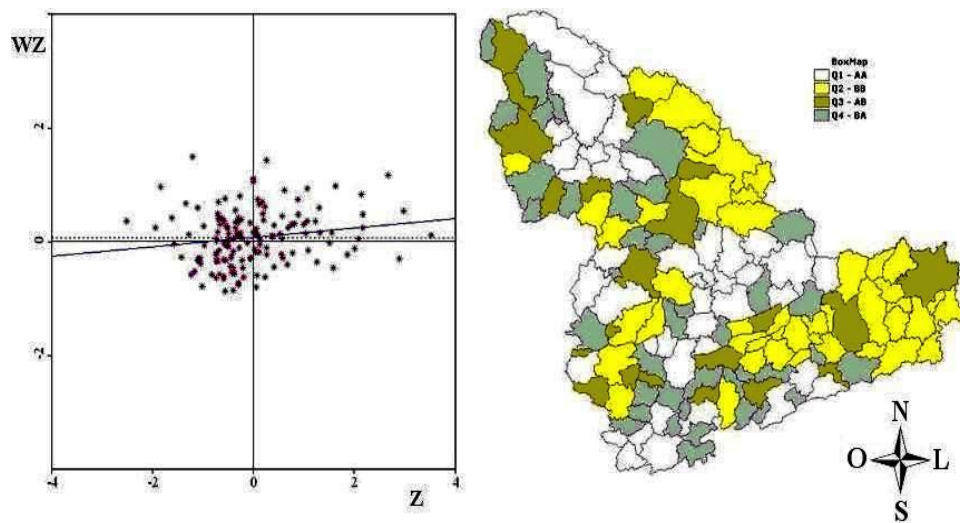


FIGURA 22 À esquerda, diagrama de espalhamento de Moran da variável IPIB. À direita, *Box Map* da variável IPIB.
 Fonte: Construído por elaboração própria com base nos softwares GeoDA (Anselin, 2004) e TerraView (INPE, 2007).

Na Figura 22, observou-se também que, a maior parte dos municípios da mesorregião Sul/Sudeste de Minas estão localizados nos quadrantes Alto-Alto (28,1%) e Baixo-Baixo (30,1%), comprovando a existência de autocorrelação espacial positiva. Pelo Box Map, percebeu-se similaridade entre os municípios com valores altos do índice PIB, localizados no noroeste e centro da mesorregião, e houve similaridade entre os municípios com valores baixos do índice PIB, localizados no norte e leste da mesorregião.

Nota-se que o índice de Moran global fornece um único valor como medida

de associação espacial para todos os valores do IPIB dos municípios, para cada mesorregião em estudo. Para examinar padrões numa escala de maior detalhes, sobre a variável em estudo, utilizou-se o índice de Moran local e o *Moran Map*.

4.1.5 Índice de Moran local e *Moran Map*

As Figura 23 e 24 mostram mapas contendo os valores positivos e negativos do índices de Moran local, e mapas com valores simulados, usando-se o teste da permutação aleatória (9999 permutações), para o índice de Moran local apresentados pelo *Moran Map*, para cada mesorregião.

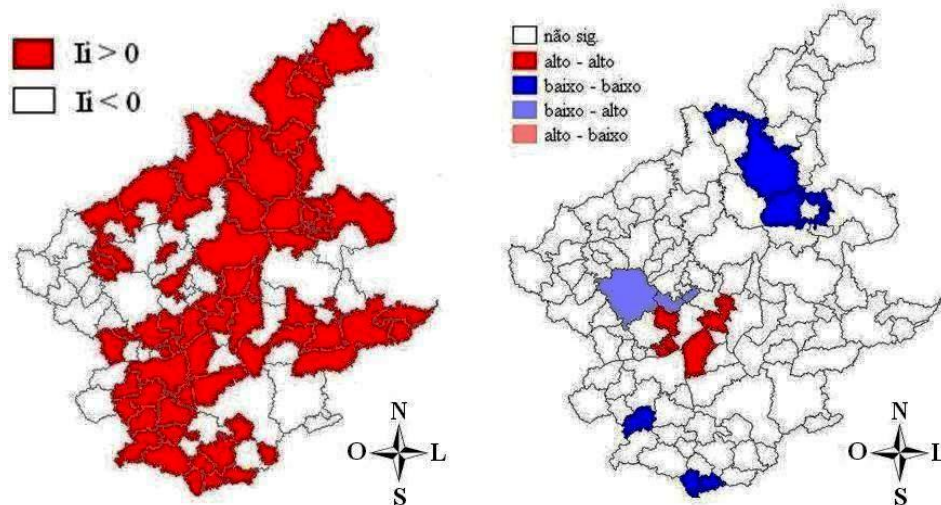


FIGURA 23 Mapa do índice de Moran local da variável IPIB na mesorregião Metropolitana de Belo Horizonte-MG, 2004. À esquerda, valores positivos e negativos do índice de Moran local. À direita, *Moran Map*.

Fonte: Construído por elaboração própria com base nos softwares TerraView (INPE, 2007) e GeoDA (Anselin, 2004).

Na Figura 23, no mapa da esquerda, observou-se que a maioria dos municípios (61%) apresentaram valores positivos para o índice de Moran Local, indicando similaridade (alto-alto ou baixo-baixo) da variável IPIB, de norte a sul, passando pelos municípios centrais, e alguns municípios do lado leste da mesorre-

gião Metropolitana de Belo Horizonte. Enquanto o restante dos municípios (39%) apresentaram valores negativos para o índice de Moran Local, indicando dissimilaridade (alto-baixo ou baixo-alto) da variável IPIB, entre alguns municípios do lado leste e oeste da mesorregião Metropolitana de Belo Horizonte. E, no mapa da direita, está representado a simulação realizada para os índices de Moran Locais pelo teste da permutação aleatória. Constatou-se similaridade entre cinco municípios com valores altos do índice do PIB (Contagem, Sabará, Sarzedo, Ibirité e Nova Lima), localizados no centro da mesorregião. Houve também similaridade entre seis municípios de baixos valores do índice do PIB, em que três municípios estão localizados no norte (Conceição do Mato Dentro, Morro do Pilar e São Sebastião do Rio Preto) e três municípios (Casa Grande, Cristiano Ottoni e Piedade dos Gerais) no sul da mesorregião. Sendo que, a maior parte dos municípios não apresentaram valores significativos.

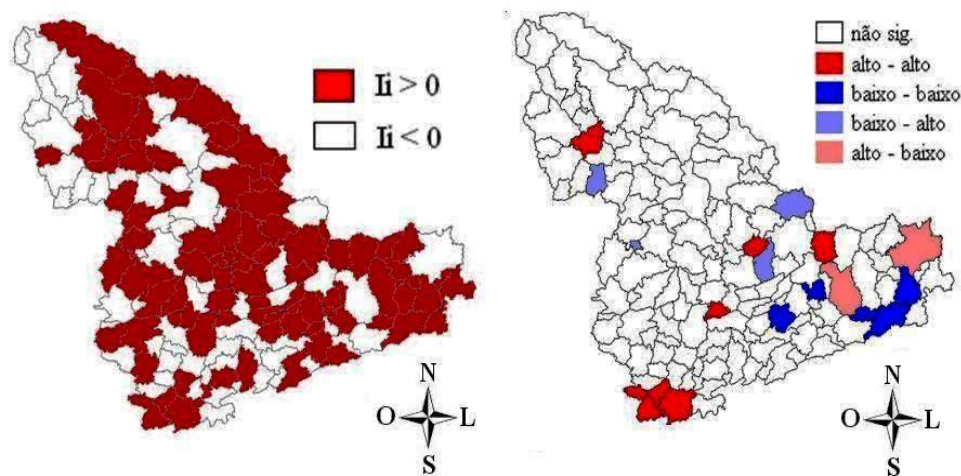


FIGURA 24 Mapa do índice de Moran local da variável IPIB na mesorregião Sul/Sudeste de Minas-MG, 2004. À esquerda, valores positivos e negativos do índice de Moran local. À direita, *Moran Map*.

Fonte: Construído por elaboração própria com base nos softwares TerraView (INPE, 2007) e GeoDA (Anselin, 2004).

Pela Figura 24, observou-se, no mapa da esquerda, que a maior parte dos municípios (58,2%) apresentaram para o índice de Moran Local valores positivos,

indicando similaridade (alto-alto ou baixo-baixo) da variável IPIB, em vários lugares na mesorregião Sul/Sudeste de Minas. Enquanto o restante dos municípios (41,8%) apresentaram valores negativos para o índice de Moran Local, indicando dissimilaridade (alto-baixo ou baixo-alto) da variável IPIB, também em vários lugares na mesorregião Sul/Sudeste de Minas. E, no mapa da direita, mostrou a simulação feita para os índices de Moran Locais pelo teste da permutação aleatória. Constatou-se similaridade entre oito municípios com valores altos do índice do PIB, em que quatro municípios estão localizados no sul (Camanducaia, Extrema, Itapeva e Toledo), um no norte (Jacuí) e três no leste (Monsenhor Paulo, São Sebastião da Bela Vista e São Tomé das Letras) da mesorregião. Houve similaridade entre cinco municípios de baixos valores do índice do PIB (Alagoa, Bocaina de Minas, Cristina, Liberdade e Soledade de Minas), localizados no leste da mesorregião. Nessa mesorregião, também, a maior parte dos municípios não apresentaram valores significativos.

4.2 Estudo da variável IVAA na mesorregião Metropolitana de Belo Horizonte e na mesorregião Sul/Sudeste de Minas

4.2.1 Análise exploratória

À princípio, analisou-se o mapa de intervalos iguais com gráfico de barras para a variável IVAA dos municípios da mesorregião Metropolitana de Belo Horizonte, e da mesorregião Sul/Sudeste de Minas, visualizados nas Figuras 25 e 26, respectivamente.

De acordo com a Figura 25, os três municípios da mesorregião Metropolitana de Belo Horizonte que se destacam com os maiores índices do VAA foram: Santa Maria do Itabira (0,3113), Pará de Minas (0,2654) e São José da Lapa (0,2202), sendo que a maioria dos municípios apresentaram índices do VAA entre 0,0306 e 0,0868.

Percebeu-se, na Figura 26, que os dois maiores índices do VAA da mesorregião Sul/Sudeste de Minas estão localizados nos municípios de Itanhandu (0,3246) e Passos (0,2082). Observou-se também vários municípios com baixos índices do VAA, entre 0,0264 e 0,0861.

A seguir, comparou-se o mapa de visualização por quintil com o mapa da

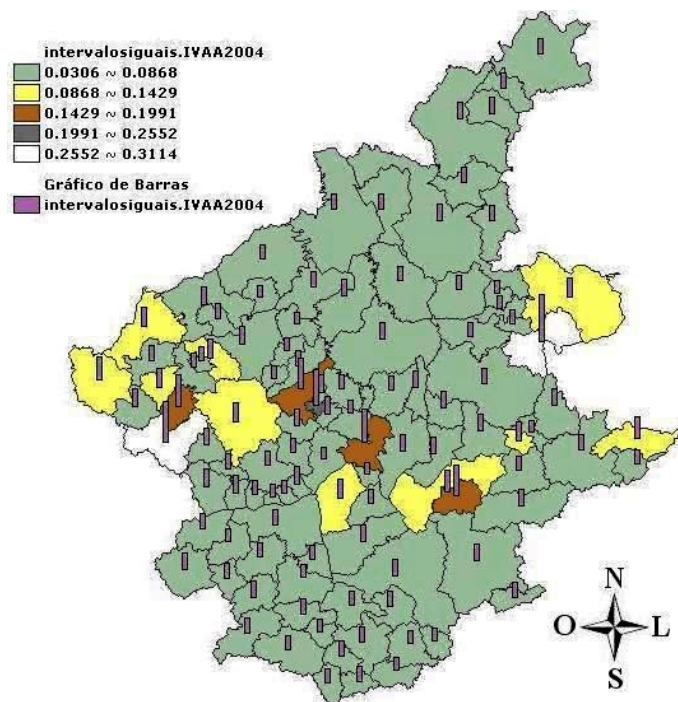


FIGURA 25 Mapa de intervalos iguais com gráfico de barras para a variável IVAA na mesorregião Metropolitana de Belo Horizonte - MG, 2004.

Fonte: Construído por elaboração própria com base no software TerraView (INPE, 2007).

média móvel para observar possíveis padrões espaciais.

4.2.2 Média móvel local

Criou-se o mapa de visualização por quintil da variável IVAA da mesorregião Metropolitana de Belo Horizonte. Em seguida, comparou-se esse mapa com o mapa de visualização da média móvel local da variável IVAA como mostra a Figura 27.

Comparando-se os dois mapas da Figura 27, a média móvel local forneceu um alisamento na distribuição da variável IVAA, isto é, o valor mínimo que era de 0,030699 passou a ser de 0,050159 e o máximo que era de 0,311301 passou a ser de 0,149481. Observou-se também, a formação de vários *clusters* espaciais do

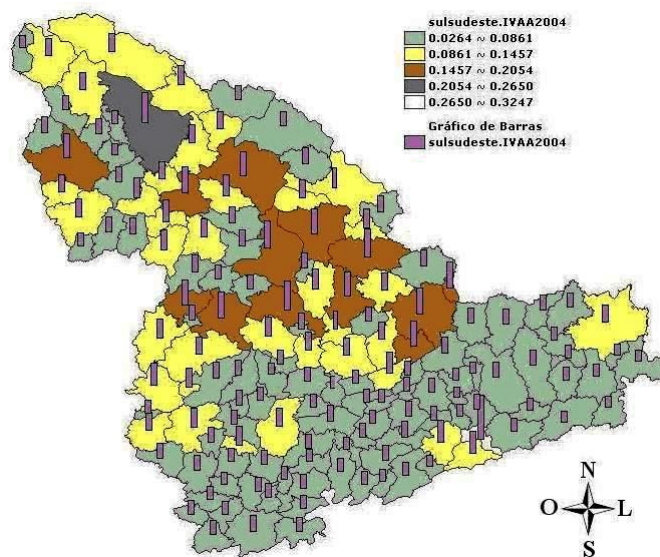


FIGURA 26 Mapa de intervalos iguais com gráfico de barras para a variável IVAA na mesorregião Sul/Sudeste de Minas - MG, 2004.

Fonte: Construído por elaboração própria com base no software TerraView (INPE, 2007).

IVAA em diferentes lugares da mesorregião.

De forma similar, estudou-se a variável IVAA na mesorregião Sul/Sudeste de Minas. Construiu-se o mapa da variável IVAA de sua mesorregião com a visualização por quintil, e comparou-se com o mapa de visualização da média móvel local da variável IVAA como mostra a Figura 28.

Ao comparar os dois mapas da Figura 28, a média móvel local, assim como na mesorregião Metropolitana de Belo Horizonte, forneceu um alisamento na distribuição da variável IVAA, ou seja, o valor mínimo que era de 0,026499 passou a ser de 0,042566 e o máximo que era de 0,32461 passou a ser de 0,165168. Observou-se, também, por meio da média móvel local, a formação de um novo *cluster* espacial de valores baixos do IVAA no sul da mesorregião.

De forma semelhante ao da variável IPIB, o uso da média móvel local para a variável IVAA, trouxe também redução da variância amostral para cada mesorregião e mostrou-se eficiente na descrição espacial e na identificação de *clusters*.

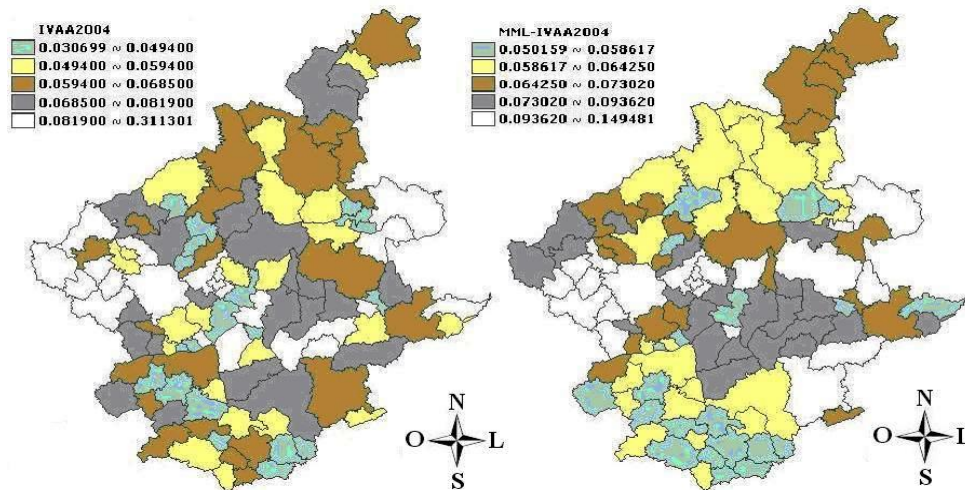


FIGURA 27 Mapa da distribuição da variável IVAA da mesorregião Metropolitana de Belo Horizonte - MG, 2004. À esquerda, apresentação da variável por quintil. À direita, média móvel local.

Fonte: Construído por elaboração própria com base no software TerraView (INPE, 2007).

4.2.3 Índice de Moran global e o teste de permutação aleatória

Os valores do índice I de Moran global para a variável IVAA nas mesorregiões Metropolitana de Belo Horizonte e Sul/Sudeste de Minas foram, respectivamente, de 0,1254 e 0,2545. Tem-se que esses valores são baixos para concluir se houve, ou não, uma autocorrelação espacial positiva. Porém, foi feito o teste de permutação aleatória do nível de significância de I , para cada uma das mesorregiões, sob a hipótese nula de ausência de autocorrelação espacial entre os municípios de cada uma das mesorregiões em estudo. Os resultados obtidos do teste com a variável IVAA, na mesorregião Metropolitana de Belo Horizonte estão representados pela Figura 29, e na mesorregião Sul/Sudeste de Minas estão representados pela Figura 30.

Na Figura 29, foram feitas 8000 permutações e obteve-se um valor- $p = 0,0269$, então rejeitou-se a hipótese nula de ausência de autocorrelação espacial entre os municípios da mesorregião Metropolitana de Belo Horizonte com relação a variá-

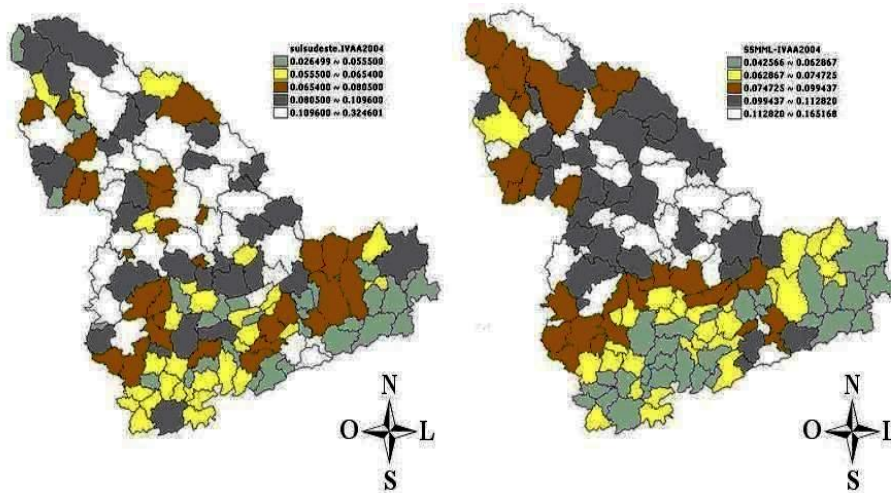


FIGURA 28 Mapa da distribuição da variável IVAA na mesorregião Sul/Sudeste de Minas - MG, 2004. À esquerda, apresentação da variável por quintil. À direita, média móvel local.

Fonte: Construído por elaboração própria com base no software TerraView (INPE, 2007).

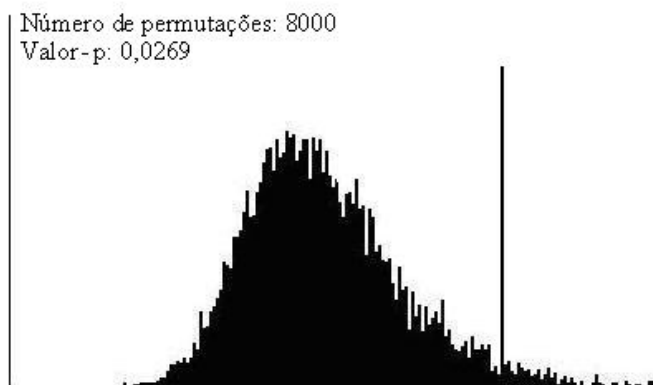


FIGURA 29 Teste de significância para o índice de Moran global da variável IVAA sob a hipótese nula de ausência de autocorrelação espacial.

Fonte: Construído por elaboração própria com base no software GeoDA (Anselin, 2004).

vel IVAA ao nível de 5% de significância.

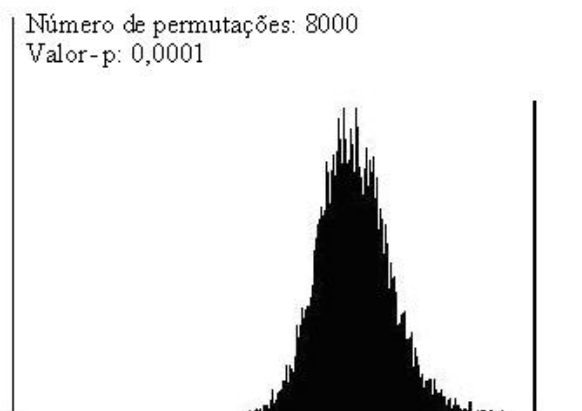


FIGURA 30 Teste de significância para o índice de Moran global da variável IVAA sob a hipótese nula de ausência de autocorrelação espacial. Fonte: Construído por elaboração própria com base no software GeoDA (Anselin, 2004).

Na Figura 30, também foram feitas 8000 permutações e obteve-se um valor-p = 0,0001, então rejeitou-se a hipótese nula de ausência de autocorrelação espacial entre os municípios da mesorregião Sul/Sudeste de Minas com relação à variável IVAA ao nível de 5% de significância.

Como em ambas as mesorregiões, o valor de I foi positivo e o teste foi significativo a 5% de probabilidade, existindo, então, autocorrelação espacial positiva entre os municípios de cada mesorregião, ou seja, o valor da variável IVAA de um município dentro de sua mesorregião, tende a ser semelhante aos valores dos seus vizinhos.

Como ferramentas para visualizar a autocorrelação existente entre os municípios de cada mesorregião, usou-se o diagrama de espalhamento de Moran e o *Box Map*.

4.2.4 Gráfico de espalhamento de Moran e *Box Map*

As Figuras 31 e 32 mostram o gráfico de espalhamento de Moran e o *Box Map* da variável IVAA, para as mesorregiões Metropolitana de Belo Horizonte e

Sul/Sudeste de Minas, respectivamente.

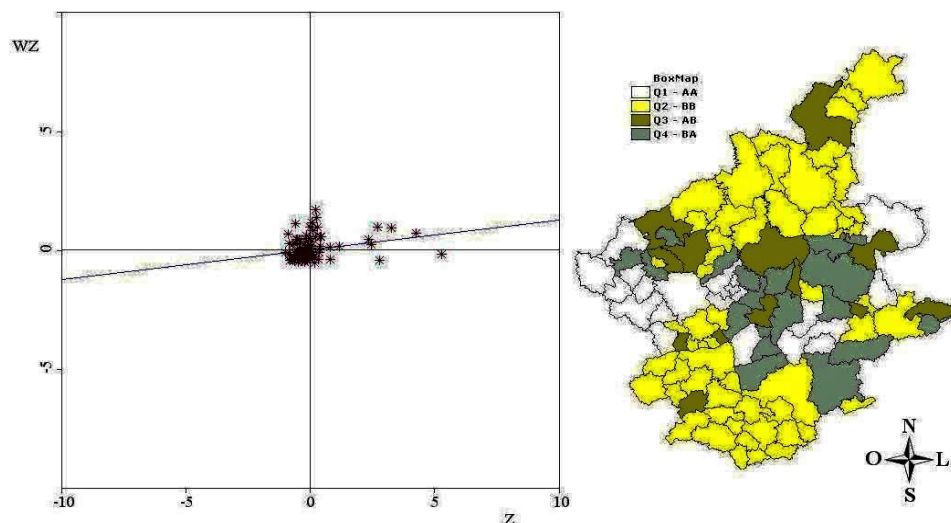


FIGURA 31 À esquerda, diagrama de espalhamento de Moran da variável IVAA. À direita, *Box Map* da variável IVAA.

Fonte: Construído por elaboração própria com base nos softwares GeoDA (Anselin, 2004) e TerraView (INPE, 2007).

Na Figura 31, teve-se que a maior parte dos municípios da mesorregião Metropolitana de Belo Horizonte localizou-se nos quadrantes Alto-Alto (19,0%) e Baixo-Baixo (50,5%), comprovando-se a existência de autocorrelação espacial positiva: valores do índice VAA dos municípios são similares aos verificados por seus vizinhos. Percebeu-se, pelo Box Map, que houve similaridade entre os municípios de valores altos do índice VAA, localizados no oeste e leste da mesorregião, e houve similaridade entre os municípios com valores baixos do índice VAA, localizados no norte e sul da mesorregião.

Na Figura 32, a maior parte dos municípios da mesorregião Sul/Sudeste de Minas localizou-se nos quadrantes Alto-Alto (30,1%) e Baixo-Baixo (41,8%), também comprovando-se a existência de autocorrelação espacial positiva: valores do índice VAA dos municípios são similares aos verificados por seus vizinhos. Pelo Box Map, houve similaridade entre vários municípios de valores altos do índice VAA, localizados em sua maioria no norte, centro, oeste e noroeste da mesorre-

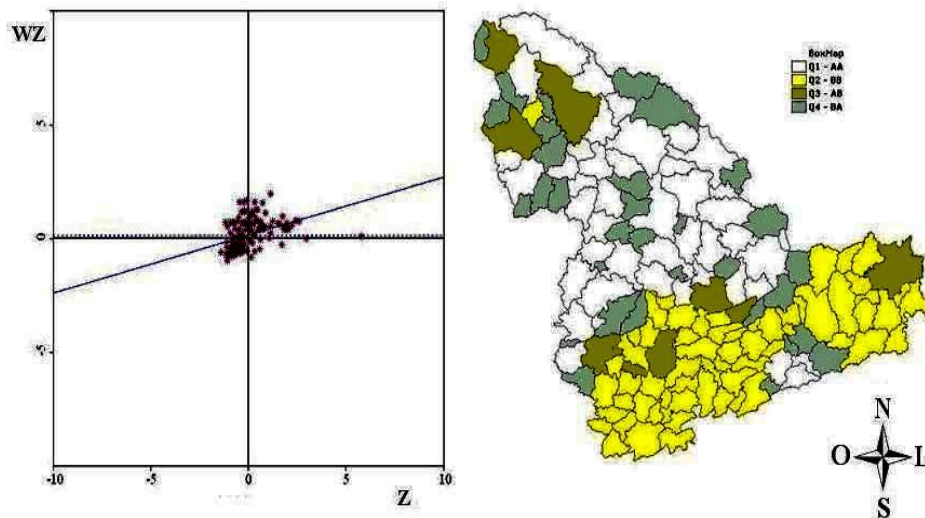


FIGURA 32 À esquerda, diagrama de espalhamento de Moran da variável IVAA. À direita, *Box Map* da variável IVAA.

Fonte: Construído por elaboração própria com base nos softwares GeoDA (Anselin, 2004) e TerraView (INPE, 2007).

ção, e houve similaridade entre vários municípios de valores baixos do índice VAA, localizados de sul a leste, passando pelos municípios do sudeste da mesorregião.

Para examinar padrões numa escala de maior detalhes, sobre a variável IVAA, utilizou-se o índice de Moran local e o *Moran Map* para cada mesorregião.

4.2.5 Índice de Moran local e *Moran Map*

Nas Figura 33 e 34, mostram-se mapas contendo os valores positivos e negativos do índices de Moran local, e mapas com valores simulados, utilizando-se o teste da permutação aleatória (9999 permutações), para o índice de Moran local, apresentado pelo *Moran Map*, para cada mesorregião.

Na Figura 33, no mapa da esquerda, observou-se que a maior parte dos municípios (69,5%) apresentaram valores positivos para o índice de Moran Local, indicando similaridade (alto-alto ou baixo-baixo) da variável IVAA, de oeste a sul, passando pelos municípios do sudoeste, vários municípios do norte e alguns

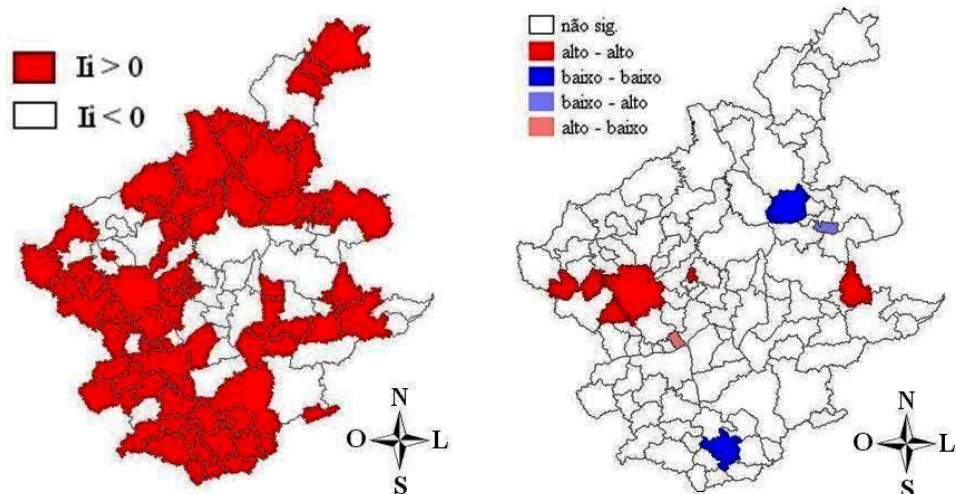


FIGURA 33 Mapa do índice de Moran local da variável IVAA na mesorregião Metropolitana de Belo Horizonte - MG, 2004. À esquerda, valores positivos e negativos do índice de Moran local. À direita, *Moran Map*.

Fonte: Construído por elaboração própria com base nos softwares TerraView (INPE, 2007) e GeoDA (Anselin, 2004).

municípios do lado leste da mesorregião Metropolitana de Belo Horizonte. Enquanto, a outra parte dos municípios (30,5%) apresentaram valores negativos para o índice de Moran Local, indicando dissimilaridade (alto-baixo ou baixo-alto) da variável IVAA, localizados em vários lugares da mesorregião, onde a maioria dos municípios estão localizados ao leste. E, no mapa da direita, está representada a simulação realizada para os índices de Moran Locais pelo teste da permutação aleatória. Constatou-se similaridade entre seis municípios com valores altos do índice do VAA, dos quais cinco (Esmeraldas, Florestal, Onça de Pitangui, Piedade dos Gerais e São José da Varginha) estão localizados no oeste e um (Nova Era) localizado a leste da mesorregião. Sendo que, a maior parte dos municípios não apresentou valores significativos.

Na Figura 34, observou-se, no mapa da esquerda, que a maioria dos municípios (71,9%) apresentaram valores positivos para o índice de Moran Local, indicando similaridade (alto-alto ou baixo-baixo) da variável IVAA, localizados em vários lu-

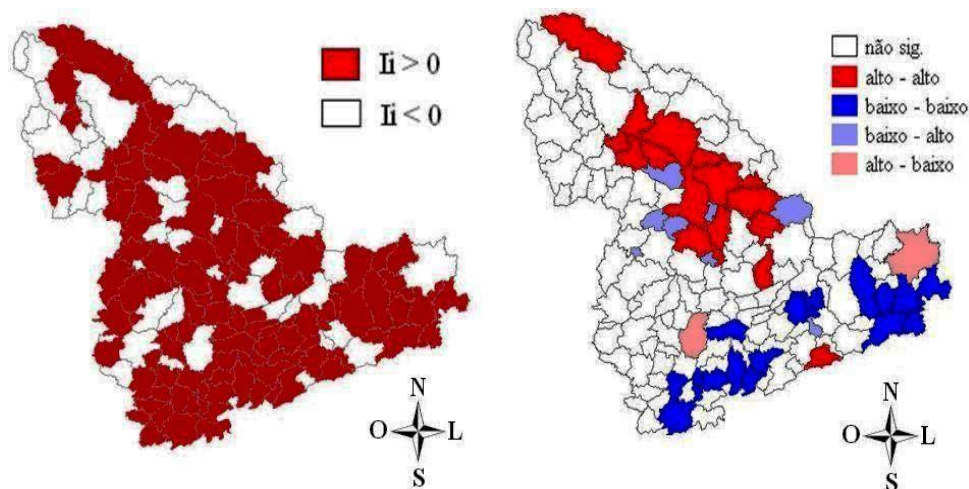


FIGURA 34 Mapa do índice de Moran local da variável IVAA na mesorregião Sul/Sudeste de Minas - MG, 2004. À esquerda, valores positivos e negativos do índice de Moran local. À direita, *Moran Map*.
 Fonte: Construído por elaboração própria com base nos softwares TerraView (INPE, 2007) e GeoDA (Anselin, 2004).

gares na mesorregião Sul/Sudeste de Minas. Enquanto o restante dos municípios (28,1%) apresentaram valores negativos para o índice de Moran Local, indicando dissimilaridade (alto-baixo ou baixo-alto) da variável IVAA, também localizados em vários lugares na mesorregião. E, no mapa da direita, está representada a simulação realizada para os índices de Moran Locais pelo teste da permutação aleatória. Percebeu-se similaridade entre dezesseis municípios de valores altos do índice do VAA, dos quais quinze (Alfenas, Alpinópolis, Bom Jesus da Penha, Conceição da Aparecida, Campanha, Campo do Meio, Campo Gerais, Carmo do Rio Claro, Cordislândia, Delfinópolis, Nova Resende, Machado, Paraguaçu, Santana da Vargem e Varginha), localizados no centro e norte da mesorregião, e um (Passa Quatro) localizado no sudeste da mesorregião. Constatou-se também similaridade entre dezesseis municípios de valores baixos do índice do VAA (Aiuruoca, Bocaína de Minas, Bom Jardim de Minas, Brasópolis, Camanducaia, Cambuí, Consolação, Carmo de Minas, Carvalhos, Itajubá, Liberdade, Piranguçu, Paraisópolis, Passa-Vinte, Santa Rita do Sapucaí e Soledade de Minas), localizados de sul a leste,

passando pelo sudeste da mesorregião. Sendo que, a maioria dos municípios não apresentou valores significativos.

5 CONCLUSÕES

Os resultados obtidos pelo presente estudo permitem concluir que:

5.1 Para a variável IPIB

5.1.1 Quanto ao padrão espacial

Em ambas as mesorregiões, o uso da média móvel local para a variável IPIB, trouxe benefício na redução da variância amostral. E, também, foi eficiente na descrição espacial e identificação de *clusters*.

5.1.2 Quanto à autocorrelação espacial

Em ambas as mesorregiões, constatou-se autocorrelação espacial positiva em termos globais, pelo índice de Moran global e o teste de permutação aleatória, indicando que existe similaridade entre os municípios para a variável IPIB dentro de cada mesorregião. E constatou-se também autocorrelação espacial em termos locais, pelo índice de Moran local e o *Moran Map*, indicando também similaridade entre os municípios para a variável IPIB dentro de cada mesorregião.

5.2 Para a variável IVAA

5.2.1 Quanto ao padrão espacial

Em ambas as mesorregiões, o cálculo da média móvel local para a variável IVAA, trouxe benefício na redução da variância amostral. E, também, foi eficiente na descrição espacial e identificação de *clusters*.

5.2.2 Quanto à autocorrelação espacial

Em ambas as mesorregiões, apresentou-se autocorrelação espacial positiva em termos globais, pelo índice de Moran global e o teste de permutação aleatória, indicando que existe similaridade entre os municípios para a variável IVAA dentro de cada mesorregião. E constatou-se também autocorrelação espacial em termos locais, pelo índice de Moran local e o *Moran Map*, indicando também similaridade entre os municípios para a variável IVAA, dentro de cada mesorregião.

5.3 Trabalhos futuros

Estudos posteriores podem ser aplicados em ambas as mesorregiões para as variáveis IPIB e IVAA, por meio de modelos com efeitos espaciais globais tais como:

- (i) Modelo espacial autoregressivo misto: considera a dependência espacial, por meio da adição ao modelo de regressão de um novo termo na forma de uma relação espacial para a variável dependente. Formalmente, isso é expresso como:

$$Y = \rho WY + X\beta + \varepsilon,$$

em que W é a matriz de proximidade espacial, e o produto WY expressa a dependência espacial em Y e ρ é o coeficiente espacial autoregressivo. A hipótese nula para a não existência de autocorrelação é que $\rho = 0$. A idéia básica nesse modelo é incorporar a autocorrelação espacial como componente do modelo.

- (ii) Modelo do erro espacial: considera que os efeitos espaciais são um ruído, ou seja, fator que precisa ser removido. Nesse caso, os efeitos da autocorrelação espacial são associados ao termo de erro ε e o modelo pode ser expresso por:

$$Y = X\beta + \varepsilon,$$

sendo que

$$\varepsilon = \lambda W + \xi,$$

em que λW é a componente do erro com efeitos espaciais, λ é o coeficiente autoregressivo e ξ é a componente do erro com variância constante e não correlacionada. A hipótese nula para a não existência de autocorrelação é que $\lambda = 0$, ou seja, o termo de erro não é espacialmente correlacionado.

REFERÊNCIAS BIBLIOGRÁFICAS

ANDRIOTTI, J. L. S. **Fundamentos de estatística e geoestatística**. São Leopoldo: UNISINOS, 2004. 168 p.

ANSELIN, L. **Exploring spatial data with *GeoDa*TM**: a workbook. Champaign: University of Illinois, 2005. 244 p. Disponível em: <<http://www.csiss.org/clearinghouse/GeoDa/>>. Acesso em: 5 fev. 2009.

ANSELIN, L. ***GeoDa*TM software for geodata analysis**: version GeoDa 0.95i-(Beta). Champaign: University of Illinois, 2004. Disponível em: <<http://www.csiss.org/clearinghouse/GeoDa/>>. Acesso em: 5 fev. 2009.

ANSELIN, L. Local indicators of spatial association. **Geographical Analysis**, Columbus, v. 27, n. 1, p. 93-115, Jan. 1995.

ANSELIN, L. The Moran scatterplot as an ESDA tool to assess local instability in spatial association. In: FISHER, M.; SCHOLTEN, H. J.; UNWIN, D. (Ed.). **Spatial analytical perspectives on GIS**. London: Taylor & Francis, 1996. p.111-125.

ASSUNÇÃO, R. M. **Estatística espacial com aplicações em epidemiologia, economia, sociologia**. São Carlos: Associação Brasileira de Estatística, 2001. 131p. Disponível em: <<http://www.est.ufmg.br/assuncao>>. Acesso em: 20 nov. 2008.

BAILEY, T. C.; GATRELL, A. C. **Interactive spatial data analysis**. Essex: Longman Scientific, 1995. 413 p.

CÂMARA, G.; MONTEIRO, A. M. V.; CARVALHO, M. S.; FUKS, S. D. **Análise espacial de dados geográficos**. São José dos Campos: INPE, 2001. 209p. Disponível em: <<http://www.dpi.inpe.br/gilberto/livro>>. Acesso em: 20 nov. 2008.

CARVALHO, D. M.; SILVA M. A. S. Uma análise exploratória da distribuição espacial do PIB per capita em Sergipe (1999-2004). **Revista da Fapese**, Sergipe, v.4, n.2, p.149-160, jul./dez. 2008.

CLIFF, A. D.; ORD, J. K. **Spatial processes: models and applications**. London: Pion, 1981. 266 p.

DRAPER, N. R.; SMITH, H. **Applied regression analysis**. 3. ed. New York: J. Wiley, 1998. 706 p.

ECONOMIC AND SOCIAL RESEARCH INSTITUTE. Redlands, 2009. Disponível em: <<http://www.esri.com/>>. Acesso em: 9 dez. 2009.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Base cartográfica digital**. Rio de Janeiro, 2009. Disponível em: <<http://www.mapas.ibge.gov.br/>>. Acesso em: 20 maio 2009.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS. **TerraView**. Rio de Janeiro: Divisão de Processamento, 2007. Disponível em: <<http://www.dpi.inpe.br/terraview/index.php>>. Acesso em: 20 jun. 2009.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS. **Tutorial TerraView**. Rio de Janeiro: Divisão de Processamento, 2009. Disponível em: <http://www.dpi.inpe.br/terraview/php/docs.php?body=Tutorial_i>. Acesso em: 20 jun. 2009.

KREMPI, A. P. **Explorando recursos de estatística espacial para análise da acessibilidade na cidade de Bauru**. 2004. 82p. Dissertação (Mestrado em Engenharia Civil) - Universidade de São Paulo, São Paulo.

LANDIM, P. M. B. **Análise estatística de dados geológicos**. São Paulo: UNESP, 1998. 226 p.

MAGALHÃES, M. N. **Probabilidade e variáveis aleatórias**. 2. ed. São Paulo: Instituto de Matemática e Estatística da USP, 2006. 411 p.

MORAN, P. A. F. Notes on continuous stochastic phenomena. **Biometrika**, London, v. 37, n.1, p. 17-23, Jun. 1950.

OLIVEIRA, F. J. P.; NASCIMENTO, L. F. C. Análise espacial da distribuição de doenças diarréicas nos municípios do Vale do Paraíba Paulista. **Ambi-Água**, Taubaté, v.4, n.2, p. 115-123, ago. 2009.

SANTOS, S. M.; BARCELLOS, C. (Org.). **Abordagens espaciais na saúde pública**. Brasília: Ministério da Saúde, 2006. v. 1.

SANTOS, S. M.; SOUZA, W. V. (Org.). **Introdução à estatística espacial para a saúde pública**. Brasília: Ministério da Saúde, 2007. v. 3.

SCOLFORO, J. R. S.; OLIVEIRA, A. D.; CARVALHO, L. M. T. **Zoneamento ecológico-econômico do Estado de Minas Gerais**: componentes geofísico e biótico. Lavras: UFLA, 2008a. 161 p.

SCOLFORO, J. R. S.; OLIVEIRA, A. D.; CARVALHO, L. M. T. **Zoneamento ecológico-econômico do Estado de Minas Gerais**: componente sócioeconômico. Lavras: UFLA, 2008b. 195 p.

SCOLFORO, J. R. S.; OLIVEIRA, A. D.; CARVALHO, L. M. T. **Zoneamento ecológico-econômico do Estado de Minas Gerais**: zoneamento e cenários exploratórios. Lavras: UFLA, 2008c. 136 p.

ANEXOS

ANEXO A Relação dos nomes dos 105 municípios da mesorregião Metropolitana de Belo Horizonte

1. Alvinópolis
2. Alvorada de Minas
3. Araçáí
4. Baldim
5. Barão de Cocais
6. Bela Vista de Minas
7. Belo Horizonte
8. Belo Vale
9. Betim
10. Bom Jesus do Amparo
11. Bonfim
12. Brumadinho
13. Cachoeira da Prata
14. Caetanópolis
15. Caeté
16. Capim Branco
17. Casa Grande
18. Catas Altas
19. Catas Altas da Noruega

20. Conceição do Mato Dentro
21. Confins
22. Congonhas
23. Congonhas do Norte
24. Conselheiro Lafaiete
25. Contagem
26. Cordisburgo
27. Cristiano Ottoni
28. Crucilândia
29. Desterro de Entre Rios
30. Diogo de Vasconcelos
31. Dionísio
32. Dom Joaquim
33. Entre Rios de Minas
34. Esmeraldas
35. Ferros
36. Florestal
37. Fortuna de Minas
38. Funilândia
39. Ibirité
40. Igarapé
41. Inhaúma

42. Itabira
43. Itabirito
44. Itaguara
45. Itambé do Mato Dentro
46. Itatiaiuçu
47. Itaverava
48. Jaboticatubas
49. Jeceaba
50. Jequitibá
51. João Monlevade
52. Juatuba
53. Lagoa Santa
54. Maravilhas
55. Mariana
56. Mário Campos
57. Mateus Leme
58. Matozinhos
59. Moeda
60. Morro do Pilar
61. Nova Era
62. Nova Lima
63. Nova União

64. Onça de Pitangui
65. Ouro Branco
66. Ouro Preto
67. Papagaios
68. Pará de Minas
69. Paraopeba
70. Passabém
71. Pedro Leopoldo
72. Pequi
73. Piedade dos Gerais
74. Pitangui
75. Prudente de Moraes
76. Queluzito
77. Raposos
78. Ribeirão das Neves
79. Rio Acima
80. Rio Manso
81. Rio Piracicaba
82. Rio Vermelho
83. Sabará
84. Santa Bárbara
85. Santa Luzia

86. Santa Maria de Itabira
87. Santana de Pirapama
88. Santana do Riacho
89. Santana dos Montes
90. Santo Antônio do Itambé
91. Santo Antônio do Rio Abaixo
92. São Brás do Suaçuí
93. São Domingos do Prata
94. São Gonçalo do Rio Abaixo
95. São Joaquim de Bicas
96. São José da Lapa
97. São José da Varginha
98. São José do Goiabal
99. São Sebastião do Rio Preto
100. Sarzedo
101. Serra Azul de Minas
102. Serro
103. Sete Lagoas
104. Taquaraçu de Minas
105. Vespasiano

ANEXO B Relação dos nomes dos 146 municípios da mesorregião Sul/Sudeste de Minas

1. Aiuruoca
2. Alagoa
3. Albertina
4. Alfenas
5. Alpinópolis
6. Alterosa
7. Andradas
8. Andrelândia
9. Arantina
10. Arceburgo
11. Areado
12. Baependi
13. Bandeira do Sul
14. Boa Esperança
15. Bocaina de Minas
16. Bom Jardim de Minas
17. Bom Jesus da Penha
18. Bom Repouso
19. Borda da Mata
20. Botelhos

21. Brasópolis
22. Bueno Brandão
23. Cabo Verde
24. Cachoeira de Minas
25. Caldas
26. Camanducaia
27. Cambuí
28. Cambuquira
29. Campanha
30. Campestre
31. Campo do Meio
32. Campos Gerais
33. Capetinga
34. Capitólio
35. Careaçú
36. Carmo da Cachoeira
37. Carmo de Minas
38. Carmo do Rio Claro
39. Carvalhópolis
40. Carvalhos
41. Cássia
42. Caxambu

43. Claraval
44. Conceição da Aparecida
45. Conceição das Pedras
46. Conceição do Rio Verde
47. Conceição dos Ouros
48. Congonhal
49. Consolação
50. Coqueiral
51. Cordislândia
52. Córrego do Bom Jesus
53. Cristina
54. Cruzília
55. Delfim Moreira
56. Delfinópolis
57. Divisa Nova
58. Dom Viçoso
59. Elói Mendes
60. Espírito Santo do Dourado
61. Estiva
62. Extrema
63. Fama
64. Fortaleza de Minas

65. Gonçalves
66. Guapé
67. Guaranésia
68. Guaxupé
69. Heliódora
70. Ibiraci
71. Ibitiúra de Minas
72. Ilicínea
73. Inconfidentes
74. Ipuiúna
75. Itajubá
76. Itamogi
77. Itamonte
78. Itanhandu
79. Itapeva
80. Itaú de Minas
81. Jacuí
82. Jacutinga
83. Jesuânia
84. Juruaia
85. Lambari
86. Liberdade

87. Machado
88. Maria da Fé
89. Marmelópolis
90. Minduri
91. Monsenhor Paulo
92. Monte Belo
93. Monte Santo de Minas
94. Monte Sião
95. Munhoz
96. Muzambinho
97. Natércia
98. Nova Resende
99. Olímpio Noronha
100. Ouro Fino
101. Paraguaçu
102. Paraisópolis
103. Passa Quatro
104. Passa-Vinte
105. Passos
106. Pedralva
107. Piranguçu
108. Piranguinho

109. Poço Fundo
110. Poços de Caldas
111. Pouso Alegre
112. Pouso Alto
113. Pratápolis
114. Santa Rita de Caldas
115. Santa Rita do Sapucaí
116. Santana da Vargem
117. São Bento Abade
118. São Gonçalo do Sapucaí
119. São João Batista do Glória
120. São João da Mata
121. São José da Barra
122. São José do Alegre
123. São Lourenço
124. São Pedro da União
125. São Sebastião da Bela Vista
126. São Sebastião do Paraíso
127. São Sebastião do Rio Verde
128. São Thomé das Letras
129. São Tomás de Aquino
130. São Vicente de Minas

131. Sapucaí-Mirim
132. Senador Amaral
133. Senador José Bento
134. Seritinga
135. Serrania
136. Serranos
137. Silvianópolis
138. Soledade de Minas
139. Tocos do Moji
140. Toledo
141. Três Corações
142. Três Pontas
143. Turvolândia
144. Varginha
145. Virgínia
146. Wenceslau Braz