

UNIVERSIDADE DE SÃO PAULO  
ESCOLA DE ENGENHARIA DE SÃO CARLOS  
DEPARTAMENTO DE TRANSPORTES

**EFEITOS DA DEPENDÊNCIA ESPACIAL EM MODELOS  
DE PREVISÃO DE DEMANDA POR TRANSPORTE**

**Arq<sup>a</sup> SIMONE BECKER LOPES**

**ORIENTADOR: Prof. Associado Antônio Nélon Rodrigues da Silva**

Dissertação apresentada ao Departamento de Transportes, da Escola de Engenharia de São Carlos da Universidade de São Paulo, como parte dos requisitos para a obtenção do título de Mestre em Engenharia Civil – Área de Concentração: Planejamento e Operação de Sistema de Transportes.

São Carlos

2005

Ficha catalográfica preparada pela Seção de Tratamento  
da Informação do Serviço de Biblioteca – EESC/USP

L864e           Lopes, Simone Becker  
                  Efeitos da dependência espacial em modelos de  
previsão de demanda por transporte / Simone Becker  
Lopes. -- São Carlos, 2005.

Dissertação (Mestrado) -- Escola de Engenharia de  
São Carlos-Universidade de São Paulo, 2005.

Área: Planejamento e Operação de Sistema de  
Transportes

Orientador: Prof. Assoc. Antônio Néelson Rodrigues  
da Silva.

1. Estatística espacial. 2. Análise de demanda por  
transportes. 3. Autocorrelação espacial. 4. SIG.  
I. Título.

## DEDICATÓRIA

---

Dedico este trabalho às minhas filhas Julia e Bruna e ao Léo. Com todo meu amor, minha profunda gratidão pelo estímulo, apoio, ajuda e compreensão nos momentos difíceis.

## AGRADECIMENTOS

---

A Deus, por ter me dado forças nas horas mais difíceis.

A toda minha família, em especial aos meus pais Miguel e Loiva (*in memoriam*), e à minha irmã Cristina, minha grande incentivadora.

Ao Prof. Associado Antônio Néelson Rodrigues da Silva, pela orientação, dedicação e paciência.

A todos os colegas da EPTC – POA, em especial ao Ernani Fagundes, Humberto Kasper, Fabiane Moscarelli, Gilmar Cardoso, Francisco Horbe e Marco Aurélio Santa Helena, pela amizade, incentivo, apoio, fazendo com que fosse possível meu afastamento temporário e sempre atendendo com prestezas minhas solicitações.

A todos os professores do Departamento de Transportes da EESC-USP pelos conhecimentos compartilhados e pela amizade.

A todo pessoal da Administração do Departamento de Transportes, em especial à Magali, à Suely, à Heloisa e à Beth, obrigada pelo apoio, carinho e amizade.

À Prof<sup>a</sup> Nair Cristina Brondino, da UNESP de Bauru, por ter participado de muitas etapas de análise no decorrer deste trabalho e por ter compartilhado seu conhecimentos de estatística, sempre com paciência e dedicação.

Ao LASTRAN-UFRGS, em especial aos Professores Helena Beatriz Betella Cybis, Luis Antonio Lindau e Orlando Strambi (Poli-USP), e aos Engenheiros Davi Ribeiro e Christine Tessele Nodari pela amizade, incentivo e presteza no atendimento das minhas solicitações, sempre que eu precisei de dados e material de Porto Alegre.

À Nívea Peixoto, da METROPLAN-POA, por ter compartilhado seu conhecimento e experiência na análise das pesquisas de Porto Alegre.

A todos os colegas do Departamento de Transportes, em especial: Márcia, Vitor, Marcela, Cláudio, Deise, Ana Paula Krempi, Vanessa, Alexandra, Renata, Fabiana, Cira, Fábio, Fred, Marcos, Eduardo, Karênina, Márcia Lika, Josiane, Adriane, Ana Paula, Cida, Dani, Rogério, Lucas e Andréa Júlia, que compartilharam momentos de trabalho, estudo, alegria, diversão, esporte e, principalmente, amizade, o que foi muito importante durante estes dois anos.

Ao CNPq pela bolsa de estudos concedida.

A todos aqueles que colaboraram para a conclusão desta dissertação, mesmo que indiretamente.

# SUMÁRIO

---

LISTA DE FIGURAS	v
LISTA DE TABELAS	vii
RESUMO	x
ABSTRACT	xi
1. INTRODUÇÃO	1
1.1. CARACTERIZAÇÃO DO PROBLEMA	1
1.2. JUSTIFICATIVA	3
1.3. OBJETIVOS	3
1.4. ESTRUTURA DO TRABALHO	4
2. PLANEJAMENTO DE TRANSPORTES	5
2.1. INTRODUÇÃO	5
2.2. MODELOS DE PREVISÃO DE DEMANDA POR TRANSPORTES	8
2.3. MODELO DE QUATRO ETAPAS	9
2.4. SELEÇÃO DE VARIÁVEIS PARA ANÁLISES DE GERAÇÃO DE VIAGENS	12
2.5. CRÍTICAS AOS MODELOS DE DEMANDA TRADICIONAIS	14
2.6. CONSIDERAÇÕES FINAIS	15
3. ANÁLISE ESPACIAL	16
3.1. INTRODUÇÃO	16
3.2. TIPOS DE DADOS EM ANÁLISE ESPACIAL	19
3.3. ANÁLISE ESPACIAL DE ÁREAS	20
3.4. ANÁLISE ESPACIAL DE SUPERFÍCIES	21
3.5. FERRAMENTAS DE ANÁLISE ESPACIAL	21

3.6. ESTATÍSTICA ESPACIAL	23
3.7. DEPENDÊNCIA ESPACIAL E AUTOCORRELAÇÃO ESPACIAL	23
3.8. ANÁLISE EXPLORATÓRIA DE DADOS ESPACIAIS	24
3.8.1. Matriz de proximidade espacial (W)	25
3.8.2. Média móvel espacial (Wz)	26
3.8.3. Indicadores de dependência espacial: medidas de autocorrelação	27
3.8.4. Visualização da dependência espacial	30
3.9. REGRESSÃO ESPACIAL	35
3.9.1. Modelos com efeitos espaciais globais	38
3.9.2. Modelos com efeitos espaciais locais	39
3.9.3. Diagnóstico de modelos com efeitos espaciais	41
3.10. CONSIDERAÇÕES FINAIS	42
4. MÉTODO	43
4.1. ETAPA 1 - ANÁLISE DOS DADOS DISPONÍVEIS	45
4.2. ETAPA 2 - DEFINIÇÃO DA ÁREA DE ESTUDOS	45
4.3. ETAPA 3 - PREPARAÇÃO DA BASE DE DADOS	45
4.4. ETAPA 4 - ANÁLISE DAS VARIÁVEIS INDEPENDENTES	46
4.5. ETAPA 5 - ANÁLISE DA AUTOCORRELAÇÃO ESPACIAL	48
4.6. ETAPA 6 - ESPECIFICAÇÃO E CALIBRAÇÃO DOS MODELOS TRADICIONAIS	51
4.7. ETAPA 7 - ESPECIFICAÇÃO E CALIBRAÇÃO DE MODELOS ALTERNATIVOS	55
4.8. ETAPA 8 – ESTIMATIVAS FUTURAS	58
4.9. ETAPA 9 - AVALIAÇÃO DOS MODELOS	58
5. APLICAÇÃO DO MÉTODO E ANÁLISE DOS RESULTADOS	59
5.1. ANÁLISE DOS DADOS DISPONÍVEIS	60
5.1.1. Pesquisa O-D de 1974 (EDOM 1974)	61
5.1.2. Pesquisa O-D de 1986 (EDOM 1986)	61
5.1.3. Pesquisa O-D de 1997 (EDOM 1997)	61
5.1.4. Pesquisa O-D de 2003 (EDOM 2003)	62
5.1.5. Síntese das pesquisas	62

5.2. DEFINIÇÃO DA ÁREA DE ESTUDOS	63
5.3. PREPARAÇÃO DA BASE DE DADOS E ANÁLISE PARA COMPATIBILIZAÇÃO DAS ZTS	65
5.3.1. Digitalização e análise para compatibilização das ZTs	66
5.3.2. Migração de dados (TransCAD – SPRING)	67
5.4. ANÁLISE DAS VARIÁVEIS INDEPENDENTES	68
5.4.1. Dados da EDOM de 1974	69
5.4.2. Dados da EDOM de 1986	70
5.4.3. Variáveis explicativas para VPBD citadas na literatura e na EDOM 1974	70
5.4.4. Análise da correlação entre as variáveis (Fase 2)	71
5.4.5. Definição de variáveis derivadas (taxas)	72
5.5. ANÁLISE DA AUTOCORRELAÇÃO ESPACIAL	74
5.5.1. Análises da Fase 1	74
5.5.2. Análises da Fase 2	77
5.5.3. Análises da Fase 3	82
5.6. ESPECIFICAÇÃO, CALIBRAÇÃO E DIAGNÓSTICO DOS MODELOS TRADICIONAIS	89
5.6.1. Modelos tradicionais da Fase 1	90
5.6.2. Modelos tradicionais da Fase 2	93
5.6.3. Modelos tradicionais da Fase 3	97
5.7. ESPECIFICAÇÃO, CALIBRAÇÃO E DIAGNÓSTICO DOS MODELOS QUE CONSIDERAM A DEPENDÊNCIA ESPACIAL	100
5.7.1. Modelos alternativos da Fase 1	100
5.7.2. Modelos alternativos da Fase 2	102
5.7.3. Modelos alternativos da Fase 3	109
5.8. ESTIMATIVAS PARA O ANO META – VALIDAÇÃO DOS MODELOS	118
5.8.1. Estimativas para o ano meta - Fase 1	118
5.8.2. Estimativas para o ano meta - Fase 2	119
5.8.3. Estimativas para o ano meta - Fase 3	120
5.9. ANÁLISE COMPARATIVA DOS RESULTADOS	121
6. CONCLUSÕES E SUGESTÕES PARA TRABALHOS FUTUROS	128
6.1. CONCLUSÕES	128

6.2. SUGESTÕES PARA TRABALHOS FUTUROS	131
---------------------------------------	-----

REFERÊNCIAS BIBLIOGRÁFICAS	133
----------------------------	-----

ANEXOS	138
--------	-----

ANEXO A

ANEXO B

ANEXO C

ANEXO D

ANEXO E

ANEXO F

## LISTA DE FIGURAS

---

Figura 2.1: Alternativas para quebrar o círculo vicioso do automóvel e do transporte público por ônibus - adaptado de Ortúzar e Willumsen (1994)	7
Figura 2.2: Atração e produção de viagens BD e BND – adaptado de Ortúzar e Willumsen (1994)	10
Figura 3.1: Inter-relação entre os quatro grupos de ferramentas de análise espacial – adaptado de Anselin (1992) e Teixeira (2003)	22
Figura 3.2: Matriz de proximidade espacial de primeira ordem, normalizada pela linha (Câmara <i>et al.</i> , 2002a)	26
Figura 3.3: Gráfico de espalhamento de Moran	31
Figura 3.4 – Diagrama de espalhamento de Moran para o índice de exclusão/inclusão social de São Paulo, Censo de 1991. Fonte: Câmara <i>et al.</i> (2002a)	32
Figura 3.5: <i>Box Map</i> da variável <i>porcentagem de viagens de ônibus</i> , para a cidade de Bauru. Fonte: Krempi (2004)	33
Figura 3.6: <i>LISA Map</i> da variável <i>porcentagem de viagens de ônibus</i> , para a cidade de Bauru. Fonte: Krempi (2004)	34
Figura 3.7: <i>Moran Map</i> da variável <i>porcentagem de viagens por ônibus</i> , para a cidade de Bauru. Fonte: Krempi (2004)	34
Figura 3.8: Gráfico de barras do Mapa de Exclusão/Inclusão Social de São Paulo, 2000 Fonte: <a href="http://www.dpi.inpe.br/geopro/exclusao/mapas.html">http://www.dpi.inpe.br/geopro/exclusao/mapas.html</a>	35
Figura 4.1: Esquema do método aplicado, com destaque para as etapas de uso de técnicas da AEDE (Análise Exploratória de Dados Espaciais)	44
Figura 5.1: Mapa de Porto Alegre - ZTs da EDOM 1974	64
Figura 5.2: Mapa de Porto Alegre - ZTs da EDOM 1986	65

Figura 5.3: Mapa de Porto Alegre com a sobreposição das ZTs de 1974 e 1986, com destaque para as áreas iguais	67
Figura 5.4: <i>Box Map</i> para as variáveis <i>DENSI</i> e <i>REND</i> A em 1974	75
Figura 5.5: <i>Box Map</i> para as variáveis <i>DENSI</i> e <i>REND</i> A em 1986	76
Figura 5.6: <i>Box Map</i> e <i>Moran Map</i> para a variável <i>DENSI</i> em 1974	78
Figura 5.7: <i>Box Map</i> e <i>Moran Map</i> para a variável <i>REND</i> A em 1974	79
Figura 5.8: <i>Box Map</i> e <i>Moran Map</i> para a variável <i>DENSI</i> em 1986	80
Figura 5.9: <i>Box Map</i> e <i>Moran Map</i> para a variável <i>REND</i> A em 1986	82
Figura 5.10: <i>Box Map</i> e <i>Moran Map</i> para a variável <i>VPBD</i> em 1974	84
Figura 5.11: Percentual de ZTs com quadrantes coincidentes entre as variáveis socioeconômicas e <i>VPBD</i> : análise dos <i>Box Map</i> , dados de 1974	88
Figura 5.12: Percentual de ZTs com quadrantes coincidentes entre as variáveis socioeconômicas e <i>VPBD</i> - análise dos <i>Moran Map</i> , dados de 1974	89
Figura 5.13: Rotina pré-estabelecida para escolha das variáveis na especificação, calibração e seleção dos modelos tradicionais da Fase 3	98
Figura 5.14: Análise comparativa dos resultados dos <i>modelos alternativos</i> da Fase 1 tomando-se como referência os <i>modelos tradicionais</i> 3T e 4T	102
Figura 5.15: Análise comparativa dos resultados dos <i>modelos alternativos</i> da Fase 2 tomando-se como referência o <i>modelos tradicional</i> 5T <sub>p</sub>	108
Figura 5.16: Análise comparativa dos resultados dos <i>modelos alternativos</i> da Fase 3 tomando-se como referência o <i>modelos tradicional</i> 5T <sub>p</sub>	115
Figura 5.17: Análise comparativa dos melhores <i>modelos alternativos</i> de cada fase, tomando-se como referência o <i>modelo tradicional</i> correspondente	122
Figura 5.18: <i>Moran Map</i> para os resíduos das estimativas para o ano meta com o modelo tradicional 5T <sub>p</sub> e o modelo alternativo 5AGL <sub>p</sub>	124
Figura 5.19: Análise comparativa dos dois melhores <i>modelos alternativos</i> , tomando-se como referência o <i>modelo tradicional</i> 5T <sub>p</sub>	126
Figura 5.20: <i>Moran Map</i> para os resíduos das estimativas para o ano base, com o modelo tradicional 5T <sub>p</sub> e com o modelo alternativo F3AGL <sub>p</sub>	127

## LISTA DE TABELAS

---

Tabela 4.1: Resumo das características das etapas de análise espacial e <i>software</i> que podem ser aplicados em cada uma delas	49
Tabela 5.1: Principais características das pesquisas O-D realizadas em Porto Alegre - RS e Região Metropolitana	63
Tabela 5.2: <i>Matriz de correlação</i> entre as variáveis socioeconômicas do ano base e a variável dependente <i>VPBD</i>	72
Tabela 5.3: <i>Matriz de correlação</i> entre as variáveis socioeconômicas do ano base e a variável dependente <i>VPBD</i> , incluindo as variáveis taxas	73
Tabela 5.4: <i>Índices de correlação</i> entre as variáveis independentes candidatas e as variáveis dependentes em forma de taxas, para os dados do ano base	74
Tabela 5.5: <i>Índices de autocorrelação espacial (I de Moran Global)</i> das variáveis socioeconômicas e da variável dependente <i>VPBD</i> para os dados do ano base (1974) e do ano meta (1986)	83
Tabela 5.6: Resumo dos quadrantes verificados nas análises do <i>Box Map</i> e do <i>Moran Map</i> para a variável dependente <i>VPBD</i> , para os dados de 1974	85
Tabela 5.7: Análises do <i>Box Map</i> e do <i>Moran Map</i> para as variáveis que representam totais de contagens e relação com os quadrantes da <i>VPBD</i> (EDOM 1974)	86
Tabela 5.8: Análises do <i>Box Map</i> e do <i>Moran Map</i> para as variáveis que representam taxas derivadas e relação com os quadrantes da <i>VPBD</i> (EDOM 1974)	87
Tabela 5.9: Valores de dólar e salário mínimo adotados para a conversão da variável <i>RENDA</i> em $RENDA_D$ e $RENDA_{SM}$	90
Tabela 5.10: Resumo dos modelos tradicionais especificados na Fase 1	91

Tabela 5.11: Resumo dos modelos tradicionais especificados na Fase 1, excluindo-se as ZTs que representavam <i>outliers</i>	92
Tabela 5.12: Resumo dos 8 modelos tradicionais inicialmente selecionados na Fase 2 e análise da significância das variáveis através dos valores de <i>T stat</i>	94
Tabela 5.13: Resumo das características dos modelos tradicionais 5T e 6T	94
Tabela 5.14: Resumo das características dos modelos tradicionais selecionados na Fase 2 e resultados obtidos	95
Tabela 5.15: Resumo das características do modelo 5T <sub>p</sub> e resultados obtidos para o ano base	96
Tabela 5.16: Resumo dos modelos analisados através da rotina pré-estabelecida	99
Tabela 5.17: Resumo dos <i>modelos alternativos</i> especificados na Fase 1, excluindo-se as ZTs que representavam <i>outliers</i>	101
Tabela 5.18: Resumo das características do <i>modelo alternativo</i> 5AG <sub>p1</sub>	103
Tabela 5.19: Resumo das características do <i>modelo alternativo</i> LISA inicial (5AL <sub>p1</sub> ) e do <i>modelo alternativo</i> LISA válido selecionado na Fase 2 (5AL <sub>p</sub> )	105
Tabela 5.20 – Resumo das características e resultados obtidos com o modelo 5AL <sub>p</sub>	105
Tabela 5.21: Resumo das características do <i>modelo alternativo</i> misto inicial (5AGL <sub>p1</sub> ) e do <i>modelo alternativo misto</i> válido selecionado na Fase 2 (5AGL <sub>p</sub> )	106
Tabela 5.22: Resumo das características e resultados obtidos com o modelo 5AGL <sub>p</sub>	107
Tabela 5.23: Resumo dos resultados obtidos com o modelo tradicional e com os modelos alternativos analisados na Fase 2	107
Tabela 5.24: Resumo das características do <i>modelo alternativo</i> inicial F3AG <sub>p</sub>	110
Tabela 5.25: Resumo das características e resultados obtidos com o modelo F3AG <sub>p</sub>	111
Tabela 5.26: Resumo das características do <i>modelo alternativo</i> inicial F3AL <sub>p1</sub>	112
Tabela 5.27: Resumo das características e resultados obtidos com o modelo F3AL <sub>p</sub>	112
Tabela 5.28: Resumo das características do <i>modelo alternativo</i> inicial F3ALG <sub>p1</sub>	113
Tabela 5.29: Resumo das características e resultados obtidos com o modelo F3AGL <sub>p</sub>	114
Tabela 5.30: Resumo dos resultados obtidos com o <i>modelo tradicional</i> e com os <i>modelos alternativos</i> analisados na Fase 3	115
Tabela 5.31: Resumo e análise das características dos modelos especificados e calibrados através do método <i>stepwise</i>	117
Tabela 5.32: Resumo das características e resultados das estimativas (ano base e ano meta) com os modelos selecionados na Fase 1	119

Tabela 5.33: Resumo das características e resultados das estimativas (ano base e ano meta) com os modelos selecionados na Fase 2	120
Tabela 5.34: Resumo das características e resultados das estimativas (ano base e ano meta) com o modelo alternativo selecionados na Fase 3	121
Tabela 5.35: Modelos ( <i>tradicionais</i> e <i>alternativos</i> ) selecionados em cada fase e resultados das estimativas (ano base e ano meta)	122

## RESUMO

---

A dependência espacial para análise de dados de demanda por transportes, que está entre as principais questões analítico-espaciais consideradas na análise de transportes, constituiu o foco deste trabalho. Ignorar questões de análise espacial pode invalidar os resultados da análise, levar a previsões inadequadas e, conseqüentemente, a um planejamento ineficiente. Em virtude disso, admitiu-se que a introdução de indicadores de dependência espacial na modelagem de demanda por transportes deveria produzir resultados mais precisos e, desta forma, mais confiáveis dos que os obtidos com modelos tradicionais. Neste sentido, o principal objetivo deste trabalho foi comparar a projeção de demanda por transportes, especificamente na fase de previsão de viagens produzidas de base domiciliar, realizada através de modelos convencionais e de modelos alternativos, que introduzem indicadores para medir a dependência espacial. O trabalho é todo desenvolvido em ambiente SIG (Sistemas de Informações Geográficas), através de ferramentas de análise e estatística espacial, assim como ferramentas de planejamento de transportes de um SIG-T (SIG para Transportes). As ferramentas de análises espaciais serviram tanto para produzir os indicadores de dependência espacial (variáveis espaciais) como para avaliar os resultados dos modelos. Aplica-se o método, que avalia a introdução de indicadores globais e locais de dependência espacial nos modelos alternativos, através de um estudo de caso na cidade de Porto Alegre - RS, que tem por base dados de pesquisa de origem e destino (O-D) obtidos através de entrevista domiciliar (EDOM) em dois períodos distintos (1974 e 1986). Estas informações correspondem aos dados necessários do “ano base”, que foram utilizados na calibração dos modelos, e do “ano meta”, que constituíram as informações necessárias para análise dos resultados de estimativas futuras. Conclui-se que a introdução de variáveis espaciais é importante, uma vez que os melhores resultados foram obtidos com modelos alternativos, tanto na etapa de calibração e diagnóstico dos modelos como na etapa de validação (estimativas futuras). No entanto, a dinâmica apresentada pelo desenvolvimento urbano, como é o caso de Porto Alegre, acarreta alterações nas relações entre as diferentes variáveis com o fenômeno estudado, modificando, inclusive, os padrões espaciais. Esta conclusão é dada pelo fato que, o modelo mais ajustado para os dados do ano base não foi o que apresentou os melhores resultados para estimativas futuras. Isto conduz à hipótese, a ser explorada em trabalhos futuros, de que a análise desta dinâmica e o estudo de formas de considerá-la nos modelos de demanda por transportes pode produzir resultados ainda melhores.

Palavras-chave: estatística espacial; análise de demanda por transportes; autocorrelação espacial; SIG.

## ABSTRACT

---

The degree of spatial data dependence, which is among the issues of spatial analysis that should be considered in transportation planning, is the focus of this study. Ignoring this particularity of data can: produce wrong estimates; jeopardize the results of analyses; and, as a consequence, lead to unsuccessful planning. Therefore, the basic assumption of this work was that the inclusion of spatial dependence indicators can produce more accurate and reliable estimates than those obtained with traditional model structures. In order to test this hypothesis, the main objective of this study was to compare demand predictions produced by traditional models with those produced by alternative models that include indicators of spatial dependence. The study was limited to home-based production trip models, which are part of the trip generation phase of the traditional four-step modeling approach. All work was conducted in a GIS (Geographic Information System) environment, making use of spatial statistics and analysis tools, as well as transportation planning tools available in a GIS-T (i.e., a dedicated GIS for Transportation). Spatial analyses tools were used to generate the spatial dependence indicators and to evaluate the results of the application. A case study was carried out in the city of Porto Alegre, which is the capital of the Brazilian State of Rio Grande do Sul, for evaluating the impacts of the addition of global and local indicators of spatial dependence in the models. Two O-D surveys carried out in the years 1974 and 1986 provided the data needed for calibration and validation. The first one was taken as the “base year” and the second one as the “goal year”. The results of the application showed that the performance of the models can be improved in both calibration and validation phases with the insertion of spatial variables. However, the urban growth observed in a very dynamic context, such as in the city studied, may dramatically change the relationships between variables, including their spatial patterns. That aspect was responsible for the fact that the model with the best performance in the calibration phase was not the one producing the most accurate forecasts. It raised the hypothesis, to be explored in future research, that the analysis of those dynamic processes and their consideration into transportation demand models are also needed to improve even further the performance of the models.

Key-words: spatial statistics; analysis of transportation demand; spatial autocorrelation; GIS.

# **1. INTRODUÇÃO**

---

Este capítulo descreve o problema estudado, que se refere à necessidade de consideração da dependência espacial na análise de modelos de demanda por transportes, bem como apresenta a justificativa, o objetivo e a estrutura do trabalho.

## **1.1. CARACTERIZAÇÃO DO PROBLEMA**

Muitos pesquisadores (Durrett, 1994; Griffith, 1996; Miller, 1999) enfatizam que a maioria dos métodos e teorias estatísticas assume que as observações são independentes, condição, no entanto, raramente encontrada quando são analisados dados espaciais. Tobler (1970), citando a Primeira Lei da Geografia, destaca que todos os objetos no espaço estão relacionados, porém objetos mais próximos no espaço estão mais relacionados, o que o levou a concluir que os resultados de análises espaciais dependem da localização dos objetos que estão sendo considerados.

Existem muitas questões analítico-espaciais que devem ser consideradas na análise de transportes e, entre elas, as que têm sido mais levantadas por diversos pesquisadores abrangem a definição de unidades de área modificáveis (zonas de tráfego), os problemas de contorno e amostragem espacial, a dependência e heterogeneidade espacial e as representações alternativas de ambientes geográficos.

O foco deste trabalho concentra-se na dependência espacial (também referida como autocorrelação espacial) para análise de dados de demanda por transportes, que para muitos autores não deve ser ignorada, sob pena de ocorrerem erros de estimativa, comuns em modelos lineares (Anselin e Griffith, 1988), ou estimativas inconsistentes de parâmetros em modelos não lineares (Case, 1992).

Na análise de demanda de viagens, a área de estudos é frequentemente segmentada em diversas zonas e cada uma delas representa unidades espaciais nas quais dados são agregados, de onde se assume que viagens são produzidas e para onde viagens são atraídas.

Os atributos temporais e espaciais das trocas de viagens entre pares de zonas são tipicamente estimados por modelos de viagens em um nível individual ou domiciliar, ou por modelos de atividades e técnicas de amostragem apropriadas, que são aplicadas para estimar o fluxo de viagens.

O conceito de autocorrelação espacial mede o nível de interdependência geográfica entre as variáveis e a natureza e a força desse relacionamento. O índice de Moran (Câmara *et al.*, 2002a), por exemplo, testa se as áreas conectadas apresentam maior semelhança quanto ao indicador estudado do que o esperado num padrão aleatório, variando de -1 a 1. O grau de autocorrelação é positivo para correlação direta, negativo quando inversa e nulo quando o indicador se distribui ao acaso entre as áreas, sem relação com a posição.

O aumento da disponibilidade de bases digitais de dados geográficos e a capacidade de manuseio de informações geográficas têm resultado em medidas que capturam os efeitos de autocorrelação espacial, como as referidas por Anselin (1995) como indicadores locais de associação espacial ou estatística LISA (*Local Indicators of Spatial Association*).

Nas últimas três décadas muitos pesquisadores têm desenvolvido trabalhos sobre o tema, embora no Brasil, apesar da crescente utilização das ferramentas de um SIG, pouco tenha sido pesquisado sobre o assunto para a modelagem de transportes (Krempi

*et al.*, 2002; Teixeira, 2003; Queiroz, 2003; Krempi, 2004, Lopes e Silva, 2004 e Henrique, 2004).

## **1.2. JUSTIFICATIVA**

Ignorar questões de análise espacial como, por exemplo, a dependência espacial, pode invalidar os resultados da análise de demanda por transportes, levar a uma previsão inadequada e, conseqüentemente, a um planejamento ineficiente. A técnica largamente usada de análise de regressão perde precisão sob os efeitos da dependência espacial, pois os parâmetros estimados não são eficientes e a significância dos testes é irreal na presença de autocorrelação espacial (Anselin e Griffith, 1988).

Em virtude disso, a introdução de indicadores de dependência espacial na modelagem de demanda por transportes deve produzir resultados mais precisos e, desta forma, mais confiáveis dos que os obtidos com modelos tradicionais. Visa-se, assim, contribuir para uma evolução do processo de modelagem e, adicionalmente, para a redução de custos, vital para um país em desenvolvimento, através do melhor aproveitamento de ferramentas já introduzidas no planejamento de transportes.

## **1.3. OBJETIVOS**

O objetivo principal deste trabalho é avaliar, mediante o uso de ferramentas de planejamento de transportes e de análise espacial em ambiente SIG (Sistemas de Informações Geográficas), os efeitos da utilização de indicadores de dependência espacial em modelos de previsão de demanda por transportes, ou seja, verificar se a proposição de modelos alternativos, com variáveis espaciais, produz melhores resultados que os obtidos por modelos estatísticos tradicionais.

Têm-se, também, os seguintes objetivos complementares:

- 1.** Analisar diferentes variáveis socioeconômicas obtidas em pesquisas Origem-Destino realizadas em Porto Alegre-RS, em 1974 e 1986, visando a seleção das que serão consideradas na modelagem de demanda por transportes e a análise dos índices de autocorrelação espacial;

2. Estudar diferentes métodos de inclusão dos índices de dependência espacial, visando a proposição de um critério que melhor atenda às necessidades dos modelos de previsão de demanda por transportes;
3. Explicar os resultados obtidos com os modelos tradicionais e com os modelos propostos, que incluem indicadores de dependência espacial, considerando-se os dados reais de viagens das referidas pesquisas O-D.

#### **1.4. ESTRUTURA DO TRABALHO**

Este trabalho é composto por seis capítulos. O Capítulo 2 apresenta uma revisão sobre planejamento de transportes, onde são abordados problemas de transporte e modelos de previsão de demanda, além de algumas críticas aos modelos tradicionais. O terceiro capítulo aborda o tema análise espacial, onde são apresentados os conceitos básicos, formas de representação dos dados espaciais e ferramentas de estatística espacial, como as medidas de autocorrelação espacial, a análise exploratória de dados espaciais e a regressão espacial. No Capítulo 4 é apresentado o método adotado nesta pesquisa, em que são descritos, detalhadamente, os dados disponíveis para o estudo de caso assim como os modelos e as análises propostas. O Capítulo 5 apresenta os resultados obtidos no estudo de caso e as análises dos resultados. Finalizando, as conclusões e sugestões para trabalhos futuros são apresentadas no sexto e último capítulo.

## **2. PLANEJAMENTO DE TRANSPORTES**

---

Este capítulo é baseado em uma revisão bibliográfica sobre planejamento de transportes, onde são abordados problemas de transporte e modelos de previsão de demanda, além de algumas críticas aos modelos tradicionais. Apresenta-se, também, ao final do capítulo, um breve comentário sobre as potenciais contribuições de uso de ferramentas dos Sistemas de Informações Geográficas (SIG) para a análise do tema.

### **2.1. INTRODUÇÃO**

O equacionamento adequado da questão de transporte urbano é, atualmente, um dos grandes desafios do Brasil e do mundo. Maximizar a eficiência de sistemas de transportes multimodais e integrados é o paradigma do transporte para o futuro. O objetivo do planejamento de transportes é providenciar as informações necessárias para a tomada de decisões sobre “quando” e “onde” melhorias no sistema deverão ser feitas e, desta forma, promover padrões de viagens e desenvolvimento urbano de acordo com os anseios da comunidade.

Vasconcellos (1996) analisa a formação recente das cidades nos países em desenvolvimento, destacando três formas de intervenção intimamente interligadas e relacionadas diretamente aos transportes:

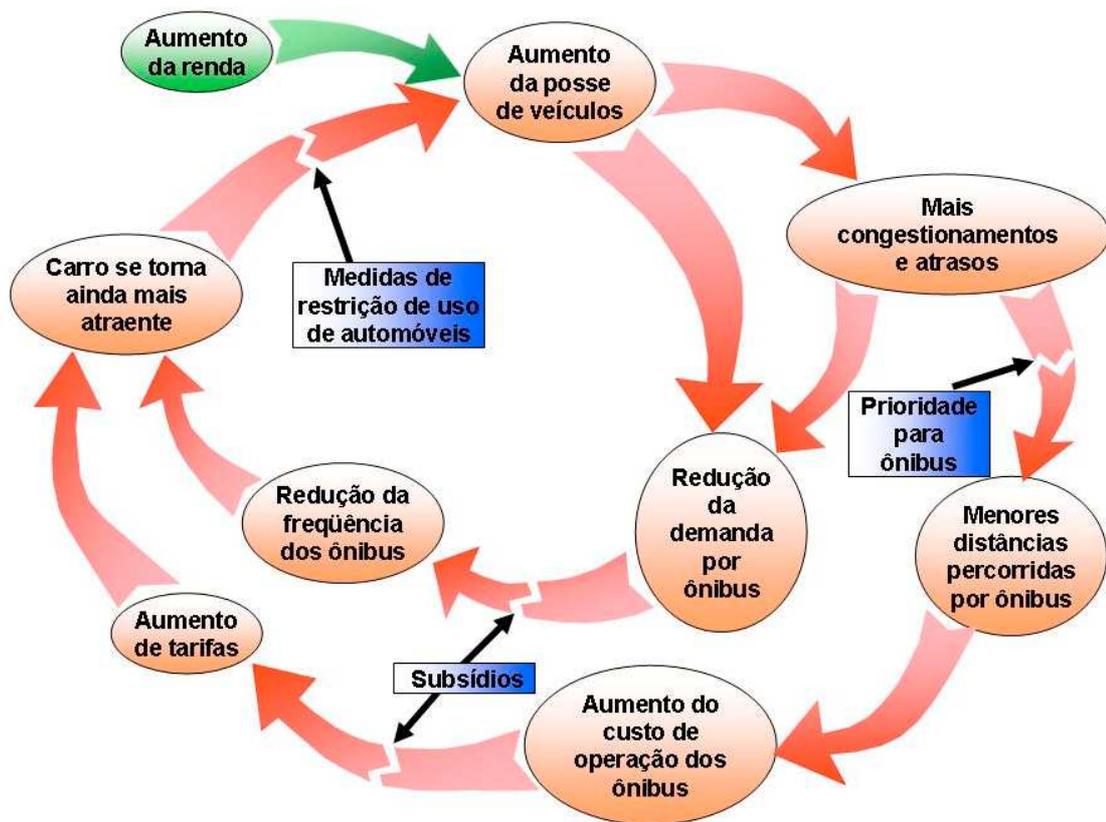
- **Planejamento urbano:** define a forma como o espaço deve ser ocupado e usado, para os mais diversos fins. Ao definir limites para a propriedade privada da terra, ele constitui uma arena política altamente conflituosa. Seus produtos são códigos e leis que definem os usos e ocupações desejados e permitidos. A sua existência não implica, necessariamente, em respeito às suas determinações e, principalmente nos países em desenvolvimento, acabam se formando duas cidades, uma “formal” e outra “informal”;
- **Planejamento de transportes:** define a infra-estrutura de circulação, que vai permitir o deslocamento de pessoas e mercadorias, bem como os veículos e serviços que serão ofertados;
- **Planejamento da circulação:** define como a estrutura viária será utilizada pelas pessoas e veículos. Envolve quatro atividades relacionadas: a legislação, que define as regras de utilização das vias (Código de Trânsito); a engenharia de tráfego, que define os esquemas de circulação; a educação, que define como as pessoas serão treinadas para usar o sistema viário; e a fiscalização, que controla o respeito às leis de trânsito.

Segundo Ortúzar e Willumsen (1994), os problemas e as técnicas de planejamento de transporte sofreram profundas mudanças a partir de 1980. Com a continuidade dos problemas dos anos 60 e 70 (congestionamento, poluição, acidentes, déficits de financiamento etc.) foi possível aprender que um longo período de planejamento deficiente de transportes, investimentos limitados, ênfase no curto prazo e falta de credibilidade na modelagem de transportes e na tomada de decisões estratégicas não são resolvidos com tentativas tímidas de melhoria da gerência de tráfego. Além disso, o advento da computação de alta capacidade e baixo custo eliminou dificuldades que anteriormente existiam na modelagem de transportes.

Ortúzar e Willumsen (1994), a esse respeito, apresentam o círculo vicioso do automóvel e transporte público e também sugestões para quebrá-lo (Figura 2.1). O crescimento econômico fornece o primeiro ímpeto para o aumento da posse de veículos. A migração do transporte público para o carro significa menos passageiros para o transporte público, o que pode levar o operador ao aumento de tarifas ou à redução da frequência (nível de serviço), ou ambos. Tais medidas tornam o uso do carro ainda mais atrativo e induzem mais pessoas a comprar carros, acelerando o processo. Após alguns

círculos (anos), os modos de transporte público passam a enfrentar índices elevados de congestionamento e, além dos atrasos, ficam cada vez mais caros e com menor frequência, mostrando que o acúmulo de decisões sensivelmente individuais resulta em um estado final em que quase todos os usuários são prejudicados. Dentre as alternativas para desacelerar ou reverter o círculo vicioso, têm-se: prioridade para ônibus, subsídios e medidas de restrição ao uso do automóvel.

Deve-se destacar que, no contexto de países em desenvolvimento, o crescimento da população mantém a demanda por transporte público por mais tempo do que em países industrializados. A experiência atual de alguns países em desenvolvimento é de fluxos de ônibus extremamente altos, atingindo 400 a 600 ônibus por hora, por sentido, ao longo de alguns corredores (Ortúzar e Willumsen, 1994).



**Figura 2.1:** Alternativas para quebrar o círculo vicioso do automóvel e do transporte público por ônibus - adaptado de Ortúzar e Willumsen (1994)

## 2.2. MODELOS DE PREVISÃO DE DEMANDA POR TRANSPORTES

Para a previsão da demanda de transportes deve-se, inicialmente, realizar um inventário detalhado das condições atuais. A cidade pode ser dividida em zonas de tráfego (ZTs) e a partir daí são determinados os movimentos realizados entre cada par de zonas. O resultado é uma tabela de origens e destinos (Matriz O-D), que tem íntima relação com a atração e produção de viagens.

Zonas de tamanho reduzido nem sempre são a melhor opção, pois a complexidade aumenta com o número de zonas e há dificuldade para se utilizar os dados de população e outros em zonas demasiadamente pequenas.

As características referentes à população, postos de trabalho, número de veículos e uso do solo podem ser obtidas de estudos específicos ou provir de estatísticas municipais. O mapeamento das atividades econômicas é elaborado a partir de dados existentes. Ao final, são estabelecidas fórmulas que relacionam a atração e produção de viagens com as características das zonas que influenciam significativamente o processo.

Os modelos de demanda são usados para previsão de mudanças nas viagens e na utilização do sistema de transportes em função do desenvolvimento da região, de alterações demográficas e de mudanças na oferta. A previsão da demanda é uma tarefa desafiadora, mas indispensável para o planejamento racional dos sistemas de transportes.

A aplicação de modelos de transportes sempre se baseou em processos computacionais. Em seus estágios iniciais, os modelos de previsão de demanda constituíram uma das aplicações pioneiras dos computadores de grande porte. Por outro lado, os modelos de demanda são inerentemente espaciais e, portanto, dependentes da manipulação de variáveis espaciais: vetores para produção e atração de viagens; matrizes de fluxo e impedância de viagens; redes para as vias e para o transporte público.

A utilização de um SIG facilita a aplicação dos modelos de demanda. No modelo de quatro etapas, por exemplo, uma ou mais camadas contêm informações das

ZTs e vetores e matrizes representam as informações do fluxo. As ferramentas de visualização gráfica de um SIG aumentam o entendimento dos dados de transportes e dos resultados dos modelos, ajudando a identificar e corrigir erros eventuais. Gráficos podem ser integrados com mapas temáticos, ilustrando as previsões de demanda.

### **2.3. MODELO DE QUATRO ETAPAS**

Um dos modelos largamente empregado no planejamento de transporte é o modelo de quatro etapas (ou seqüencial), que permite a estimativa de viagens das pessoas entre as diversas zonas de tráfego (matriz O-D) com base na consideração de que a realização de uma viagem depende de um processo desenvolvido em fases:

- Geração: quantas viagens têm origem ou destino nas zonas consideradas?
- Distribuição: quantas viagens são realizadas entre as zonas consideradas?
- Divisão Modal: quais os modos de transporte utilizados?
- Alocação de Tráfego: qual o caminho utilizado pelas pessoas (para modos de transporte que permitem a escolha da rota)?

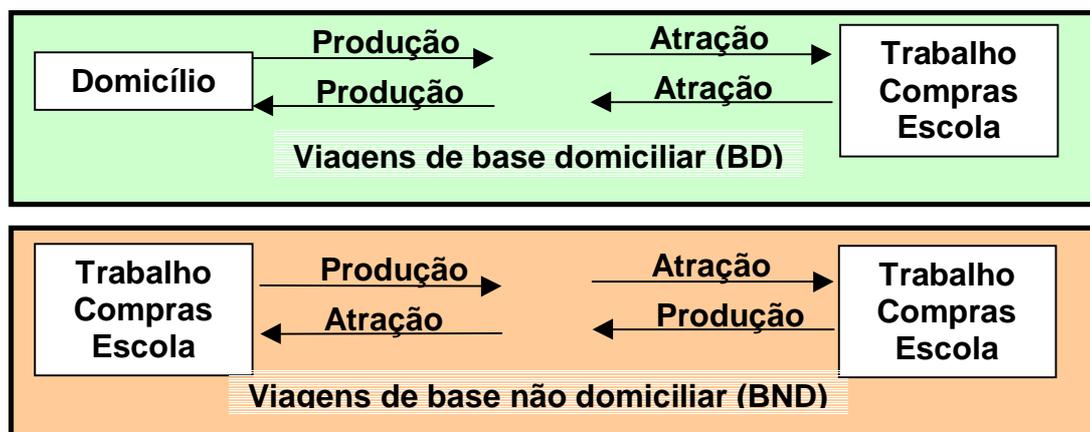
O objetivo da geração de viagens, que será objeto deste estudo, é a previsão do número de viagens que são produzidas e/ou atraídas para cada zona de tráfego da área em estudo. É um estágio do planejamento de transportes preocupado apenas com o número de viagens que começam ou terminam em cada zona de tráfego, sem necessariamente se importar com as conexões entre origens e destinos. Os métodos de geração de viagens fazem a previsão da produção (origens) e da atração (destinos) e, posteriormente, zonas de tráfego são arranjadas em uma matriz origem-destino (O-D).

A análise de geração de viagens consiste na estimativa do número total de viagens, por diferentes motivos, produzidas ou atraídas para cada zona. A geração de viagens relaciona o número ou freqüência de viagens a características dos indivíduos/domicílios, da zona e da rede de transportes. Das análises resultam tantos modelos quantos forem os motivos adotados nos estudos:

- Viagem de base domiciliar: casa/trabalho, casa/estudo e casa/outros;
- Viagem de base não domiciliar: todas as viagens que não têm a residência como origem ou destino.

Para a correta análise de geração de viagens é importante, inicialmente, a compreensão de alguns conceitos básicos usados na modelagem de transporte, que são definidos por Ortúzar e Willumsen (1994), como mostra a Figura 2.2:

- **Viagem ou jornada:** movimento de um único sentido de um ponto de origem para um ponto de destino. Embora o interesse seja especificamente por viagens veiculares, deslocamentos a pé a partir de uma determinada distância definida pelo estudo (normalmente 300 m ou 3 quarteirões) são geralmente considerados. Viagens de crianças menores de 5 anos de idade são geralmente ignoradas nas pesquisas;
- **Viagens de base domiciliar (BD):** são as viagens em que o domicílio do indivíduo que a realiza é a origem ou o destino;
- **Viagens de base não domiciliar (BND):** são as viagens em que a origem ou o destino não é o domicílio do indivíduo que a realiza;
- **Produção de viagens:** refere-se à extremidade domiciliar (origem ou destino) de uma viagem BD ou a origem de uma viagem BND;
- **Atração de viagens:** refere-se à extremidade não domiciliar (origem ou destino) de uma viagem BD ou o destino de uma viagem BND;
- **Geração de viagens:** refere-se ao número total de viagens geradas pelos domicílios em uma ZT, sejam elas BD ou BND.



**Figura 2.2:** Atração e produção de viagens BD e BND – adaptado de Ortúzar e Willumsen (1994)

Os modelos de geração de viagens mais utilizados são:

- **Classificação Cruzada:** separa a população de uma área urbana em grupos relativamente homogêneos, de acordo com certas características socioeconômicas. Posteriormente, taxas médias de produção de viagens por domicílio ou indivíduo são estimadas para cada grupo;
- **Regressão:** podem ser usados dados agregados ao nível da zona de tráfego, com o número médio de viagens por domicílio sendo a variável dependente e características da zona de tráfego sendo as variáveis explanatórias (independentes). Quando são utilizados dados desagregados, a variável dependente continua sendo o número de viagens feitas por domicílio ou indivíduo, mas as variáveis independentes passam a ser características dos indivíduos ou domicílios;
- **Escolha Discreta:** usa dados desagregados para estimar a probabilidade de viagens por domicílio ou indivíduo, cujos resultados são, posteriormente, agregados para fornecer a previsão do número de viagens produzidas.

A geração de viagens do transporte de passageiros pode ser quantificada em termos de viagens de pessoas ou de viagens de veículos. Considerando-se o aspecto comportamental, há vantagens em se considerar o passageiro: são os indivíduos que decidem sobre a frequência de viagens, não os veículos.

A geração de viagens pode ser baseada nos domicílios ou indivíduos. Cálculos baseados nos domicílios são mais comuns, mas se houver disponibilidade de dados mais detalhados, o uso de taxas associadas aos indivíduos pode resultar em maior precisão, por considerar os aspectos comportamentais determinantes.

Os motivos das viagens são extremamente variados, refletindo as diversas atividades de trabalho e lazer desenvolvidas pelas pessoas. Para o propósito de análise, no entanto, as viagens são agrupadas em categorias ou motivos, em número que depende do estudo que se pretende fazer e dos dados disponíveis. Segundo Ortúzar e Willumsen (1994), a prática tem mostrado que com a divisão em categorias por motivo obtêm-se melhores modelos de geração de viagens. São apresentadas cinco categorias geralmente aplicadas para o caso de viagens BD:

- **Trabalho;**
- **Estudo;**
- **Compras;**
- **Sociais ou de lazer;**
- **Outras viagens.**

As viagens de trabalho e estudo são chamadas de viagens compulsórias e são, geralmente, as mais numerosas, enquanto que as demais são consideradas viagens opcionais. A última categoria engloba as viagens menos rotineiras como, por exemplo, as por motivo de saúde, necessidades burocráticas etc. As viagens BND normalmente não são divididas em categorias, pois representam apenas de 15 a 20% de todas as viagens.

As viagens podem ainda ser classificadas por período do dia (horários de pico ou fora de pico). Os horários de pico variam de cidade para cidade e, muitas vezes, de região para região dentro de uma mesma cidade. Os motivos de viagem neste período são na maioria para trabalho ou estudo. Outra classificação importante, apresentada por Ortúzar e Willumsen (1994), é por características pessoais (nível de renda, posse de veículo ou estrutura e tamanho do domicílio), uma vez que o comportamento individual de viagens é altamente dependente de atributos socioeconômicos.

#### **2.4. SELEÇÃO DE VARIÁVEIS PARA ANÁLISES DE GERAÇÃO DE VIAGENS**

A escolha de variáveis para a previsão das taxas de geração de viagens é, de longa data, uma preocupação dos planejadores de transportes. É necessário fazer a distinção de quatro grupos principais, onde diferentes variáveis serão consideradas para a modelagem da geração de viagens. São eles:

- Viagens de pessoas (produção ou atração);
- Viagens de carga (produção ou atração).

Os grupos que englobam as viagens de pessoas são os de interesse deste trabalho e os fatores que afetam a produção, destacados por Ortúzar e Willumsen (1994), são:

- Renda;
- Posse de veículo;
- Estrutura do domicílio;
- Tamanho da família;
- Valor do terreno;
- Densidade residencial;
- Acessibilidade.

Para a atração de viagens de pessoas, outros fatores são considerados, como, por exemplo, espaço disponível para indústria, comércio e serviços e, também, o número de empregos por ZT. Deve-se destacar, ainda, a necessidade de verificação de dois aspectos relacionados aos parâmetros de geração de viagens:

- **Estabilidade Temporal:** os modelos de transporte, em geral, são desenvolvidos para auxiliar a formulação e avaliação de planos e projetos de transporte. Embora muitas vezes sejam utilizados dados estatísticos, obtidos durante um longo período de coleta, a maioria dos casos envolve a hipótese de que os parâmetros do modelo permanecerão constantes (estáveis) entre a data de seu desenvolvimento e os anos do período de análise. Copley e Lowe (1981)<sup>1</sup> *apud* Ortúzar e Willumsen (1994) mostraram que as taxas de viagens de ônibus para algumas categorias de domicílios apresentam-se estáveis ao longo do tempo, enquanto as taxas de viagens de carro são altamente dependentes do preço dos combustíveis;
- **Estabilidade Geográfica:** é um importante atributo em todo modelo de demanda, que pode permitir a redução dos custos de pesquisas em diferentes áreas de uma região metropolitana. Nem todas as características podem ser transferidas entre áreas ou cidades, pois a duração média de uma viagem até o serviço depende das dimensões das áreas, suas formas e distribuições dos locais de trabalho e das zonas residenciais no espaço. Mas, considerando-se que as viagens são relacionadas a grupos homogêneos de pessoas, as taxas de geração de viagens podem, em muitos casos, permanecer estáveis geograficamente.

---

<sup>1</sup> Copley, G. and Lowe, S.R. (1981) The temporal stability of trip rates: some findings and implications. *Proceedings 9<sup>th</sup> PTRC Summer Annual Meeting*, University of Warwick, July 1981, England.

## 2.5. CRÍTICAS AOS MODELOS DE DEMANDA TRADICIONAIS

Os modelos de demanda são desenvolvidos e calibrados em função do comportamento das pessoas e das condições do sistema de transportes que prevalecem no local e na época do estudo. Os valores sociais podem mudar com o tempo, resultando em erros nas projeções de demanda, além de que problemas específicos existentes no sistema de transporte em análise também podem comprometer os resultados.

Saraiva (2000) salienta que a magnitude do problema é ainda maior em países em desenvolvimento como é o caso do Brasil, porque, enquanto o planejamento de transportes em países desenvolvidos trabalha com uma situação em que prevalecem populações e áreas urbanas relativamente estáveis, as análises de demanda em cidades brasileiras incluem até mesmo áreas desocupadas.

Hoover e Altshuler (1977)<sup>2</sup> *apud* Vasconcelos (1996) criticam os modelos de demanda tradicionais por terem uma visão limitada do processo de desenvolvimento urbano, com um enfoque orientado prioritariamente à definição de projetos e para facilitar a projeção futura das tendências existentes, quando deveriam tratá-las como políticas a serem analisadas pelo processo de planejamento. Comentam, também, sobre a falta de interdisciplinaridade dos enfoques e a visão estreita de que o planejamento de sistemas está limitado à preparação de grandes planos.

Stopher e Meyburg (1975)<sup>3</sup> *apud* Vasconcelos (1996) apontam possíveis fontes de erros nas previsões de demanda em todas as etapas do processo, desde a coleta de dados, passando pela definição da amostra, até a etapa de calibração dos modelos.

Taco (1997) critica os modelos tradicionais pelo caráter estático, pois não apresentam recursos que possibilitem a captação, com rapidez, das mudanças urbanas relativas ao uso do solo. Além disso, necessitam de grande quantidade de dados, o que impossibilita a sua atualização rápida e contínua, em razão dos elevados custos operacionais.

---

<sup>2</sup> Hoover, J. H. and Altschuler, A. A. (1977) *Involving cities in Metropolitan Region Transportation Planning* EUA, U.S. Dept of Transportation.

<sup>3</sup> Stopher, P. R. and Meyburg, A. H.(1975) *Urban transportation modeling an planning*, EUA.

Vasconcelos (1996) resume os principais problemas dos modelos de demanda utilizados no Brasil, destacando a falta de dados confiáveis, o uso de modelos importados, que não reproduzem adequadamente as condições dos países em desenvolvimento, e as altas taxas de mudança demográfica, econômica e social, que resultam em grandes estruturas de transporte ociosas ou desnecessárias.

Uma alternativa para solução dos problemas apresentados pelos modelos tradicionais é a utilização de técnicas que reconheçam e incorporem características espaciais no processo, o que acentua a potencialidade do uso de Sistemas de Informações Geográficas (SIG), particularmente de suas ferramentas de análise espacial.

## **2.6. CONSIDERAÇÕES FINAIS**

O Planejamento de Transportes em uma cidade ou região tem como objetivo prover as informações necessárias para subsidiar a definição da infra-estrutura, dos veículos e dos serviços adequados para permitir o deslocamento de pessoas e mercadorias e, conseqüentemente, promover um desenvolvimento urbano que atenda aos anseios da comunidade. Por esta razão, deve estar integrado ao Planejamento Urbano e ao Planejamento da Circulação.

A partir dos anos 80 avanços tecnológicos intensificaram e aprimoraram as técnicas de modelagem de demanda. Existem, no entanto, várias críticas aos modelos tradicionais. Algumas delas dizem respeito à grande quantidade de dados necessários, o que dificulta o processo; à falta de consideração de mudanças urbanas relativas ao uso do solo e à falta de incorporação de características espaciais aos modelos.

Uma vez que os modelos de demanda são inerentemente espaciais, a utilização de um SIG facilita sua aplicação no que diz respeito ao armazenamento e manipulação de dados. Além disso, suas ferramentas de visualização e de análise espacial permitem um melhor entendimento das variáveis, dos resultados e também ajudam na identificação e correção de erros eventuais, tanto de diagnóstico como de previsão.

### **3. ANÁLISE ESPACIAL**

---

Este capítulo aborda o tema de Análise Espacial e os principais conceitos do subcampo chamado Estatística Espacial. Baseado em uma revisão bibliográfica sobre o assunto, inicia com um breve resumo sobre conceitos básicos e o papel dos SIGs como propulsor da introdução deste tema nos estudos de transportes. Apresenta também, as formas de representação dos dados espaciais e as ferramentas de estatística espacial, concluindo com os temas de interesse deste estudo, que são: a autocorrelação espacial e a dependência espacial, a análise exploratória de dados espaciais (ESDA) e a regressão espacial.

#### **3.1. INTRODUÇÃO**

O objetivo da análise espacial é mensurar propriedades e relacionamentos, levando em conta a localização espacial do fenômeno em estudo de forma explícita. Ou seja, a idéia central é incorporar o espaço à análise que se deseja fazer. Pode-se definir a análise espacial como o estudo, exploração e modelagem de processos que se expressam através de uma distribuição no espaço. Tais processos são também denominados fenômenos geográficos (Câmara *et al.*, 2002b). O conjunto amplo de técnicas, que incluem métodos estatísticos e que procuram descrever a variação espacial do fenômeno em estudo, a partir de amostras disponíveis, é chamado de “Estatística Espacial”.

No século XIX, mais precisamente em 1854, John Snow, intuitivamente, incorporou a categoria espaço às análises, realizadas em Londres, de uma das várias epidemias de cólera trazidas das Índias. As análises espaciais ajudaram a identificar uma das bombas de água que abasteciam a cidade como epicentro da epidemia. Para Câmara *et al.* (2002b), esta é uma situação típica onde a relação espacial entre os dados contribuiu significativamente para o avanço na compreensão do fenômeno, sendo um dos primeiros exemplos da análise espacial.

Conforme citado por Krempi (2004), vários estudos foram realizados até que avanços no campo da estatística começassem a ser transferidos para o que hoje se conhece como estatística espacial. Em muitos trabalhos publicados sobre métodos de estatística tradicionais já se observava o reconhecimento dos problemas causados pela dependência espacial, mas foram Moran (1948) e Geary (1954) que apresentaram os primeiros índices formais para detectar a presença da autocorrelação espacial em um conjunto de dados.

Krempi (2004) ainda salienta que este desenvolvimento vem acontecendo de forma mais intensa a partir dos anos oitenta e noventa do século XX, trazendo avanços notáveis para o campo do conhecimento hoje denominado econometria espacial. O avanço é devido, entre outros aspectos, ao aumento do interesse na área, à crescente disponibilidade de bases de dados socioeconômicas georreferenciadas e à expansão de tecnologias eficientes e de baixo custo, onde hoje já se poderia incluir tanto os SIG como os programas específicos para análise de dados espaciais, que permitem tratar dados referenciados espacialmente.

Para muitos autores, a introdução da análise espacial nos estudos de transportes deu-se com o advento da aplicação de Sistemas de Informações Geográficas nos seus variados ramos de análise e planejamento. Para Miller (1999), esse tradicional subcampo da geografia foi revitalizado com o uso dos SIGs, resultando em particular benefício para o planejamento de transportes, através do aprimoramento das análises de fenômenos intrinsecamente espaciais, como é o caso, por exemplo, da demanda por transportes. Para Teixeira (2003), com o uso mais freqüente de SIG, o termo “análise espacial” passou a fazer parte do cotidiano do profissional de transportes, embora pouco tenha sido de fato discutido sobre suas bases teóricas, seus conceitos e aplicações.

Se, por um lado, a comunidade de transportes reconhece os grandes benefícios oferecidos pelo SIG como um "Sistema de Gerenciamento de Banco de Dados Espaciais" e como um "Sistema de Suporte a Decisões Espaciais", por outro lado os problemas analítico-espaciais, inerentes à análise de transportes, são menos reconhecidos. Miller (1999) chama a atenção para o fato que a capacidade de análise espacial de um SIG poderia ser melhor explorada, aprimorando a solução de problemas de transportes já inseridos em ambiente SIG.

Para Câmara *et al.* (2002b), grande parte dos usuários limita seu uso de SIG a operações de visualização, tirando conclusões intuitivas. Os processos de análise espacial incluem métodos não só de visualização, mas que possibilitam traduzi-los em considerações objetivas, ou seja, métodos que permitem a investigação do padrão dos dados, isto é, se os dados apresentam uma agregação definida ou se a distribuição é aleatória e, também, se esta distribuição pode ser associada a causas mensuráveis. Os autores citam, ainda, que o grande problema com padrões de áreas (caso deste estudo) é que estes apresentam discontinuidades. Deste modo, é necessário lançar mão de instrumentos que permitam estabelecer a continuidade espacial entre as diferentes áreas (as matrizes de continuidade espacial) e de indicadores que possam medir a correlação espacial entre as ocorrências.

Assunção (2001) salienta que a visualização é um fator importante da estatística espacial, mas complementar à descrição, comparação e interpretação dos dados. A estatística descritiva ou inferencial é, portanto, eficaz e essencial para tornar explícito o que está apenas implícito em mapas ou para analisá-los, chamando a atenção para características que não eram perceptíveis apenas pela visualização.

Miller (1999) aponta várias questões analítico-espaciais que devem ser consideradas na análise de transportes e, dentre elas, está o caso da dependência espacial, que é o foco deste trabalho, mais precisamente na análise de demanda por transportes. Desta forma, por se tratar de um tema amplo, a revisão bibliográfica sobre Análise Espacial e Estatística Espacial está mais focalizada nas características e conceitos de interesse deste estudo.

### 3.2. TIPOS DE DADOS EM ANÁLISE ESPACIAL

Para Câmara *et al.* (2002b), a classificação mais utilizada para caracterizar problemas de análise espacial considera três tipos de dados:

- **Eventos ou padrões pontuais** – fenômenos expressos através de ocorrências identificadas como pontos localizados no espaço, denominados processos pontuais. São exemplos: localização de crimes, ocorrência de doenças, localização de espécies vegetais e, também, um tema mais relacionado aos estudos de transporte, os acidentes de trânsito;
- **Superfícies contínuas** – estimadas a partir de um conjunto de amostras de campo, que podem estar regularmente ou irregularmente distribuídas. Usualmente esse tipo de dado é resultante de levantamento de recursos naturais, apresentados em mapas geológicos, topográficos, ecológicos, fitogeográficos e pedológicos;
- **Áreas com contagens e taxas agregadas** – tratam de dados associados a levantamentos populacionais e de transportes, como censos, estatísticas de saúde e pesquisas O-D, que originalmente se referem a indivíduos localizados em pontos específicos do espaço. Por razões de confidencialidade ou por não ser vantajoso se trabalhar com a localização exata dos eventos, esses dados são agregados em unidades de análise, usualmente delimitadas por polígonos fechados (setores censitários, zonas de endereçamento postal, zonas de tráfego, municípios).

Alguns autores, como Teixeira (2003) e Carvalho (1997), dentre outros, ressaltam ainda um quarto tipo, um pouco menos citado na literatura de análise espacial:

**Análise de Redes** - os objetos analisados são entidades lineares, de grande utilidade quando se pretende estudar, por exemplo, o acesso e o fluxo entre regiões, otimizando trajetórias e estudando a localização de equipamentos urbanos, o que é amplamente discutido em todas as áreas de transportes.

### 3.3. ANÁLISE ESPACIAL DE ÁREAS

O uso de métodos de análise de dados espaciais, cuja localização está associada a áreas delimitadas por polígonos, ocorre com muita frequência quando se está lidando com eventos agregados por municípios, bairros, setores censitários ou zonas de tráfego, onde não se dispõe da localização exata dos eventos, mas de um valor por área. Nessas análises, procura-se identificar padrões espaciais de distribuição, aglomerados e sinais de dependência espacial, porém deve-se ter em mente alguns problemas que podem ocorrer (Miller, 1999). Um deles é a questão da descontinuidade das fronteiras, pois, em geral, valores próximos à fronteira, em áreas contíguas, tendem a ser semelhantes, embora cada área seja representada por sua média, o que distorce a análise nas regiões fronteiriças, apresentando uma “quebra” de valor que não se aplica na realidade. O outro é o Problema de Unidade de Área Modificável (*Modifiable Areal Unit Problem – MAUP*), cuja evidência foi apontada pela primeira vez por Genkle e Biehl (1934)<sup>4</sup> *apud* Miller (1999) e depois por Robinson<sup>5</sup> (1950) *apud* Miller (1999), e que passou a ser cada vez mais estudado, por exemplo, por pesquisadores como Cressie (1995, 1996) e Miller (1999). O que se verifica é que o nível de agregação das subáreas, assim como a escala do mapa, pode conduzir a erros no processo de análise.

Miller (1999) salienta, ainda, que uma área do planejamento de transportes onde o problema da MAUP é crítico é a análise de previsão de demanda, em que a agregação espacial em ZTs, para facilitar o tratamento computacional dos dados e também a disponibilidade dos mesmos, pode afetar substancialmente os resultados das análises. Muitos pesquisadores têm trabalhado no desenvolvimento de técnicas para reduzir o problema, como Moellerring e Tobler (1972), Batty (1976), Openshaw (1978) e Batty e Sikdar (1982).

Câmara *et al.* (2002a) chamam a atenção para o fato de que, principalmente em países com grandes contrastes sociais como o Brasil, é frequente que grupos sociais distintos estejam agregados em uma mesma região de coleta (ZTs, setores censitários

---

<sup>4</sup> Gehlke, C.E. e Biehl, K. (1934) “Certain effects of grouping upon the size of the correlation coefficient in census tract material”, *Journal of the American Statistical Association Supplement*, 29, 169-170.

<sup>5</sup> Robinson, W.S. (1950) “Ecological correlation and the behavior of individuals”, *American Sociological Review*, 15, 351-357.

etc.) – favela e áreas nobres – resultando em indicadores calculados que representam a média entre populações diferentes. Ou seja, a suposição de haver homogeneidade interna nas áreas delimitadas por polígonos fechados e de que mudanças só ocorrem nas fronteiras, muitas vezes, não é verdadeira. Neste caso, tanto a apresentação em mapas cloropléticos como o cálculo de indicadores populacionais pode levar a distorções, sendo necessária a utilização de técnicas de ajuste de distribuição.

### 3.4. ANÁLISE ESPACIAL DE SUPERFÍCIES

Na análise de superfície, os atributos e variáveis observados são analisados de forma contínua no espaço (Landim, 1998). Como se admite uma variação contínua do atributo analisado, procura-se identificar nos mapas as características principais, como a intensidade da ocorrência do evento e a variabilidade espacial das observações.

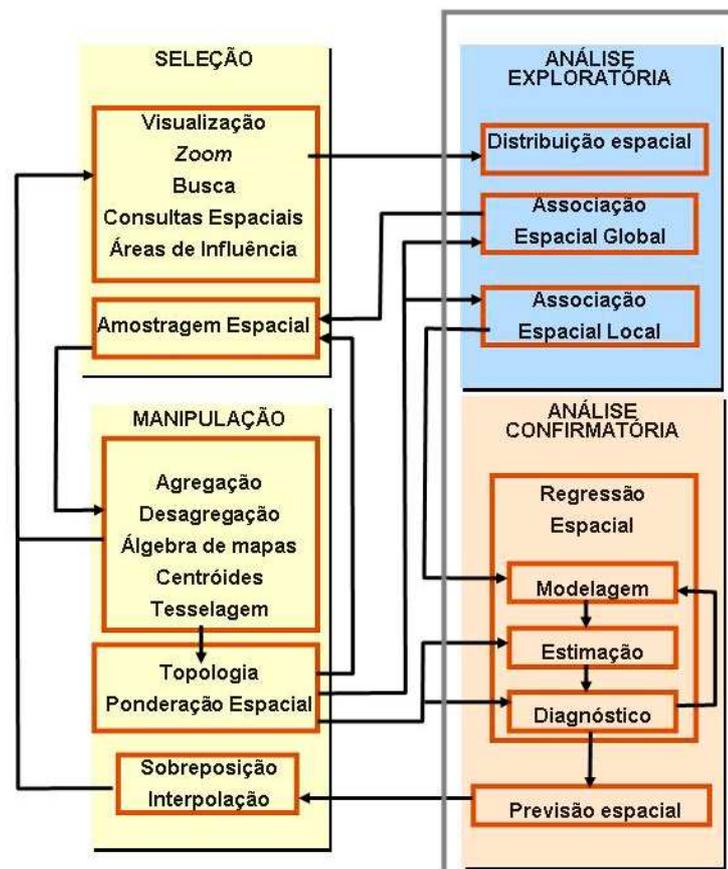
A variabilidade é medida pela isotropia, ou seja, se o fenômeno observado varia de forma uniforme em todas as direções, diz-se que é isotrópico, e se varia de forma irregular para direções distintas é anisotrópico. Dessa forma, a análise de superfície se mostra superior à análise por áreas nos casos onde o problema de descontinuidade nas fronteiras se mostra significativo, caso contrário, para áreas com características homogêneas, a análise agregada é muito mais simples e requer menor esforço computacional (Teixeira, 2003).

### 3.5. FERRAMENTAS DE ANÁLISE ESPACIAL

Uma subdivisão das ferramentas de análise espacial em quatro grupos (Figura 3.1) é proposta por Anselin (1992). São elas:

- **Seleção** - envolve o processo mais simples de consultas a banco de dados, com procedimentos simples de amostragem e agrupamentos de dados analisados e a apresentação dos dados em mapas temáticos, além de métodos gráficos e numéricos para sumarizar dados, como histogramas, diagramas de dispersão e estatística descritiva;
- **Manipulação** – permite a criação de novos dados espaciais usando propriedades e relacionamento entre entidades espaciais, destacando-se a função de agregação de dados espaciais e a geração e análise de áreas de influência;

- **Análise exploratória** – tem como objetivo visualizar e descrever distribuições espaciais, identificar padrões de associação espacial (aglomerados espaciais ou *clusters*), identificar observações atípicas (valores extremos ou *outliers*) ou a existência de instabilidades espaciais (não-estacionaridade). Wise *et al.* (1998) salientam que, neste grupo, os métodos são descritivos e não confirmatórios, desejando-se detectar padrões, elaborar hipóteses e estimar modelos espaciais; as técnicas são visuais e resistentes a valores discrepantes; poucas transformações são efetuadas sobre os dados originais. Os autores salientam, ainda, que é importante o uso de instrumental computacional gráfico, por serem mais intuitivos que os métodos estatísticos e numéricos;
- **Análise confirmatória** - agrupa os processos quantitativos de modelagem, estimação e validação necessários à análise de componentes espaciais. Destacam-se, neste grupo, o ferramental disponibilizado pela estatística espacial e pela econometria espacial.



**Figura 3.1:** Inter-relação entre os quatro grupos de ferramentas de análise espacial – adaptado de Anselin (1992) e Teixeira (2003)

### 3.6. ESTATÍSTICA ESPACIAL

Os dois últimos grupos de ferramentas de análises espaciais apresentados no item anterior e destacados na Figura 3.1 são os objetos de estudo da estatística espacial. Pode-se dizer que a estatística espacial abrange todo o ferramental matemático destinado a fornecer ao planejador a capacidade de estabelecer critérios quantitativos de análise exploratória e confirmatória em conjuntos de dados espaciais (Teixeira, 2003).

Para salientar a diferença da estatística espacial dos ramos mais tradicionais da estatística, Bivand (1998), fazendo um paralelo com o exemplo clássico da probabilidade, que diz que a retirada de uma bola colorida de uma urna não influencia a cor da próxima, destaca que, do ponto de vista da estatística espacial, o exemplo mais correto seria o da retirada de uma uva de um cacho, onde o fato de se retirar uma uva azeda sugere que as próximas também o serão. Ou seja, na estatística espacial parte-se do pressuposto de que os dados são *espacialmente dependentes*, o que impossibilita a utilização de muitos dos modelos estatísticos tradicionais, que exigem como atributo básico a *independência* entre os eventos observados.

Portanto, o objetivo principal da estatística espacial é caracterizar padrões espaciais entre os dados. Segundo Anselin (1992), esses padrões espaciais causam problemas de mensuração, conhecidos como efeitos espaciais, tais como dependência espacial e heterogeneidade espacial, que afetam a validade dos métodos estatísticos tradicionais.

### 3.7. DEPENDÊNCIA ESPACIAL E AUTOCORRELAÇÃO ESPACIAL

Para Câmara *et al.* (2002b), um conceito importante para a compreensão e análise dos fenômenos espaciais é a *dependência espacial*. Tal afirmação se baseia na citação de Tobler (1970), referindo-se à Primeira Lei da Geografia, de que todos os objetos no espaço estão relacionados, porém objetos mais próximos no espaço estão mais relacionados, e também na afirmação de Cressie (1991)<sup>7</sup> *apud* Câmara *et al.*

---

<sup>7</sup> Cressie, N. (1991) *Statistics for Spatial Data*. Chichester, John Wiley.

(2002b), de que a dependência espacial está presente em todas as direções, mas fica mais fraca à medida que aumenta a dispersão na localização dos dados. Isto significa que a maior parte das ocorrências, sejam naturais ou sociais, apresentam entre si uma relação que depende da distância. Câmara *et al.* (2002b) cita como exemplo o fato de que se for encontrada poluição num trecho de um lago é provável que locais próximo à amostra também estejam poluídos.

No caso mais específico de transportes, Miller (1999) afirma que os modelos de regressão utilizados para previsão de demanda por transportes utilizam variáveis explicativas, tais como uso do solo (no caso de ZTs) e fatores demográficos (no caso de domicílios). Uma vez que essas variáveis são dados geográficos, conseqüentemente sofrerão os efeitos da dependência espacial. O autor salienta que a questão ainda não é bem reconhecida na literatura de modelagem de transportes, apesar de constituir um problema potencial.

A expressão computacional do conceito de dependência espacial é a *autocorrelação espacial* (Câmara *et al.*, 2002b), termo derivado do conceito estatístico de *correlação*, utilizado para mensurar o relacionamento entre duas variáveis aleatórias. A preposição “auto” indica que a medida de correlação é realizada com a mesma variável aleatória, medida em locais distintos do espaço. Diz-se que a autocorrelação é *positiva* se a ocorrência de um dado evento influencia para que outro semelhante aconteça ao seu redor, o que implica em uma distribuição aglomerada de eventos. Se a ocorrência de um mesmo evento dificulta ou impede a ocorrência de outros semelhantes em seu entorno, diz-se que a autocorrelação é *negativa*, o que resulta em uma distribuição aproximadamente equidistante dos eventos (Teixeira, 2003).

### **3.8. ANÁLISE EXPLORATÓRIA DE DADOS ESPACIAIS**

Como já mencionado, as técnicas de análise exploratória aplicadas a dados espaciais são essenciais ao desenvolvimento das etapas da modelagem estatística espacial. Consistem em adaptações de ferramentas usuais e são referidas por Anselin (1995) como ESDA (*Exploratory Spatial Data Analysis*), podendo ser consideradas uma extensão da EDA (*Exploratory Data Analysis*) (Messner e Anselin, 2002).

A EDA consiste em ferramentas estatísticas descritivas e gráficas com a intenção de detectar padrões nos dados e sugerir hipóteses, através da imposição de um mínimo de estrutura possível (Tukey, 1977<sup>8</sup>, *apud* Messner e Anselin, 2002). Já a ESDA está focalizada em técnicas que descrevem e visualizam a distribuição espacial, identificam localizações atípicas ou *outliers* espaciais, procuram descobrir padrões de associação espacial, agrupamentos de valores semelhantes (*clusters*) e sugerir regimes espaciais ou outras formas de heterogeneidade espacial.

As técnicas ESDA são essenciais no estudo de processos de possível difusão, por estarem baseadas no conceito de autocorrelação espacial, gerando como resultados índices globais e locais. Os índices das técnicas ESDA medem a associação espacial (Índice de Moran, por exemplo) e podem ser visualizados através de gráfico de espalhamento (*Moran Scatterplot*) e mapas (*Box map*). Auxiliam na identificação de agrupamentos de objetos cujos atributos possuem valores altos e baixos, áreas de transição e casos atípicos. Em toda técnica ESDA são encontrados três elementos básicos: matriz de proximidade espacial ( $W$ ), vetor dos desvios ( $Z$ ) e vetor de médias ponderadas ( $W_z$ ).

### 3.8.1. Matriz de proximidade espacial ( $W$ )

Quando se trata do estudo da dependência espacial de dados de áreas, o elemento chave desta análise é o conceito da matriz de proximidade espacial. A matriz de proximidade espacial ( $W$ ) estima a variabilidade espacial de dados de áreas, 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 regiões  $i$  e  $j$ . Na matriz  $W$  ( $n \times n$ ) cada elemento  $w_{ij}$  representa uma medida de proximidade entre  $A_i$  e  $A_j$ , sendo  $A_i$  e  $A_j$  as zonas que estão sendo analisadas.

O valor  $W$  pode ser baseado em algumas medidas de adjacência, tais como:

- $w_{ij} = 1$ , se o centróide de  $A_i$  está a uma determinada distância de  $A_j$ , e  $w_{ij} = 0$ , caso contrário;

---

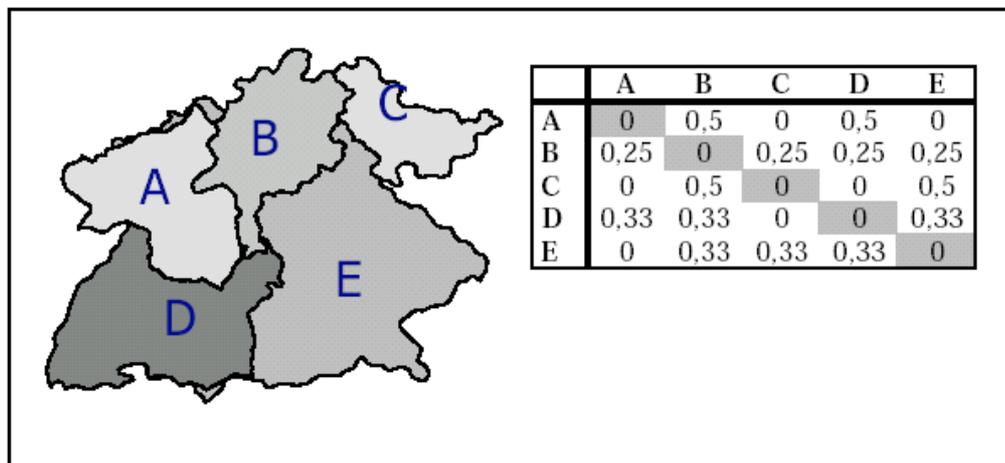
<sup>8</sup> Tukey, J. (1977) *Exploratory Data Analysis*. Reading: Addison Wesley.

- $w_{ij} = 1$ , se  $A_i$  compartilha um lado comum com  $A_j$ , e  $w_{ij} = 0$ , caso contrário;
- $w_{ij} = l_{ij}/l_i$ , onde  $l_{ij}$  é o comprimento da fronteira entre  $A_i$  e  $A_j$  e  $l_i$  é o perímetro de  $A_i$ .

Como a matriz de proximidade é utilizada em cálculo de indicadores na fase de análise exploratória, é muito útil normalizar suas linhas, para que a soma dos pesos de cada linha seja igual a 1:

$$w_i = \sum_j w_{ij} = 1 \quad (1)$$

Câmara *et al.* (2002a) salientam que a normalização simplifica vários cálculos de índice de autocorrelação espacial. Um exemplo de matriz de proximidade espacial normalizada pelas linhas é apresentado na Figura 3.2. Muitas vezes torna-se importante especificar o grau de proximidade de diferentes ordens, denominadas por faixas de distâncias (ou “lags”), onde a proximidade de primeira ordem é obtida pela conexão de duas zonas, a de segunda ordem por zonas que não se conectam diretamente mas são contíguas, e assim por diante.



**Figura 3.2:** Matriz de proximidade espacial de primeira ordem, normalizada pela linha (Câmara *et al.*, 2002a)

### 3.8.2. Média móvel espacial ( $W_z$ )

Para o cálculo da média móvel espacial é necessário conhecer o vetor dos desvios ( $Z$ ), calculado a partir da média geral ( $\mu$ ). Cada elemento deste vetor é calculado subtraindo-se o valor de  $\mu$  do valor do atributo para cada região:

$$z_i = y_i - \mu \quad (2)$$

onde:

- $z_i$  = vetor de desvios;
- $y_i$  = valores do atributo;
- $\mu$  = média geral.

A média móvel espacial ( $W_z$ ), também chamada de vetor de médias ponderadas ou média dos valores dos vizinhos, é uma medida útil para o cálculo da variação da tendência espacial. A estimativa da média móvel espacial ( $\hat{\mu}_i$ ) é obtida pela seguinte expressão:

$$\hat{\mu}_i = \frac{\sum_{j=1}^n w_{ij} \times y_j}{\sum_{j=1}^n w_{ij}} \quad (3)$$

Segundo Câmara *et al.* (2002a), um aspecto fundamental na análise exploratória espacial é a caracterização da dependência espacial, que mostra como os valores estão correlacionados no espaço. A utilização da média móvel espacial ( $W_z$ ) permite a identificação de padrões e tendências espaciais. Pode ser empregada para estudar a dependência espacial do valor de determinado atributo através da verificação da análise da autocorrelação dos valores, de acordo com sua localização no espaço, advindo daí o conceito de autocorrelação espacial. As funções utilizadas para estimar quanto o valor observado de um atributo de uma região é dependente dos valores dessa mesma variável nas localizações vizinhas são a *autocorrelação espacial* e o *variograma*.

### 3.8.3. Indicadores de dependência espacial: medidas de autocorrelação

A informação que se busca através do cálculo da autocorrelação espacial é de quanto o valor de uma variável em uma área é parecido com o do vizinho mais próximo e quão diferente do vizinho mais distante. A autocorrelação espacial ocorre quando observações organizadas no espaço influenciam-se mutuamente. Segundo Câmara *et al.* (2002a), a autocorrelação espacial mede o quanto o valor observado de um atributo numa região é dependente dos valores desta mesma variável nas localizações vizinhas.

### 3.8.3.1. Índice I de Moran (global)

Uma das formas de medir a autocorrelação espacial é através do índice I de Moran, cujo valor varia de  $-1$  a  $+1$  e indica quanto cada área analisada é semelhante à sua vizinhança imediata:

$$I = \frac{Z' \times W_Z}{Z' \times Z} \quad (4)$$

onde:

- $I$  = índice de autocorrelação espacial;
- $Z'$  = vetor transposto dos desvios;
- $W_Z$  = média móvel espacial.

Valores de  $I$  próximos de zero indicam a inexistência de autocorrelação espacial significativa entre os valores dos objetos e seus vizinhos. Valores positivos para o índice indicam autocorrelação espacial positiva, ou seja, o valor do atributo de um objeto tende a ser semelhante aos valores dos seus vizinhos. Valores negativos para o índice indicam autocorrelação espacial negativa. Ou seja, o Índice I de Moran é uma das formas de detecção de similaridade entre as áreas.

Segundo Câmara *et al.* (2002a), este indicador é uma medida de correlação espacial usada para detectar afastamentos de uma distribuição espacial aleatória, o que permite observar a existência de padrões espaciais, isto é, permite detectar a similaridade de áreas de primeira ordem. Os desvios em relação à média são multiplicados pelos desvios da vizinhança. O índice de Moran testa se as áreas vizinhas apresentam maior semelhança quanto ao indicador estudado do que o esperado num padrão aleatório. Tendo em mãos estes resultados, a maior preocupação deve ser com sua validade estatística. Para verificar se a correlação obtida é significativa, duas abordagens são possíveis:

- **Teste de pseudo-significância:** são geradas diferentes permutações dos valores de atributos associados às regiões, onde cada permutação gera um novo arranjo espacial, pois os valores são redistribuídos entre as áreas. Como apenas um dos arranjos corresponde à situação observada, é construída uma situação empírica

de I. Caso o valor de I corresponda ao extremo da distribuição simulada, tratar-se-á de um evento com significância;

- **Distribuição Aproximada:** para um número suficiente de subregiões e supondo que as variáveis aleatórias associadas a cada localização do atributo são independentes e normalmente distribuídas, assume-se que o índice I possui distribuição aproximadamente normal.

Conforme mencionado por Câmara *et al.* (2002a), a hipótese implícita do cálculo do índice de Moran é a estacionaridade de primeira e segunda ordem, e o índice perde sua validade ao ser calculado para dados não estacionários. A estacionariedade existe quando a autocorrelação espacial apresenta comportamento homogêneo na região de estudos. Quando existir não-estacionariedade de primeira ordem (tendência), os vizinhos tenderão a ter valores mais parecidos que áreas distantes, pois cada valor é comparado à média global, inflacionando o índice. Da mesma forma, se a variância não é constante, nos locais de maior variância o índice será mais baixo e vice-versa. Quando o dado é não-estacionário, a função de autocorrelação continua decaindo mesmo após ultrapassar a distância onde há influências locais. Algumas variações deste modelo são o teste C de Geary e o teste Ipop.

### 3.8.3.2. Teste C de Geary

O Teste C de Geary difere do teste I de Moran por utilizar a diferença entre os pares, enquanto que Moran utiliza a diferença entre cada ponto e a média global. Assim, o indicador C de Geary assemelha-se ao variograma, e o I de Moran ao correlograma.

A obtenção do índice global de associação espacial através da estatística C de Geary dá-se pela expressão (5), podendo ser utilizada qualquer definição da matriz de proximidade espacial, sendo habitual normalizá-la.

$$C = \frac{(N-1) \sum_{ij} w_{ij} (x_i - x_j)^2}{2S_0 \sum_{i=1}^N (x_i - \bar{x})^2} \quad i \neq j \quad (5)$$

onde:

- $x_i$ : valor da variável  $x$  na região  $i$ ;
- $x_j$ : valor da variável  $x$  na região  $j$ ;

- $w_{ij}$ : elementos da matriz de proximidade espacial;
- $\bar{x}$ : média amostral;
- $S_0 = \sum_i \sum_j w_{ij}$ .

### 3.8.3.3. Índice Local de Moran (Estatística LISA)

Enquanto os indicadores globais, como o índice I de Moran, fornecem um único valor como medida de associação espacial para todo o conjunto de dados, os indicadores locais produzem um valor específico para cada área, permitindo, assim, a identificação de agrupamentos de áreas com valores de atributos semelhantes (*clusters*), valores anômalos (*outliers*) e de mais de um regime espacial. Anselin (1995) refere-se a estes indicadores como estatística LISA (*Local Indicators of Spatial Association*), os quais devem satisfazer dois requisitos básicos:

- Permitir a identificação de padrões de associação espacial significativos;
- A soma dos indicadores LISA para todas as observações deve ser proporcional ao índice global de associação espacial.

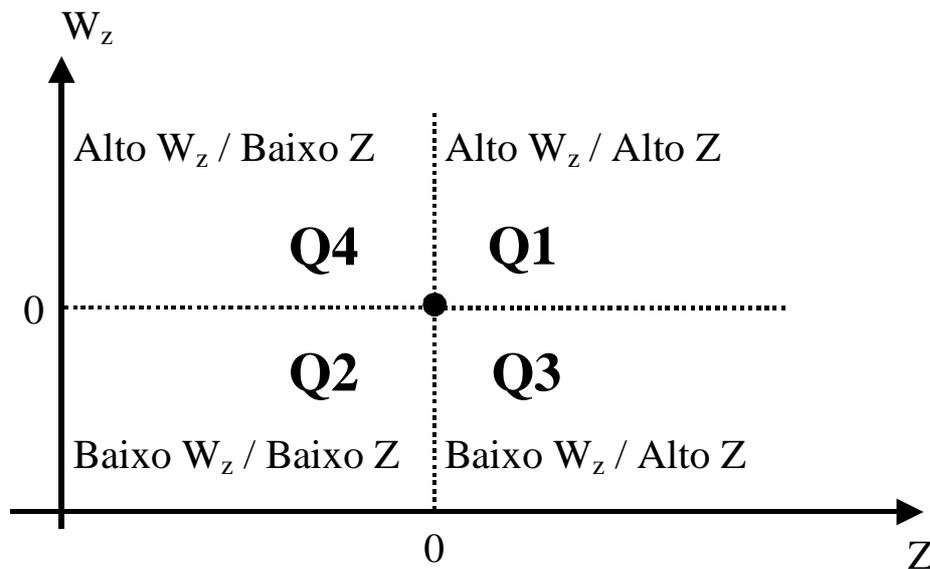
A significância estatística do uso do índice de Moran local é computada de forma similar ao caso do índice global. Depois de calculado o índice para cada área, permuta-se aleatoriamente o valor das demais áreas, até ser obtida uma pseudo-distribuição para a qual se possa computar os parâmetros de significância. É útil gerar mapas (*LISA Map* ou *Moran Map*) indicando as regiões que apresentam correlação local significativamente diferente do resto dos dados, pois são áreas com dinâmica espacial própria, ou seja, “bolsões” de não-estacionaridade e merecem análise detalhada. Autocorrelação com significância superior a 95% indica, a princípio, áreas com grande semelhança em relação a seus vizinhos.

### 3.8.4. Visualização da dependência espacial

Como comentado anteriormente é muito útil gerar gráficos e mapas para visualizar a dependência espacial, auxiliando, desta forma, as análises. Atualmente existem *software* livres na internet, como é o caso do *software* Spring, que realiza estas operações, o que torna estas análises mais diretas e facilitam a geração de mapas e gráficos.

### 3.8.4.1. Diagrama de espalhamento de Moran

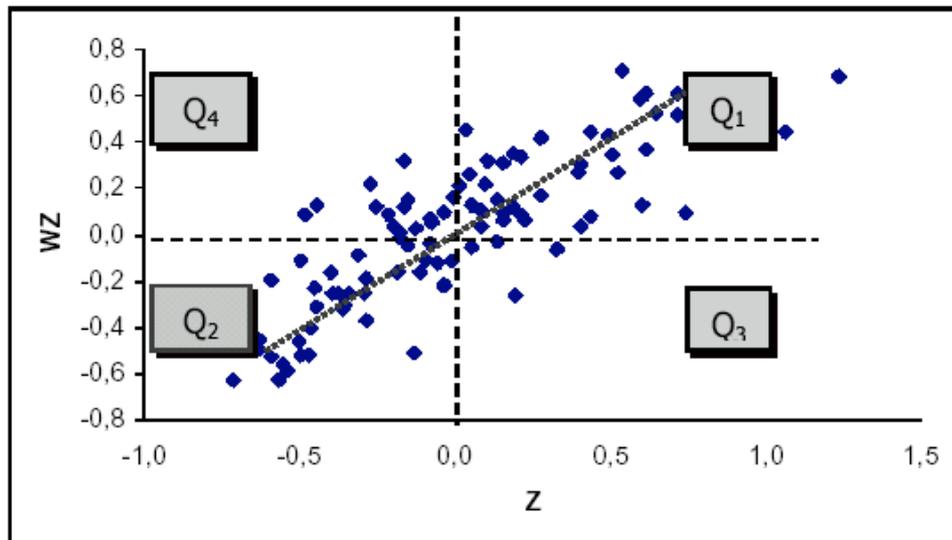
Uma maneira de visualizar a dependência espacial é através do diagrama de espalhamento de Moran (Figura 3.3). Construído com base nos valores normalizados ( $Z$ ) e comparados com a média dos vizinhos ( $W_z$ ), em um gráfico bidimensional dividido em quatro quadrantes, permite analisar o comportamento da variabilidade espacial.



**Figura 3.3:** Gráfico de espalhamento de Moran

Um exemplo é mostrado na Figura 3.4, que apresenta o diagrama de espalhamento de Moran para o índice de exclusão/inclusão social de São Paulo, Censo de 1991 (Câmara *et al.*, 2002a). O índice  $I$  de Moran é equivalente ao coeficiente de regressão linear, que indica a inclinação da reta de regressão ( $\alpha$ ) de  $W_z$  em  $Z$ , e os quadrantes podem ser interpretados como:

- **Q1** (valores positivos, médias positivas) e **Q2** (valores negativos, médias negativas): indicam pontos de associação espacial positiva, no sentido que uma localização possui vizinhos com valores semelhantes.
- **Q3** (valores positivos e médias negativas) e **Q4** (valores negativos e médias positivas): indicam pontos de associação espacial negativa, no sentido que uma localização possui vizinhos com valores distintos.



**Figura 3.4** – Diagrama de espalhamento de Moran para o índice de exclusão/inclusão social de São Paulo, Censo de 1991. Fonte: Câmara *et al.* (2002a)

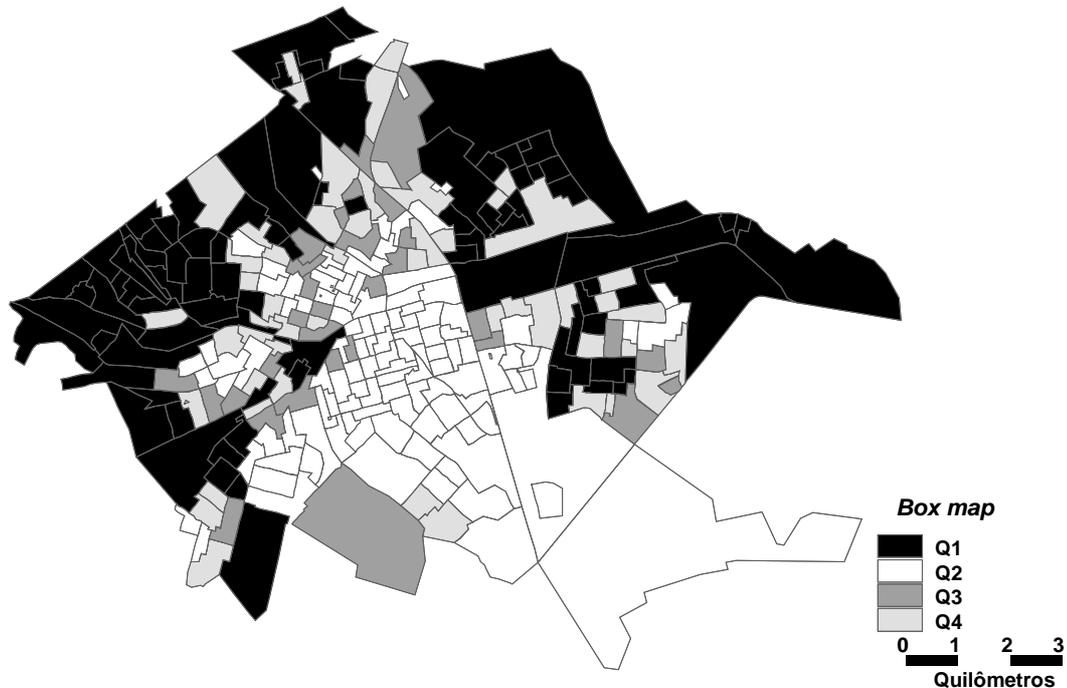
#### 3.8.4.2. Box Map

O diagrama de espalhamento de Moran também pode ser apresentado na forma de um mapa temático bidimensional (ou *Box Map*), no qual cada polígono é apresentado indicando-se seu quadrante no diagrama de espalhamento. Para Serrano e Valcarce (2000), através do *Box Map* pode-se identificar *outliers* espaciais para dados de área. Um exemplo pode ser visto na Figura 3.5, que analisa a variável *porcentagem de viagens de ônibus* na cidade de Bauru, SP. O estudo apresentado por Krempi (2004) procura explorar a união de ferramentas de estatística espacial e SIG com o propósito de analisar as relações entre aspectos da distribuição espacial de características da população de uma cidade média brasileira e os diversos níveis de acessibilidade por diferentes modos de transporte nela observados.

#### 3.8.4.3. LISA Map

Para análise dos Indicadores Locais de Associação Espacial, o mapa denominado *LISA Map* indica as regiões que apresentam correlação local significativamente diferente do resto dos dados. Essas áreas possuem dinâmica espacial própria e merecem análise detalhada. Na geração do mapa os índices são classificados em quatro grupos (não significativos, com significância de 95 %, de 99 % e de 99,9 %), como mostra a Figura 3.6, que apresenta um exemplo de *Lisa Map* para a análise da variável *porcentagem de viagens de ônibus* (Krempi, 2004), onde os percentuais são

classificados em função da significância dos valores de seus índices locais. Trata-se de uma ferramenta que identifica os agrupamentos com valores dos indicadores de associação espacial local significativos, sem, contudo, diferenciá-los em agrupamentos de altos ou baixos valores.

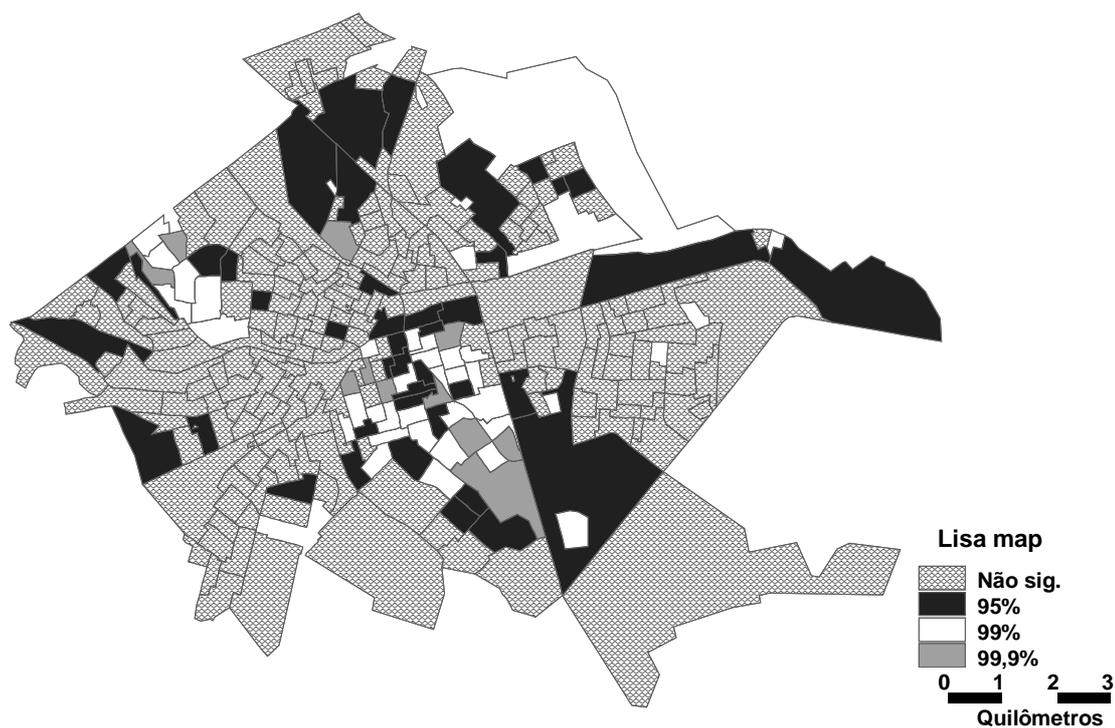


**Figura 3.5:** *Box Map* da variável *porcentagem de viagens de ônibus*, para a cidade de Bauru. Fonte: Krempi (2004)

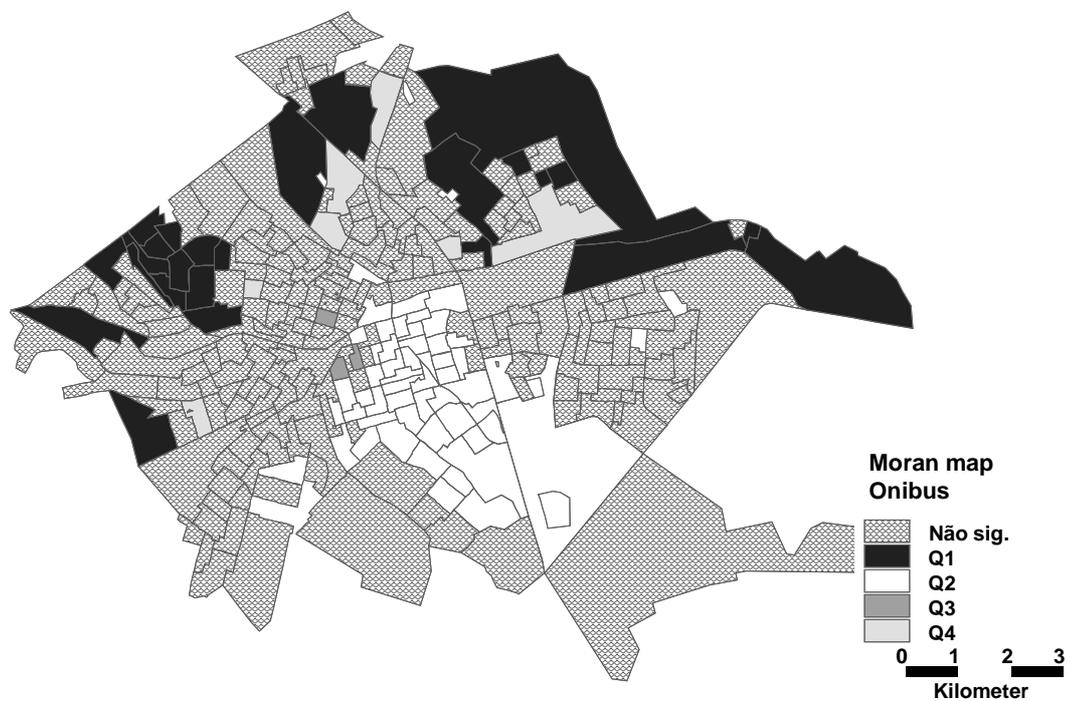
Na geração do *LISA Map*, através do *software* Spring, por exemplo, a avaliação da significância é feita comparando-se os valores de LISA com uma série de valores obtidos por meio de permutações dos valores dos atributos dos vizinhos (número definido pelo usuário, podendo chegar a 9999 permutações), sob a hipótese nula de não existência de autocorrelação espacial.

#### 3.8.4.4 Moran Map

No *Moran Map*, de forma semelhante ao *LISA Map*, somente são apresentados os objetos para os quais os valores de LISA foram considerados significativos ( $p < 0,05$ ), porém, com classificação em quatro grupos, conforme os quadrantes aos quais pertençam no gráfico de espalhamento. Os demais objetos ficam classificados como "sem significância". Um exemplo é apresentado na Figura 3.7, também para a análise da variável *porcentagem de viagens de ônibus* para a cidade de Bauru (Krempi, 2004).



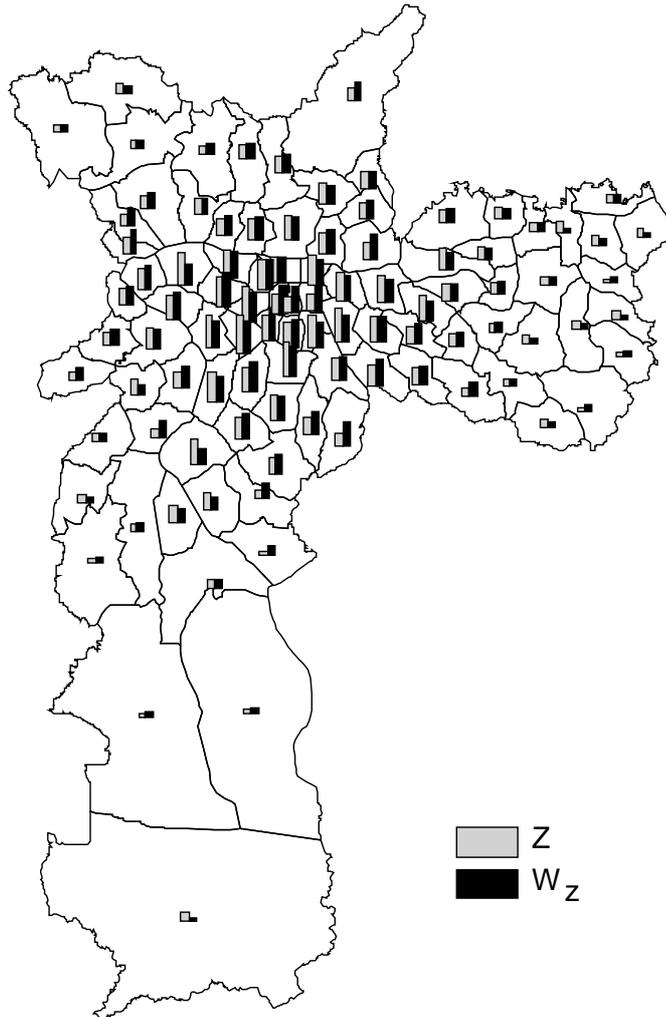
**Figura 3.6:** *LISA Map* da variável *porcentagem de viagens de ônibus*, para a cidade de Bauru. Fonte: Krempi (2004)



**Figura 3.7:** *Moran Map* da variável *porcentagem de viagens por ônibus*, para a cidade de Bauru. Fonte: Krempi (2004)

#### 3.8.4.5. Gráfico de Barras

O gráfico de barras  $Z \times W_z$  da Figura 3.6, plotado no mapa sobre as áreas correspondentes, permite ainda visualizar simultaneamente o valor do atributo ( $Z$ ) e do valor correspondente à sua vizinhança ( $W_z$ ). A altura das barras é proporcional aos valores de  $Z$  e  $W_z$  de um determinado atributo. Um exemplo é apresentado na Figura 3.8 para o índice de Exclusão/Inclusão Social de São Paulo (INPE, 2004).



**Figura 3.8:** Gráfico de barras do Mapa de Exclusão/Inclusão Social de São Paulo, 2000

Fonte: <http://www.dpi.inpe.br/geopro/exclusao/mapas.html>

### 3.9. REGRESSÃO ESPACIAL

Modelos de regressão são amplamente utilizados na etapa de geração de viagens da análise de demanda por transportes. São ferramentas estatísticas que utilizam o relacionamento existente entre duas ou mais variáveis, de maneira que uma delas possa

ser descrita ou seu valor estimado a partir das demais. No entanto, na situação de dados espaciais, quando está presente a autocorrelação espacial, as estimativas do modelo devem incorporar essa estrutura espacial, uma vez que a dependência entre as observações altera o poder explicativo do modelo. Segundo Câmara *et al.* (2002a), a significância dos parâmetros é usualmente superestimada e a existência de variações de larga escala pode até mesmo induzir à presença de associações espúrias.

Para um melhor entendimento dos modelos de regressão espacial, Fotheringham *et al.* (2000) apresentam, inicialmente, os modelos de regressão linear em termos de  $n$  observações de uma variável dependente (vetor  $Y$ ) e  $n$  observações de  $k - 1$  variáveis explicativas contidas na matriz  $X$  ( $n \times k$ ), conforme expresso na equação matricial:

$$Y = X \times \beta + \varepsilon, \varepsilon \approx MVN(0, \sigma^2 I) \quad (6)$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & \dots & X_{1k-1} \\ 1 & X_{21} & \dots & X_{2k-1} \\ \dots & \dots & \dots & \dots \\ 1 & X_{n1} & \dots & X_{nk-1} \end{bmatrix} \times \begin{bmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_{k-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{bmatrix}$$

onde:

- $Y$ : variável dependente, composta de um vetor ( $n \times 1$ ) de observações tomadas em cada uma das  $n$  áreas;
- $X$ : matriz ( $n \times k$ ) com  $k - 1$  variáveis explicativas, também tomadas nas  $n$  áreas;
- $\beta$ : vetor ( $k \times 1$ ) com os coeficientes de regressão;
- $\varepsilon$ : vetor ( $n \times 1$ ) de erros aleatórios ou resíduos;
- $I$ : matriz identidade;
- $0$ : vetor zero;
- $MVN$ : distribuição normal multivariável, com vetor médio  $M$  e matriz variância-covariância  $S$ .

Tipicamente, quando se faz uma análise de regressão, procura-se alcançar dois objetivos (Câmara *et al.*, 2002a):

- a) Encontrar um bom ajuste entre os valores preditos pelo modelo e os valores observados da variável dependente;
- b) Descobrir quais das variáveis explicativas contribuem de forma significativa para o relacionamento linear.

A hipótese padrão é que as observações não são correlacionadas e, portanto, os resíduos  $\varepsilon_i$  do modelo são independentes e não-correlacionados com a variável dependente, têm variância constante e apresentam distribuição normal com média zero. O problema decorrente da consideração de que a variância de  $\varepsilon$  é  $\sigma^2 I$  conduz à necessidade de ampliação para matrizes variância-covariância (C) que permitam a não-independência dos resíduos (Fotheringham *et al.*, 2000), passando-se a assumir a forma mais geral:

$$\varepsilon \approx MVN(0, C) \quad (7)$$

No caso de dados espaciais, portanto, onde está presente a dependência espacial, é muito pouco provável que a hipótese padrão de observações não correlacionadas seja verdadeira. No caso mais comum, os resíduos continuam apresentando a autocorrelação espacial presente nos dados, que pode se manifestar por diferenças regionais sistemáticas nas relações do modelo ou, ainda, por uma tendência espacial contínua.

A investigação dos resíduos da regressão, em busca de sinais de estrutura espacial, é o primeiro passo em uma regressão espacial. As ferramentas usuais de análise gráfica e o mapeamento de resíduos podem dar as primeiras indicações de que os valores observados estão mais correlacionados do que seria esperado sob uma condição de independência (Fotheringham *et al.*, 2000). Neste caso, utilizar os testes de autocorrelação espacial – Moran e Geary – nos resíduos da regressão informa sobre sua presença. Em caso de existir autocorrelação, deve-se especificar um modelo que considere a interferência causada pela mesma.

A análise de regressão em dados espaciais incorpora, na modelagem, a dependência espacial entre os dados, melhorando o poder preditivo do modelo. Primeiramente, faz-se a análise exploratória com intuito de identificar a estrutura de dependência nos dados, visando a definição da forma de incorporação dessa dependência ao modelo de regressão. Câmara *et al.* (2002a) apresentam dois tipos básicos de modelagem que permitem incorporar o efeito espacial: as de forma Global e as de forma Local.

### 3.9.1. Modelos com efeitos espaciais globais

Os modelos de forma Global capturam a estrutura espacial através de um único parâmetro que é adicionado ao modelo de regressão tradicional (Câmara *et al.*, 2002a). Os modelos mais simples são: modelo espacial auto-regressivo misto (*Spatial Auto Regressive* = SAR ou *Spatial Lag Model*) e modelo do erro espacial (*Conditional Auto Regressive* = CAR ou *Spatial Error Model*).

#### 3.9.1.1. SAR

A autocorrelação espacial ignorada é atribuída à variável dependente  $Y$ . Considera-se a dependência espacial através da adição, ao modelo de regressão, de um novo termo na forma de uma relação espacial para a variável dependente. Formalmente, Anselin (2002) apresenta o modelo SAR por:

$$Y = \rho WY + X\beta + \varepsilon \quad (8)$$

onde:

- $Y$  = variável dependente;
- $X$  = variáveis independentes;
- $\beta$  = coeficientes de regressão;
- $\varepsilon$  = erros aleatórios com média zero e variância  $\sigma^2$ ;
- $W$  = matriz de vizinhança espacial ou matriz de ponderação espacial;
- $\rho$  = coeficiente espacial autoregressivo.

A hipótese nula para a não existência de autocorrelação é que  $\rho = 0$ . A idéia básica é incorporar a autocorrelação espacial como componente do modelo.

#### 3.9.1.2. CAR

O segundo tipo de modelo de regressão espacial com parâmetros globais considera que os efeitos espaciais são um ruído, ou perturbação, ou seja, fator que precisa ser removido. Neste caso, os efeitos da autocorrelação espacial são associados ao termo de erro  $\varepsilon$  e o modelo pode ser expresso por:

$$Y = X\beta + \varepsilon, \quad \varepsilon = \lambda W\varepsilon + \xi \quad (9)$$

onde:

- $W\varepsilon$  = erros com efeito espacial;

- $\xi$  = erros aleatórios com média zero e variância  $\sigma^2$ ;
- $\lambda$  = coeficiente autoregressivo.

A hipótese nula para a não existência de autocorrelação é que  $\lambda = 0$ , ou seja, o termo de erro não é espacialmente correlacionado.

Câmara *et al.* (2002a) salientam que, na prática, a distinção entre os dois tipos de modelos de regressão espacial com parâmetros globais é difícil, pois, apesar da diferença nas suas motivações, eles são muito próximos em termos formais. Esses modelos estão incluídos em ambientes de estatística espacial avançados, como nos softwares **SpaceSat**<sup>TM</sup>, **S-Plus**<sup>TM</sup> e **R**, de domínio público.

É importante salientar, ainda, que os modelos de regressão espacial com efeitos globais partem do princípio de que o processo espacial subjacente aos dados analisados é estacionário. Isso implica que os padrões de autocorrelação espacial existentes nos dados podem ser capturados num único parâmetro. Na prática, para conjuntos de dados censitários de médio e grande porte, a natureza dos processos espaciais é tal que diversos padrões de associação espacial podem estar presentes. Essa hipótese, que se aplica também às variáveis socioeconômicas normalmente utilizadas na modelagem de demanda por transporte, pode ser verificada, por exemplo, pelos indicadores locais de autocorrelação espacial. Nesse caso, deve-se aplicar os modelos que variam no espaço, chamados de *modelos com efeitos espaciais locais*.

### 3.9.2. Modelos com efeitos espaciais locais

Quando o processo espacial é não-estacionário, os coeficientes de regressão precisam refletir a heterogeneidade espacial (Câmara *et al.*, 2002a). Para tanto, há duas grandes alternativas: (a) modelar a tendência espacial de forma contínua, com parâmetros variantes no espaço; (b) modelar a variação espacial de forma discreta, ao dividir o espaço em sub-regiões estacionárias, chamadas de regimes espaciais.

#### 3.9.2.1. Regimes Espaciais

Este modelo é o mais indicado para modelagem de demanda por transportes, onde os dados estão agrupados em áreas (ZTs) e na maior parte são não-estacionários,

visto que a idéia de regimes espaciais é dividir a região de estudo em sub-regiões, cada uma com seu padrão espacial próprio, e realizar regressões em separado, uma para cada região. As observações são classificadas em dois ou mais subconjuntos, a partir de uma variável por indicação, a saber:

$$\begin{aligned} Y_1 &= X_1\beta_1 + \varepsilon_1, & ind &= 1 \\ Y_2 &= X_2\beta_2 + \varepsilon_2, & ind &= 2 \end{aligned} \quad (10)$$

onde:

- $\mathbf{Y}$  = variável dependente;
- $\mathbf{X}$  = variáveis independentes;
- $\beta$  = coeficientes de regressão;
- $\varepsilon$  = erros aleatórios com média zero e variância  $\sigma^2$ ;
- $ind$  = índice dos regimes espaciais.

Apesar de cada regime possuir os seus próprios valores de coeficientes, estes valores são estimados conjuntamente, ou seja, todo o conjunto de observações disponível é utilizado na regressão. Para a determinação dos regimes espaciais, as técnicas de análise exploratória apresentadas no início do capítulo são muito úteis, especialmente o mapa de espalhamento de Moran e os indicadores locais de autocorrelação espacial.

Conforme Câmara *et al.* (2002a), na prática, para os dados socioeconômicos típicos de cidades brasileiras, o modelo de regimes espaciais tende a apresentar resultados melhores que os modelos de regressão simples ou de regressão espacial com efeitos globais. Isso ocorre em função das fortes desigualdades sociais no Brasil, que ocasionam descontinuidades abruptas nos fenômenos estudados, como no caso do recorte entre favelas e áreas ricas, como é freqüente nas grandes cidades do país.

### 3.9.2.2. Efeitos Espaciais Contínuos

Essa classe de modelos procura modelar fenômenos não-estacionários. Diferentemente do modelo por regimes espaciais, os efeitos espaciais são modelados de forma contínua, com duas hipóteses: (a) a existência de uma variação suave em larga escala, sem efeitos locais significativos; (b) a existência de variações locais contínuas, sem uma forte tendência global. Como não correspondem ao interesse específico deste

estudo, não cabe aqui um aprofundamento maior. Mais detalhes podem ser encontrados em Câmara *et al.* (2002a).

### 3.9.3. Diagnóstico de modelos com efeitos espaciais

A análise gráfica dos resíduos é o primeiro passo para a avaliação da qualidade do ajuste da regressão. Mapear os resíduos é uma etapa importante no diagnóstico do modelo, buscando indícios de ruptura dos pressupostos de independência. Uma alta concentração de resíduos positivos (ou negativos) numa parte do mapa é um bom indicador da presença de autocorrelação espacial. Para um teste quantitativo, o mais comum é utilizar o índice *I de Moran* sobre os resíduos.

Como os estimadores e os diagnósticos tradicionais de regressão não levam em conta os efeitos espaciais, as inferências, como, por exemplo, as indicações de qualidade de ajuste baseadas no  $R^2$  (coeficiente de determinação), serão incorretas. As conseqüências são similares às que acontecem quando uma variável explicativa significativa é omitida do modelo de regressão. Quando se quer comparar um ajuste obtido por um modelo de regressão padrão com um ajuste obtido por um dos modelos cuja especificação considera a autocorrelação espacial, uma medida como o  $R^2$  não é mais confiável.

O método mais usual de seleção de modelos de regressão baseia-se nos valores de máxima verossimilhança dos diferentes modelos, ponderando pela diferença no número de parâmetros estimados. Nos modelos com estrutura de dependência – espacial ou temporal – utilizam-se os critérios de informação onde a avaliação do ajuste é penalizada por uma função do número de parâmetros. Cabe observar que é necessário ainda levar em conta o número de parâmetros independentes ao se incluir funções espaciais nos modelos. Para cada nova variável em modelo de regressão, acrescenta-se um parâmetro.

Usualmente a comparação de modelos é feita utilizando o logaritmo da máxima verossimilhança, que é o que possui melhor ajuste para os dados observados. O Critério de Informação de Akaike (AIC) é expresso por:

$$AIC = -2 \times LIK + 2k \quad (11)$$

onde:

- LIK: é o log de verossimilhança maximizado;
- k: é o número de coeficientes de regressão.

Segundo esse critério, o melhor modelo é o que possui menor valor de AIC (Câmara *et al.*,2002a). Diversos outros critérios de informação estão disponíveis, a maior parte dos quais são variações do AIC, com mudanças na forma de penalização de parâmetros ou observações.

### **3.10. CONSIDERAÇÕES FINAIS**

Câmara *et al.* (2002a) salientam que as técnicas de análise espacial podem ampliar consideravelmente a capacidade de compreensão dos padrões espaciais associados a dados de área, especialmente quando se trata de indicadores sociais, que apresentam autocorrelação espacial global e local. Técnicas exploratórias, como os indicadores de Moran e os mapas de espalhamento de Moran são muito úteis para mostrar as agregações espaciais e indicar áreas prioritárias em termos de política pública. Métodos de estimativa bayesiana para taxas permitem a correção de efeitos associados a pequenas populações.

Modelos de regressão espacial permitem o estabelecimento das relações entre as variáveis, levando em conta os efeitos espaciais, o que pode aumentar o poder explicativo dos modelos. A geração de superfícies é uma maneira eficiente de apreensão visual dos padrões espaciais. Em resumo, estudiosos de dados socioeconômicos podem se beneficiar substancialmente das técnicas de análise espacial de áreas.

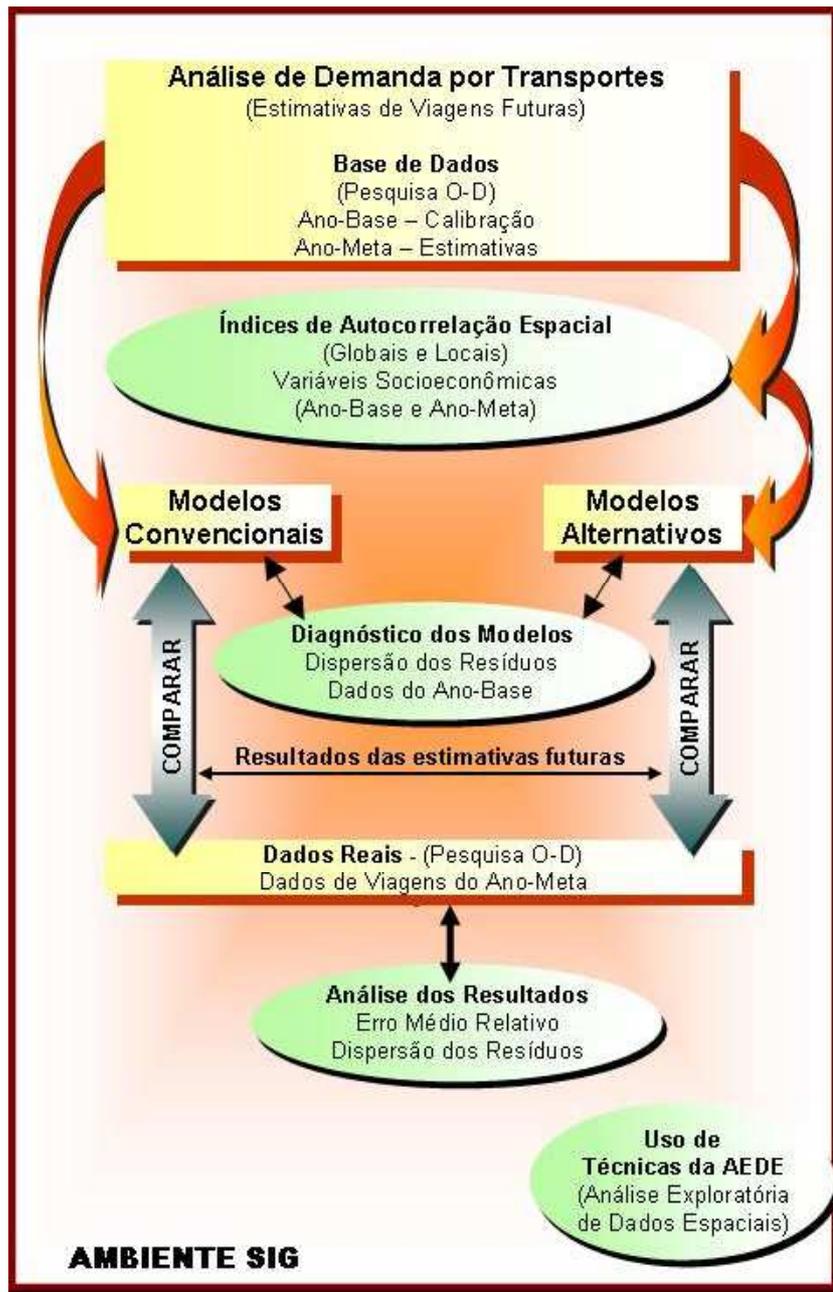
## 4. MÉTODO

---

Este capítulo descreve o método aplicado para atingir os objetivos que foram apresentados no primeiro capítulo desta dissertação. É composto por nove etapas, sendo apresentado, de forma esquemática, na Figura 4.1, em que se dá destaque para as etapas que envolvem o uso de ferramentas de Análise Exploratória de Dados Espaciais (AEDE, ou ESDA, na sigla em inglês). É importante salientar que praticamente todo o processo dá-se em ambiente SIG.

Inicialmente, deve-se ressaltar que este estudo tem por base dados de pesquisa de origem e destino (O-D) em dois períodos distintos, denominados de “**ano base**” e “**ano meta**”:

- Dados da pesquisa do “**ano base**” – são utilizados para calibração de modelos de demanda por transportes e também para teste dos melhores modelos, ou seja, aqueles que são usados para estimativas de viagens futuras, tanto por métodos tradicionais (**modelos tradicionais**) como através de métodos que introduzem variáveis que medem o grau de dependência espacial (**modelos alternativos**).
- Dados da pesquisa do “**ano meta**” – são os dados da pesquisa do ano de projeção das viagens, que servem de base para a avaliação da qualidade das estimativas futuras de viagens obtidas através da aplicação de cada um dos métodos analisados: **modelos tradicionais** e **modelos alternativos**.



**Figura 4.1:** Esquema do método aplicado, com destaque para as etapas de uso de técnicas da AEDE (Análise Exploratória de Dados Espaciais)

Deve-se salientar, também, que os efeitos da consideração de indicadores de dependência espacial em modelos de previsão de demanda por transportes verificados nesta pesquisa se restringem apenas à fase de **Viagens Produzidas de Base Domiciliar**, que corresponde a apenas uma parte da primeira etapa do tradicional **Modelo de Quatro Etapas**. Além disto, as viagens estão agregadas em **todos os modos e motivos**, por falta de informação mais desagregada nos relatórios da pesquisa do **ano base**. Desta forma o método aplicado não pretende esgotar o assunto, apenas dar início a uma

investigação que se insere no extenso e ainda pouco explorado tema de Análise Espacial, apresentando um exemplo de aplicação e verificando sua relevância para o Planejamento de Transportes.

Conforme apresentado na Figura 4.1, as técnicas da AEDE são aplicadas não somente para obtenção de indicadores de dependência espacial, necessários para a especificação dos modelos alternativos, como, também, no diagnóstico dos modelos (tradicionais e alternativos), auxiliando na seleção de modelos para as estimativas futuras, e na análise dos resultados, apresentados no **Capítulo 5**.

#### **4.1. ETAPA 1 - ANÁLISE DOS DADOS DISPONÍVEIS**

Esta etapa consiste na verificação de pesquisas O-D disponíveis para o local de estudo e análise das características de cada uma delas para definição das duas mais apropriadas. São necessárias duas pesquisas em períodos distintos (**ano base** e **ano meta**), com características compatíveis em termos de metodologia de coleta de dados, tipo de dados, zoneamento adotado etc.

É importante verificar, também, a forma de disponibilidade dos dados, pois para o estudo é necessário que ambas estejam ou possam ser georreferenciadas e facilmente transferíveis para um mesmo ambiente SIG.

#### **4.2. ETAPA 2 - DEFINIÇÃO DA ÁREA DE ESTUDOS**

Consiste na definição da abrangência da área de estudos com base na disponibilidade de dados analisados na etapa anterior, em termos de limites (municipais, metropolitanos etc) e número de ZTs. São consideradas apenas as ZTs pesquisadas, ou seja, as que apresentam dados coletados.

#### **4.3. ETAPA 3 - PREPARAÇÃO DA BASE DE DADOS**

A preparação da base de dados engloba atividades preliminares de digitalização e importação dos dados para o ambiente SIG, caso não se encontrem georreferenciados, e a posterior verificação do zoneamento adotado nas duas pesquisas selecionadas,

através de uma análise comparativa para definição da melhor forma de compatibilização das mesmas.

Como mais de um *software* SIG é utilizado nesta pesquisa, uma vez que ainda não existe nenhum que contemple todas as ferramentas de análise espacial e de planejamento envolvidas neste estudo, é necessário, primeiramente, verificar quais deles serão utilizados e desenvolver um método para facilitar a migração da base georreferenciada e dos dados entre os mesmos, uma vez que, devido a diferentes sistemas utilizados pelos *software*, o processo não é direto.

#### 4.4. ETAPA 4 - ANÁLISE DAS VARIÁVEIS INDEPENDENTES

Uma vez definida a variável dependente, que, no caso deste estudo, são as **Viagens Produzidas de Base Domiciliar (VPBD)**, é necessário verificar quais são os dados socioeconômicos incluídos nas pesquisas selecionadas que podem ser possíveis variáveis explicativas para os modelos de demanda. Alguns critérios devem ser levados em consideração:

1. Variáveis citadas na literatura para estimativas de **VPBD**;
2. Variáveis, entre as disponíveis na base de dados do **ano base**, que apresentam correlação linear significativa com a variável dependente **VPBD**;
3. Variáveis comuns às duas pesquisas (**ano base** e **ano meta**), uma vez que devem entrar nos modelos na fase de calibração (**ano base**) e posteriormente para as estimativas (**ano meta**).

A análise de “diagramas de dispersão” é o primeiro passo para verificar a existência de relações lineares entre a variável dependente e as variáveis independentes. No caso de ser verificada nenhuma relação explícita, o próximo passo é a verificação da possibilidade de transformação das variáveis.

É importante salientar que a análise de correlação linear, mencionada no segundo critério, pode ser realizada em ambiente SIG. Alguns programas de computador, como é o caso do TransCAD, apresentam ferramentas de estatística que permitem verificar o valor do coeficiente de correlação **R** (que varia de **-1** a **1**) entre

tantos campos (ou variáveis), de uma camada do mapa ou de um *dataview*, quanto se deseja analisar.

No caso do TransCAD, os resultados são apresentados em uma matriz de correlação simétrica, que pode ser utilizada para a análise de dados de entrada para um modelo linear:

- Valores de **R** próximos de **1** entre duas variáveis indicam que existe **alta correlação positiva**, ou seja, que os valores de ambas tendem a crescer ou decrescer conjuntamente;
- Valores de **R** próximos de **-1** entre duas variáveis indicam que existe **alta correlação negativa**, ou seja, que os valores de uma variável tendem a crescer enquanto os valores da outra decrescem.

Dessa forma, uma correlação de 1 (ou -1) indica que as duas variáveis são **linearmente dependentes** e uma correlação baixa (próxima de zero) indica que não existe relação linear previsível entre as duas variáveis, ou seja, são **linearmente independentes**. Alta correlação é desejável entre as variáveis independentes e a variável dependente, porém é indesejável entre duas variáveis independentes, uma vez que o acréscimo de uma concomitantemente com o da outra, além de não contribuir com nova informação para o modelo, pode também interferir na habilidade do modelo de discernir sobre a relativa importância das variáveis.

A definição de variáveis derivadas (**taxas**) também pode ser incluída nesta etapa. Geralmente, nas pesquisas O-D, os dados socioeconômicos são fornecidos em forma agregada (por ZT, por exemplo), com variação espacial discreta, cujos atributos são os totais das contagens das diferentes variáveis. Há, conseqüentemente, desvantagens inerentes aos dados de áreas (a MAUP e o problema de descontinuidade das fronteiras, apresentados no **Capítulo 3**, por exemplo), o que pode ocasionar distorções significativas nas médias zonais.

Uma alternativa, apresentada por Teixeira (2003), para o problema da variação espacial discreta, é supor que os dados apresentam variação espacial contínua. Dessa forma, as contagens agregadas podem ser transformadas em taxas ou indicadores, pois o que varia continuamente no espaço são as taxas e não as contagens agregadas.

É importante verificar, também, se a pesquisa do **ano base** não inclui algum estudo de planejamento com estimativas futuras. Em caso afirmativo, deve-se verificar quais as variáveis são utilizadas e considerá-las nos modelos de previsão em estudo.

#### 4.5. ETAPA 5 - ANÁLISE DA AUTOCORRELAÇÃO ESPACIAL

Esta etapa, em que são aplicadas as técnicas da AEDE, é o ponto chave da pesquisa, pois, conforme apresentado no início deste capítulo (Figura 4.1), ela tem relação com diversas outras etapas de fundamental importância para o estudo, tais como:

1. Definição das **variáveis espaciais (Globais e Locais)** dos modelos alternativos – através da análise espacial das variáveis socioeconômicas e verificação das que apresentam autocorrelação espacial significativa e também correlação significativa com a distribuição espacial da variável dependente;
2. Diagnóstico dos modelos (tradicionais e alternativos) – através da análise da dispersão dos resíduos das estimativas para o **ano base**;
3. Análise dos resultados finais – através da análise da dispersão dos resíduos resultantes de estimativas futuras (**ano meta**).

Existem ferramentas de análise espacial embutidas em diversos *software* SIG, porém o nível de detalhe e aprofundamento das análises são diferentes em cada um. O *software* TransCAD, por exemplo, permite realizar a análise da autocorrelação espacial global fornecendo como resultado apenas o ***I de Moran Global***. Através de uma extensão do *software* ArcView é possível visualizar também o *Diagrama de Espalhamento de Moran* e os quadrantes em que se encontram cada ZT (*Box Map*), porém não é possível verificar os índices locais de autocorrelação espacial (***I de Moran Local***). Essa análise é possível através do *software* SPRING, que é, provavelmente, dentre os que foram explorados neste trabalho, o mais robusto em termos de ferramentas de análise espacial. Portanto, deve ser verificado o nível de detalhe que se pretende previamente à definição sobre a necessidade de migração de dados entre as diferentes bases georreferenciadas.

A Tabela 4.1 apresenta um resumo das características das etapas consideradas nesta pesquisa, em termos de análise espacial, e onde cada um dos *software* citados pode ser aplicado. É importante salientar que, para análise de indicadores globais (possível com qualquer dos três *software*), os resultados obtidos para um mesmo atributo podem apresentar pequenas diferenças, uma vez que cada um utiliza metodologias diferentes na geração da *matriz de proximidade espacial*. Além disso, o *software* SPRING e o *software* ArcView apresentam inversão na definição dos quadrantes 3 e 4. As apresentadas neste estudo referem-se às utilizadas pelo SPRING.

**Tabela 4.1:** Resumo das características das etapas de análise espacial e *software* que podem ser aplicados em cada uma delas

<b>Etapas de análise espacial</b>	<b>Análises e indicadores necessários</b>	<b>SIG que pode ser aplicado na análise</b>
<b>Análise da Autocorrelação Espacial das variáveis socioeconômicas</b>	<i>I de Moran Global</i> <i>Diagrama de Espalhamento de Moran</i> <i>Box Map</i>	ArcView SPRING
<b>Produção de indicadores Globais</b> – (Variáveis Espaciais Globais)	<i>Diagrama de Espalhamento de Moran</i> (quadrantes para cada ZT)	ArcView SPRING
<b>Produção de Indicadores Locais</b> – (Variáveis Espaciais Locais)	<i>I de Moran Local</i>	SPRING
<b>Diagnóstico dos modelos</b> – Distribuição espacial dos resíduos das estimativas para o ano base	<i>I de Moran Global</i>	TransCAD ArcView SPRING
<b>Análise dos resultados das previsões futuras</b> – Distribuição espacial dos resíduos das estimativas para o ano meta	<i>I de Moran Global</i> <i>Box Map</i> <i>LISA Map</i> <i>Moran Map</i>	SPRING

Conforme apresentado no item 3.8, o primeiro passo para a análise dos dados de áreas é o cálculo da *matriz de proximidade espacial*, ponto de partida para os demais cálculos e elemento chave da análise de dependência espacial. A *matriz de proximidade espacial* (**W**) estima a variabilidade espacial de dados de áreas, sendo uma ferramenta muito útil para descrever o arranjo espacial dos objetos. Alguns programas possuem rotinas pré-programadas para o cálculo da matriz, bastando apenas selecioná-las. Não é necessário que o cálculo seja feito separadamente, pois o próprio programa entende as relações de vizinhança entre áreas, bastando apenas que a topologia esteja correta, de tal

modo que o programa “entenda” os polígonos e a conectividade entre as áreas.

As técnicas de análise exploratória de dados espaciais implementadas no **SPRING** estão relacionadas direta e indiretamente ao **índice de Moran (local e global)**. Essas técnicas, combinadas com funções de visualização de atributos de objetos, formam um conjunto de ferramentas para a análise exploratória. Para o atributo em análise, além de apresentar o valor do índice ***I de Moran Global***, são geradas colunas na tabela de objetos correspondentes, contendo as informações necessárias para os módulos de visualização apresentados no **item 3.8.4 do Capítulo 3**.

Dessa forma, além dos valores, para cada área (no caso, as ZTs), do *vetor de desvios (Z)*, da *média móvel espacial (WZ)* e dos *índices locais de associação espacial LISA (IMORAN)*, são geradas também outras três colunas, denominadas:

- **BOXMAP** – apresenta os números dos quadrantes:
  - **1** = **Z** positivo / **WZ** positivo (Obs: notação adotada no **SPRING**);
  - **2** = **Z** negativo / **WZ** negativo;
  - **3** = **Z** positivo / **WZ** negativo;
  - **4** = **Z** negativo / **WZ** positivo.
  
- **LISAMAP** – apresenta os quatro grupos conforme o grau de significância dos **índices locais LISA**:
  - **0** = não significante;
  - **1** = com significância de 95,0%;
  - **2** = com significância de 99,0%;
  - **3** = com significância de 99,9%.
  
- **MORANMAP** – apresenta os quadrantes apenas para as áreas cujos índices foram considerados significativos no **LISAMAP**. As demais áreas recebem valor zero, classificadas como não significativas.

Como os índices e dados necessários para os módulos de visualização (mapas e gráficos) são apresentados em uma tabela e automaticamente é gerado um banco de dados na planilha eletrônica **Access**, os mapas e gráficos podem tanto ser apresentados no próprio *software* **SPRING** como podem ser exportados em formatos (extensão dbf,

por exemplo) que podem ser facilmente executados em planilhas eletrônicas (**MS-Excel**, por exemplo) e também no **SIG-T TransCAD** (desde que já contenha a base georreferenciada no mesmo). A realização deste procedimento é importante, não somente pela maior flexibilidade de escolha de cores e padrões na geração de mapas temáticos oferecida pelo *software TransCAD*, mas também por serem necessários esses dados (variáveis espaciais) para a modelagem de demanda realizada no SIG-T.

Uma das maiores preocupações com relação ao *Índice de Moran* é estabelecer a significância estatística, o que o *software SPRING* avalia através do teste de pseudo-significância. Este teste gera diferentes permutações, sendo possível escolher o número desejado no *software*, entre as opções 99, 999 e 9999. Cada permutação produz um novo arranjo espacial dos valores do atributo. Se o valor do *Índice I* encontrado nos dados originais corresponder ao extremo da distribuição simulada trata-se de um evento com significância estatística.

Aplicam-se técnicas de análise espacial em diferentes etapas da pesquisa, através dos métodos e *software* adequados a cada uma delas (Tabela 4.1). Como estas análises servem para a definição das variáveis espaciais para os *modelos alternativos*, a partir dos resultados das análises espaciais das variáveis socioeconômicas, é importante, também, executar a análise espacial da variável dependente. Após, devem ser estabelecidos critérios de análise comparativa entre a distribuição espacial da variável dependente e das demais variáveis, para verificação do nível de correlação existente e definição das variáveis espaciais que serão introduzidas nos *modelos alternativos*.

#### **4.6. ETAPA 6 - ESPECIFICAÇÃO E CALIBRAÇÃO DOS MODELOS TRADICIONAIS**

Esta etapa consiste na determinação dos melhores modelos que utilizam variáveis tradicionais para fins de comparação com os resultados das estimativas feitas com os modelos alternativos correspondentes. Para a especificação dos modelos tradicionais (modelos **T**) devem ser considerados os resultados obtidos com as análises da etapa anterior. Deve-se, também, calibrar vários modelos, levando-se em consideração alguns critérios na escolha das variáveis, e verificar, através de medidas estatísticas, em que grau o fenômeno estudado pode ser explicado pelo modelo adotado.

Além disso, como um dos objetivos do trabalho é realizar o estudo em ambiente SIG, é importante verificar as ferramentas disponíveis nos SIG-T (SIG específicos para planejamento de transportes) e avaliar em que nível elas permitem especificar e avaliar os diferentes modelos testados.

O SIG-T TransCAD, por exemplo, apresenta ferramentas que permitem estimar e aplicar vários tipos de modelos matemáticos que tentam produzir relações estatísticas entre os campos de dados de uma tabela. O modelo de regressão linear múltipla é, talvez, a ferramenta mais amplamente utilizada para determinar relações entre campos de dados. São definidas relações entre a medida de interesse (variável **dependente**) e as outras medidas das quais a medida de interesse depende (variáveis **independentes**). Dessa forma, fornecido um conjunto de dados para a variável dependente e para as variáveis independentes, o TransCAD determina os parâmetros que permitem modelar a melhor combinação dos valores encontrados na amostra.

São gerados relatórios para cada regressão, onde além dos valores dos coeficientes de cada variável e da constante, é possível verificar, quantitativamente, em que grau o fenômeno considerado pode ser explicado pelo modelo adotado, através de tabela de análise da variância, do *coeficiente de determinação* ( $R^2$ ) e do *coeficiente de determinação ajustado* ( $R^2$  ajustado).

O  $R^2$ , por exemplo, mede o poder explicativo do modelo como um todo, ou seja, mede a correlação entre os valores estimados através do modelo e o valor real da variável dependente, ou seja, quanto mais próximo de “1” melhor é a previsão do modelo. Como, na regressão múltipla, o  $R^2$  aumenta quando um regressor (variável) é adicionado, a comparação somente dos valores de  $R^2$  de modelos com número de variáveis diferente não é confiável, uma vez que o aumento do valor do coeficiente de determinação pode apenas estar relacionado ao número maior de variáveis e não à melhoria de ajuste do modelo. Uma forma de correção é deflacionar ou reduzir o  $R^2$  em algum fator, determinando-se o  $R^2$  ajustado (Stock e Watson, 2004).

Outra medida, também fornecida no relatório da análise de regressão do TransCAD, é a *estatística T* (*T stat*) para cada variável incluída. O *T stat* é um teste de significância dos parâmetros, sendo os valores em torno de dois ou mais considerados

representativos de uma significância com 95% de confiança. Dos modelos podem ser excluídas as variáveis que apresentam *T stat* menor que dois, com posterior avaliação dos novos resultados.

No entanto ainda se fazem necessárias a análise e interpretação dos coeficientes das variáveis resultantes do processo de calibração. Por exemplo, uma variável que apresenta coeficiente com sinal negativo e cuja correlação com a variável dependente é positiva é indício de algum problema no modelo ajustado. Este fenômeno pode ocorrer quando duas ou mais variáveis independentes apresentam correlação entre si. Segundo Hutchinson (1979), a colinearidade entre variáveis independentes consiste no erro mais comum na estimação de modelos de regressão para geração de viagens. Embora casos mais evidentes de múltipla colinearidade possam ser detectados por algumas rotinas de estimação de modelos, outros menos explícitos devem ser verificados pelo analista.

Além da análise dessas medidas, é importante, também, avaliar o desempenho da aplicação dos modelos para estimativas. As rotinas para especificação de modelos do SIG-T TransCAD produzem um arquivo de texto para cada modelo calibrado, com informações que podem ser utilizadas para avaliação dos modelos através de um conjunto de dados com campos semelhantes. Dessa forma, para utilizar os procedimentos de avaliação basta escolher as variáveis contidas nos campos da tabela de dados e também escolher o campo (criado anteriormente pelo usuário) que vai receber os valores das estimativas.

Avaliam-se, assim, os resultados de estimativas para o **ano base**, onde os valores de referência são os próprios dados utilizados para calibrar os modelos. Esta análise é importante também para verificar a presença de *outliers*, possibilitando excluir áreas que apresentam resultados discrepantes, quando de novas calibrações. Para avaliar os resultados das estimativas é necessário calcular os valores dos *erros relativos médios* (**ERM**), através das seguintes equações:

$$ERM = \frac{\sum_{i=1}^n |ER_i|}{n} \quad (12)$$

$$ER_i = \frac{R_i}{VPBD_i} \quad (13)$$

$$R_i = E\_VPBD_i - VPBD_i \quad (14)$$

Onde:

- $ER_i$  = erro relativo da estimativa para a ZTi;
- $R_i$  = resíduo da estimativa para a ZTi;
- $E\_VPBD_i$  = valor estimado pelo modelo para as viagens produzidas de base domiciliar para a ZTi;
- $VPBD$  = valor “real” de viagens produzidas de base domiciliar obtido na pesquisa O-D;
- $n$  = número de ZTs da pesquisa.

Os cálculos podem ser realizados no próprio SIG, mediante criação de campos de fórmula nas tabelas (Equações 13 e 14) e geração de relatórios estatísticos, que fornecem, além dos valores máximos e mínimos de cada atributo, as médias (Equação 3) e o desvio padrão.

Finalmente, faz-se a análise da *distribuição espacial dos resíduos*, que é um importante parâmetro de análise do desempenho do modelo. Avalia-se o resultado do índice **I de Moran (global)**, que pode ser calculado através de ferramentas de análise espacial do próprio TransCAD, para os resíduos resultantes das estimativas de cada modelo. Quanto mais próximo de zero os valores do índice I de Moran (global) para os resíduos, melhores são os resultados das estimativas com o modelo.

Em síntese, para especificação dos modelos, consideram-se as variáveis resultantes das análises da etapa anterior. Um primeiro modelo é testado incluindo-se todas as variáveis. Os critérios para retirada de variáveis e especificação de outros modelos são:

- Análise dos coeficientes de correlação **R** em relação à variável dependente (**VPBD**) – quanto menor o valor, menor a relação e menos importante é a variável;
- Análise dos coeficientes de correlação **R** em relação às variáveis independentes – se duas variáveis independentes apresentam alta correlação, é excluída a que apresenta menor valor de **R** em relação à **VPBD**;
- Análise da estatística **T** (*T stat*) – são excluídas as variáveis com valores

menores que dois (em módulo);

- Interpretação dos coeficientes das variáveis – deve ser excluída a variável cujo coeficiente é incoerente com a relação da mesma com a variável dependente.

Se, entre os modelos estimados, em mais de um se verificar todos os critérios mencionados acima, a escolha dos melhores modelos deve ser através da comparação dos valores do  $R^2$ , ou do  $R^2$  ajustado (no caso de modelos com diferentes números de variáveis), do ERM e do I de Moran (Global) dos resíduos das estimativas para o ano base. São destacados, então, os melhores *modelos tradicionais*, que servem de base para o teste dos *modelos alternativos* correspondentes, nos três métodos considerados nesta pesquisa: modelos **Globais**, modelos **LISA** e modelos **Mistos**, que são apresentados no item 4.7.

#### 4.7. ETAPA 7 - ESPECIFICAÇÃO E CALIBRAÇÃO DE MODELOS ALTERNATIVOS

A especificação dos modelos alternativos é baseada nos resultados das **Etapas 5** e **6**. São utilizadas três diferentes abordagens para especificação dos diferentes tipos de modelos alternativos, chamados aqui de:

- 1- **Modelos Globais ou AG** (Introdução de indicadores globais);
- 2- **Modelos LISA ou AL** (Introdução de indicadores locais);
- 3- **Modelos Mistos ou AGL** (Introdução de indicadores globais e locais).

O que diferencia cada um dos modelos alternativos são as variáveis espaciais incluídas. É importante, neste caso, apresentar primeiramente a forma com que cada tipo de variável é definida:

- **Variáveis Espaciais Globais** – são definidas através dos quadrantes resultantes do *Diagrama de Espalhamento de Moran* obtidos nas análises de autocorrelação espacial, conforme método descrito na **Etapa 5**. São variáveis binárias, ou seja, apresentam valores **0** (zero) ou **1** (um) conforme o quadrante a que pertence cada ZT, para os diferentes atributos analisados. Nas análises realizadas através do *software* SPRING, por exemplo, os números referentes aos quadrantes (1, 2, 3 e 4) são armazenados na coluna denominada **BOXMAP**. A partir destes dados, para cada variável socioeconômica analisada e que for considerada significativa

como variável espacial global para explicar o fenômeno em estudo, são definidas três variáveis binárias correspondentes, representando os quadrantes **Q1**, **Q2** e **Q3**. Neste caso, para uma determinada variável “**X**” são criadas três *variáveis espaciais globais* (**X\_Q1**, **X\_Q2** e **X\_Q3**). Desta forma, na tabela de atributos no SIG-T TransCAD são criadas três novas colunas. Se uma determinada ZT pertencer ao quadrante um (**Q1**), por exemplo, recebe valor “**1**” para a variável “**X\_Q1**” e valor “**0**” para as variáveis **X\_Q2** e **X\_Q3**, e assim sucessivamente. O quadrante quatro (**Q4**) estará representado pelo valor “**0**” nas três variáveis.

- **Variáveis Espaciais Locais** – são definidas através dos *índices locais* obtidos na análise de autocorrelação espacial no *software* SPRING, para cada variável socioeconômica que apresentou autocorrelação espacial significativa e que foi considerada significativa como variável espacial local para explicar o fenômeno em estudo. Neste caso, usando o mesmo exemplo, para a variável hipotética “**X**”, é criada uma *variável espacial local* (**X\_LISA**). Desta forma, na tabela de atributos, cada ZT receberá o valor que foi armazenado na coluna **IMORAN** pelo *software* SPRING, resultante da análise espacial da variável **X**.

É importante observar que estes atributos (**IMORAN** e **BOXMAP**) são alterados automaticamente pelo *software* SPRING, conforme as variáveis vão sendo analisadas. Neste caso, é necessário exportar os dados da tabela imediatamente após cada análise e identificar o arquivo de forma que possa ser posteriormente formatado e importado pelo *software* TransCAD, onde os modelos serão especificados e calibrados.

Os *modelos alternativos* são especificados a partir dos *modelos tradicionais* (**T**). Desta forma, os melhores *modelos tradicionais* são selecionados e os *modelos alternativos Globais* (**AG**) são especificados, acrescentando-se ao modelo, inicialmente, todas as *variáveis espaciais globais* selecionadas para análise. O mesmo procedimento é realizado para especificação dos *modelos alternativos LISA* (**AL**), substituindo-se, neste caso, as *variáveis espaciais globais* pelas *variáveis espaciais locais*. Da mesma forma, os *modelos alternativos Mistos* (**AGL**) são especificados, inicialmente, introduzindo-se todas as *variáveis espaciais (globais e locais)*.

É importante, no entanto, analisar os resultados nos relatórios de regressão e verificar quais variáveis estão contribuindo e quais não são significativas ou que podem estar atrapalhando o poder explicativo do modelo. Os métodos de análise e diagnóstico dos *modelos alternativos* são os mesmos aplicados para os modelos tradicionais e já apresentados no **item 4.6 (Etapa 6)**.

Uma forma complementar às especificações e avaliações dos modelos através de ferramenta SIG, uma vez que são muitas variáveis em análise, é utilizar o método *stepwise*, contido em ferramentas estatísticas, para especificar um modelo de regressão e, através do resultado, avaliar a significância das variáveis. No entanto, também é necessária a avaliação do modelo resultante pelo analista, uma vez que, em alguns casos, estas ferramentas podem não conseguir interpretar o problema de colinearidade.

Nesta abordagem, todas as variáveis (*tradicionais, globais e locais*) são incluídas inicialmente. As equações de regressão são desenvolvidas através de uma seqüência de etapas, onde são examinados os relacionamentos entre a variável dependente com as variáveis independentes em separado, a fim de identificar não-linearidade; se for verificada, o relacionamento deve ser linearizado pela transformação da variável independente, da dependente ou de ambas. No processo, é avaliada, também, a matriz de correlação para detectar as variáveis independentes que têm associação estatística com a variável dependente e fontes potenciais de colinearidade entre pares de variáveis independentes.

O diagnóstico dos *modelos alternativos* inclui, ainda, os resultados de estimativas para o **ano base**, que são avaliados pelo mesmo método descrito para os *modelos tradicionais* (**item 4.6**), ou seja, através do cálculo dos valores dos *erros relativos médios* (**ERM**), utilizando as mesmas equações (12, 13 e 14), e também é analisada a *distribuição espacial dos resíduos*, através do índice **I de Moran (global)**. Os próximos passos são a aplicação dos modelos para estimativas futuras (**ano meta**) (**ETAPA 8**) e a análise comparativa dos resultados obtidos com cada modelo (**ETAPA 9**), que são descritas nos **itens 4.8 e 4.9**, respectivamente.

#### 4.8. ETAPA 8 – ESTIMATIVAS FUTURAS

Nesta etapa, aplicam-se os modelos para previsão das viagens futuras utilizando-se os dados socioeconômicos (variáveis independentes) da pesquisa do **ano meta**, para aplicação dos modelos calibrados na **Etapa 6** e **Etapa 7** (Tradicionais e Alternativos, respectivamente). Os arquivos texto produzidos pelas rotinas de especificação de modelos podem ser utilizados para avaliar os mesmos em outro conjunto de dados com campos semelhantes. Desta forma, uma vez existente o mesmo conjunto de variáveis na base de dados do **ano meta**, os modelos calibrados podem ser aplicados, bastando escolher as variáveis contidas nos campos da tabela de dados e também escolher o campo (criado anteriormente pelo usuário) que vai receber os valores das estimativas futuras.

Os resultados de estimativas futuras são avaliados da mesma forma descrita no **item 4.6.**, ou seja, através do cálculo dos valores dos *erros relativos médios (ERM)*, utilizando as mesmas equações (12, 13 e 14). Também é analisada a *distribuição espacial dos resíduos*, através do índice **I de Moran (global)**.

#### 4.9. ETAPA 9 - AVALIAÇÃO DOS MODELOS

Uma vez que os melhores modelos (*tradicionais e alternativos*) são definidos e são aplicados para estimativas futuras, faz-se a análise comparativa entre os mesmos, através dos resultados obtidos para os modelos e, principalmente, das estimativas para o **ano base** e para o **ano meta (ERM e I\_Mor\_Res)**. Faz-se, ainda, a análise comparativa dos *Moran Map* para os resíduos das estimativas obtidas através do melhor *modelo tradicional* e do melhor *modelo alternativo*.

## 5. APLICAÇÃO DO MÉTODO E ANÁLISE DOS RESULTADOS

---

Este capítulo apresenta a aplicação do método descrito no Capítulo 4 em um estudo de caso na cidade de Porto Alegre - RS, que tem por base dados de pesquisa de origem e destino (O-D) obtidos através de entrevista domiciliar (EDOM) em dois períodos distintos, que correspondem aos dados necessários do “**ano base**” e do “**ano meta**” mencionados anteriormente. São apresentados, também, os resultados obtidos em cada uma das nove etapas que compõem o estudo.

Cabe ressaltar que o estudo ocorreu em **três fases**. A **primeira fase** foi um estudo preliminar onde apenas algumas etapas foram desenvolvidas, e de forma não tão aprofundada. Esta fase teve o intuito de verificar a validade da hipótese básica, ou seja, que a introdução de indicadores de dependência espacial no processo de modelagem de demanda por transportes poderia produzir resultados mais acurados e, desta forma, mais confiáveis que os obtidos com modelos tradicionais. Este estudo preliminar, resumido em Lopes e Silva (2004), além de comprovar a hipótese, serviu para definir algumas diretrizes para a segunda fase.

O estudo exploratório inicial considerou apenas indicadores globais de dependência espacial e apenas para duas variáveis socioeconômicas, mas permitiu verificar *outliers*. Indicou, também, a necessidade de pesquisas adicionais, de análise e inclusão de outras medidas de dependência espacial nos modelos da demanda. Uma das alternativas, a estatística LISA (indicadores locais de autocorrelação espacial), é analisada na **segunda fase**. Outro aspecto levantado, a necessidade de uma análise mais detalhada dos modelos, como, por exemplo, a significância das variáveis incluídas, foi realizada através de ferramentas de SIG-T e também de técnicas estatísticas tradicionais (em particular o método de regressão *stepwise*).

Na **terceira e última fase**, outras variáveis foram geradas e incluídas nos modelos, sob a forma de “taxas”. Foram adotados critérios mais rigorosos na especificação e diagnóstico dos modelos e também na análise das variáveis explicativas. Para cada uma das etapas do estudo de caso, apresentadas nos itens 5.1 ao 5.9, são descritos os detalhes de cada uma das fases, além da análise dos resultados obtidos.

## 5.1. ANÁLISE DOS DADOS DISPONÍVEIS

A análise dos dados disponíveis para o Município de Porto Alegre foi o primeiro passo para o início do estudo, tendo sido realizada totalmente na **primeira fase**. Foram verificadas as pesquisas de origem e destino (O-D) disponíveis para este município e analisadas as características de cada uma delas, para definição das duas mais apropriadas ao estudo proposto. As pesquisas O-D têm sido realizadas a cada dez anos, aproximadamente, em uma amostra de domicílios da Região Metropolitana de Porto Alegre (RMPA). Existem três pesquisas já realizadas e outra em estágio de finalização, cujas características serão apresentadas a seguir.

As pesquisas O-D, também conhecidas como Entrevistas Domiciliares (EDOM) nos órgãos de planejamento porto-alegrenses, têm por objetivo comum obter os dados relacionados aos deslocamentos efetuados pelos indivíduos e às características socioeconômicas dos moradores das residências selecionadas, necessários para o planejamento de transportes, além de colher outras informações sobre a situação de moradia da população, visando outros estudos de âmbito regional.

Os dados das três pesquisas já realizadas (e concluídas) em Porto Alegre - RS (1974, 1986, 1997) estão arquivados na Fundação de Planejamento de Transporte Metropolitano e Regional (METROPLAN), no Grupo Técnico de Programação e Suporte de Informática – Diretoria Administrativa (Peixoto, 2002). Os dados da pesquisa de 2003 estão em fase de processamento eletrônico pela agência contratada, o que inclui a digitação, georreferenciamento, verificação de consistência e sistematização (desenvolvimento de ferramenta de busca automática e selecionada) e deverão estar disponíveis até o final de 2004. Após a finalização dos trabalhos, os dados serão entregues para a EPTC/SMT (Empresa Pública de Transporte e Circulação / Secretaria Municipal de Transportes), órgão contratante da pesquisa.

#### **5.1.1. Pesquisa O-D de 1974 (EDOM 1974)**

A primeira pesquisa O-D realizada em Porto Alegre foi nos anos de 1974 e 1975 (2º semestre de 1974 na capital e 1º semestre de 1975 no restante da região metropolitana), tendo originado o Plano Diretor de Transportes Urbanos na RMPA – PLAMET/PA – 1976, objeto do convênio celebrado entre a METROPLAN e a Empresa Brasileira de Planejamento de Transportes (GEIPOT). O escopo do trabalho visava à elaboração de um plano global de transportes para a RMPA, com horizontes de curto prazo (1980), médio prazo (1985) e longo prazo, tendo o ano de 1974 como referência.

#### **5.1.2. Pesquisa O-D de 1986 (EDOM 1986)**

A segunda pesquisa foi realizada pela METROPLAN, em 1986, com o objetivo de obter as informações necessárias à já prevista atualização do PLAMET/PA. Peixoto (2002) fez uma análise detalhada da EDOM 86, concluindo que a pesquisa foi devidamente finalizada e apresenta consistência nos resultados encontrados.

#### **5.1.3. Pesquisa O-D de 1997 (EDOM 1997)**

A pesquisa realizada em 1997, conhecida como EDOM III, foi contratada pela Secretaria da Coordenação e Planejamento, contando com a participação da METROPLAN no fornecimento de dados para o planejamento e fiscalização dos trabalhos. A finalidade principal foi a caracterização dos deslocamentos da população

da RMPA, bem como a obtenção de informações socioeconômicas. Peixoto (2002) ressalta que foi feita a revisão dos limites das zonas de tráfego da região metropolitana, com agregação de algumas zonas pertencentes a um mesmo município, principalmente em Porto Alegre, em função de limitações financeiras para a realização das entrevistas e o limitado interesse para o conhecimento das viagens de forma mais detalhada do ponto de vista espacial. Portanto, apesar de envolver um número maior de municípios (24) do que as anteriores e também corresponder a uma maior população, ocorreu uma redução significativa do número de domicílios e indivíduos entrevistados. Peixoto (2002) salienta, ainda, que uma série de problemas ocorridos durante o planejamento, trabalhos de campo e digitação da pesquisa resultaram em um conjunto de informações com muitas e variadas deficiências.

#### **5.1.4. Pesquisa O-D de 2003 (EDOM 2003)**

Os problemas com a EDOM de 1997 levaram a Secretaria Municipal de Transporte – SMT, através de convênio com a Universidade Federal do Rio Grande do Sul (UFRGS), a investir na preparação das condições necessárias para a utilização do programa *Simulation and Assignment of Traffic to Urban Road Networks* (SATURN), um simulador do comportamento dos usuários da rede viária. Os dados de demanda, necessários para a utilização do programa, estavam muito defasados, o que motivou a contratação da EDOM 2003 pelo Programa de Segurança Viária do Município de Porto Alegre, com o objetivo principal de obter informações atualizadas sobre os deslocamentos realizados no Município de Porto Alegre, com um nível de detalhe compatível com as necessidades do SATURN. A área da pesquisa inicial abrangia a região mais densamente urbanizada do Município de Porto Alegre e dotada de infraestrutura viária de importância estrutural (70 das 95 ZTs definidas pela EDOM 86). O programa foi ampliado antes do início da pesquisa, passando abranger a quase totalidade do município.

#### **5.1.5. Síntese das pesquisas**

Algumas características das pesquisas O-D realizadas na RMPA são sintetizadas na Tabela 5.1. A utilização da EDOM 1974 e da EDOM 1986 neste trabalho deve-se aos problemas apresentados pela EDOM 1997, principalmente por ter um zoneamento

muito diferente dos anteriores, devido à agregação das zonas, e uma amostragem reduzida, o que poderia prejudicar a análise espacial. Os dados da EDOM 2003, por outro lado, ainda não se encontram disponíveis.

**Tabela 5.1:** Principais características das pesquisas O-D realizadas em Porto Alegre - RS e Região Metropolitana

	EDOM 1974		EDOM 1986		EDOM 1997		EDOM 2003	
<b>Abrangência</b>	RMPA 14 municípios		RMPA 14 municípios		RMPA 24 municípios		Município de Porto Alegre	
<b>Nº de ZTs pesquisadas/total</b>	<b>RMPA</b>	139/232	<b>RMPA</b>	149	<b>RMPA</b>	100	<b>RMPA</b>	-
	<b>PA</b>	69/97	<b>PA</b>	75/95	<b>PA</b>	28/68	<b>PA</b>	91/95
<b>Nº de Domicílios Pesquisados</b>	<b>RMPA</b>	40.844	<b>RMPA</b>	28.926	<b>RMPA</b>	19.322	<b>RMPA</b>	-
	<b>PA</b>	20.869	<b>PA</b>	11.382	<b>PA</b>	N/C	<b>PA</b>	16300
<b>Nº de Moradores Pesquisados</b>	<b>RMPA</b>	176.664	<b>RMPA</b>	113.995	<b>RMPA</b>	67.933	<b>RMPA</b>	-
	<b>PA</b>	85.036	<b>PA</b>	42.009	<b>PA</b>	N/C	<b>PA</b>	N/C
<b>Amostras</b>	<ul style="list-style-type: none"> <li>• Amostragem seqüencial/ZT</li> <li>• Cadastro de domicílios da CEEE (Companhia Estadual de Energia Elétrica)</li> <li>• Variância das viagens/domicílio da pesquisa piloto</li> <li>• Em torno de 10 %/ZT</li> <li>• Erro máximo: 10 %</li> <li>• Nível de confiança: 95 %</li> </ul>		<ul style="list-style-type: none"> <li>• Amostra aleatória simples/ZT</li> <li>• Cadastro de domicílios do IBGE</li> <li>• Variância das viagens/pessoa da EDOM 74</li> <li>• Erro máximo:10 %</li> <li>• Nível de confiança: 90 %</li> </ul>		<ul style="list-style-type: none"> <li>• Amostra aleatória simples/ZT</li> <li>• Cadastro de domicílios do IBGE</li> <li>• Variância das viagens/pessoa da EDOM 86</li> <li>• Erro máximo:10 %</li> <li>• Nível de confiança: 90 %</li> </ul>		<ul style="list-style-type: none"> <li>• Distribuição proporcional aos domicílios do Censo 2000</li> <li>• Cadastro de domicílios do DMAE (Departamento Municipal de Águas e Esgoto)</li> <li>• Mínimo 40 entrevistas por ZT</li> </ul>	
<b>Pesquisas Complementares</b>	<ul style="list-style-type: none"> <li>• Contagens volumétricas</li> <li>• Pesquisas na pista</li> </ul>		<ul style="list-style-type: none"> <li>• Contagens volumétricas</li> <li>• Frequência e ocupação visual em ônibus urbanos e interurbanos</li> <li>• Pesquisa de placas (<i>screen line</i>)</li> </ul>		-		<ul style="list-style-type: none"> <li>• Contagens volumétricas</li> <li>• Frequência e ocupação visual em ônibus urbanos e interurbanos</li> <li>• Pesquisa na linha de contorno</li> </ul>	

## 5.2. DEFINIÇÃO DA ÁREA DE ESTUDOS

Após a verificação dos dados disponíveis nas pesquisas, definiu-se como abrangência do estudo os limites do Município de Porto Alegre, desconsiderando-se os demais municípios da Região Metropolitana. Além de se considerar que uma área com a extensão e características deste município seria suficiente para os propósitos deste trabalho, também foi decisiva a disponibilidade de base georreferenciada das ZTs da EDOM de 1986 apenas para Porto Alegre.

O Município de Porto Alegre corresponde a uma área de aproximadamente 500 km<sup>2</sup>, possuía em torno de 1.000.000 habitantes em 1974 (data da primeira pesquisa), 1.200.000 habitantes aproximadamente em 1986 (data da segunda pesquisa) e, no último censo, o de 2000, 1.361.773 habitantes (IBGE, 2000). Foi dividido em 97 ZTs para a pesquisa de 1974 (Figura 5.1) e em 95 ZTs para a pesquisa de 1986, porém somente 75 destas foram pesquisadas e apresentam resultados (Figura 5.2).

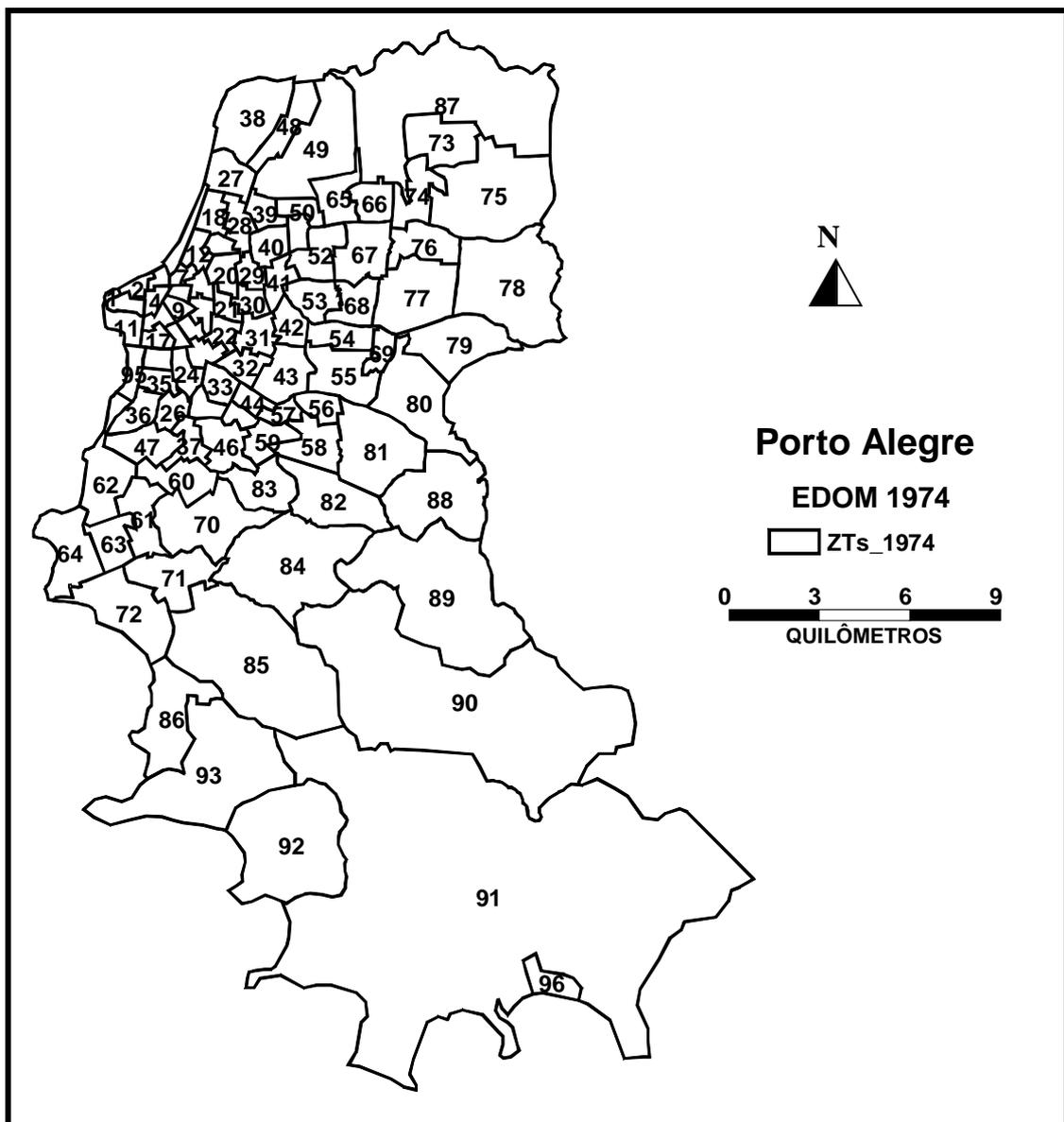


Figura 5.1: Mapa de Porto Alegre - ZTs da EDOM 1974

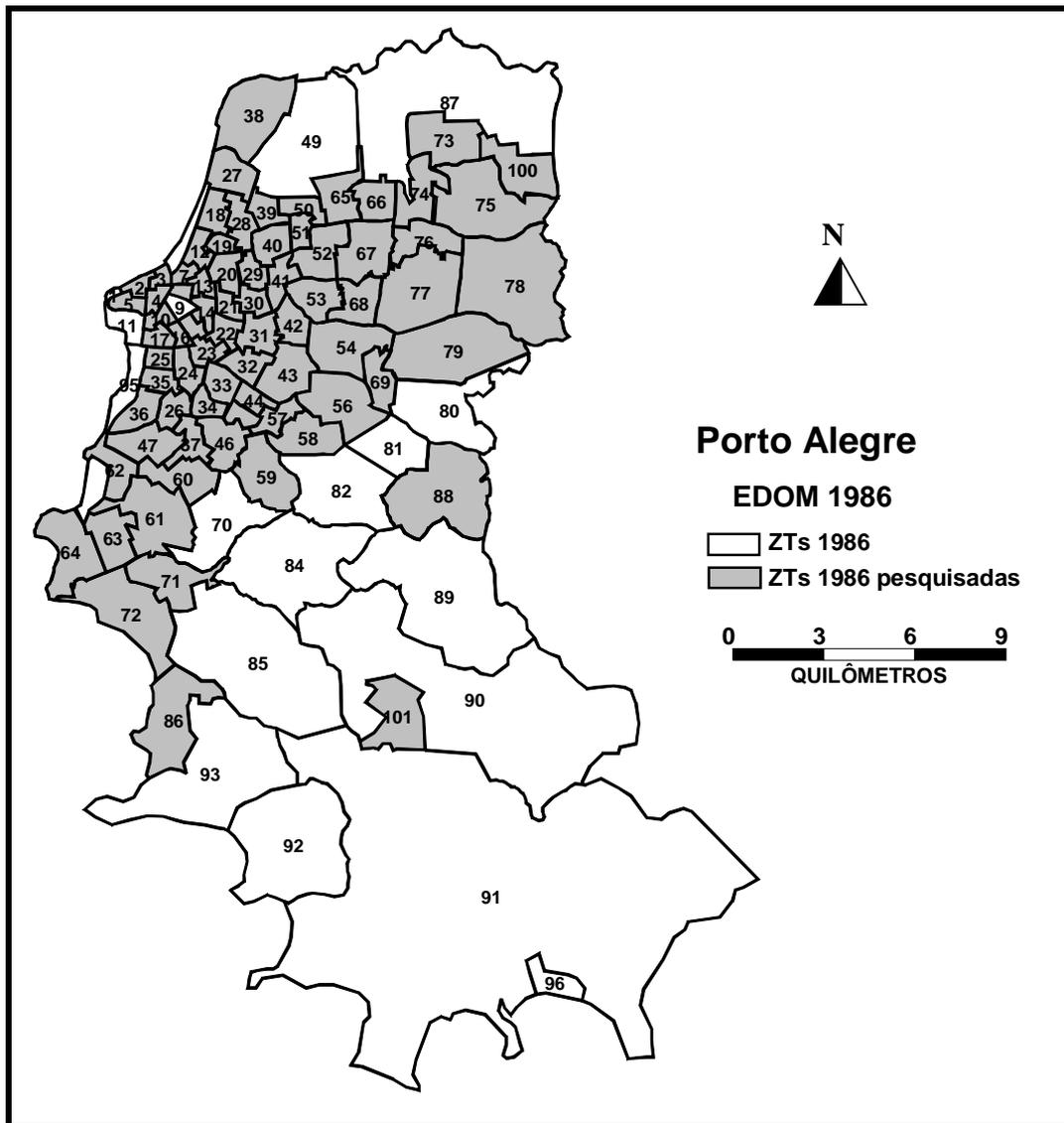


Figura 5.2: Mapa de Porto Alegre - ZTs da EDOM 1986

### 5.3. PREPARAÇÃO DA BASE DE DADOS E ANÁLISE PARA COMPATIBILIZAÇÃO DAS ZTS

Apresenta-se o método aplicado para preparação da base de dados, que engloba atividades preliminares de importação dos dados para o ambiente SIG e a posterior verificação do zoneamento adotado nas duas pesquisas selecionadas, através de uma análise comparativa para definição da melhor forma de compatibilização das mesmas. São apresentados, também, métodos desenvolvidos para migração de dados entre diferentes bases georreferenciadas (*software* TransCAD e SPRING). Esta etapa envolve atividades tanto da **primeira fase** (item 5.3.1) quanto da **segunda fase** (item 5.3.2).

### 5.3.1. Digitalização e análise para compatibilização das ZTs

Os dados da EDOM de 1986 já estavam em formato eletrônico, com algumas informações de todas as 95 ZTs armazenadas em um SIG. Como os dados da EDOM de 1974 não estavam em formato eletrônico, a alternativa foi inserir no TransCAD uma imagem (arquivo TIF) de um mapa com as 97 ZTs de 1974.

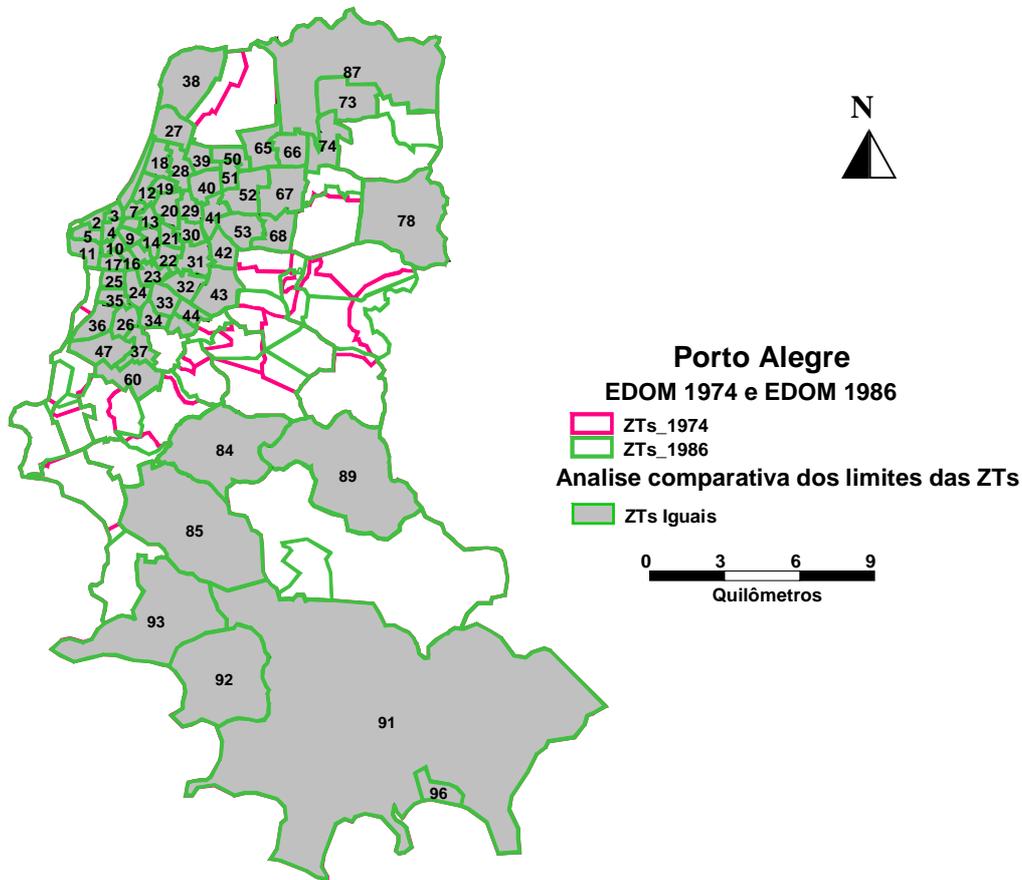
Em seguida, passou-se à tentativa de ajuste de escala entre os mapas, com a verificação de quais ZTs seriam iguais nas duas pesquisas. Esta análise foi importante para a avaliação do grau de compatibilidade entre os zoneamentos e definição da melhor forma de digitalização da base de dados de 1974: se através do aproveitamento da base georreferenciada de 1986, executando as alterações necessárias, ou digitalizando completamente sobre a imagem importada.

Foram observadas 64 ZTs aparentemente iguais e 5 que deixaram alguma dúvida se a diferença na forma seria problema de desenho ou seriam limites realmente diferentes. As 26 restantes apresentavam, nitidamente, limites diferentes. Para verificar os limites das ZTs que deixaram dúvidas, foi feita análise de compatibilidade das vias e de comparação entre áreas. As áreas das ZTs de 1974 foram obtidas de tabelas dos relatórios, enquanto as de 1986 foram obtidas da base de dados georreferenciada.

Chegou-se à conclusão de que existem 65 zonas que possuem limites iguais nas duas pesquisas e 30 (na EDOM 1986) ou 31 (na EDOM 1974) que são diferentes (Figura 5.3). A ZT 97, que corresponde à região das ilhas, não consta como área pesquisada em 1986, portanto não foi considerada no zoneamento de 1974, para efeito deste trabalho. Decidiu-se, então, pelo aproveitamento da base das ZTs de 1986 (através da ferramenta de exportação) para criação da base de 1974, com utilização da ferramenta de edição de áreas. Concluiu-se que este procedimento seria melhor para garantir a perfeita sobreposição das zonas iguais.

As informações da pesquisa de 1974 (dados socioeconômicos e de viagens), que se encontravam em tabelas nos relatórios, foram transformadas para o formato digital através de *scanner* e salvas como arquivos de texto (com extensão TXT). Depois de transformados em planilha eletrônica e corrigidas algumas distorções, os dados foram

devidamente formatados e salvos como arquivo com extensão DBF, sendo as informações, então, incluídas na base de dados georreferenciada.



**Figura 5.3:** Mapa de Porto Alegre com a sobreposição das ZTs de 1974 e 1986, com destaque para as áreas iguais

Com esses procedimentos concluiu-se a preparação da base de dados de 1974 (ano base). Com relação aos dados de 1986, apesar de já estarem em forma digital e o zoneamento em ambiente SIG, algumas informações da pesquisa encontram-se, em arquivos de um gerenciador de banco de dados, na forma desagregada, ou seja, as informações de cada viagem individual constituem uma linha de uma planilha. Tais dados, porém, não foram necessários neste estudo.

### 5.3.2. Migração de dados (TransCAD – SPRING)

Para as análises da **segunda fase** era necessário que os dados das duas pesquisas fossem importados para o *software* SPRING. As grandes diferenças de sistemas, no que

diz respeito a configurações iniciais de banco de dados, coordenadas, formatos de armazenamento, importação, exportação e leitura de dados, adotados por cada um dos *software*, torna este processo complexo e trabalhoso. Foi necessário, desta forma, desenvolver uma rotina para possibilitar esta migração.

Inicialmente, fez-se a preparação do SPRING para receber os dados e as rotinas adotadas para exportação do TransCAD e importação para o SPRING (**Anexo A**). Foi desenvolvido, também, um método de exportação dos dados analisados, do SPRING para o TransCAD, que se fez necessário porque os resultados das análises espaciais realizadas no SPRING deveriam ser, posteriormente, processados no TransCAD, tanto para a especificação e aplicação dos modelos como também para a criação de mapas temáticos.

#### **5.4. ANÁLISE DAS VARIÁVEIS INDEPENDENTES**

Em uma análise preliminar dos relatórios das pesquisas de 1974 e 1986, foram verificados dados disponíveis em formas não compatíveis com as necessidades do estudo (muito agregados, no caso da EDOM 1974) ou não disponíveis diretamente (muito desagregados, no caso da EDOM 1986). Desta forma, fez-se necessário, em primeiro lugar, verificar as informações disponíveis na EDOM 1974 que estavam agregadas em ZTs, pois seriam estas que iriam definir o nível de detalhamento do estudo, por ser mais viável (e confiável), apesar de trabalhosa, a obtenção de dados agregados a partir de desagregados do que o contrário.

Esta análise permitiu a definição da variável dependente a ser analisada neste estudo, **Viagens Produzidas de Base Domiciliar (VPBD)**, para todos os modos. Esta definição se deu pela verificação dos dados disponíveis em cada uma das pesquisas, que são apresentados nos **itens 5.4.1 e 5.4.2** e no **Anexo B**.

O próximo passo foi verificar quais são os dados socioeconômicos incluídos nas pesquisas selecionadas que podem ser possíveis variáveis explicativas para os modelos de demanda. Os critérios citados no **Capítulo 4, item 4.4**, foram levados em consideração para essas análises, que ocorreram nas diferentes fases do estudo.

### 5.4.1 Dados da EDOM de 1974

As informações por ZT estão divididas em dados socioeconômicos e dados de viagem, como seguem:

- **Dados socioeconômicos**
  - Referentes à população
    - Número de domicílios
    - População residente
      - Total; Ativa; Estudantil
  - Renda média domiciliar mensal (Unidade = Cr\$ 1,00)
  - Número de veículos
  - Matrículas escolares
    - Pré-primário; 1º Grau; 2º Grau; Outros; Superior e Total
  - Empregos
    - Indústria; Comércio; Serviços privados; Serviços públicos e Total
  - Superfície da ZT
- **Dados de viagens**
  - Produzidas (em um dia útil médio – todos os modos)
    - Base-domiciliar e Base não domiciliar
  - Atraídas (em um dia útil médio – todos os modos)
    - Base-domiciliar e Base não domiciliar

Os diferentes modos pesquisados na EDOM 1974 estão agregados em totais de viagem, para a área central de Porto Alegre e para o restante do município. São eles:

- A pé
- Condutores de automóvel
- Passageiros
  - Automóvel
  - Táxi
  - Ônibus
  - Mistos
  - Outros

### 5.4.2 Dados da EDOM de 1986

As informações que estão apresentadas por ZT e de forma direta são as listadas a seguir, também divididas em dados socioeconômicos e dados de viagem:

- **Dados socioeconômicos**
  - Referentes à população
    - Número de domicílios
    - População residente
      - Total e Ativa
  - Renda média domiciliar mensal (Unidade = CZ\$ 1,00)
  - Consumo de energia elétrica
  - Número de veículos
    - Álcool; Gasolina; Diesel; Motos e Bicicletas
  - Superfície da ZT
- **Dados de viagens**
  - Produzidas (em um dia útil médio – todos os modos)
    - Base-domiciliar e Base não domiciliar

Outros dados, como motivo do deslocamento, modo do deslocamento (12 modos, menos a pé), grau de instrução (9 categorias), base do deslocamento (domiciliar e não domiciliar) e atividade (14 categorias), estão disponíveis por indivíduo entrevistado.

### 5.4.3 Variáveis explicativas para VPBD citadas na literatura e na EDOM 1974

Após estudo das variáveis citadas na literatura como explicativas para Viagens Produzidas de Base Domiciliar (**VPBD**), podem ser destacadas, dentre as disponíveis nas duas pesquisas, as seguintes:

- *Renda média domiciliar mensal;*
- *População residente total;*
- *População residente ativa;*
- *Número de veículos;*

- *Densidade populacional* (que é uma variável derivada, obtida a partir do dado da população e da superfície da ZT).

Nos relatórios da EDOM 1974 são mencionadas as variáveis *população residente*, *densidade populacional* e *renda média domiciliar* como potenciais variáveis explicativas para VPBD, porém não fica claro se elas foram utilizadas nos modelos para estimativa de demanda futura nos planos de curto e médio prazo. Ainda assim, foram as variáveis analisadas no estudo preliminar deste trabalho (**Fase 1**).

#### 5.4.4 Análise da correlação linear entre as variáveis (Fase 2)

Na **Fase 2** do estudo foi realizada a análise da correlação linear das variáveis. Foram analisados *digramas de dispersão* e verificou-se a existência de relação linear, não necessitando transformação de variáveis. A *matriz de correlação* (**Tabela 5.2**) foi gerada através de ferramenta de estatística do *software* TransCAD, utilizando-se a base de dados da EDOM 1974 (**ano base**). Foram analisados os valores dos *coeficientes de correlação R* entre as variáveis socioeconômicas (potenciais variáveis explicativas) e as *viagens produzidas de base domiciliar VPBD* (variável dependente).

Foram selecionadas sete variáveis, consideradas importantes para a modelagem das *viagens produzidas de base domiciliar VPBD*, que são:

1. *AREA*: corresponde à superfície de cada ZT, em km<sup>2</sup>;
2. *DOMIC*: corresponde ao número total de domicílios por ZT;
3. *POP\_T*: corresponde à população total de cada ZT;
4. *POP\_A*: corresponde à população ativa (que trabalha) de cada ZT;
5. *FROTA*: corresponde ao número de veículos particulares por ZT;
6. *DENS*: corresponde à densidade populacional de cada ZT, em hab/ km<sup>2</sup>;
7. *RENDA*: corresponde à média da renda mensal dos domicílios em cada ZT (em dólares).

Pode-se observar, através dos valores da primeira coluna da matriz, que as variáveis *DOMIC*, *POP\_T*, *POP\_A* e *FROTA* apresentam altos índices de correlação linear com a variável dependente *VPBD*, com destaque para as duas primeiras. Não se pode afirmar, no entanto, que as demais não irão contribuir para os modelos.

**Tabela 5.2:** *Matriz de correlação* entre as variáveis socioeconômicas do **ano base** e a variável dependente **VPBD**

	<b>VPBD</b>	<b>AREA</b>	<b>DOMIC</b>	<b>POP_T</b>	<b>POP_A</b>	<b>FROTA</b>	<b>DENSI</b>	<b>RENDA</b>
<b>VPBD</b>	1,00	-0,28	0,90	0,91	0,87	0,84	0,49	0,26
<b>AREA</b>	-0,28	1,00	-0,23	-0,21	-0,21	-0,26	-0,35	-0,22
<b>DOMIC</b>	<b>0,90</b>	-0,23	1,00	0,96	0,97	0,77	0,59	0,16
<b>POP_T</b>	<b>0,91</b>	-0,21	<b>0,96</b>	1,00	0,98	0,68	0,43	0,03
<b>POP_A</b>	0,87	-0,21	<b>0,97</b>	<b>0,98</b>	1,00	0,64	0,47	-0,02
<b>FROTA</b>	0,84	-0,26	0,77	0,68	0,64	1,00	0,62	0,60
<b>DENSI</b>	0,49	-0,35	0,59	0,43	0,47	0,62	1,00	0,50
<b>RENDA</b>	0,26	-0,22	0,16	0,03	-0,02	0,60	0,50	1,00

São apresentados, também, os coeficientes de correlação linear entre os pares de variáveis socioeconômicas (demais colunas da tabela). Percebe-se forte correlação entre as variáveis **DOMIC**, **POP\_T**, **POP\_A** (acima de 0,90). Portanto, estas não devem ser incluídas em uma mesma equação, concomitantemente, por ser indesejável a existência de correlação entre as variáveis independentes de um modelo de regressão.

A Tabela 5.2 destaca, ainda, os valores de **R** acima de 0,50, critério adotado neste estudo e que serviu apenas para auxiliar nas análises no momento da especificação dos modelos tradicionais.

#### 5.4.5 Definição de variáveis derivadas (taxas)

Foram definidas novas variáveis (**taxas**) para serem incluídas nas análises da **Fase 3** do estudo, derivadas a partir das variáveis **AREA**, **DOMIC**, **POP\_T** e **FROTA**:

**X21**: corresponde ao número de domicílios por km<sup>2</sup>;

**X32**: corresponde ao número de habitantes por domicílio;

**X51**: corresponde ao número de veículos particulares por km<sup>2</sup>;

**X52**: corresponde ao número de veículos por domicílio;

**X53**: corresponde ao número de veículos por habitantes.

Esta definição é baseada na hipótese de que as variáveis em forma de “taxas” poderiam amenizar os problemas inerentes aos dados de áreas (**MAUP**), minimizando as distorções nas médias das zonas, conforme mencionado no **item 4.4** do **Capítulo 4**.

Uma nova *matriz de correlação* foi gerada e foram analisadas as relações destas novas variáveis com a variável **VPBD** e também com as demais (Tabela 5.3). Percebe-

se que, assim como as variáveis *DENSI* e *RENDA*, que também podem ser consideradas **taxas**, uma vez que não representam totais de contagens e sim médias por área e por domicílio, respectivamente, as novas variáveis em forma de taxas não apresentam índices muito altos de correlação com a variável dependente *VPBD*. Nota-se, no entanto, forte correlação das variáveis *X21* e *X51* com a variável *DENSI* e também entre si (acima de 0,90). As variáveis *X52* e *X53* também estão altamente correlacionadas entre si e ambas apresentam, ainda, índices razoáveis de correlação (acima de 0,80) com a variável *RENDA*.

Definem-se, além das variáveis independentes, outras variáveis dependentes em forma de taxas, derivadas a partir da variável dependente *VPBD* e das variáveis independentes *AREA*, *DOMIC* e *POP\_T*:

*Y11*: corresponde ao número de *VPBD* por km<sup>2</sup>;

*Y12*: corresponde ao número de *VPBD* por domicílio;

*Y13*: corresponde ao número de *VPBD* por habitante.

**Tabela 5.3:** Matriz de correlação entre as variáveis socioeconômicas do ano base e a variável dependente *VPBD*, incluindo as variáveis **taxas**

	VPBD	AREA	DOMIC	POP_T	FROTA	DENSI	RENDA	X21	X32	X51	X52	X53
VPBD	1,00	-0,28	0,90	0,91	0,84	0,49	0,26	0,42	-0,22	0,43	0,28	0,30
AREA	-0,28	1,00	-0,23	-0,21	-0,26	-0,35	-0,22	-0,31	0,16	-0,30	-0,19	-0,21
DOMIC	<b>0,90</b>	-0,23	1,00	0,96	0,77	0,59	0,16	0,56	-0,33	0,46	0,09	0,17
POP_T	<b>0,91</b>	-0,21	<b>0,96</b>	1,00	0,68	0,43	0,03	0,37	-0,10	0,29	0,00	0,03
FROTA	<b>0,84</b>	-0,26	0,77	0,68	1,00	0,62	0,60	0,59	-0,49	0,69	0,64	0,68
DENSI	0,49	-0,35	0,59	0,43	0,62	1,00	0,50	0,98	-0,66	0,93	0,30	0,46
RENDA	0,26	-0,22	0,16	0,03	0,60	0,50	1,00	0,52	-0,55	0,69	0,84	0,89
X21	0,42	-0,31	0,56	0,37	0,59	<b>0,98</b>	0,52	1,00	-0,73	0,93	0,28	0,47
X32	-0,22	0,16	-0,33	-0,10	-0,49	-0,66	-0,55	-0,73	1,00	-0,72	-0,39	-0,62
X51	0,43	-0,30	0,46	0,29	0,69	<b>0,93</b>	0,69	<b>0,93</b>	-0,72	1,00	0,52	0,68
X52	0,28	-0,19	0,09	0,00	0,64	0,30	0,84	0,28	-0,39	0,52	1,00	0,96
X53	0,30	-0,21	0,17	0,03	0,68	0,46	0,89	0,47	-0,62	0,68	<b>0,96</b>	1,00

São analisados, então, os *índices de correlação* (Tabela 5.4) das variáveis independentes candidatas com estas novas variáveis dependentes. Pode-se perceber, de imediato, que, salientando-se os valores de **R** acima de 0,50 e destacando-se, ainda, entre estes, os acima de 0,90, a variável dependente *Y11* é a que apresenta maior número de variáveis independentes com alta correlação. Porém, é importante perceber que as variáveis *DENSI*, *X21* e *X51*, que apresentam relação mais forte com a variável

*Y11*, também são as que apresentam forte correlação entre si na análise da Tabela 5.3, o que não é desejável entre variáveis independentes de um mesmo modelo.

**Tabela 5.4:** Índices de correlação entre as variáveis independentes candidatas e as variáveis dependentes em forma de taxas, para os dados do ano base

	<i>Y11</i>	<i>Y12</i>	<i>Y13</i>
<i>AREA</i>	-0,34	-0,23	-0,31
<i>DOMIC</i>	0,53	-0,08	0,10
<i>POP_T</i>	0,38	-0,02	0,04
<i>FROTA</i>	0,67	0,25	0,51
<i>DENSI</i>	0,96	-0,09	0,29
<i>RENDA</i>	0,57	0,34	0,62
<i>X21</i>	0,92	-0,17	0,24
<i>X32</i>	-0,65	0,21	-0,34
<i>X51</i>	0,95	0,02	0,43
<i>X52</i>	0,43	0,53	0,72
<i>X53</i>	0,56	0,38	0,70

## 5.5. ANÁLISE DA AUTOCORRELAÇÃO ESPACIAL

As análises de *autocorrelação espacial* ocorreram em três fases distintas.

### 5.5.1 Análises da Fase 1

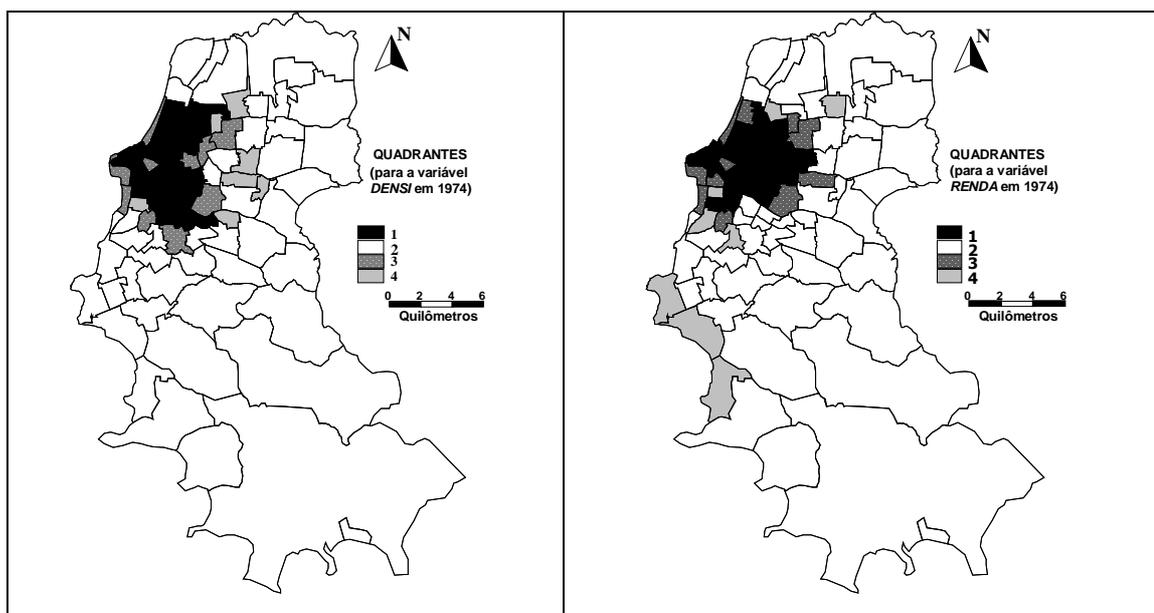
A **primeira fase**, que corresponde ao estudo preliminar, mencionado no início deste capítulo, foi menos aprofundada. Sendo assim, foram determinados apenas os **índices globais** de autocorrelação espacial e somente para duas variáveis socioeconômicas. Foi utilizado, nesta etapa, o *software* ArcView, pois além de estar compatível com o nível de detalhe da análise desejada, apresenta melhor compatibilidade, em termos de transferência de arquivos, com o *software* TransCAD.

As variáveis analisadas nesta fase, que são a *densidade populacional* (*DENSI*) e a *renda domiciliar média mensal* (*RENDA*), foram escolhidas por serem as mais utilizadas nos modelos de previsão de demanda por transportes e também por terem sido as indicadas, nos relatórios do **Plano Diretor de Transportes** originado da **Pesquisa**

**O-D de 1974 (EDOM 1974)**, como explicativas para as estimativas de *viagens produzidas de base domiciliar (VPBD)* para os planos de curto e médio prazos.

Foram aplicadas as ferramentas da ESDA para os dados das 96 ZTs do **ano base (1974)** e das 75 ZTs do **ano meta** (ZTs pesquisadas na EDOM 1986). As análises indicaram alto nível de *autocorrelação espacial* para as variáveis **DENSI** e **RENDA**, apresentando, respectivamente, valores de 0,48 e 0,42 para os índices *I de Moran* para o ano de 1974 e índices de 0,60 e 0,43 para ano de 1986.

Os mapas (*Box Map*) apresentados na Figura 5.4 indicam as ZTs, conforme o quadrante obtido no *gráfico de espalhamento de Moran*, para as duas variáveis (**DENSI** e **RENDA**) avaliadas com os dados do ano de 1974. Como pode ser observado, existem muitas ZTs agrupadas nos quadrantes Q1 e Q2, indicando a existência de semelhança entre os atributos destas zonas e de seus vizinhos. As zonas com valores altos (Alto-Alto) estão mais concentradas na área central da cidade e proximidades, indicando que elas têm, tanto para a variável **DENSI** como para a variável **RENDA**, média superior à média medida para toda a cidade. As zonas com valores baixos (Baixo-Baixo) encontram-se na periferia, estendendo-se ao sul do município, o que significa que as médias dos atributos são menores do que a média da cidade.

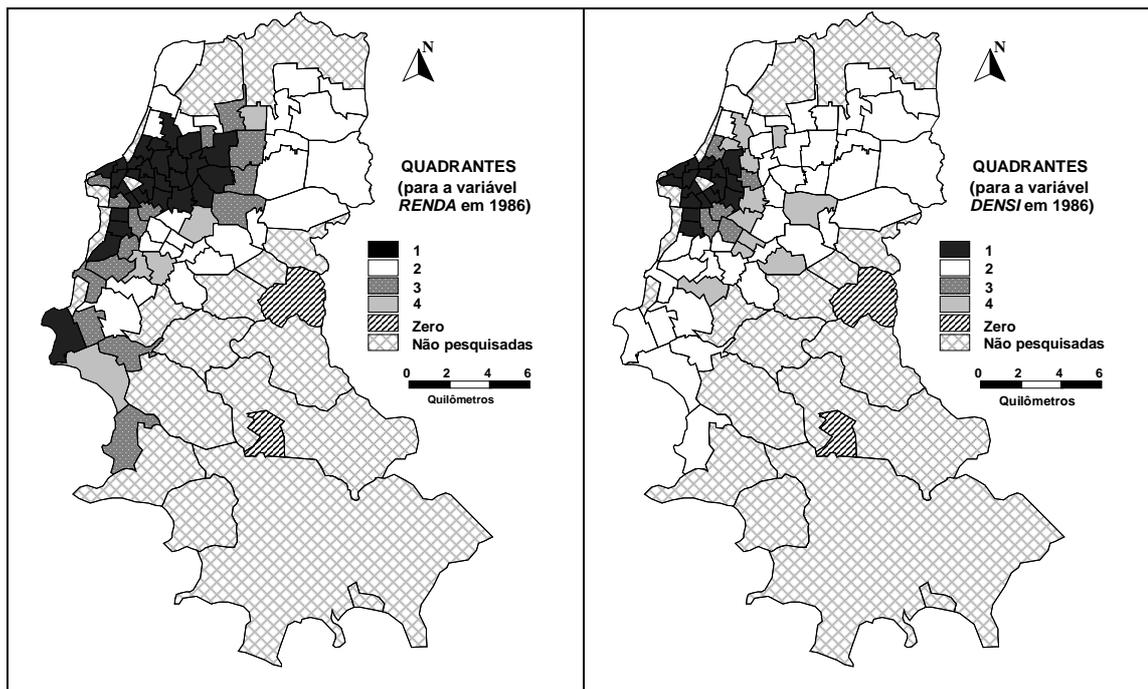


**Figura 5.4:** *Box Map* para as variáveis **DENSI** e **RENDA** em 1974

Existem poucas zonas de transição, ou seja, zonas cujo atributo é alto com

vizinhos cujos atributos são baixos (Alto-Baixo) ou o contrário (Baixo-Alto). Deve-se salientar que existem algumas zonas atípicas entre as zonas da área central da cidade, como, por exemplo, a região portuária e os parques da Redenção e Marinha do Brasil. Estas zonas, que apresentam baixa densidade populacional e baixa renda, inseridas em uma região de atributos altos, foram classificadas como zonas de transição (Baixo-Alto). Análises de autocorrelação excluindo-se algumas zonas atípicas, como as acima citadas, resultam em valores da estatística *I de Moran* mais elevados.

Analisando-se os *Box Map* para as mesmas variáveis em 1986 (Figura 5.5) percebe-se a mesma tendência de concentração de ZTs nos quadrantes Q1 (Alto-Alto) na área central da cidade, enquanto as zonas com quadrantes Q2 (Baixo-Baixo) localizam-se na periferia.



**Figura 5.5:** *Box Map* para as variáveis *DENSÍ* e *RENDA* em 1986

No entanto, existem diferenças nos quadrantes de algumas ZTs em relação a 1974. Isto pode ter ocorrido como decorrência de mudanças nos padrões de algumas zonas de um período para o outro, mas, provavelmente, esta diferença também se dá pela redução na área de abrangência da análise (de 96 para 75 zonas).

Cabe ressaltar que, para duas das 75 ZTs pesquisadas na EDOM de 1986, o *I de Moran* não pode ser calculado e elas não foram classificadas em nenhum quadrante, recebendo valor zero. As demais 21 ZTs foram classificadas como “não pesquisadas”.

### 5.5.2 Análises da Fase 2

Na **segunda fase da Etapa 5**, as análises espaciais para as variáveis *DENSI* e *RENDA* foram realizadas através do *software* SPRING, para que fosse possível a obtenção dos índices *I de Moran locais*, ou seja, as *variáveis espaciais locais* para testar novos *modelos alternativos*. Também foi necessária a verificação dos novos índices globais, uma vez que as ZTs que representavam *outliers* foram excluídas da base de 1974.

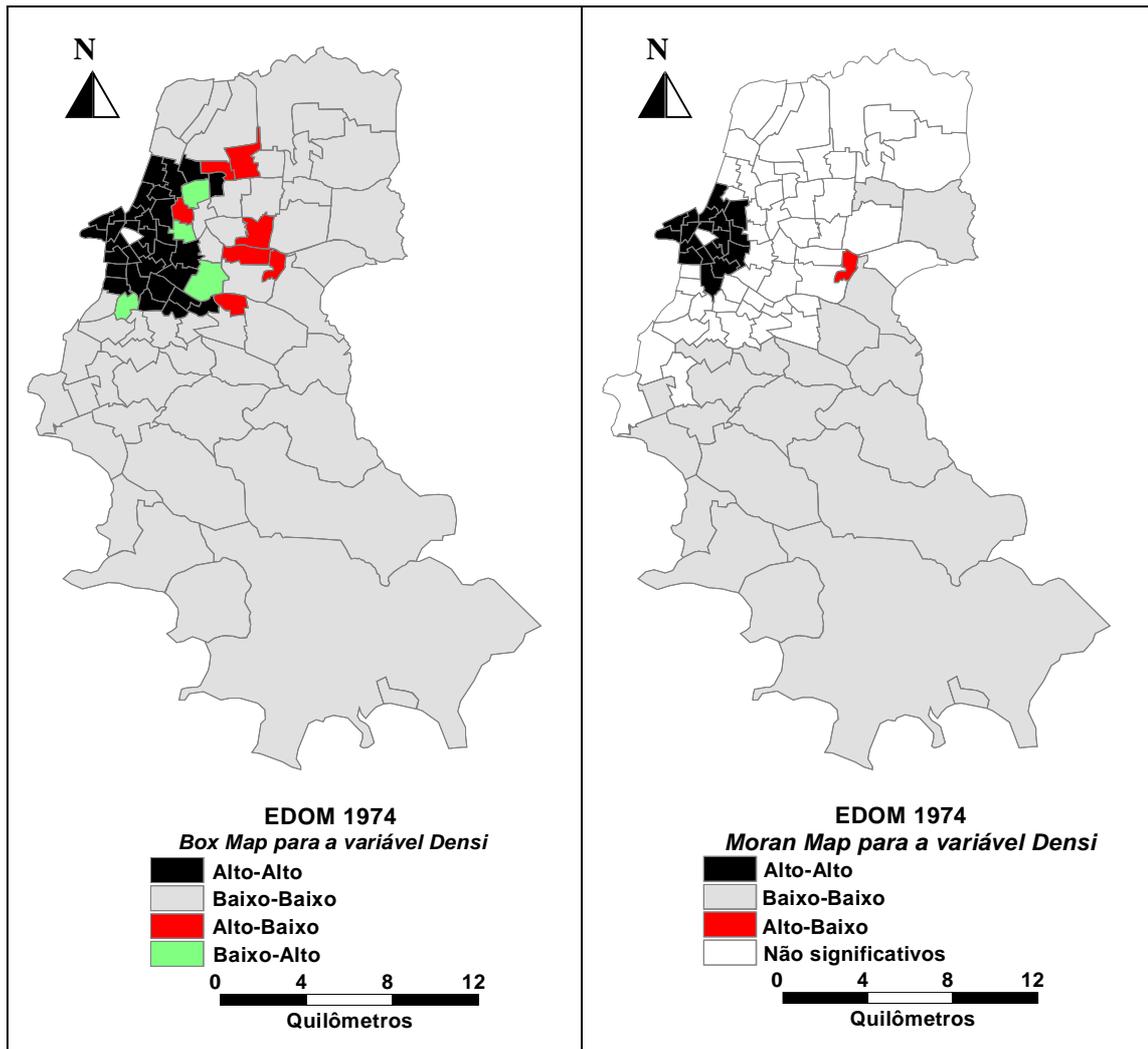
Os resultados obtidos para os índices *I de Moran Globais* foram 0,73 e 0,59, respectivamente para as variáveis *DENSI* e *RENDA*, para os dados do **ano base**, o que representa que houve aumentos significativos em relação aos valores resultantes das análises da **fase 1** (0,48 e 0,42, respectivamente). Tal aumento já era esperado, devido à exclusão de seis zonas atípicas.

Os valores resultantes das análises para os dados de 1986 foram 0,57 para *DENSI* e 0,44 para *RENDA*, também apresentando diferenças em relação aos valores verificados na primeira análise (0,60 e 0,43, respectivamente). Neste caso, porém, as diferenças não são significativas e são devidas às diferentes metodologias aplicadas por cada um dos *software* (ArcView e SPRING), principalmente no que diz respeito aos critérios para geração da *matriz de proximidade espacial*.

A partir das informações nas colunas criadas automaticamente pelo *software* SPRING na tabela de objetos é possível gerar os mapas e gráficos necessários para a análise visual dos resultados. As Figuras 5.6 (a e b) e 5.7 (a e b) apresentam os *Box Map* e *Moran Map* para as variáveis analisadas (*DENSI* e *RENDA*, respectivamente), para os dados de 1974, enquanto as Figuras 5.8 (a e b) e 5.9 (a e b) apresentam as mesmas análises para os dados de 1986.

Analisando-se o *Box Map* apresentado na Figura 5.6 (a), para variável *DENSI*

em 1974 observa-se a mesma concentração (verificada na **fase 2**) de zonas com valores altos (Alto-Alto) na área central da cidade (32 ZTs), enquanto que os agrupamentos de valores baixos de densidade (Baixo-Baixo) são observados na periferia, ao Sul, ao Norte e ao Leste do centro da cidade, distribuídos em 47 ZTs. Porém, é verificado um número menor de zonas intermediárias, 11 ao todo: 7 ZTs de densidade alta com vizinhos de densidade baixa (Alto-Baixo) e 4 ZTs de densidade baixa com vizinhos cujo atributo é alto (Baixo-Alto).



**Figura 5.6:** *Box Map* e *Moran Map* para a variável *DENSI* em 1974

O *Moran Map* apresenta apenas as zonas cujos *índices locais de Moran* foram considerados significativos (acima de 95 % de significância), conforme o quadrante a que pertence. Portanto, na Figura 5.6 (b) pode-se observar um agrupamento importante de 32 ZTs com densidades populacionais altas (Alto-Alto) no centro da cidade e um outro grande agrupamento (16 zonas na região Sul) de baixos índices de densidade (Baixo-Baixo).

Na Zona Leste observam-se duas ZTs de atributo baixo altamente relacionados com os atributos de seus vizinhos. Apenas uma zona intermediária, de atributo alto próxima de zonas de atributo baixo (Alto-Baixo), apresenta significância acima de 95 % e 50 das 90 ZTs estão abaixo de 95 % de significância na análise do índice local de Moran.

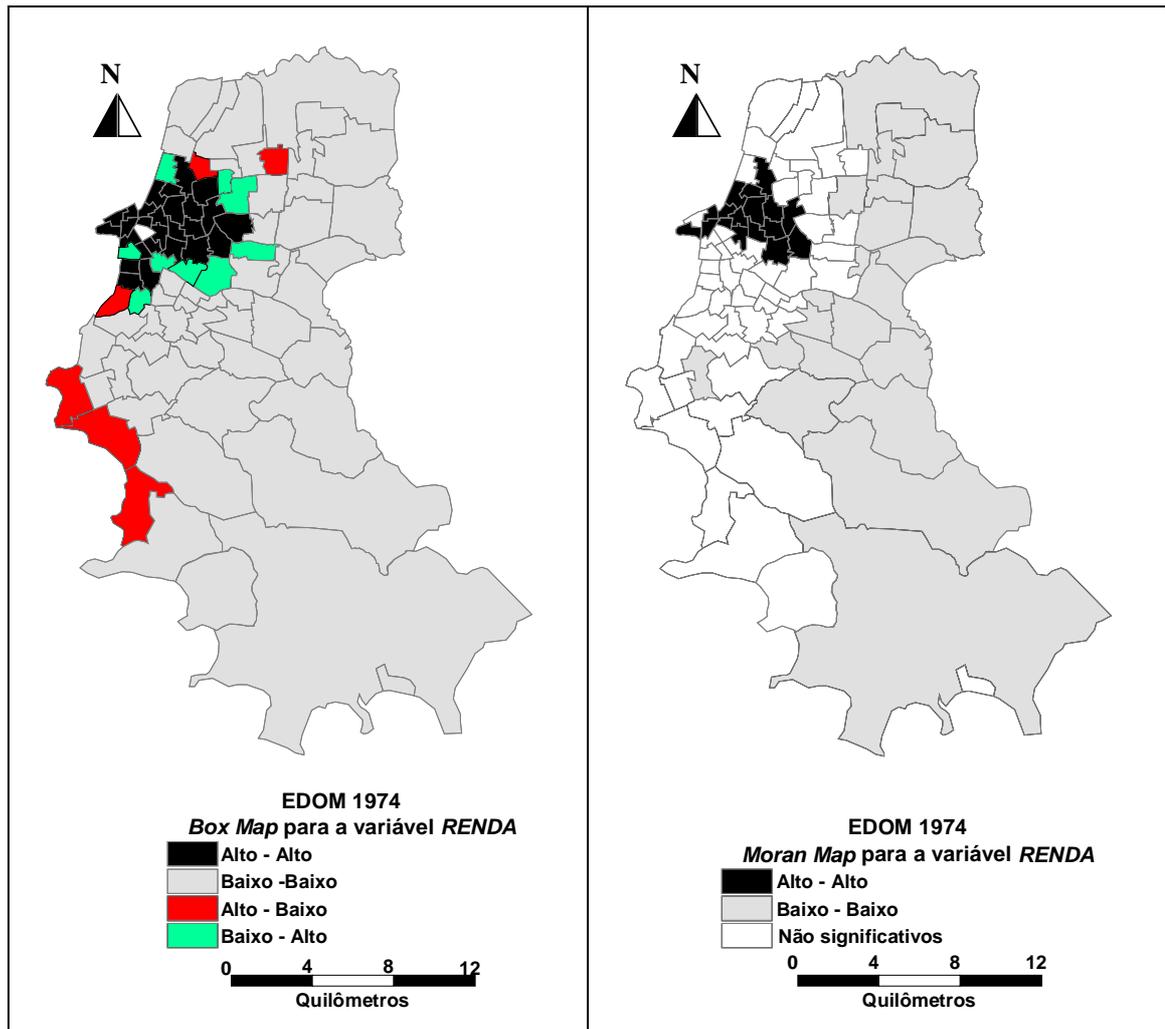
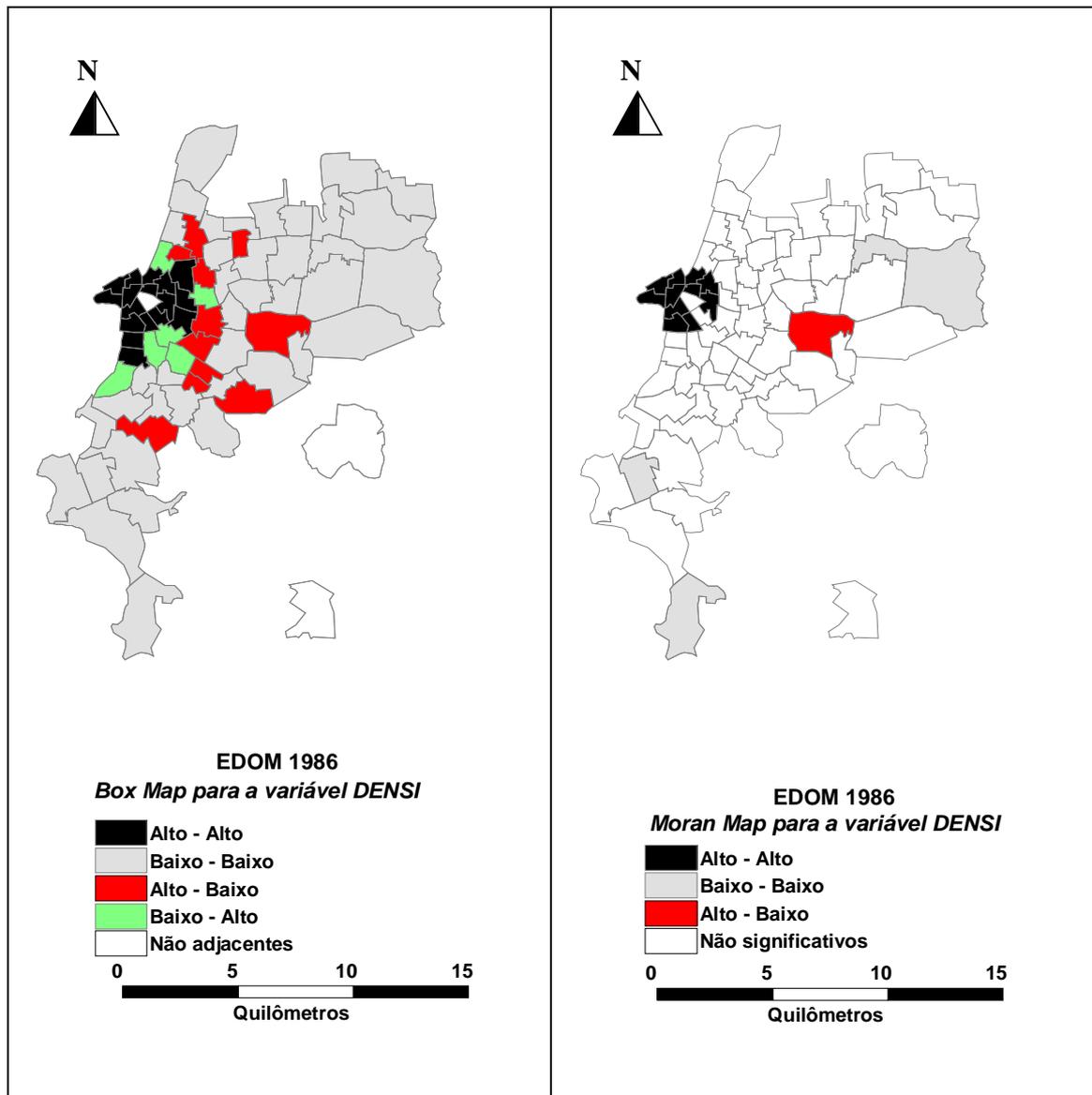


Figura 5.7: Box Map e Moran Map para a variável RENDA em 1974

Na análise da variável **RENDA** para os dados de 1974 verifica-se, no *Box Map* da Figura 5.7 (a), uma concentração de 27 zonas com atributo alto no centro da cidade, das quais 17 apresentam *índices locais de Moran* significativos, conforme observado no *Moran Map* da Figura 5.7 (b).

Um grande grupo de zonas de renda baixa (48 ZTs) está localizado na periferia

ao Norte, Leste e Sul do centro da cidade, conforme Figura 5.7 (a), havendo uma faixa com 22 zonas, que vai de Norte a Sul ao longo da fronteira Leste, representando um agrupamento com significância maior que 95 %. Das 15 zonas de transição, 6 são classificadas como “Alto-Baixo” e 9 como “Baixo-Alto” e nenhuma apresenta *índice local de Moran* significativo. São verificadas 52 zonas com índices não significativos.



**Figura 5.8:** Box Map e Moran Map para a variável *DENSITY* em 1986

Para as análises de 1986, cabe lembrar que somente 75 ZTs apresentam dados e que 2 das ZTs não são adjacentes, o que impossibilita a determinação do quadrante porque não se pode calcular a média dos vizinhos. Conseqüentemente, são ao todo apenas 73 ZTs analisadas, o que reduz a abrangência da área de estudos em relação às

análises de 1974.

O *Box Map* da Figura 5.8 (a) apresenta um agrupamento de 17 ZTs de densidade alta na área central da cidade (Alto-Alto), 39 ZTs cuja densidade é baixa, distribuídas na periferia da área de estudos, e 19 ZTs intermediárias, onde 11 são zonas de densidade alta com vizinhos de densidade baixa (quadrante 3, Alto-Baixo) e 8 são zonas de densidade baixa cujos vizinhos apresentam densidade alta (quadrante 4, Baixo-Alto).

Na Figura 5.8 (b), é possível observar, através do *Moran Map*, que apenas 16 zonas apresentam significância maior que 95 % nos *índices locais de Moran*. Entre estas, 12 são zonas de densidade alta, das quais 11 estão localizadas no centro da cidade (quadrante 1, Alto-Alto), e uma zona de transição (quadrante 3, Alto-Baixo), localizada na periferia, a leste da área central. Existem apenas 4 zonas no quadrante 2 (Baixo-Baixo), nenhuma no quadrante 4 (Baixo-Alto) e 59 são classificadas como não significantes.

Analisando-se a distribuição espacial da variável **RENDA** para 1986, na Figura 5.9 (a), observa-se a existência, ao todo, de 33 zonas com atributos altos, sendo que 28 estão localizadas também entre vizinhos de renda alta (Alto-Alto), 24 das quais no centro e estendendo-se a leste, 3 nas imediações do centro em direção ao sul e uma única, mais afastada, também em direção à Zona Sul. As outras 5 ZTs estão localizadas em regiões de renda baixa (quadrante 3, Alto-Baixo), na periferia.

As zonas de renda baixa são 40 ao todo, estando 14 delas inseridas entre zonas de renda alta (Baixo-Alto), nas imediações do centro da cidade e também na região Sul, sendo as outras 26 (Baixo-Baixo) localizadas nas zonas periféricas a leste e norte do centro da cidade e, ainda, em uma área intermediária entre a Zona Sul e o Centro.

Os agrupamentos significativos, que podem ser observados no *Moran Map*, Figura 5.9 (b), consistem em 11 zonas de renda alta (Alto-Alto) nas imediações do centro, estendendo-se a leste, uma zona intermediária de renda alta (Alto-Baixo), ao sul do centro, e 7 zonas de renda baixa (Baixo-Baixo), divididas em 3 agrupamentos na periferia da área de estudos, a nordeste, leste e sudeste. Existem, no total, 56 zonas classificadas como não significantes para esta análise.

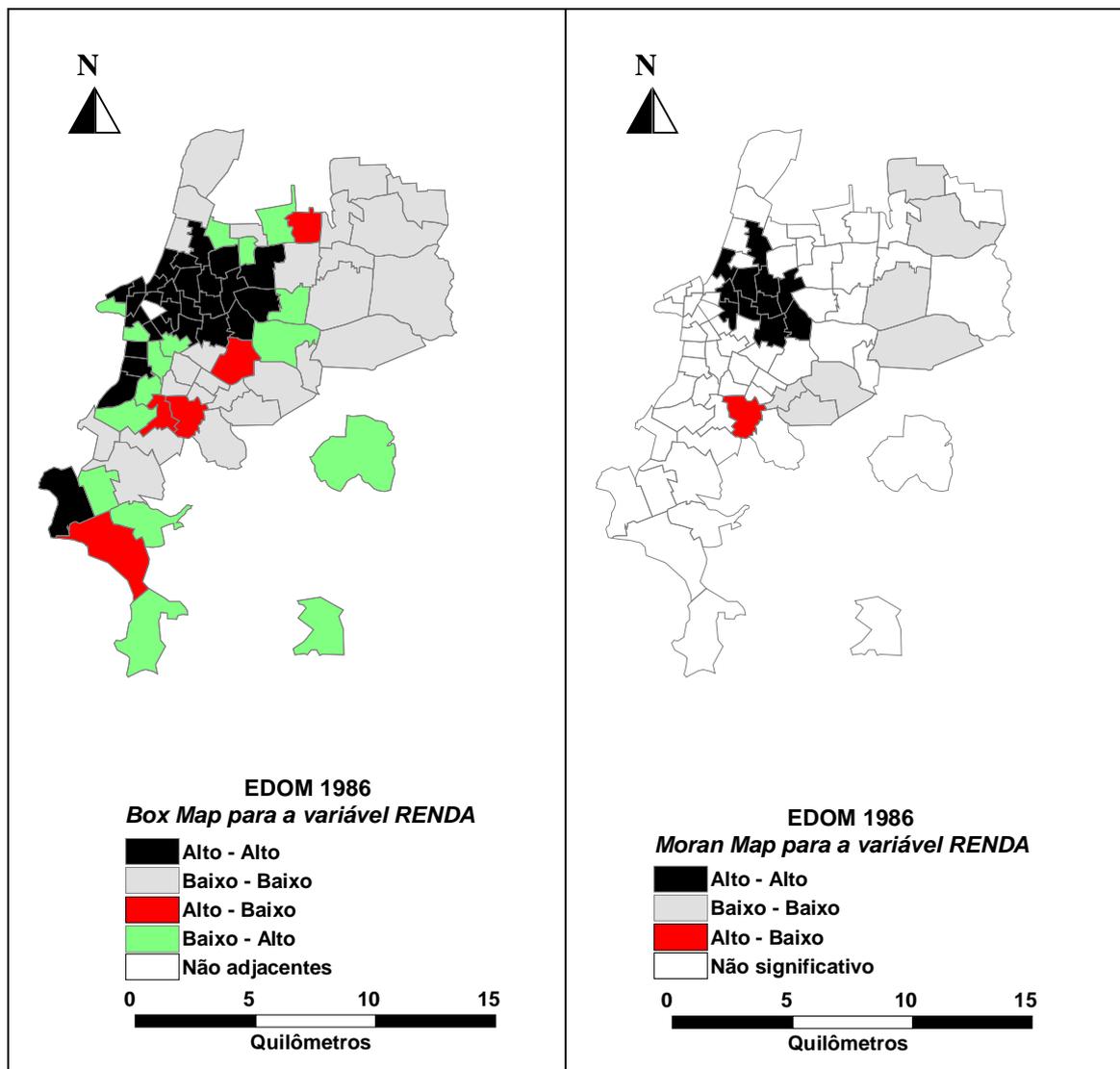


Figura 5.9: *Box Map* e *Moran Map* para a variável *RENDA* em 1986

### 5.5.3 Análises da Fase 3

As análises espaciais da **fase 3** incluíram as variáveis “taxas” (*X21*, *X32*, *X51*, *X52* e *X53*) e também as variáveis socioeconômicas *POP\_T*, *DOMIC* e *FROTA*, uma vez que estas são as que mais se correlacionam com a variável dependente *VPBD*, que também foi analisada. O objetivo principal é obter outros indicadores de dependência espacial (além dos já estudados nas fases anteriores) para introduzir nos modelos de demanda alternativos que serão analisados nesta fase.

Portanto, faz-se necessário, além de verificar o nível de dependência espacial de cada variável através da análise do índice *I Global de Moran* (Tabela 5.5), também

analisar como a distribuição espacial de cada variável socioeconômica se relaciona com a distribuição espacial da variável dependente *VPBD*. Este procedimento, que foi realizado através de levantamentos numéricos dos quadrantes (Tabelas 5.6, 5.7 e 5.8) e da relação destes com os quadrantes da variável dependente (Figuras 5.10, 5.11 e 5.12), orientou a decisão de quais indicadores de dependência espacial, dentre os obtidos com as variáveis socioeconômicas analisadas, seriam as potenciais *variáveis espaciais* para os modelos alternativos.

**Tabela 5.5:** Índices de autocorrelação espacial (*I de Moran Global*) das variáveis socioeconômicas e da variável dependente *VPBD* para os dados do **ano base** (1974) e do **ano meta** (1986)

Variável dependente	Descrição da Variável	I de Moran Global (1974)	I de Moran Global (1986)
<b>VPBD</b>	Viagens produzidas de base domiciliar	0,3467	0,2849
<b>Variáveis independentes (contagem)</b>			
<b>DOMIC</b>	Número total de domicílios	0,1881	0,1216
<b>POP_T</b>	Número total de habitantes	<b>0,2352</b>	<b>0,2379</b>
<b>FROTA</b>	Número total de veículos particulares	0,2300	0,1990
<b>Variáveis independentes (taxas derivadas)</b>			
<b>RENDA</b>	Renda média mensal por domicílio	<b>0,5881</b>	0,4404
<b>DENSI</b>	Densidade (habitantes por km <sup>2</sup> )	<b>0,7330</b>	<b>0,5681</b>
<b>X21</b>	Número de domicílios por Km <sup>2</sup>	<b>0,7860</b>	<b>0,6428</b>
<b>X32</b>	Habitantes por domicílio	0,3784	<b>0,5349</b>
<b>X51</b>	Veículos particulares por km <sup>2</sup>	<b>0,7543</b>	<b>0,6279</b>
<b>X52</b>	Veículos particulares por domicílio	0,4469	0,2584
<b>X53</b>	Veículos particulares por habitante	0,2997	<b>0,4321</b>

Nos índices apresentados na Tabela 5.5 observa-se que as variáveis de taxas derivadas apresentam, em geral, maiores valores de autocorrelação espacial do que as variáveis que representam os totais de contagens, com destaque para as variáveis *X21* e *X51*, seguidas pelas variáveis *DENSI* e *RENDA*, para os dados de 1974. Para os dados de 1986, os maiores valores são verificados também para as variáveis *X21* e *X51*, seguidas pelas variáveis *DENSI* e *X32*. Dentre as variáveis que representam totais de contagens por ZT, os maiores índices são verificados para a variável *POP\_T*, tanto para

os dados do **ano base** como para os do **ano meta**, conforme está destacado na tabela.

A variável dependente *VPBD* também apresenta índices significativos de autocorrelação espacial (0,3467 e 0,2849 para o **ano base** e **ano meta**, respectivamente), indicando que a variável é espacialmente dependente. Percebe-se, também, que, em geral, os índices de autocorrelação espacial são mais elevados para o **ano base** do que para o **ano meta**, porém uma comparação direta entre os mesmos não é prudente, uma vez que a abrangência da área de estudos é diferente.

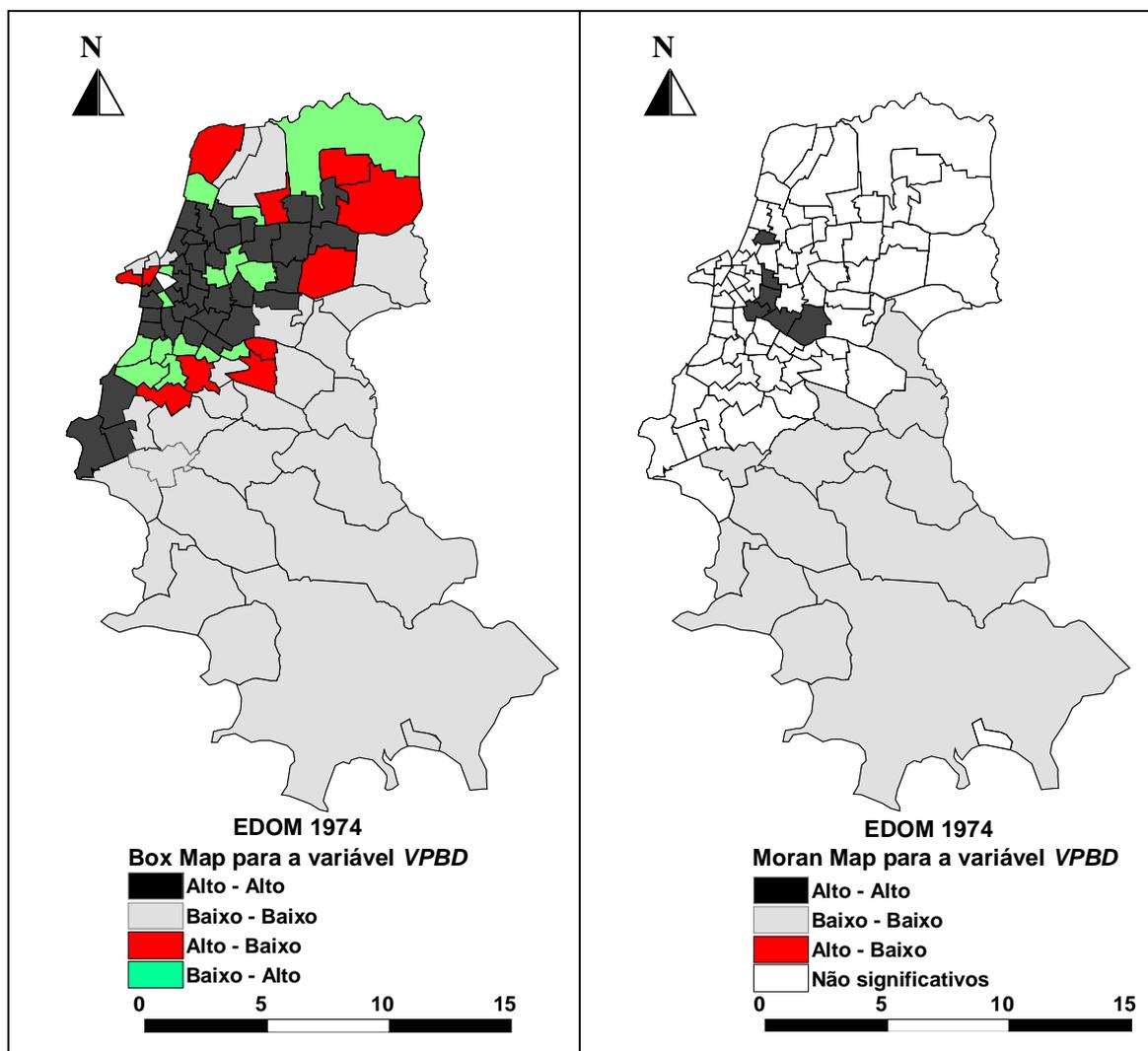


Figura 5.10: *Box Map* e *Moran Map* para a variável *VPBD* em 1974

A Figura 5.10 apresenta a distribuição espacial da variável dependente *VPBD* para 1974, através do *Box Map* e *Moran Map*, com o resultado dos quadrantes verificados sendo apresentado na Tabela 5.6. A coluna “**Box Map**” mostra, para o

período de 1974, o número de ZTs em cada quadrante, enquanto que a coluna “**Moran Map**” apresenta o número de ZTs, em cada quadrante, cujos índices locais de Moran apresentam significância acima de 95 %, para o mesmo período.

**Tabela 5.6:** Resumo dos quadrantes verificados nas análises do *Box Map* e do *Moran Map* para a variável dependente *VPBD*, para os dados de 1974

		EDOM 1974	
		BOX MAP	MORAN MAP
VPBD	Q1 (Alto-Alto)	36	6
	Q2 (Baixo-Baixo)	28	12
	Q3 (Alto-Baixo)	11	0
	Q4 (Baixo-Alto)	15	0
	Não significativa		71

Desta forma, nos resultados da análise espacial da variável dependente *VPBD*, para os dados de 1974 (Figura 5.10 e Tabela 5.6), percebe-se, através da coluna “**Box Map**” da Tabela 5.6, um maior número de zonas que apresentam números altos de viagens com vizinhos também com o atributo alto (36 ZTs no Q1). Estas zonas estão divididas em dois grupos, como pode ser observado na Figura 5.10: um grande grupo de 33 ZTs, que se estende da área central para a região nordeste, e um grupo menor (3 ZTs), localizado a sudoeste do centro, em uma área da cidade chamada de “Zona Sul” (Bairros Tristeza e Ipanema).

Um número um pouco menor (28 ZTs) é observado para as zonas de baixo número de viagens com vizinhos cujo atributo também é baixo (Q2). No entanto, como pode ser observado na coluna “**Moran Map**” (Tabela 5.6), as que apresentam *índices locais de Moran* significativos são, na maior parte, as de atributos baixos (12 ZTs no Q2), que representam um grande agrupamento nas regiões sul e sudeste da cidade. Entre as ZTs localizadas no quadrante 1 para *VPBD*, apenas 6 são significativas e, das 26 zonas de transição verificadas no *Box Map* (Q3 e Q4), nenhuma apresenta significância maior do que 95 % no índice *I de Moran Local*.

As Tabelas 5.7 e 5.8 apresentam a análise espacial para as variáveis socioeconômicas de totais de contagens e taxas derivadas, respectivamente, conforme pode ser verificado, também, nos *Box Map* e *Moran Map* das Figuras 5.6 e 5.7, para a variável *DENSI* e *RENDA*, e **Figuras C1 a C8 do Anexo C**, para as demais variáveis.

Percebe-se que, em geral, para as variáveis de totais de contagens (Tabela 5.7), o maior número de zonas é verificado no quadrante 1 (Alto-Alto), enquanto que o contrário (maior número de zonas no **Q2**) é verificado para as variáveis de taxas derivadas (Tabela 5.8), conforme está destacado nas colunas “**Box Map**”. No entanto, excluindo-se as variáveis **FROTA**, **X52** e **X53**, o maior número de ZTs em agrupamentos significativos é verificado para as zonas pertencentes ao quadrante 2 (Baixo-Baixo).

A análise comparativa se dá através da verificação da coincidência entre os quadrantes das variáveis socioeconômicas com os quadrantes verificados para **VPBD** (Figuras 5.11 e 5.12), que é apresentada, também, nas colunas “**comum**”, à direita das colunas “**Box Map**” e “**Moran Map**”, nas Tabelas 5.7 e 5.8.

**Tabela 5.7:** Análises do *Box Map* e do *Moran Map* para as variáveis que representam totais de contagens e relação com os quadrantes da **VPBD** (EDOM 1974)

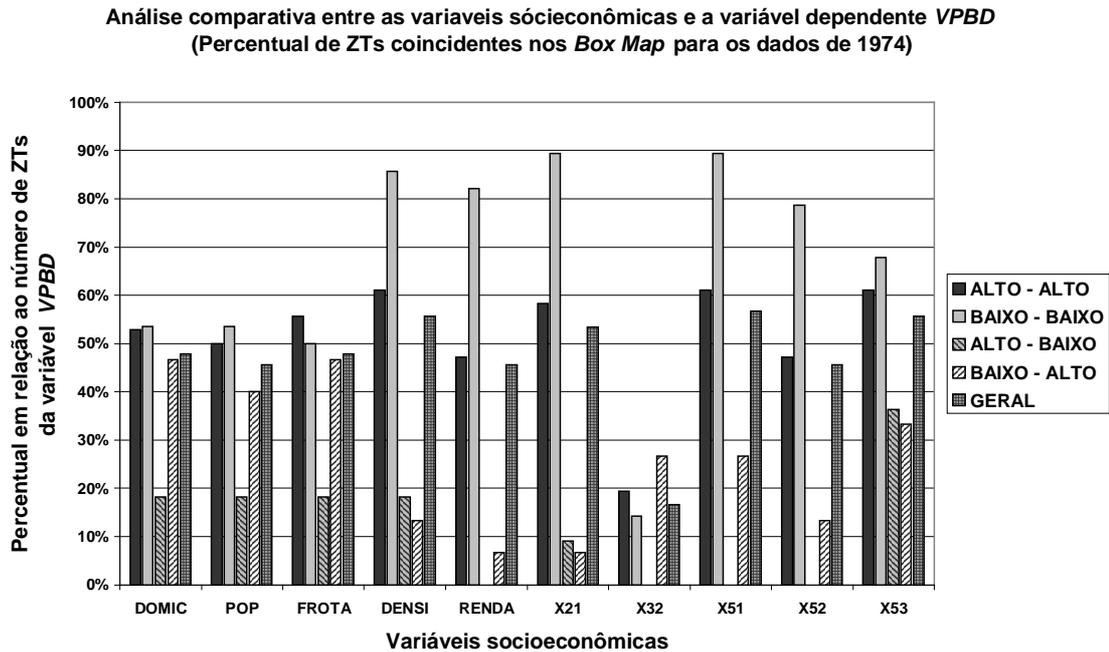
		<b>BOX MAP</b>	<b>Comum</b>	<b>MORAN MAP</b>	<b>Comum</b>
<b>POP_T</b>	<b>Q1 (Alto-Alto)</b>	32	21	3	0
	<b>Q2 (Baixo-Baixo)</b>	27	20	9	5
	<b>Q3 (Alto-Baixo)</b>	11	2	1	0
	<b>Q4 (Baixo-Alto)</b>	22	6	1	0
	<b>Não significativa</b>			78	
<b>DOMIC</b>	<b>Q1 (Alto-Alto)</b>	29	19	5	1
	<b>Q2 (Baixo-Baixo)</b>	24	15	6	5
	<b>Q3 (Alto-Baixo)</b>	18	2	0	0
	<b>Q4 (Baixo-Alto)</b>	19	7	3	0
	<b>Não significativa</b>			76	
<b>FROTA</b>	<b>Q1 (Alto-Alto)</b>	31	20	8	4
	<b>Q2 (Baixo-Baixo)</b>	29	14	7	6
	<b>Q3 (Alto-Baixo)</b>	14	2	3	0
	<b>Q4 (Baixo-Alto)</b>	16	7	1	0
	<b>Não significativa</b>			71	

O gráfico da Figura 5.11 apresenta o percentual de coincidência entre os quadrantes das ZTs verificados para cada variável socioeconômica na análise dos *Box Map*, para os dados de 1974, tomando como base os quadrantes verificados para a variável dependente **VPBD**, no mesmo período. Percebe-se que os maiores percentuais de coincidências são para o quadrante 2 (**Q2**), com destaque para as variáveis **DENSI**,

**RENDA**, **X21** e **X51**, cujos índices de relação estão entre 80 % e 90 %. As variáveis **DENSI** e **X51**, juntamente com a variável **X53**, destacam-se entre os maiores percentuais de coincidência para os quadrantes **Q1** (pouco acima de 60 %), seguidas pela variável **X21** (pouco abaixo de 60 %). Na análise geral, onde não é feita distinção entre quadrantes, destacam-se, mais uma vez, as variáveis **DENSI**, **X21**, **X51** e **X53**.

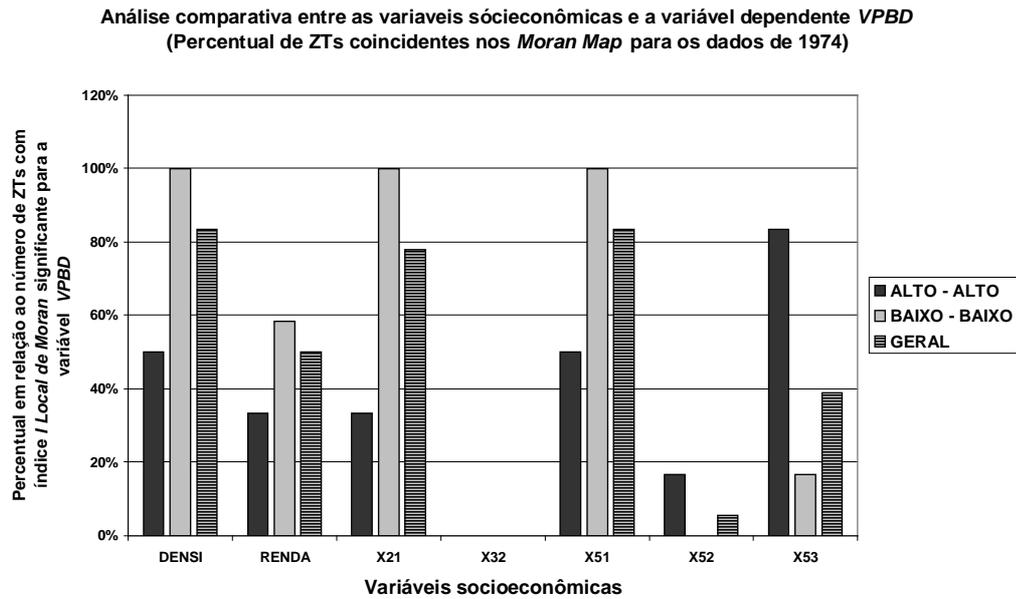
**Tabela 5.8:** Análises do *Box Map* e do *Moran Map* para as variáveis que representam taxas derivadas e relação com os quadrantes da *VPBD* (EDOM 1974)

		<b>BOX MAP</b>	<b>Comum</b>	<b>MORAN MAP</b>	<b>Comum</b>
<b>DENSI</b>	<b>Q1 (Alto-Alto)</b>	32	22	18	3
	<b>Q2 (Baixo-Baixo)</b>	47	24	21	12
	<b>Q3 (Alto-Baixo)</b>	7	2	1	0
	<b>Q4 (Baixo-Alto)</b>	4	2	0	0
	<b>Não significativa</b>				50
<b>RENDA</b>	<b>Q1 (Alto-Alto)</b>	27	17	17	2
	<b>Q2 (Baixo-Baixo)</b>	48	23	22	7
	<b>Q3 (Alto-Baixo)</b>	6	0	0	0
	<b>Q4 (Baixo-Alto)</b>	9	1	0	0
	<b>Não significativa</b>				51
<b>X21</b>	<b>Q1 (Alto-Alto)</b>	31	21	15	2
	<b>Q2 (Baixo-Baixo)</b>	52	25	23	11
	<b>Q3 (Alto-Baixo)</b>	4	1	0	0
	<b>Q4 (Baixo-Alto)</b>	3	1	0	0
	<b>Não significativa</b>				52
<b>X32</b>	<b>Q1 (Alto-Alto)</b>	35	7	5	0
	<b>Q2 (Baixo-Baixo)</b>	35	4	9	0
	<b>Q3 (Alto-Baixo)</b>	0	0	0	0
	<b>Q4 (Baixo-Alto)</b>	20	4	7	0
	<b>Não significativa</b>				69
<b>X51</b>	<b>Q1 (Alto-Alto)</b>	29	22	17	3
	<b>Q2 (Baixo-Baixo)</b>	51	25	27	12
	<b>Q3 (Alto-Baixo)</b>	4	0	0	0
	<b>Q4 (Baixo-Alto)</b>	6	4	0	0
	<b>Não significativa</b>				46
<b>X52</b>	<b>Q1 (Alto-Alto)</b>	24	17	12	1
	<b>Q2 (Baixo-Baixo)</b>	48	22	1	0
	<b>Q3 (Alto-Baixo)</b>	7	0	3	0
	<b>Q4 (Baixo-Alto)</b>	11	2	3	0
	<b>Não significativa</b>				71
<b>X53</b>	<b>Q1 (Alto-Alto)</b>	32	22	14	5
	<b>Q2 (Baixo-Baixo)</b>	37	19	6	2
	<b>Q3 (Alto-Baixo)</b>	11	4	2	0
	<b>Q4 (Baixo-Alto)</b>	10	5	1	0
	<b>Não significativa</b>				67



**Figura 5.11:** Percentual de ZTs com quadrantes coincidentes entre as variáveis socioeconômicas e *VPBD*: análise dos *Box Map*, dados de 1974

Na Figura 5.12 é apresentada análise semelhante à anterior, porém considerando-se apenas as ZTs em agrupamentos significativos. O cálculo do percentual de ZTs verificadas em cada quadrante dos *Moran Map* para as variáveis socioeconômicas tem por base o número total, em cada quadrante, de ZTs em agrupamentos significativos na análise da variável dependente. Nesta análise também se verificam maiores coincidências no quadrante **Q2**, com destaque mais uma vez para as variáveis *DENSI*, *X21*, *X51*, que apresentaram 100 % de coincidência com as ZTs verificadas, neste quadrante, na mesma análise para a variável *VPBD*. Para as ZTs coincidentes no quadrante 1 (**Q1**), destaca-se a variável *X53*, com percentual acima de 80 %. A variável *X32* não apresenta ZTs em agrupamentos significantes, concomitantemente com a variável dependente.



**Figura 5.12:** Percentual de ZTs com quadrantes coincidentes entre as variáveis socioeconômicas e *VPBD* - análise dos *Moran Map*, dados de 1974

## 5.6. ESPECIFICAÇÃO, CALIBRAÇÃO E DIAGNÓSTICO DOS MODELOS TRADICIONAIS

Para a especificação de modelos tradicionais, que ocorreu nas três fases, foram levados em consideração, entre as informações disponíveis na base de dados de 1974 (**ano base**), cinco critérios para a escolha das variáveis independentes:

1. Variáveis que eram comuns às duas pesquisas (**ano base** e **ano meta**), uma vez que deveriam entrar nos modelos na fase de calibração (**ano base**) e, posteriormente, para as estimativas (**ano meta**), conforme análise dos dados das pesquisas, apresentada nos **itens 5.4.1** e **5.4.2**;
2. Variáveis indicadas nos relatórios do Plano Diretor de Transportes, originado pela Pesquisa O-D de 1974 (EDOM 1974), como explicativas para as estimativas de **viagens produzidas de base domiciliar (VPBD)** para os planos de curto e médio prazos incluídos no estudo, citadas no **item 5.4.3**;
3. Variáveis citadas na literatura para estimativas de **VPBD**, também citadas no **item 5.4.3**;
4. Variáveis que apresentaram correlação significativa com a variável dependente **VPBD** na análise realizada na **Etapa 4**, citada no **item 5.4.4**;

5. Hipótese de que as variáveis “taxas” poderiam amenizar os problemas inerentes aos dados de áreas (MAUP). Ver **item 4.4 do Capítulo 4**, e **item 5.4.5**.

### 5.6.1 Modelos tradicionais da Fase 1

No estudo preliminar, foram analisados modelos cujas variáveis independentes eram diferentes combinações de *população residente total (POP\_T)*, *densidade populacional (DENSI)* e *renda média mensal (RENDA)*, atendendo primeiramente aos critérios “1” e “2”.

Foram especificados quatro modelos tradicionais, cuja diferença básica entre eles foi a inclusão ou não da variável *DENSI* e a consideração da variável *RENDA* em unidades de dólar (*RENDA<sub>D</sub>*) ou em unidades de salário mínimo (*RENDA<sub>SM</sub>*). A conversão da variável *RENDA* foi necessária, uma vez que a base de dados de cada pesquisa (**ano base** e **ano meta**) apresenta os valores em unidades da moeda corrente nas épocas de cada pesquisa (Cr\$ e Cz\$, respectivamente em 1974 e 1986). Os valores de salários mínimos e de dólares, adotados para conversão, são apresentados na **Tabela 5.9**.

**Tabela 5.9:** Valores de dólar e salário mínimo adotados para a conversão da variável *RENDA* em *RENDA<sub>D</sub>* e *RENDA<sub>SM</sub>*

	Dólar	Salário mínimo
<b>1974 (ano base)</b>	Cr\$ 6,34	Cr\$ 376,80
<b>1986 (ano meta)</b>	Cz\$ 13,84	Cz\$ 804,00

O estudo da **Fase 1** foi essencial para definir algumas diretrizes para a pesquisa e chamou a atenção para a necessidade de uma análise mais detalhada dos modelos. Um resumo destes modelos, que apresenta os coeficientes para a constante e para cada variável incluída em cada modelo analisado, os valores do  $R^2$  dos modelos e os resultados das estimativas para o **ano base**, através dos valores dos *erros relativos médios (ERM)*, podem ser observados na **Tabela 5.10**.

Uma conclusão importante foi a verificação de 6 zonas atípicas (*outliers*), que foram excluídas posteriormente. Já havia sido detectada a existência de algumas zonas

atípicas na etapa de análise espacial. As áreas correspondentes ao Parque Farroupilha (ZT 9) e as que correspondem a uma região onde, atualmente, se encontra o Parque Marinha do Brasil (ZT 94 e ZT 95), não constituíam áreas residenciais e, portanto, a população era nula. Os outros 3 *outliers* (ZT 1, ZT 6 e ZT 11) foram detectados na aplicação dos modelos em estimativas para o **ano base**. São áreas que apresentam valores bem abaixo da média para a população residente e, principalmente, para o número de viagens produzidas de base domiciliar.

**Tabela 5.10:** Resumo dos modelos tradicionais especificados na **Fase 1**

	<b>1T</b>	<b>2T</b>	<b>3T</b>	<b>4T</b>
<b>Constante</b>	-1698,40	-1698,40	-1690,46	-1690,46
<b>POP_T</b>	1,00967	1,00967	1,00859	1,00859
<b>DENSI</b>	-0,001955	-0,001955		
<b>RENDA<sub>SM</sub></b>	318,124		316,479	
<b>RENDA<sub>D</sub></b>		5,35273		5,32505
<b>R<sup>2</sup></b>	0,8980	0,8980	0,8980	0,8980
<b>ERM (ano base)</b>	80 %	80 %	80 %	80 %

Pode-se observar, através da análise da **Tabela 5.10**, que o valor do coeficiente de determinação (**R<sup>2</sup>**) não é alterado com a inclusão ou não da variável **DENSI** ou alternando-se a variável **RENDA** para unidades de dólar (**RENDA<sub>D</sub>**) ou salários mínimos (**RENDA<sub>SM</sub>**). Quanto aos resultados das estimativas, os valores dos ERM das projeções para o **ano base** permanecem iguais (80 %) para todos os modelos.

Percebe-se que os valores de **ERM** são muito altos para as estimativas para o **ano base**, indicando que existe algum problema com os modelos, o que requer uma investigação mais detalhada. Foram excluídas as zonas que representavam os *outliers* (ZTs que não constavam da base de dados da pesquisa de 1986) quando da calibração e avaliação dos modelos, o que reduziu consideravelmente os **ERM** resultantes das estimativas para o **ano base** (de 80 % para 15 %), como pode ser verificado nos novos resultados apresentados na **Tabela 5.11**.

A **Tabela 5.11** apresenta o **resumo das características dos modelos** especificados e calibrados, utilizando-se os dados de 90 das 96 ZTs da base de dados de 1974. Na análise dos resultados das estimativas para o **ano base** foram verificados, além dos valores de **ERM**, também os valores dos índices *I global de Moran* resultantes da

análise da distribuição espacial dos resíduos (**I\_Mor\_Res**).

**Tabela 5.11:** Resumo dos modelos tradicionais especificados na **Fase 1**, excluindo-se as ZTs que representavam *outliers*

	<b>1T</b>	<b>2T</b>	<b>3T</b>	<b>4T</b>
<b>Constante</b>	-2144,70	-2144,70	-2084,63	-2084,63
<b>POP_T</b>	1,02665	1,02665	1,01966	1,01966
<b>DENSI</b>	<b>-0,0120629</b>	<b>-0,0120629</b>		
<b>RENDA<sub>SM</sub></b>	356,504		345,54	
<b>RENDA<sub>D</sub></b>		5,9985		5,81402
<b>R<sup>2</sup></b>	0,884	0,884	0,883	0,883
<b>R<sup>2</sup><sub>aj</sub></b>	0,880	0,880	0,881	0,881
<b>ERM (ano base)</b>	<b>15 %</b>	<b>15 %</b>	<b>15 %</b>	<b>15 %</b>
<b>I_Mor_Res</b>	0,39	0,39	0,39	0,39

Comparando-se os resultados apresentados na **Tabela 5.11** com os resultados da análise anterior (**Tabela 5.10**), a diferença mais relevante é a redução dos **ERM** das estimativas (de 80 para 15 %). Percebe-se, também, uma redução dos valores de **R<sup>2</sup>**, o que já era esperado, uma vez que se reduziu o número de observações (de 96 ZTs para 90 ZTs). Verificam-se, ainda, pequenas alterações nos valores dos coeficientes.

Na análise comparativa entre os diferentes modelos, através da análise dos valores da **Tabela 5.11**, não se verificam grandes alterações nas características e resultados obtidos. Apenas é verificada uma pequena redução do valor do **R<sup>2</sup>** nos modelos sem a variável **DENSI** (modelos **3T** e **4T**), o que é justificado pelo número menor de variáveis incluídas, no entanto os valores de **R<sup>2</sup><sub>aj</sub>**, cuja análise é mais indicada quando comparam-se modelos com diferentes números de variáveis, apresentaram uma pequena melhora nos modelos sem a variável **DENSI** (de 0,880 para 0,881).

Quanto aos resultados das estimativas, os valores verificados são os mesmos para todos os modelos. Porém, os resíduos apresentam índices de autocorrelação espacial (**I\_Mor\_Res**) razoavelmente altos (**0,39**).

Conclui-se, assim, que o ponto crítico destes modelos tradicionais analisados na **Fase 1**, que devem ser melhorados, são os altos índices de autocorrelação espacial dos resíduos (**I\_Mor\_Res**). Outra conclusão relevante é que não se justifica a inclusão da variável **DENSI**, ou, mais ainda, que sua inclusão não é indicada, uma vez que,

analisando-se os coeficientes, percebe-se que o sinal negativo é incoerente com a correlação positiva que esta variável apresenta com a variável dependente. Os modelos tradicionais que se destacam (“válidos”) são o **3T** e o **4T**.

### 5.6.2 Modelos tradicionais da Fase 2

Em um segundo momento, levando-se em consideração as conclusões do estudo preliminar e dando atenção aos critérios 3 e 4 do início do item 5.6, foram consideradas as sete variáveis e as análises apresentadas no **item 5.4.4**. O intuito foi testar novos modelos tradicionais, que produzissem melhores resultados que os obtidos com os modelos analisados na **Fase 1**.

Foram testados 21 modelos através de diferentes combinações das variáveis *AREA*, *DOMIC*, *POP\_T*, *POP\_A*, *FROTA*, *DENSI* e *REND*, para a modelagem das viagens produzidas de base domiciliar (*VPBD*), usando-se os dados de 90 das 96 ZTs da pesquisa O-D de 1974, através do *software* TransCAD. O diagnóstico dos modelos incluiu, primeiramente, além da análise dos valores de  $R^2$ , também os valores do  $R^2_{aj}$  ( $R^2$  ajustado), uma vez que envolvia a comparação de modelos com diferentes números de variáveis. Sendo assim, entre os 21 modelos analisados destacaram-se 8, que apresentaram valores de  $R^2$  acima de 0,90 ou que o valor do  $R^2_{aj}$  era maior que o observado nos modelos da **Fase 1**.

O segundo passo foi analisar a significância das variáveis incluídas, o que foi realizado através da verificação dos valores da *estatística T* (*T stat*) para cada variável, considerando-se aceitáveis valores acima de |2|. Estes dados são fornecidos pelo *software* nos relatórios gerados para cada modelo calibrado.

A **Tabela 5.12** apresenta os 8 modelos pré-selecionados com os valores de *T stat*, para cada variável incluída, destacando-se os menores que de |2|. Destacam-se, ainda, os modelos que não apresentaram nenhuma variável considerada sem significância, através desta análise, assim como as variáveis que apresentaram valores significativos para todos os modelos em que foram incluídas. Os modelos estão numerados conforme a ordem em que foram sendo especificados e calibrados, sem nenhum critério pré-estabelecido.

**Tabela 5.12:** Resumo dos 8 modelos tradicionais inicialmente selecionados na **Fase 2** e análise da significância das variáveis através dos valores de *T stat*

Modelos selecionados entre os 21 analisados								
Variáveis	9	10	13	14	15	16	18	21
<i>AREA</i>						-1,44	-1,62	-1,44
<i>DOMIC</i>		-3,12	-2,90	-2,93	-2,52	-2,72	-3,16	-2,46
<i>POP_T</i>	14,69	8,34	7,17	7,06		6,78	7,79	4,65
<i>POP_A</i>					5,16			0,45
<i>DENSI</i>			0,71	0,84		0,45		0,40
<i>RENDA</i>				-0,53	-1,18	-0,60	-0,49	-0,57
<i>FROTA</i>	9,68	10,06	9,84	7,22	7,45	7,22	7,26	6,88
$R^2$	0,916	0,925	0,925	0,925	0,902	0,927	0,927	0,9272
$R^2_{aj}$	0,914	0,922	0,921	0,920	0,897	0,922	0,923	0,9210

Através da análise dos dados apresentados na **Tabela 5.12** foram, então, selecionados os modelos **9** e **10** desta série de modelos analisados. Estes modelos foram os que apresentaram os valores de *T stat* considerados significativos para todas as variáveis incluídas. Destacam-se, ainda, as variáveis *POP\_T*, *FROTA* e *DOMIC* como sendo as mais significativas para explicar o fenômeno em estudo, uma vez que estas apresentaram sempre valores acima de  $|2|$  para a estatística T.

Os 2 modelos selecionados serão aqui chamados de **5T** e **6T**, dando seqüência à numeração de modelos tradicionais da **Fase 1**. Após essas análises, o diagnóstico também incluiu análises dos coeficientes das variáveis, que podem ser observados na **Tabela 5.13**

**Tabela 5.13:** Resumo das características dos modelos tradicionais 5T e 6T

	5T		6T	
	Coefficientes	<i>T stat</i>	Coefficientes	<i>T stat</i>
<b>Constante</b>	118,398		39,772	
<i>POP_T</i>	0,707	14,690	1,088	8,335
<i>DOMIC</i>			-1,917	-3,919
<i>FROTA</i>	3,315	9,675	4,047	10,061
$R^2$		0,916		0,924
$R^2_{aj}$		0,914		0,922

Na análise comparativa entre os dois modelos selecionados (modelos **5T** e **6T**) um aspecto que se verifica no resumo apresentado na **Tabela 5.13** é que, apesar de todas as variáveis aparentemente apresentarem significância, verificando-se que todos os valores da *estatística T (T stat)* estão acima de |2|, a variável **DOMIC**, no modelo **6T**, apresenta coeficiente negativo, indicando, assim, a possível presença de colinearidade entre as variáveis do modelo, tendo em vista a incoerência deste coeficiente. Seleciona-se o modelo **5T** como o melhor modelo tradicional, entre os analisados na **Fase 2**, apesar do modelo **6T** apresentar resultados ligeiramente melhores que o modelo **5T** para os valores de  $R^2$  e  $R^2_{Aj}$ .

Da mesma forma que nas análises da **Fase 1**, os resultados das estimativas para o **ano base**, utilizando-se o modelo **5T**, foram avaliados através dos valores de **ERM** e **I\_Mor\_Res**, sendo os resultados obtidos apresentados na **Tabela 5.14**.

**Tabela 5.14:** Resumo das características dos modelos tradicionais selecionados na **Fase 2** e resultados obtidos

		5T
<b>Resultados</b>	<b>Modelo</b>	$R^2$ 0,916
		$R^2_{aj}$ 0,914
	<b>Estimativas (ano base)</b>	<b>ERM</b> 12 %
		<b>I Mor Res</b> 0,39

Comparando-se os modelos das duas fases, são observados, além do valor mais alto para o  $R^2$  (em torno de 3 %), um melhor resultado para o **ERM** da estimativa (também 3 %), mantendo-se o mesmo resultado para a distribuição espacial dos resíduos.

A grande amplitude dos valores das variáveis incluídas pode ser a causa das estimativas para o **ano meta** apresentarem maiores erros do que as estimativas para o **ano base**. No caso em estudo, por exemplo, a variável **POP\_T** apresenta uma variação de 500 habitantes (para a **ZT** de menor população) a valores acima de 23.000 (para as **ZTs** mais populosas), enquanto a variável **FROTA** varia de valores em torno de 40 veículos a valores em torno de 3500 veículos.

Os coeficientes ajustados podem não conseguir representar de forma adequada

as variações que ocorrem nestas variáveis de um período para outro. Uma forma de tentar corrigir este problema é através da padronização das variáveis. Este procedimento coloca todos os valores das variáveis em um mesmo patamar de grandeza, uma vez que são levados em consideração no cálculo, a **média** e o **desvio padrão** das distribuições, transformando-as em distribuições reduzidas, ou padronizadas, de média zero e desvio padrão igual a um. As variáveis padronizadas são calculadas através da seguinte equação:

$$X_p = \frac{X_i - \bar{X}}{s} \quad (15)$$

Onde:

- $X_p$  – variável padronizada;
- $X_i$  – corresponde ao valor da variável na ZT<sub>i</sub>;
- $\bar{X}$  – média da variável considerando todas as ZTs incluídas na análise;
- $s$  – desvio padrão da variável.

Com base nestas observações, foi calibrado um novo modelo, similar ao modelo **5T**, porém com as variáveis padronizadas. As características e resultados obtidos com o modelo **5T<sub>p</sub>** são apresentados na **Tabela 5.15**.

**Tabela 5.15:** Resumo das características do modelo **5T<sub>p</sub>** e resultados obtidos para o ano base

		<b>5T<sub>p</sub></b>		
		<b>Coefficientes</b>	<b>T stat</b>	
	<b>Constante</b>	12195,4		
<b>Variáveis independentes</b>	<b>POP_T<sub>p</sub></b>	3911,26	14,69	
	<b>FROTA<sub>p</sub></b>	2576,07	9,68	
<b>Resultados</b>	<b>Modelo</b>	<b>R<sup>2</sup></b>	0,916	
		<b>R<sup>2</sup><sub>aj</sub></b>	0,914	
	<b>Estimativas (ano base)</b>	<b>EMR</b>	12 %	
		<b>I Mor Res</b>	<b>0,39</b>	

O resultado referente à estimativa para o **ano base** permanece igual ao apresentado pelo modelo **5T** (sem as variáveis padronizadas), assim como os valores de **T stat** para cada variável e o **R<sup>2</sup>** e **R<sup>2</sup><sub>aj</sub>** para o modelo. Os coeficientes, no entanto, são bem mais elevados que os dos modelos anteriormente analisados, o que é justificado pela redução nos valores das observações, uma vez que foram padronizadas. Como o

efeito da padronização tende a ser mais significativo quando da análise das estimativas futuras (**item 5.8**), admite-se que o modelo  $5T_p$  é superior ao modelo  $5T$ , apesar de ainda apresentar um índice muito alto de autocorrelação espacial dos resíduos das estimativas para o **ano base**.

### 5.6.3 Modelos tradicionais da Fase 3

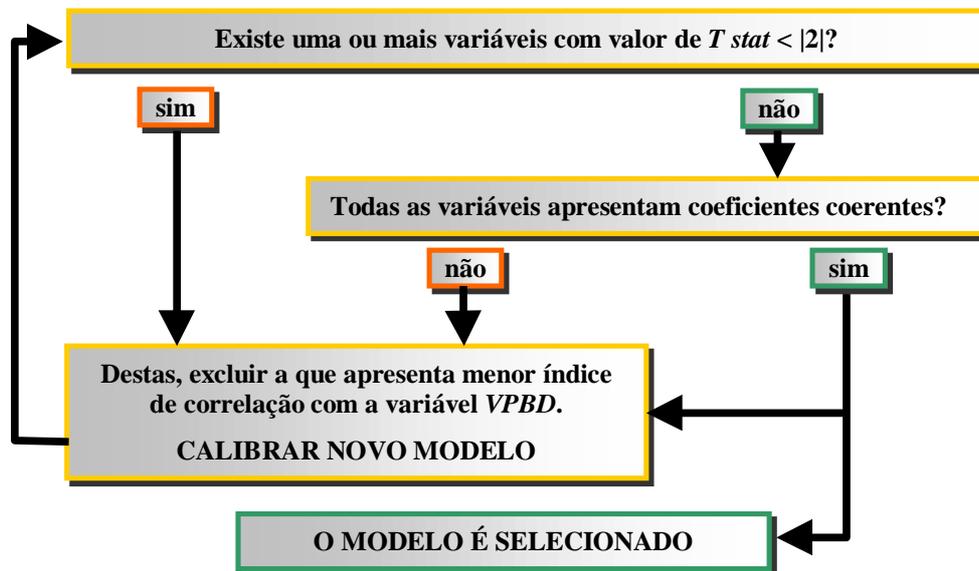
Para a especificação dos modelos tradicionais da Fase 3, introduzem-se as variáveis derivadas (**taxas**) definidas na **Etapa 4**, cujas análises estão apresentadas no **item 5.4.5**. Esta fase atende ao quinto e último critério dos estabelecidos no início da **Etapa 6**, que definiram os fatores a serem considerados na escolha das variáveis tradicionais a serem incluídas nos modelos. Desta forma, são testadas em novos modelos as variáveis  $X21$ ,  $X32$ ,  $X51$ ,  $X52$  e  $X53$ , que são as taxas derivadas das variáveis  $AREA$ ,  $DOMIC$ ,  $POP_T$  e  $FROTA$ , já definidas anteriormente (**item 5.4.5**).

Dentre as variáveis de contagens agregadas, são incluídas apenas as variáveis  $POP_T$  e  $FROTA$ , que foram identificadas como as mais significativas, entre as tradicionais disponíveis, nas análises das fases anteriores. Incluem-se, também, as variáveis  $DENSI$  e  $REND$ , uma vez que estas também representam taxas derivadas. Primeiramente, especifica-se e calibra-se, através do *software* TransCAD, um modelo que inclui todas as nove variáveis citadas acima, destacando-se que se tratam de variáveis padronizadas.

Analisa-se os resultados das características do modelo, que são apresentados nos relatórios emitidos pelo *software*. São verificados os valores de  $T_{stat}$  apresentados para as variáveis, sendo consideradas significativas as que apresentam valores acima de  $|2|$ . É verificada, também, a coerência dos coeficientes das variáveis. Para exclusão ou manutenção das variáveis no modelo, são levados em consideração os índices de correlação  $R$  das variáveis explicativas com a variável dependente, conforme a *matriz de correlação* apresentada na **Tabela 5.3 (item 5.4.5)**.

Segue-se uma rotina pré-estabelecida, apresentada na **Figura 5.13**, para exclusão de variáveis e seleção dos melhores modelos. Mais de um modelo pode ser selecionado e, então, são aplicados em estimativas para o **ano base** e para o **ano meta**. A

**Tabela 5.16** apresenta os modelos analisados, assim como os valores do *T stat* e dos *coeficientes* para cada variável incluída. A coluna à esquerda da variável apresenta a ordem em relação aos índices de correlação com a variável dependente (1 representa o maior índice).



**Figura 5.13:** Rotina pré-estabelecida para escolha das variáveis na especificação, calibração e seleção dos modelos tradicionais da **Fase 3**

De acordo com os resultados das características dos modelos analisados, apresentados na **Tabela 5.16**, conclui-se que nenhuma variável derivada (taxa), entre as analisadas, é significativa para explicar as *viagens produzidas de base domiciliar*. O modelo 6, que apresentou todos os valores de *T stat* acima de |2|, apresentou, no entanto, coeficiente negativo para a variável *X21*, o que representa uma incoerência, uma vez que a relação desta variável com a variável dependente é direta (positiva). Tal resultado indica a presença de colinearidade, o que é confirmado com a análise da *matriz de correlação* apresentada na **Tabela 5.3**, onde se verifica índice de correlação entre a variável *DENSI* e *X21* de 0,98.

No entanto, excluindo-se a variável *X21*, percebe-se que a variável *DENSI* perde a significância, que aparentemente apresentava (vide a variação do *T stat* dos outros modelos para o modelo 7). No modelo 7, verifica-se que a variável *DENSI*, além de apresentar o valor de *T stat* inferior a |2| (-1,28), também apresenta coeficiente

negativo, o que, pelo mesmo motivo que para a variável  $X2I$ , é incoerente.

**Tabela 5.16:** Resumo dos modelos analisados através da rotina pré-estabelecida

		<b>Modelos analisados conforme rotina pré-estabelecida (Fase 3)</b>						
		<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>
		Coeficientes e valores de <i>T stat</i> para cada variável incluída						
<i>Constante</i>		12195,4	12195,4	12195,4	12195,4	12195,4	12195,4	12195,4
1	<i>POP_T</i>							
	<i>Coef.</i>	3665,07	3552,89	3596,43	3652,93	3622,30	3580,55	3913,48
	<i>T stat</i>	7,84	7,71	7,95	7,98	10,36	14,64	14,75
2	<i>FROTA</i>							
	<i>Coef.</i>	2740,27	2885,34	2806,13	2700,10	2746,90	2799,60	2770,65
	<i>T stat</i>	4,33	4,62	4,62	4,41	6,61	10,32	9,06
3	<i>DENSI</i>							
	<i>Coef.</i>	5815,04	6091,30	5984,76	4629,66	4616,01	4678,11	-315,53
	<i>T stat</i>	4,29	4,54	4,53	4,14	4,19	4,53	-1,28
8	<i>RENDA</i>							
	<i>Coef.</i>	345,99	236,47					
	<i>T stat</i>	0,83	0,58					
5	<i>X2I</i>							
	<i>Coef.</i>	-6851,33	-6658,71	-6561,78	-4988,71	-4991,42	-4950,49	
	<i>T stat</i>	-5,10	-4,93	-4,92	-4,79	-4,82	-4,95	
9	<i>X32</i>							
	<i>Coef.</i>	-786,31						
	<i>T stat</i>	-1,245,17						
4	<i>X51</i>							
	<i>Coef.</i>	348,71	-227,726	-140,44	103,00	126,45		
	<i>T stat</i>	0,38	-0,28	-0,18	0,13	0,17		
7	<i>X52</i>							
	<i>Coef.</i>	8,11	-2050,46	-1941,46				
	<i>T stat</i>	0,004	-1,91	-1,85				
6	<i>X53</i>							
	<i>Coef.</i>	-621,80	2016,47	2120,10	46,63			
	<i>T stat</i>	-0,25	1,65	1,76	0,10			
$R^2$		0,940	0,939	0,939	0,936	0,902	0,927	0,927
$R^2_{aj}$		0,933	0,933	0,933	0,931	0,897	0,922	0,923

Conclui-se, com esta análise, que o melhor modelo tradicional dentre os analisados na **Fase 3** é igual ao modelo  $5T_p$ , selecionado como melhor nas análises da **Fase 2**, pois com a exclusão da variável *DENSI* do modelo **7** (**Tabela 5.16**) restam apenas as variáveis *POP\_T* e *FROTA* como significativas. As características e resultados das estimativas para o **ano base** com este modelo já foram apresentados e analisados na **Fase 2** (**item 5.6.2 e Tabela 5.15**). Uma vez selecionados os melhores *modelos tradicionais* de cada fase, a próxima etapa é a especificação, calibração e análise dos *modelos alternativos* correspondentes, para posterior análise comparativa dos resultados obtidos com os diferentes modelos de estimativas para o **ano meta**.

## 5.7. ESPECIFICAÇÃO, CALIBRAÇÃO E DIAGNÓSTICO DOS MODELOS QUE CONSIDERAM A DEPENDÊNCIA ESPACIAL

Seguindo o método descrito no **Capítulo 4, item 4.7**, a especificação dos *modelos alternativos* é baseada nos resultados das análises espaciais apresentadas no **item 5.5** e nos modelos tradicionais selecionados na **Etapa 6**, apresentados no **item 5.6**. Foram testados *modelos alternativos*, estabelecidos com base nos três métodos considerados nesta pesquisa:

- através da introdução dos indicadores globais de autocorrelação espacial (modelos globais ou “**AG**”);
- através de indicadores locais de autocorrelação espacial (modelo **LISA** ou “**AL**”);
- através da introdução de indicadores globais e locais (modelos mistos ou “**AGL**”).

Esta etapa também foi desenvolvida em 3 fases, sendo que cada tem sua correspondente nas etapas de análise da *autocorrelação espacial* e de especificação e calibração de *modelos tradicionais*.

### 5.7.1 Modelos alternativos da Fase 1

A primeira abordagem aplicada nesta pesquisa para considerar a dependência espacial nos modelos de demanda por transportes consistiu de um estudo exploratório inicial, cujo método é apresentado por Lopes e Silva (2004), através, simplesmente, da inclusão de variáveis *dummy* associadas aos quadrantes do *diagrama de Moran* (indicador Global). A definição de variáveis espaciais foi restrita apenas às variáveis analisadas na **Fase 1** da **Etapa 5** (**DENSI** e **RENDA**).

Foram geradas, desta forma, três variáveis para cada variável socioeconômica analisada, representando os quadrantes **Q1**, **Q2** e **Q3** (**item 4.7 – Variáveis Espaciais Globais** e o **Anexo D**). Nesta fase, para a especificação dos *modelos alternativos AG*, além das variáveis tradicionais, para cada um dos 2 *modelos tradicionais* selecionados foram incluídas, de uma só vez, as seis variáveis espaciais globais, gerando assim os 4 modelos alternativos correspondentes (**1AG**, **2AG**, **3AG** e **4AG**). As variáveis incluídas

são:

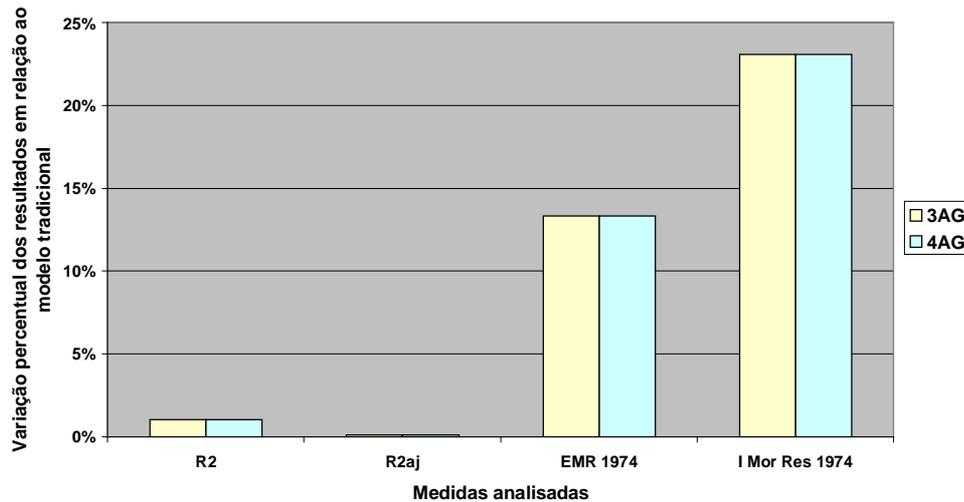
- $D\_Q1$ ,  $D\_Q2$  e  $D\_Q3$  – correspondentes aos quadrantes da variável **DENSI**;
- $R\_Q1$ ,  $R\_Q2$  e  $R\_Q3$  – correspondentes aos quadrantes da variável **RENDA**.

Desta forma, os *modelos alternativos* **AG** da **Fase 1** foram calibrados, através da ferramenta de planejamento de transportes do SIG-T TransCAD, e foram aplicados para estimativas para o **ano base** e para o **ano meta (item 5.8)**, sendo analisados os **ERM** e o **I\_Mor\_res** para cada estimativa, da mesma forma que para os *modelos tradicionais*.

A **Tabela 5.17** apresenta um resumo das características e resultados obtidos com cada um destes modelos. Observam-se reduções no **ERM** (de 15 para 13 %) e no **I\_Mor\_Res** (de 0,39 para 0,30), que indicam um efeito positivo da introdução das variáveis espaciais, com melhoras de 13 % e 23 %, respectivamente, para as medidas analisadas, se tomarmos como referência os resultados dos modelos tradicionais **3T** e **4T**, conforme pode ser observado no gráfico da **Figura 5.14**. Pouca alteração, no entanto, foi verificada para os valores de  $R^2$  e  $R^2_{aj}$ .

**Tabela 5.17:** Resumo dos *modelos alternativos* especificados na **Fase 1**, excluindo-se as ZTs que representavam *outliers*

	<b>3AG</b>	<b>4AG</b>
<b>Constante</b>	853,957	853,957
<b>POP_T</b>	0,986183	0,986183
<b>DENSI</b>		
<b>RENDA<sub>SM</sub></b>	153,850	
<b>RENDA<sub>D</sub></b>		2,58867
<b>D_Q1</b>	646,345	646,345
<b>D_Q2</b>	-278,340	-278,340
<b>D_Q3</b>	237,382	237,382
<b>R_Q1</b>	-298,305	-298,305
<b>R_Q2</b>	-1845,47	-1845,47
<b>R_Q3</b>	-484,265	-484,265
<b>R<sup>2</sup></b>	0,892	0,892
<b>R<sup>2</sup><sub>aj</sub></b>	0,881	0,881
<b>ERM (ano base)</b>	<b>13 %</b>	<b>13 %</b>
<b>I_Mor_Res</b>	0,30	0,30



**Figura 5.14:** Análise comparativa dos resultados dos *modelos alternativos* da **Fase 1** tomando-se como referência os *modelos tradicionais* **3T** e **4T**

### 5.7.2 Modelos alternativos da Fase 2

O estudo exploratório inicial, que considerou apenas indicadores globais de dependência espacial (modelos **AG**), indicou a necessidade de pesquisas adicionais de análise e inclusão de outras medidas de dependência espacial nos modelos da demanda. Uma alternativa, que é analisada nesta fase, é a estatística *LISA* (*indicadores locais de autocorrelação espacial*). Portanto, a hipótese secundária desta fase do estudo é de que a introdução de *indicadores locais de dependência espacial* nos modelos (modelos **AL**) pode produzir melhores resultados do que os obtidos com a introdução de indicadores globais (modelos **AG**). Estuda-se, ainda, a combinação de indicadores (Globais e Locais), que geram os *modelos alternativos mistos*, ou **AGL**.

Outro aspecto levantado foi a necessidade de uma análise mais detalhada dos modelos, como, por exemplo, a significância das variáveis incluídas. Sendo assim, da mesma forma que foi realizada a análise para selecionar o melhor *modelo tradicional* da **Fase 2**, a mesma técnica, que foi realizada através de ferramentas do SIG-T TransCAD, é aplicada para selecionar os melhores *modelos alternativos* de cada tipo (**AG**, **AL** e **AGL**). Cabe ressaltar que o ponto de partida para a especificação dos modelos alternativos desta fase foi o modelo tradicional **5T<sub>p</sub>**, selecionado como o melhor da Fase 2.

### 5.7.2.1 Modelos alternativos AG da Fase 2

As *variáveis espaciais globais* introduzidas nos modelos **AG** são as mesmas analisadas na **Fase 1**, ou seja, as variáveis *dummy* que representam os quadrantes de cada ZT para as variáveis **RENDA** e **DENSI**. Desta forma, especificou-se, inicialmente, um modelo que inclui as variáveis **POP<sub>T<sub>p</sub></sub>**, **FROTA<sub>p</sub>** e as seis variáveis espaciais globais, denominado **5AG<sub>p1</sub>**.

Foram analisadas as características e os resultados apresentados no relatório de regressão emitido pelo *software*, observando-se os mesmos critérios estabelecidos para a análise dos modelos tradicionais desta fase, para verificar quais as variáveis não eram significativas ou apresentavam coeficientes incoerentes com a correlação de cada uma com a variável dependente (*matriz de correlação*, **Anexo E**). Procurou-se obter um *modelo alternativo* válido, do tipo **AG**, para então ser aplicado para as estimativas (**ano base**). Um resumo das características do modelo **5AG<sub>p1</sub>** é apresentado na **Tabela 5.18**

**Tabela 5.18:** Resumo das características do *modelo alternativo 5AG<sub>p1</sub>*

<b>5AG<sub>p1</sub></b>		
	<b>Coefficientes</b>	<b>T stat</b>
<b>Constante</b>	14071,9	
<b>POP<sub>T<sub>p</sub></sub></b>	3871,52	11,64
<b>FROTA<sub>p</sub></b>	2559,87	6,04
<b>D<sub>Q1</sub></b>	-492,374	-0,48
<b>D<sub>Q2</sub></b>	-1240,52	-1,13
<b>D<sub>Q3</sub></b>	-1587,14	-1,28
<b>R<sub>Q1</sub></b>	-1413,31	-1,79
<b>R<sub>Q2</sub></b>	-849,550	-1,04
<b>R<sub>Q3</sub></b>	-725,636	-0,75
<b>R<sup>2</sup></b>	0,923	
<b>R<sup>2</sup><sub>aj</sub></b>	0,916	

Através da análise das características do **Modelo 5AG<sub>p1</sub>** percebe-se que nenhuma variável espacial é considerada significativa, uma vez que todas apresentaram valores de **T stat** menores do que |2|. Salienta-se, ainda, que testes, acrescentando-se as variáveis espaciais às tradicionais, uma a uma, foram realizados e nenhuma apresentou valor significativo para o **T stat**. Conclui-se, assim, que o modelo **5AG<sub>p1</sub>** não é um modelo válido para análise e, como não podem ser especificados *modelos alternativos* válidos a partir deste, **nenhum modelo alternativo global** é selecionado na **Fase 2**.

### 5.7.2.2 Modelos alternativos AL da Fase 2

As variáveis espaciais locais foram definidas a partir dos *índices locais de associação espacial LISA (IMORAN)*, obtidos através da análise espacial da **Fase 2**, apresentada no **item 5.4**, e conforme procedimento descrito no **Capítulo 4 (item 4.7 – Variáveis Espaciais Locais e Anexo D)**. Os modelos **AL** foram, da mesma forma descrita para os modelos **AG**, especificados a partir do modelo **5T<sub>p</sub>**.

Sendo assim, para calibração do primeiro *modelo alternativo LISA (5AL<sub>p1</sub>)* foram introduzidas, além das *variáveis tradicionais* do modelo tradicional correspondente (**5T<sub>p</sub>**), mais duas *variáveis espaciais locais*, que são:

- **LISA<sub>D</sub>** – correspondendo ao *I Local de Moran* para a variável **DENSI**;
- **LISA<sub>R</sub>** – correspondendo ao *I Local de Moran* para a variável **RENDA**.

Da mesma forma que para os modelos anteriores desta fase, a partir do modelo inicial foi analisado o relatório de regressão, excluídas as variáveis não significativas e, então, selecionado um *modelo alternativo* válido do tipo **AL**. As características do *modelo alternativo LISA* inicial (**5AL<sub>p1</sub>**) e do *modelo alternativo LISA* válido selecionado nesta fase (**5AL<sub>p</sub>**) estão resumidas na **Tabela 5.19**.

Utilizando-se os mesmos critérios aplicados anteriormente, a variável **LISA<sub>R</sub>** foi excluída por ter apresentado o valor de *T stat* menor que |2| (**Tabela 5.19**). Foi, então, especificado e calibrado um novo modelo, selecionado como o modelo do tipo **AL** válido para análise da **Fase 2** (modelo **5AL<sub>p</sub>**).

O modelo **5AL**, que inclui como variável espacial apenas a variável **LISA<sub>D</sub>**, apesar de ter apresentado pequena redução do valor de **R<sup>2</sup>** e **R<sup>2</sup><sub>aj</sub>** ( $\cong 0,2\%$ ), que não foi considerada significativa, foi selecionado por apresentar características que indicam que todas as variáveis são significativas. Este modelo foi então aplicado para estimativas para o **ano base** e os resultados são apresentados na **Tabela 5.20**.

**Tabela 5.19:** Resumo das características do *modelo alternativo LISA* inicial ( $5AL_{p1}$ ) e do *modelo alternativo LISA* válido selecionado na **Fase 2** ( $5AL_p$ )

	$5AL_{p1}$		$5AL_p$	
	<i>Coefficientes</i>	<i>T stat</i>	<i>Coefficientes</i>	<i>T stat</i>
<b>Constante</b>	12195,4		12195,4	
$POP\_T_p$	3554,66	14,60	3675,26	15,52
$FROTA_p$	2968,81	11,96	2823,00	11,90
$LISA\_D$	-822,508	-4,58	-918,409	-5,30
$LISA\_R$	-337,66	-1,77		
$R^2$	0,939		0,937	
$R^2_{aj}$	0,936		0,934	

**Tabela 5.20** – Resumo das características e resultados obtidos com o modelo  $5AL_p$ 

		$5AL_p$	
		<i>Coefficientes</i>	<i>T stat</i>
<i>Variáveis tradicionais</i>	<b>Constante</b>	12195,4	
	$POP\_T_p$	3675,26	15,52
	$FROTA_p$	2823,00	11,90
	<i>Variável espacial</i>	$LISA\_D$	-918,409
<b>Resultados</b>	<b>Modelo</b>	$R^2$	0,937
		$R^2_{aj}$	0,934
	<b>Estimativas (ano base)</b>	<b>EMR</b>	10 %
		<b>I Mor Res</b>	0,31

### 5.7.2.3 Modelos alternativos AGL da Fase 2

Para encerrar as análises dos modelos alternativos da **Fase 2**, testou-se, finalmente, o terceiro e último tipo de modelo com variáveis espaciais, que é o *modelo alternativo misto* ou do **tipo AGL**. Especificou-se um modelo inicial, tendo-se como base o modelo tradicional  $5T_p$ , e introduzindo-se as 8 variáveis espaciais testadas nesta fase: as 6 *variáveis espaciais globais* e as 2 *variáveis espaciais locais*. Da mesma forma que nas análises anteriores, o passo seguinte foi a análise das características apresentadas no relatório de regressão e a exclusão de possíveis variáveis não significativas até a seleção de um modelo do tipo **AGL** válido para análise. O resumo das características do modelo inicial  $5AGL_{p1}$  e do modelo selecionado  $5AGL_p$  são apresentadas na **Tabela 5.21**.

Percebe-se, através da análise dos valores de *T stat* (**Tabela 5.21**) apresentados para o modelo inicial especificado ( $5AGL_{p1}$ ), que, da mesma forma que para o modelo inicial do tipo AG ( $5AG_{p1}$ ) analisado anteriormente, nenhuma variável espacial global

se mostrou significativa. No entanto desta vez, ao contrário do modelo inicial do tipo AL ( $5AL_{p1}$ ), as duas variáveis LISA (inclusiva a variável  $LISA_R$ ) apresentaram valores aceitáveis para o  $t$  stat. Adotou-se, então, o mesmo critério que foi utilizado anteriormente na tentativa de encontrar um modelo válido do tipo AG, para selecionar um modelo válido do tipo AGL.

**Tabela 5.21:** Resumo das características do *modelo alternativo* misto inicial ( $5AGL_{p1}$ ) e do *modelo alternativo misto* válido selecionado na **Fase 2** ( $5AGL_p$ )

	$5AGL_{p1}$		$5AGL_p$	
	<i>Coefficientes</i>	<i>T stat</i>	<i>Coefficientes</i>	<i>T stat</i>
<b>Constante</b>	12081,4		11587,7	
$POP_{Tp}$	3932,23	14,73	3683,05	16,34
$FROTA_p$	2166,14	6,29	2479,25	9,67
$D_{Q1}$	1267,70	1,50	1709,20	4,14
$D_{Q2}$	21,3471	0,02		
$D_{Q3}$	-1013,17	-1,02		
$R_{Q1}$	430,965	0,61		
$R_{Q2}$	-650,770	-0,99		
$R_{Q3}$	-769,604	-0,91		
$LISA_D$	-1148,11	-6,28	-1079,25	-6,14
$LISA_R$	-439,705	-2,24	-356,590	-2,05
$R^2$	0,952		0,949	
$R^2_{aj}$	0,947		0,946	

Mantendo-se as 4 variáveis com valor de  $t$  stat aceitáveis (*tradicionais e espaciais locais*), testou-se vários modelos incluindo-se, uma a uma, as *variáveis espaciais globais*. Verificou-se que, diferentemente da primeira tentativa frustrada de encontrar uma *variável espacial global* com significância para o modelo, a variável  $D_{Q1}$ , combinada, desta vez, com as variáveis espaciais locais, apresentou valor de  $t$  stat igual a 4,14 e, conseqüentemente com significância para o modelo.

O modelo chamado então de  $5AGL_p$ , que apresenta 3 variáveis espaciais, além das duas variáveis tradicionais, é selecionado como o modelo do tipo AGL, válido para análise, da **Fase 2**. Aplica-se, então, o modelo para estimativas para o **ano base**. Os resultados destas estimativas, assim como as características do modelo  $5AGL_p$  estão resumidos na **Tabela 5.22**.

**Tabela 5.22:** Resumo das características e resultados obtidos com o modelo **5AGL<sub>p</sub>**

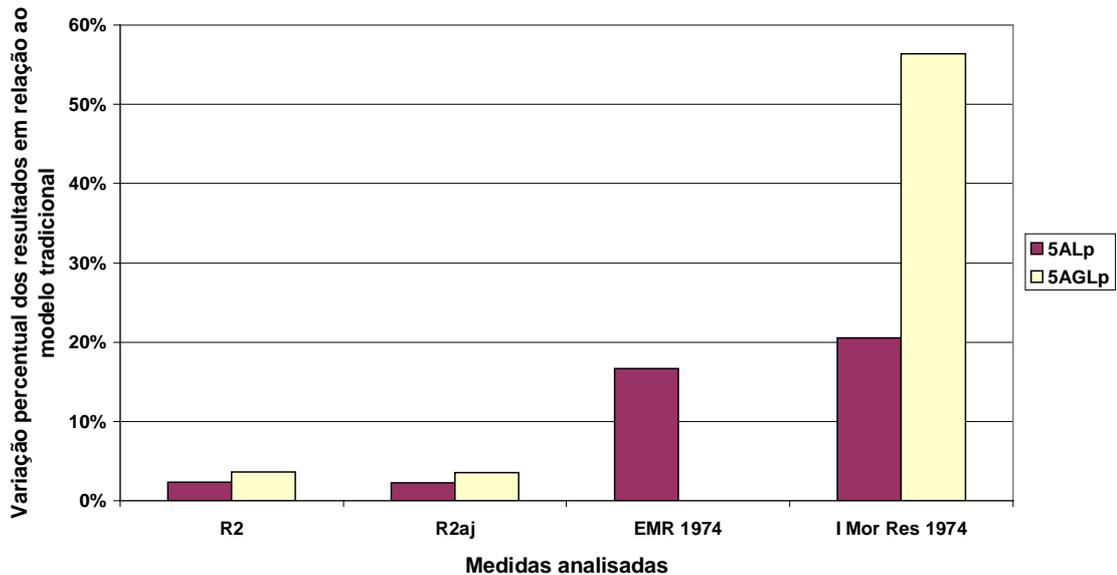
		<b>5AGL<sub>p</sub></b>	
		<b>Coefficientes</b>	<b>T stat</b>
		<b>Constante</b>	11587,7
<b>Variáveis tradicionais</b>	<b>POP_T<sub>p</sub></b>	3683,05	16,34
	<b>FROTA<sub>p</sub></b>	2479,25	9,67
	<b>D_QI</b>	1709,20	4,14
<b>Variáveis espaciais</b>	<b>LISA_D</b>	-1079,25	-6,14
	<b>LISA_R</b>	-356,590	-2,05
	<b>R<sup>2</sup></b>	0,949	
<b>Resultados</b>	<b>R<sup>2</sup><sub>aj</sub></b>	0,946	
	<b>EMR</b>	12 %	
	<b>I Mor Res</b>	0,17	
	<b>Estimativas (ano base)</b>		

Após a seleção dos *modelos alternativos* para cada tipo de abordagem analisada e do diagnóstico através das estimativas para o **ano base** com a aplicação de cada um deles, realizou-se, então, a análise comparativa dos resultados obtidos com os dois modelos selecionados (**5AL<sub>p</sub>** e **5AGL<sub>p</sub>**), verificando-se as variações em relação aos resultados obtidos com o *modelo tradicional* **5T<sub>p</sub>**.

A **Tabela 5.23** apresenta, lado a lado, os resultados obtidos para as medidas consideradas com o *modelo tradicional* e cada um dos *modelos alternativos* selecionados. Analisaram-se comparativamente os valores tanto do **R<sup>2</sup>** e **R<sup>2</sup><sub>aj</sub>** (calibração do modelo), como os valores de **ERM** e **I\_Mor\_Res** das estimativas para o **ano base** (diagnóstico dos modelo) tomando-se como base os resultados obtidos com o *modelo tradicional* (**5T<sub>p</sub>**) e verificando-se o percentual de variação, apresentado no gráfico da **Figura 5.15**.

**Tabela 5.23:** Resumo dos resultados obtidos com o modelo tradicional e com os modelos alternativos analisados na Fase 2

		<b>5T<sub>p</sub></b>	<b>5AL<sub>p</sub></b>	<b>5AGL<sub>p</sub></b>
<b>Modelo</b>	<b>R<sup>2</sup></b>	0,916	0,937	0,949
	<b>R<sup>2</sup><sub>aj</sub></b>	0,914	0,934	0,946
<b>Estimativas (ano base)</b>	<b>EMR</b>	12 %	10 %	12 %
	<b>I_Mor_Res</b>	0,39	0,31	0,17



**Figura 5.15:** Análise comparativa dos resultados dos *modelos alternativos* da **Fase 2** tomando-se como referência o *modelos tradicional 5T<sub>p</sub>*

A análise comparativa apresentada no gráfico da **Figura 5.15** permite a visualização imediata do desempenho dos modelos alternativos em relação ao modelo **5T<sub>p</sub>** tomado como o melhor *modelo tradicional*, considerando-se as análises apresentadas para a **Fase 1** e a **Fase 2**. Em outras palavras, permite verificar se a introdução das variáveis espaciais está contribuindo, atrapalhando ou não produzindo efeito no desempenho do modelo.

Verifica-se que, para a maioria das medidas analisadas, em termos gerais, houve melhora através da introdução de variáveis espaciais. Analisando-se cada *modelo alternativo* separadamente, com o modelo **5ALp** obteve-se melhora nas 4 medidas consideradas para avaliação dos modelos. Com o modelo **5AGLp** obteve-se melhora para 3 medidas e 1 medidas (**ERM**) manteve-se inalterada. Em nenhuma medida o *modelo tradicional* se mostrou melhor do que os *modelos alternativos*, salientando-se ainda que houve o modelo **5AGLp** apresentou melhora significativa para o **I\_Mor\_Res** da estimativa para o **ano base (57 %)**.

Conclui-se, com as análises da **Fase 2**, que a introdução de *variáveis espaciais* contribuiu para a melhora no desempenho dos modelos. Conclui-se ainda que a

combinação de *variáveis espaciais globais e locais* produziu melhores resultados do que quando introduzidas individualmente. Isto se baseia no fato que *variáveis espaciais globais* não apresentaram significância já na fase de calibração e que o modelo do tipo **AL** (**5AL<sub>p</sub>**), com apenas *variáveis espaciais locais* apresentou em geral, pior desempenho do que o *modelo alternativo misto* (**5AGL<sub>p</sub>**).

### 5.7.3 Modelos alternativos da Fase 3

A especificação de *modelos alternativos* da Fase 3 é baseada na análise espacial da **Fase 3**, apresentada no **item 5.5**, para onde são verificados os níveis de dependência espacial para todas as variáveis socioeconômicas (as contagens agregadas e as taxas derivadas definidas na **Fase 3** da **Etapa 4**). A análise comparativa da distribuição espacial das variáveis socioeconômicas com a variável dependente **VPBD**, revelou a existência de outras variáveis espaciais, que poderiam ser potenciais variáveis explicativas para a variável dependente **VPBD**.

Analisou-se, desta forma, a introdução de outras variáveis espaciais (*globais e locais*) referentes às variáveis socioeconômicas que apresentaram maior correlação com a *variável dependente*, conforme análises apresentadas no **item 5.5, subitem 5.5.3**. Selecionaram-se, então, as 12 *variáveis espaciais globais* e 4 variáveis espaciais locais derivadas da análise espacial das variáveis **DENSI**, **X21**, **X51** e **X53** para especificação dos modelos alternativos da **Fase 3** (**Anexo D**). O modelo tradicional que serve de base para a especificação de modelos da **Fase 3**, é o mesmo modelo **5T<sub>p</sub>** da **Fase 2**, uma vez que este foi o modelo resultante das análises para seleção de *modelos tradicionais*, também nesta fase.

#### 5.7.3.1 Modelos alternativos AG da Fase 3

As *variáveis espaciais globais* introduzidas nos modelos **AG** da **Fase 3** são as variáveis *dummy* que representam os quadrantes de cada ZT para as variáveis **DENSI**, **X21**, **X51** e **X53**. Especificou-se inicialmente, da mesma forma que na **Fase 2**, um modelo que inclui as variáveis **POP<sub>T<sub>p</sub></sub>**, **FROTA<sub>p</sub>** como *variáveis tradicionais* e as 12 *variáveis espaciais globais*, chamado desta vez de **F3AG<sub>pI</sub>**, para diferenciar dos modelos da **Fase 2**. Foram analisadas as características e resultados apresentados no relatório de regressão emitido pelo *software*, observando-se os mesmos critérios (valor

de *T stat* para as variáveis incluídas e coerência dos coeficientes, que se baseou na *matriz de correlação* (Anexo E). As características do modelo inicial  $F3AG_{pl}$  são apresentadas na **Tabela 5.24**. Para exclusão das variáveis segue-se a mesma rotina aplicada para os *modelos tradicionais* da **Fase 3** (Figura 5.13).

A análise do nível de correlação com a variável dependente, para a escolha da variável a ser excluída, é baseada nos resultados das análises espaciais da **Fase 3** (ver Figura 5.11). Para facilitar, anota-se a prioridade da variável (coluna à esquerda), que é referente ao nível de correlação espacial com *VPBD*, sendo **1** o valor para a maior correlação. Por exemplo, as variáveis  $X21\_Q2$  e  $X51\_Q2$  foram as que apresentaram o maior número de ZTs coincidentes com a variável *VPBD*.

**Tabela 5.24:** Resumo das características do *modelo alternativo* inicial  $F3AG_{pl}$

		$F3AG_{pl}$	
		<i>Coefficientes</i>	<i>T stat</i>
	<b>Constante</b>	14741,200	
	$POP\_T_p$	4007,180	12,87
	$FROTA_p$	2235,880	5,69
<b>4</b>	<b>D_Q1</b>	5840,480	1,75
<b>2</b>	<b>D_Q2</b>	1445,820	0,67
<b>7</b>	<b>D_Q3</b>	1088,570	0,45
<b>5</b>	<b>X21_Q1</b>	-5667,540	-1,67
<b>1</b>	<b>X21_Q2</b>	-1579,300	-0,74
<b>8</b>	<b>X21_Q3</b>	-1394,780	-0,53
<b>4</b>	<b>X51_Q1</b>	-1453,080	-1,25
<b>1</b>	<b>X51_Q2</b>	-2198,870	-2,01
<b>9</b>	<b>X51_Q3</b>	-1382,230	-1,02
<b>4</b>	<b>X53_Q1</b>	-1241,780	-1,67
<b>3</b>	<b>X53_Q2</b>	-621,416	-0,88
<b>6</b>	<b>X53_Q3</b>	-514,379	-0,58
	$R^2$	0,932	
	$R^2_{aj}$	0,920	

Desta forma, analisando-se as características do *modelo alternativo* inicial  $F3AG_{pl}$  (Tabela 5.24) percebe-se que a única variável espacial que apresenta *T stat* maior que |2|, é a variável  $X51\_Q2$ . Entre as restantes, as primeira a serem excluídas foram as variáveis  $X21\_Q3$  e  $X51\_Q3$  uma vez que apresentaram prioridade 8 e 9 respectivamente, ou seja, menor correlação com a variável *VPBD*. Desta forma, excluindo-se desta vez, duas a duas, tendo em vista o grande número de variáveis,

sucessivamente calibrou-se novos modelos até chegar ao modelo do **Tipo AG** definitivo, que então foi chamado de modelo **F3AG<sub>p</sub>**. O modelo selecionado foi aplicado para estimativas para o **ano base** e suas características e resultados obtidos estão apresentados na **Tabela 5.25**.

O modelo **F3AG<sub>p</sub>** apresenta apenas uma variável espacial (**X51\_Q2**), que apresentou **T stat** válido já no modelo inicial e prioridade 1 na análise de correlação com a variável dependente. Todas as outras 11 *variáveis espaciais* não se mostraram significativas na fase de calibração.

**Tabela 5.25:** Resumo das características e resultados obtidos com o modelo **F3AG<sub>p</sub>**

		<b>F3AG<sub>p</sub></b>	
		<b>Coefficientes</b>	<b>T stat</b>
	<b>Constante</b>	12799,50	
<b>Variáveis tradicionais</b>	<b>POP_T<sub>p</sub></b>	4144,85	14,66
	<b>FROTA<sub>p</sub></b>	2122,27	6,32
<b>Variável espacial</b>	<b>X51_Q2</b>	-1066,17	-2,15
<b>Modelo</b>	<b>R<sup>2</sup></b>	0,920	
	<b>R<sup>2</sup><sub>aj</sub></b>	0,917	
<b>Resultados</b>	<b>Estimativas (ano base)</b>	<b>EMR</b>	11 %
		<b>I Mor Res</b>	0,34
<b>Estimativas (ano meta)</b>		<b>EMR</b>	14 %
		<b>I Mor Res</b>	0,12

#### 5.7.3.2 Modelos alternativos AL da Fase 3

As 4 *variáveis espaciais locais* foram definidas a partir dos *índices locais de associação espacial LISA (IMORAN)*, de cada ZT obtidos através da análise espacial da **Fase 3**, para as variáveis **DENSI**, **X21**, **X51** e **X53**. A partir do modelo **5T<sub>p</sub>**, especificou-se o *modelo alternativo inicial LISA*, que contém 6 variáveis (2 *tradicionais* e 4 *espaciais*) denominado de **F3AL<sub>p1</sub>** e foram analisados os resultados da calibração que estão apresentados na **Tabela 5.26**.

Conforme se verifica nos resultados apresentados na **Tabela 5.26**, as *variáveis espaciais LISA\_X51* e *LISA\_X53* apresentaram valores não satisfatórios para o **T stat**. A coluna à esquerda das variáveis apresenta a ordem de prioridade que foi adotada para cada *variável espacial local*, tomando-se como base os resultados obtidos de correlação com a variável dependente (**VPBD**) através da análise espacial (resultados da análise

comparativa dos *Moran Map* (**Figura 5.12**)). A primeira variável excluída foi então a *variável espacial local LISA\_X53*, destacada na **Tabela 5.26**. Foram 3 modelos, do tipo AL, calibrados e analisados até chegar ao modelo definitivo **F3AL<sub>p</sub>**, apresentado na **Tabela 5.27**.

**Tabela 5.26:** Resumo das características do *modelo alternativo inicial F3AL<sub>p1</sub>*

		<b>F3AL<sub>p1</sub></b>	
		<b>Coefficientes</b>	<b>T stat</b>
	<b>Constante</b>	12195,3	
	<b>POP_T<sub>p</sub></b>	3880,03	16,81
	<b>FROTA<sub>p</sub></b>	2851,07	12,13
<b>1</b>	<b>LISA_D</b>	1860,40	2,30
<b>2</b>	<b>LISA_X21</b>	-2669,41	-3,86
<b>1</b>	<b>LISA_X51</b>	-148,90	-0,39
<b>3</b>	<b>LISA_X53</b>	-114,45	-0,66
	<b>R<sup>2</sup></b>		0,946
	<b>R<sup>2</sup><sub>aj</sub></b>		0,943

**Tabela 5.27:** Resumo das características e resultados obtidos com o modelo **F3AL<sub>p</sub>**

		<b>F3AL<sub>p</sub></b>	
		<b>Coefficientes</b>	<b>T stat</b>
	<b>Constante</b>	12195,40	
<b>Variáveis tradicionais</b>	<b>POP_T<sub>p</sub></b>	3738,59	16,73
	<b>FROTA<sub>p</sub></b>	2841,92	12,59
	<b>Variável espacial LISA_X21</b>	-1033,95	-6,28
<b>Modelo</b>	<b>R<sup>2</sup></b>		0,942
	<b>R<sup>2</sup><sub>aj</sub></b>		0,940
<b>Resultados</b>	<b>Estimativas (ano base)</b>	<b>EMR</b>	10 %
		<b>I Mor Res</b>	0,26
	<b>Estimativas (ano meta)</b>	<b>EMR</b>	12 %
		<b>I Mor Res</b>	0,07

Verifica-se, que apesar das variáveis *LISA\_D* e *LISA\_X51* apresentarem índice 1 de prioridade (**Tabela 5.26**), através do critério adotado, a variável que permaneceu no modelo foi a variável *LISA\_X21* (**Tabela 5.27**), que era a segunda mais correlacionada com a variável dependente. A exclusão da variável *LISA\_X51* foi necessária por ter sido a única a apresentar valor do *T stat* não satisfatório no segundo modelo calibrado. Finalmente, a variável *LISA\_D* apresentou sinal positivo para o coeficiente, em todos os modelos calibrados, indicando incoerência e possivelmente

presença de colinearidade. Cabe salientar que todas as *variáveis espaciais locais* apresentam correlação negativa com a variável dependente (conforme *matriz de correlação* (**Anexo E**). O modelo do tipo **AL** (**F3AL<sub>p</sub>**), selecionado na **Fase 3**, é então aplicado para as estimativas para o **ano base** e os resultados também podem ser observados na **Tabela 5.27**.

### 5.7.3.3 Modelos alternativos AGL da Fase 3

Finalmente, da mesma forma que na **Fase 2** analisa-se a introdução dos dois tipos de variáveis espaciais em um mesmo modelo (*modelo alternativo misto* ou **AGL**). O modelo inicial (**5AGL<sub>p1</sub>**), que é apresentado na **Tabela 5.28**, apresenta ao todo 18 variáveis, sendo 2 variáveis tradicionais, 12 variáveis espaciais globais e 4 variáveis espaciais locais.

**Tabela 5.28:** Resumo das características do *modelo alternativo* inicial **F3ALG<sub>p1</sub>**

<b>F3ALG<sub>p1</sub></b>		
	<b>Coefficientes</b>	<b>T stat</b>
<b>Constante</b>	13427,10	
<b>POP<sub>T<sub>p</sub></sub></b>	4328,71	17,67
<b>FROTA<sub>p</sub></b>	1889,90	6,29
<b>4 D_Q1</b>	5122,52	2,05
<b>2 D_Q2</b>	836,30	0,52
<b>7 D_Q3</b>	339,70	0,19
<b>5 X21_Q1</b>	-4510,93	-1,78
<b>1 X21_Q2</b>	-619,67	-0,39
<b>8 X21_Q3</b>	-572,82	-0,29
<b>4 X51_Q1</b>	-382,93	-0,43
<b>1 X51_Q2</b>	-1952,98	-2,38
<b>9 X51_Q3</b>	-1301,93	-1,27
<b>4 X53_Q1</b>	-204,31	-0,32
<b>3 X53_Q2</b>	-416,10	-0,72
<b>6 X53_Q3</b>	-387,76	-0,58
<b>1 LISA_D</b>	1898,13	2,37
<b>2 LISA_X21</b>	-2907,76	-4,34
<b>1 LISA_X51</b>	-163,22	-0,45
<b>3 LISA_X53</b>	-36,57	-0,18
<b>R<sup>2</sup></b>	0,965	
<b>R<sup>2</sup><sub>aj</sub></b>	0,956	

Verifica-se que o modelo **5AGL<sub>p1</sub>** apresenta apenas 6 das 18 variáveis com valor do **T stat** aceitável. Para escolha das primeiras variáveis com valor de **T stat** não aceitáveis, a serem excluídas, adota-se o mesmo critério de priorização baseado na

correlação com a variável dependente (coluna à esquerda das variáveis na **Tabela 5.28**). No entanto, devido ao grande número de variáveis, exclui-se 4 a 4, selecionando-se uma *variável espacial local* e 3 *variáveis espaciais globais* (estão destacadas na **Tabela 5.28** as primeiras variáveis excluídas). Finalmente chegou-se ao *modelo alternativo misto* (ou **AGL**) da **Fase 3 (F3AGL<sub>p</sub>)** com 3 variáveis espaciais significativas, além das 2 variáveis tradicionais, que é apresentado na **Tabela 5.29**. São apresentados, também, os resultados das estimativas para o **ano base** e através da aplicação deste modelo.

**Tabela 5.29:** Resumo das características e resultados obtidos com o modelo **F3AGL<sub>p</sub>**

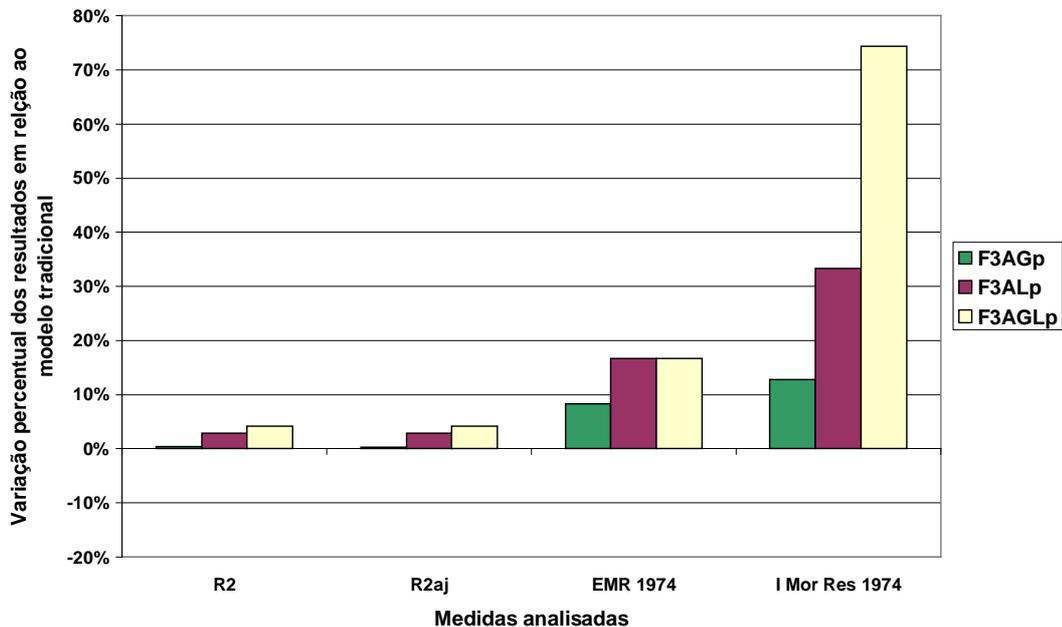
		<b>F3AGL<sub>p</sub></b>	
		<b>Coefficientes</b>	<b>T stat</b>
		<b>Constante</b>	12428.00
<b>Variáveis tradicionais</b>	<b>POP_T<sub>p</sub></b>	4053.69	18.84
	<b>FROTA<sub>p</sub></b>	2108.89	8.25
	<b>D_Q1</b>	1030.06	2.12
<b>Variáveis espaciais</b>	<b>X51_Q2</b>	-1056.87	-2.16
	<b>LISA_X21</b>	-1264.40	-8.08
	<b>Modelo</b>	<b>R<sup>2</sup></b>	0.955
<b>Resultados</b>		<b>R<sup>2</sup><sub>aj</sub></b>	0,952
	<b>Estimativas (ano base)</b>	<b>EMR</b>	10 %
		<b>I Mor Res</b>	0,10

Da mesma forma que foi realizado nas Fases 1 e 2, após a seleção dos *modelos alternativos* para cada tipo de abordagem analisada da Fase 3 e do diagnóstico através das estimativas para o **ano base** realizou-se a análise comparativa dos resultados obtidos com os três modelos **F3AG<sub>p</sub>**, **F3AL<sub>p</sub>** e **F3AGL<sub>p</sub>**, verificando-se as variações em relação aos resultados obtidos com o *modelo tradicional 5T<sub>p</sub>*.

A **Tabela 5.30** apresenta, lado a lado, os resultados obtidos para as medidas consideradas com o *modelo tradicional* e cada um dos *modelos alternativos* selecionados. Analisaram-se comparativamente os valores tanto do **R<sup>2</sup>** e **R<sup>2</sup><sub>aj</sub>** (calibração dos modelos), como os valores de **ERM** e **I\_Mor\_Res** das estimativas para o **ano base** (diagnóstico dos modelos) tomando-se como base os resultados obtidos com o *modelo tradicional (5T<sub>p</sub>)* e verificando-se os percentuais de variação, apresentados no gráfico da **Figura 5.16**.

**Tabela 5.30:** Resumo dos resultados obtidos com o *modelo tradicional* e com os *modelos alternativos* analisados na **Fase 3**

		<b>5T<sub>p</sub></b>	<b>F3AG<sub>p</sub></b>	<b>F3AL<sub>p</sub></b>	<b>F3AGL<sub>p</sub></b>
<b>Modelo</b>	<b>R<sup>2</sup></b>	0,916	0,920	0,942	0,955
	<b>R<sup>2</sup><sub>aj</sub></b>	0,914	0,917	0,940	0,952
<b>Estimativas (ano base)</b>	<b>EMR</b>	12 %	11 %	10 %	10 %
	<b>I_Mor_Res</b>	0,39	0,34	0,26	0,10

**Figura 5.16:** Análise comparativa dos resultados dos *modelos alternativos* da **Fase 3** tomando-se como referência o *modelos tradicional* **5T<sub>p</sub>**

Verifica-se que em todas as medidas analisadas houve melhora nos resultados com a introdução de variáveis espaciais nos modelos. Analisando-se cada *modelo alternativo* separadamente, com o modelo **F3AG<sub>p</sub>** obteve-se melhora quase nula para os valores de  $R^2$  e  $R^2_{aj}$  e melhoras menos significativas que com os outros dois modelos nos resultados das estimativas. O modelo **F3AGL<sub>p</sub>** foi o que mais se destacou, principalmente para o **I\_Mor\_Res** (74 %). O valor do ERM para os modelos **F3AL<sub>p</sub>** e **F3AGL<sub>p</sub>** foi o mesmo, e 17 % melhor que o *modelo tradicional*. Da mesma forma que verificou-se na **Fase 2**, em nenhuma medida o *modelo tradicional* se mostrou melhor do que os *modelos alternativos*.

#### 5.7.3.4 Modelos alternativos da Fase 3 – Método Stepwise

Uma vez que foram muitas variáveis em análise, como método complementar as técnicas aplicadas, especificou um modelo com o uso de ferramenta estatística através do método *Stepwise*. O método *Stepwise*, que se caracteriza por incluir apenas as variáveis significativas para o modelo, serviu de verificação da eficiência dos métodos adotados através do uso apenas das ferramentas de planejamento do SIG-T. No entanto, estas ferramentas nem sempre conseguem captar de forma eficaz o problema da colinearidade. Após a especificação de um modelo de regressão, é necessário que se faça a análise das variáveis incluídas através dos resultados apresentados nos relatórios do *software*. Uma forma é analisar a coerência dos coeficientes.

Incluiu-se em um modelo, para análise através do *software* SSPS, um total de 51 variáveis, entre *tradicionais* e *espaciais*, que foram:

- 11 *variáveis tradicionais*
  - 4 de totais de contagens (*AREA*, *DOMIC*, *POP\_T*, e *FROTA*);
  - 7 em forma de taxas derivadas (*RENDA*, *DENSI*, *X21*, *X32*, *X51*, *X52* e *X53*);
- 40 *variáveis espaciais*
  - 10 espaciais locais – (variáveis LISA correspondente a cada *variável tradicional*, excluindo-se a *AREA*);
  - 30 espaciais globais – (os 3 quadrantes para cada uma das 10 *variáveis tradicionais*).

O procedimento adotado pelo *software* é de seleção, primeiramente da variável considerada mais importante. Outras variáveis vão sendo sucessivamente incluídas, até chegar em um modelo considerado, pelos critérios adotados pelo método, o mais ajustado. Das 51 variáveis, foram selecionadas pelo *software*, através do método *stepwise*, apenas 8, que estão são apresentadas na **Tabela 5.31**, na ordem em que foram incluídas no modelo (modelo **SW1**). No entanto as duas variáveis (*LISA\_D* e *X51\_Q3*) que estão destacadas na tabela apresentaram valores de coeficientes incoerentes com base na *matriz de correlação* (**Anexo C**).

O procedimento adotado para eliminar o problema mencionado foi a exclusão das variáveis com coeficientes incoerentes e a execução do processo novamente para

especificação de um novo modelo. Este procedimento foi repetido 3 vezes até todas as variáveis apresentarem coeficientes considerados coerentes. Finalmente, o quarto modelo resultante da análise de regressão através do processo *Stepwise* (modelo **SW4**), que foi o primeiro em que não foram verificados problemas, foi igual ao modelo **F3AGL<sub>p</sub>** selecionado na Fase 3. Um resumo, apresentando as variáveis na ordem em que foram incluídas pelo programa e os coeficientes de cada variável dos quatro modelos analisados, é apresentado na **Tabela 5.31**. Foram destacados os coeficientes que não foram considerados coerentes.

**Tabela 5.31:** Resumo e análise das características dos modelos especificados e calibrados através do método *stepwise*

SW1		SW2		SW3		SW4	
(Constant)	13259.14	(Constant)	12650.79	(Constant)	13207.96	(Constant)	12428.04
POP_R_TOT	4105.645	POP_R_TOT	4178.623	POP_R_TOT	4283.818	POP_R_TOT	4053.714
FROTA	2257.907	FROTA	1889.094	FROTA	2081.381	FROTA	2108.874
LISA_X21	-3351.01	LISA_X21	-563.493	LISA_X21	-2986.95	LISA_X21	-1264.37
X51_Q2	-1761.48	X51_Q2	-1112.57	X51_Q2	-1786.89	X51_Q2	-1056.87
LISA_D	2307.417	POP_Q1	-682.082	LISA_D	1867.486	D_Q1	1030.045
LISA_POP	-375.199	D_Q1	5492.24				
LISA_R	-365.566	X21_Q1	-4457.11				
X51_Q3	-1475.57	LISA_DOMIC	-434.097				
		LISA_X32	-580.596				

Esta análise mostra que o método adotado para especificação e seleção de modelos com o uso das ferramentas de planejamento do SIG-T TransCAD foi eficiente, porém mais trabalhoso e, conseqüentemente, mais lento. Apesar da necessidade de análises mais detalhadas dos coeficientes dos modelos de regressão gerados através de ferramenta estatística pelo processo *stepwise*, foram apenas 4 repetições para as 51 variáveis analisadas, enquanto que um grande número de modelos precisaram ser especificados e analisados, um a um, no SIG-T, para ser selecionado o melhor.

Conclui-se, com as análises das 3 fases, que a introdução de *variáveis espaciais*, e, principalmente, a combinação dos dois tipos analisados (*globais* e *locais*), contribuiu para a melhora no desempenho dos modelos, no que diz respeito à etapa de calibração e diagnóstico. No entanto, julgou-se necessária a validação destes modelos, ou seja, a aplicação para estimativas futuras (**ano meta**), que é analisada na próxima etapa. Este procedimento é realizado para os melhores *modelos alternativos* selecionados em cada fase do estudo e também para os *modelos tradicionais* selecionados (**item 5.8**). As

análises comparativas dos resultados obtidos são apresentadas no **item 5.9**.

## 5.8. ESTIMATIVAS PARA O ANO META – VALIDAÇÃO DOS MODELOS

Os modelos calibrados com os dados do **ano base** e selecionados em cada fase da **Etapa 6** e **Etapa 7** (*tradicionais* e *alternativos*, respectivamente) foram aplicados para a previsão das viagens futuras utilizando-se os dados socioeconômicos (variáveis independentes) da pesquisa do **ano meta** e as variáveis espaciais geradas através da análise espacial para estes dados. São apresentados e analisados os resultados destas estimativas que são também comparados com os resultados das estimativas para o **ano base**, verificando a validade do modelo para estimativas futuras.

### 5.8.1 Estimativas para o ano meta - Fase 1

A **Tabela 5.32** apresenta o resumo das características e resultados obtidos nas estimativas da etapa de calibração e diagnóstico (**ano base**) e na etapa de validação (**ano meta**) com os modelos selecionados na **Fase 1** (tradicionais e alternativos).

Comparando-se os dois *modelos tradicionais* selecionados, observa-se que o valor do **ERM** nas estimativas para o **ano meta** com o modelo **3T**, cuja variável da **REND**A é em unidades de salários mínimos, é um pouco mais alto do que com o modelo **4T**, no qual a variável **REND**A é em unidades de dólar (50 % e 49 % respectivamente). Quanto ao **I\_Mor\_Res**, para a mesma estimativa o resultado observado é o mesmo para os dois modelos (0,14).

Na análise comparativa das estimativas para o **ano meta** com as do **ano base**, observa-se uma redução razoável dos valores de **I\_Mor\_Res** (de 0,39 para 0,14), indicando que os resíduos apresentam uma distribuição menos concentrada para a estimativa futura. No entanto, quanto aos **ERM**, estes valores são bem mais elevados para as estimativas (50 % e 49 %) em relação as estimativas para o **ano base** (15 %).

**Tabela 5.32:** Resumo das características e resultados das estimativas (**ano base** e **ano meta**) com os modelos selecionados na **Fase 1**

	<b>3T</b>	<b>4T</b>	<b>3AG</b>	<b>4AG</b>
<b>Constante</b>	-2084,63	-2084,63	853,957	853,957
<b>POP_T</b>	1,01966	1,01966	0,986183	0,986183
<b>DENSI</b>				
<b>RENDA<sub>SM</sub></b>	345,54		153,850	
<b>RENDA<sub>D</sub></b>		5,81402		2,58867
<b>D_Q1</b>			646,345	646,345
<b>D_Q2</b>			-278,340	-278,340
<b>D_Q3</b>			237,382	237,382
<b>R_Q1</b>			-298,305	-298,305
<b>R_Q2</b>			-1845,47	-1845,47
<b>R_Q3</b>			-484,265	-484,265
<b>R<sup>2</sup></b>	0,883	0,883	0,892	0,892
<b>R<sup>2</sup><sub>aj</sub></b>	0,881	0,881	0,881	0,881
<b>ERM (ano base)</b>	<b>15 %</b>	<b>15 %</b>	<b>13 %</b>	<b>13 %</b>
<b>I_Mor_Res</b>	0,39	0,39	0,30	0,30
<b>ERM (ano meta)</b>	50 %	49 %	44 %	44 %
<b>I_Mor_Res</b>	0,14	0,14	0,18	0,18

Quanto aos resultados dos *modelos alternativos*, também apresentados na **Tabela 5.32**, não é verificada nenhuma alteração nos resultados obtidos com os dois diferentes modelos (**3AG** e **4AG**). No entanto, na análise comparativa das estimativas para o **ano meta** com as do **ano base**, observa-se as mesmas relações verificadas para os modelos tradicionais, ou seja, uma redução razoável dos valores de **I\_Mor\_Res** (de 0,30 para 0,18) e também valores mais elevados para os ERM das estimativas do **ano meta** (**44 %**) em relação as estimativas para o **ano base** (**13 %**).

### 5.8.2 Estimativas para o ano meta - Fase 2

O *modelo tradicional* selecionados na **Fase 2** é o modelo **5T<sub>p</sub>**, no entanto, além deste, aplica-se também o modelo tradicional **5T** (sem as variáveis padronizadas) para verificar a hipótese de melhora nos resultados de estimativas futuras através do procedimento de padronização. O *modelo alternativo* selecionado é o modelo **5AGL<sub>p</sub>**, que é um modelo misto. As características dos modelos assim como os resultados das estimativas (**ano meta** e também **ano base**) são apresentados na **Tabela 5.33**.

Comparando-se, primeiramente os modelos **5T** e **5T<sub>p</sub>** verifica-se que o procedimento de padronização das variáveis foi válido, uma vez que, a única diferença

observada de resultados obtidos entre estes dois modelos é a redução significativa da estimativa do ERM para o **ano meta** (de 57 para 13 %), o que valida o modelo para previsões futuras.

Analisando-se os resultados das estimativas futuras com o modelo **5T<sub>p</sub>** em relação aos resultados das estimativas para o **ano base**, da mesma forma que verificado nos modelos da **Fase 1** os valores para o **I\_Mor\_Res** das estimativa para o **ano meta** são razoavelmente mais baixos (varia de 0,39 para 0,09). Quanto ao ERM, aumenta de 12 % (**ano base**) para 13 % nas estimativas para o **ano meta**. O que se observa com o modelo alternativo **5AGL<sub>p</sub>** são também valores mais altos de **ERM** (13 %) e mais baixos do **I\_Mor\_Res** (0,04) para estimativas futuras em relação as estimativas para o **ano base** (12 % e 0,17 respectivamente).

**Tabela 5.33:** Resumo das características e resultados das estimativas (**ano base** e **ano meta**) com os modelos selecionados na **Fase 2**

		<b>5T</b>	<b>5T<sub>p</sub></b>	<b>5AGL<sub>p</sub></b>	
		<b>Coefficientes</b>	<b>Coefficientes</b>	<b>Coefficientes</b>	
		<b>Constante</b>	118,398	12195,4	11587,7
<i>Variáveis tradicionais</i>	<i>POP_T<sub>p</sub></i>	0,707	3911,26	3683,05	
	<i>FROTA<sub>p</sub></i>	3,315	2576,07	2479,25	
<i>Variáveis espaciais</i>	<i>D_QI</i>			1709,20	
	<i>LISA_D</i>			-1079,25	
	<i>LISA_R</i>			-356,590	
<b>Modelo</b>	<b>R<sup>2</sup></b>	0,916	0,916	0,949	
	<b>R<sup>2</sup><sub>aj</sub></b>	0,914	0,914	0,946	
<b>Resultados</b>	<b>Estimativas (ano base)</b>	<b>EMR</b>	12 %	12 %	12 %
		<b>I Mor Res</b>	0,39	0,39	0,17
<b>Estimativas (ano meta)</b>	<b>EMR</b>	57 %	13 %	13 %	
	<b>I Mor Res</b>	0,09	0,09	0,04	

### 5.8.3 Estimativas para o ano meta - Fase 3

Como o modelo tradicional selecionado na **Fase 3** é o mesmo que o tradicional **5T<sub>p</sub>** da **Fase 2**, e que já foi apresentado na **Tabela 5.33**, apresenta-se na **Tabela 5.34** somente as características e resultados das estimativas com modelo alternativo **F3AGL<sub>p</sub>** selecionado nesta fase. Os resultados apresentados por este modelo diferem dos demais analisados. Apresenta os dois resultados melhores para as estimativas para o **ano base**

em relação ao **ano meta**, como pode ser observado na **Tabela 5.34**. Uma diferença maior é observada para os **ERM** que varia de valores de 10 % para 13 % para as estimativas para o **ano base** e **ano meta**, respectivamente, e uma diferença não muito significativa (de 0,10 para 0,11) para os valores de **I\_Moran\_Res**.

**Tabela 5.34:** Resumo das características e resultados das estimativas (**ano base** e **ano meta**) com o modelo alternativo selecionados na **Fase 3**

		<b>F3AGL<sub>p</sub></b>	
		<b>Coefficientes</b>	
		<b>Constante</b>	12428.00
<i>Variáveis tradicionais</i>		<i>POP_T<sub>p</sub></i>	4053.69
		<i>FROTA<sub>p</sub></i>	2108.89
		<i>D_Q1</i>	1030.06
<i>Variáveis espaciais</i>		<i>X51_Q2</i>	-1056.87
		<i>LISA_X21</i>	-1264.40
		<b>R<sup>2</sup></b>	0.955
<b>Modelo</b>		<b>R<sup>2</sup><sub>aj</sub></b>	0,952
<b>Resultados</b>	<b>Estimativas (ano base)</b>	<b>EMR</b>	10 %
		<b>I Mor Res</b>	0,10
	<b>Estimativas (ano meta)</b>	<b>EMR</b>	13 %
		<b>I Mor Res</b>	0,11

## 5.9. ANÁLISE COMPARATIVA DOS RESULTADOS

Um resumo indicando as variáveis incluídas em cada um dos 8 modelos selecionados, em cada fase nas Etapas 6 (*tradicionais*) e Etapa 7 (*alternativos*), é apresentado na **Tabela 5.35**. São apresentados, também, os resultados obtidos com os modelos na etapa de calibração e diagnóstico (Etapa 6 e 7), assim como na etapa de validação (Etapa 8). Está destacado o melhor resultado para cada medida analisada. As fases em que cada modelo foi selecionado estão indicadas na primeira linha da tabela.

Verifica-se que os melhores resultados na etapa de calibração e diagnóstico ( $R^2$ ,  $R^2_{aj}$  e estimativas para o **ano base**) são apresentados pelo modelo alternativo **F3AGL<sub>p</sub>**. No entanto, na fase de validação (estimativas para o **ano meta**), foi o modelo alternativo **5AGL<sub>p</sub>** que se destacou. Entre os modelos *tradicionais*, como se pode, também,

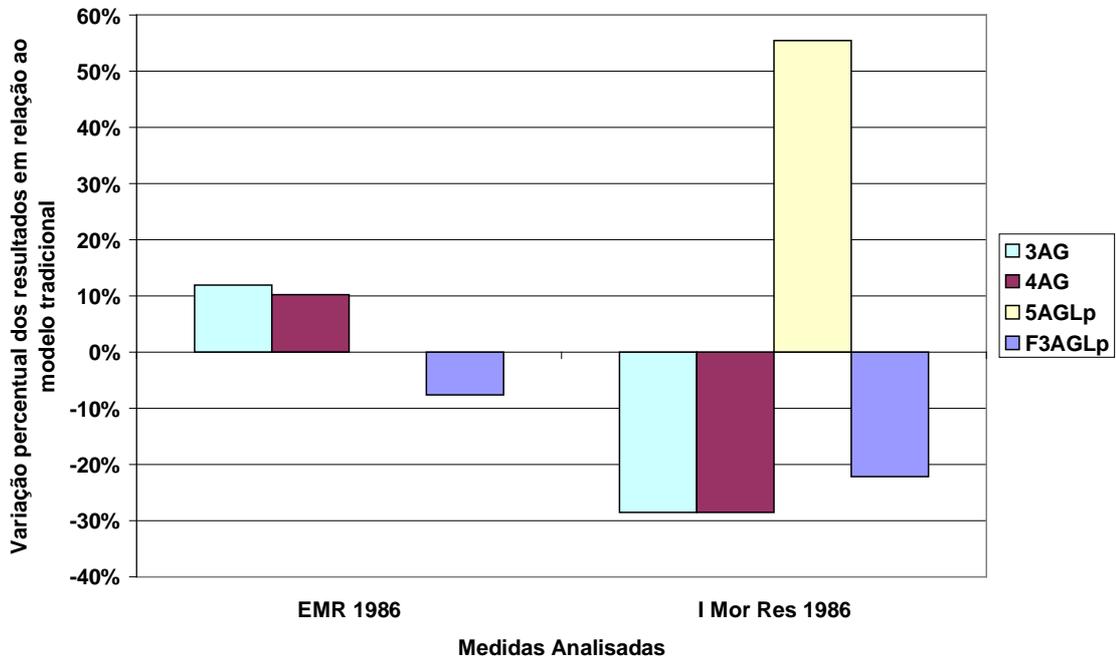
observar na **Tabela 5.35**, o que apresentou melhores resultados foi o modelo **5T<sub>p</sub>**, sendo que, o que diferencia este do modelo **5T**, é a redução significativa do **ERM** (de 57% para 13%).

**Tabela 5.35:** Modelos (*tradicionais e alternativos*) selecionados em cada fase e resultados das estimativas (**ano base e ano meta**)

	F1				F2	F2 e F3	F3	
	3T	4T	3AG	4AG	5T	5AGL <sub>p</sub>	5T <sub>p</sub>	F3AGL <sub>p</sub>
<i>POP<sub>T</sub></i>	•	•	•	•	•			
<i>FROTA</i>					•			
<i>DENSI</i>								
<i>RENDA<sub>SM</sub></i>	•		•					
<i>RENDA<sub>D</sub></i>		•		•				
<i>POP<sub>T<sub>p</sub></sub></i>						•	•	•
<i>FROTA<sub>p</sub></i>						•	•	•
<i>D<sub>Q1</sub></i>			•	•		•		•
<i>D<sub>Q2</sub></i>			•	•				
<i>D<sub>Q3</sub></i>			•	•				
<i>R<sub>Q1</sub></i>			•	•				
<i>R<sub>Q2</sub></i>			•	•				
<i>R<sub>Q3</sub></i>			•	•				
<i>X51<sub>Q2</sub></i>								•
<i>LISA<sub>D</sub></i>						•		
<i>LISA<sub>R</sub></i>						•		
<i>LISA<sub>X21</sub></i>								•
<b>R<sup>2</sup></b>	0,883	0,883	0,892	0,892	0,916	0,949	0,916	0,955
<b>R<sup>2</sup><sub>aj</sub></b>	0,881	0,881	0,881	0,881	0,914	0,946	0,914	0,952
<b>ERM (ano base)</b>	15 %	15 %	13 %	13 %	12 %	12 %	12 %	10 %
<b>I<sub>Mor_Res</sub></b>	0,39	0,39	0,30	0,30	0,39	0,17	0,39	0,10
<b>ERM (ano meta)</b>	50 %	49 %	44 %	44 %	57 %	13 %	13 %	14 %
<b>I<sub>Mor_Res</sub></b>	0,14	0,14	0,18	0,18	0,09	0,04	0,09	0,11

Apesar de já verificados os melhores resultados entre os 8 modelos selecionados, é importante analisar, primeiramente, o efeito nas estimativas com a introdução de variáveis espaciais. Desta forma, foram comparados os resultados das estimativas para o **ano meta** (**ERM** e **I<sub>Mor\_Res</sub>**) com os 4 modelos alternativos (selecionados nas 3 fases), tomando-se como base o modelo tradicional correspondente, ou seja, para a mesma fase de análise (**Figura 5.17**). Desta forma, usou-se como valor de referência para os resultados dos *modelos alternativos* **3AG** e **4AG** os resultados dos modelos

*tradicionais 3T e 4T*, respectivamente. Analogamente, para os *modelos alternativos 5AGLp e F3AGLp* usou-se como referência os resultados do *modelo tradicional 5Tp* (selecionado na Fase 2 e na Fase 3).

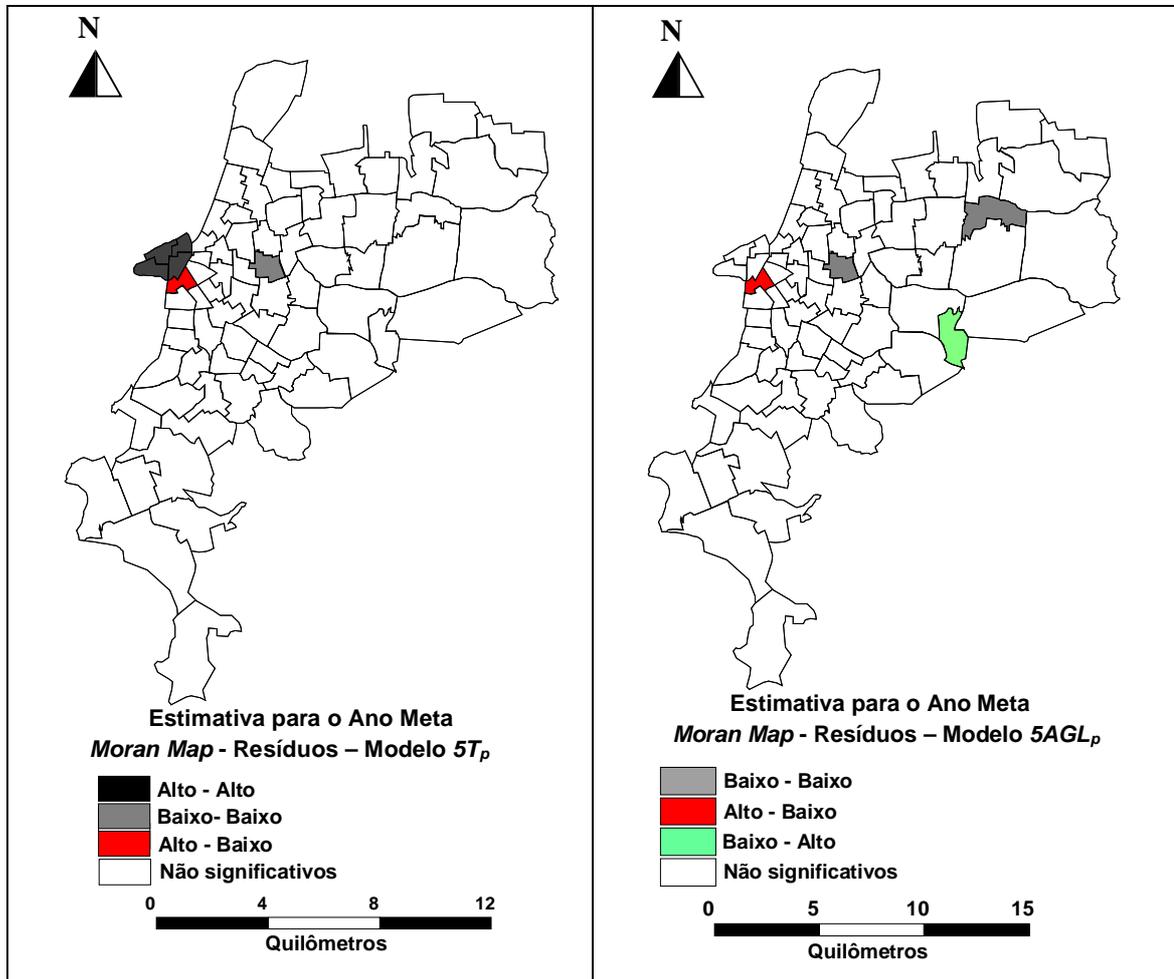


**Figura 5.17:** Análise comparativa dos melhores *modelos alternativos* de cada fase, tomando-se como referência o *modelo tradicional* correspondente

Observa-se, no gráfico da Figura 5.17, que, tanto o modelo **3AG** como o modelo **4AG** da **Fase 1**, apresentam melhores resultados para o **ERM**, no entanto, apresentam piores resultados para a distribuição espacial dos resíduos em relação aos modelos tradicionais **3T** e **4T** respectivamente. Verifica-se, também, que o *modelo alternativo F3AGL<sub>p</sub>*, que foi o modelo considerado melhor na etapa de calibração e diagnóstico, no entanto, para as estimativas para o **ano meta**, é o que apresenta piores resultados em relação ao *tradicional 5T<sub>p</sub>*, tanto para o **ERM** (-8 %) como na *distribuição espacial dos resíduos* (-22 % para **I\_Mor\_Res**).

O *modelo alternativo 5AGL<sub>p</sub>*, selecionado na **Fase 2**, é o que apresenta os melhores resultados nas estimativas futuras (**Figura 5.17**). Apesar de não apresentar variação no resultado do **ERM** em relação ao modelo tradicional **5T<sub>p</sub>**, apresenta melhoria de 56% para a *distribuição espacial dos resíduos*, o que também pode ser

verificado através de análise visual dos *Moran Map* para os resíduos das estimativas para o **ano meta** com o modelo  $5T_p$  e com o modelo  $5AGL_p$  (Figura 5.18). Destaca-se, neste caso, o modelo  $5AGL_p$  como o melhor modelo para estimativas futuras.



**Figura 5.18:** Moran Map para os resíduos das estimativas para o **ano meta**, com o modelo tradicional  $5T_p$  e com o modelo alternativo  $5AGL_p$

Observa-se, na análise do *Moran Map* (Figura 5.18), um agrupamento com significância acima de 95%, constituído por 4 ZTs no centro da cidade, que apresentaram valores altos de resíduos (Alto–Alto). Este agrupamento, que não é verificado para os resíduos com o modelo  $5AGL_p$ , representa ZTs com valores baixos de viagens em 1986 (**ano meta**) e que foram supervalorizados nas estimativas com o modelo tradicional. As outras duas ZTs em agrupamentos significativos, verificadas com o modelo  $5T_p$ , são as mesmas apresentadas para o modelo  $5AGL_p$  (sendo uma Baixo-Baixo e uma Alto-Baixo).

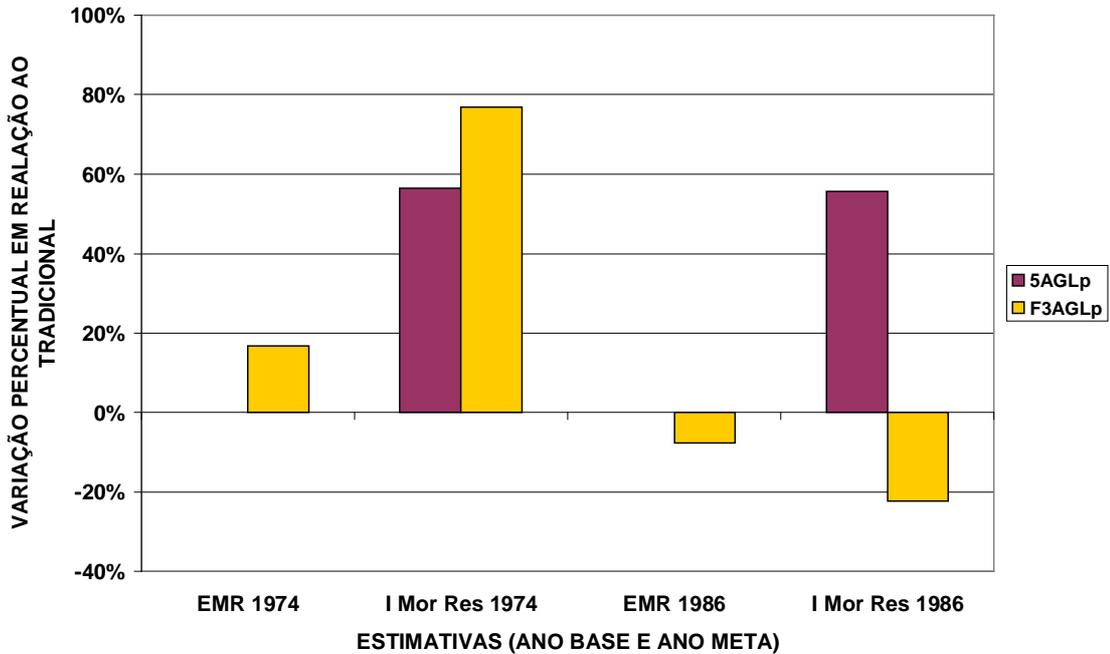
Outras duas ZTs em agrupamentos significativos são observadas para o modelo alternativo, o que não é verificado com o modelo tradicional. São ZTs que apresentaram baixo valor de resíduo, ou seja, a estimativa subestimou o número de viagens para estas ZTs. Uma delas é a ZT 76 (Baixo-Baixo), que na base de dados de 1986 é uma ZT com poucas viagens entre vizinhos com maior número de viagens, e a outra é a ZT 69 (Baixo-Alto), que para os dados reais é uma ZT com muitas viagens entre vizinhos também com muitas viagens. Neste caso, o modelo alternativo, ao subestimar esses valores, apresenta esta ZT como sendo uma zona com baixo número de viagens entre vizinhos com muitas viagens.

Realizou-se, também, análise comparativa dos resultados das duas estimativas (**ano base** e **ano meta**) com os dois melhores modelos alternativos (**5AGL<sub>p</sub>** e **F3AGL<sub>p</sub>**) (**Figura 5.19**). Como se pode verificar na **Tabela 5.35**, o primeiro foi selecionado por apresentar os melhores resultados para as estimativas futuras (validação do modelo) e o segundo apresentou melhores resultados para os dados do **ano base** (calibração e diagnóstico do modelo).

Comparando-se a distribuição espacial dos resíduos para as estimativas para o **ano base** do modelo tradicional **5T<sub>p</sub>** com o modelo alternativo **F3AGL<sub>p</sub>** (**Figura 5.20**), verifica-se que um grande agrupamento significativo (maior que 95%) de zonas com resíduos baixos (Baixo-Baixo) é observado no *Moran Map* para o modelo tradicional, o que não é observado no *Moran Map* dos resíduos com o *modelo alternativo*. Deste agrupamento de 12 ZTs, apenas duas apresentam valor do *I local de Moran* significativo para os resíduos do *modelo alternativo*. Estas zonas representam agrupamento de ZTs de muitas viagens para o ano de 1974, cujos valores foram subestimados para a estimativa com o modelo tradicional.

O gráfico da **Figura 5.19** apresenta a análise comparativa dos resultados dos dois melhores modelos alternativos (**5AGL<sub>p</sub>** e **F3AGL<sub>p</sub>**) em relação ao melhor modelo tradicional (**5T<sub>p</sub>**) das estimativas para o **ano base** e **ano meta**. Observa-se que o modelo **F3AGL<sub>p</sub>**, se destacou com os melhores resultados na fase de calibração, em relação ao modelo tradicional (17% melhor para ERM e 77% melhor para o *I\_Mor\_Res*). Esta melhoria na distribuição espacial dos resíduos das estimativas para o **ano base** com este modelo alternativo pode, também, ser observada na **Figura 5.20**, que apresenta o

**Moran Map** para os resíduos destas estimativas com os dois modelos (**5T<sub>p</sub>** e **F3AGL<sub>p</sub>**). No entanto, este mesmo *modelo alternativo* apresentou pior desempenho em relação ao outro modelo alternativo (**5AGL<sub>p</sub>**) e também em relação ao modelo tradicional **5T<sub>p</sub>**, nas estimativas para o **ano meta** (-8 e -22% para o **EMR** e **I\_Mor\_Res**, respectivamente).



**Figura 5.19:** Análise comparativa dos dois melhores *modelos alternativos*, tomando-se como referência o *modelo tradicional 5T<sub>p</sub>*

Ainda analisando-se o gráfico da **Figura 5.19**, observa-se que modelo alternativo **5AGL<sub>p</sub>** foi o que apresentou melhores resultados no geral, uma vez que apresenta melhores valores, relativamente ao tradicional, para o **I\_Mor\_Res** (56%), tanto para as estimativas para o **ano base** como para o **ano meta**, e manteve o mesmo resultado que o modelo tradicional para o **ERM**.

Também se verifica um agrupamento de 5 ZTs, na região central, com valores de resíduos altos. Estas ZTs apresentam números baixos de viagens no **ano base** e os seus valores foram superestimados pelo modelo. No entanto, no **Moran Map** para os resíduos com o modelo alternativo (**F3AGL<sub>p</sub>**), observa-se um agrupamento de *índices locais de Moran* significativos, de 5 ZTs com resíduos altos (Alto-Alto). Estas ZTs representam zonas com números altos de viagem para o ano base e que foram ainda superestimados na estimativa com o modelo tradicional.

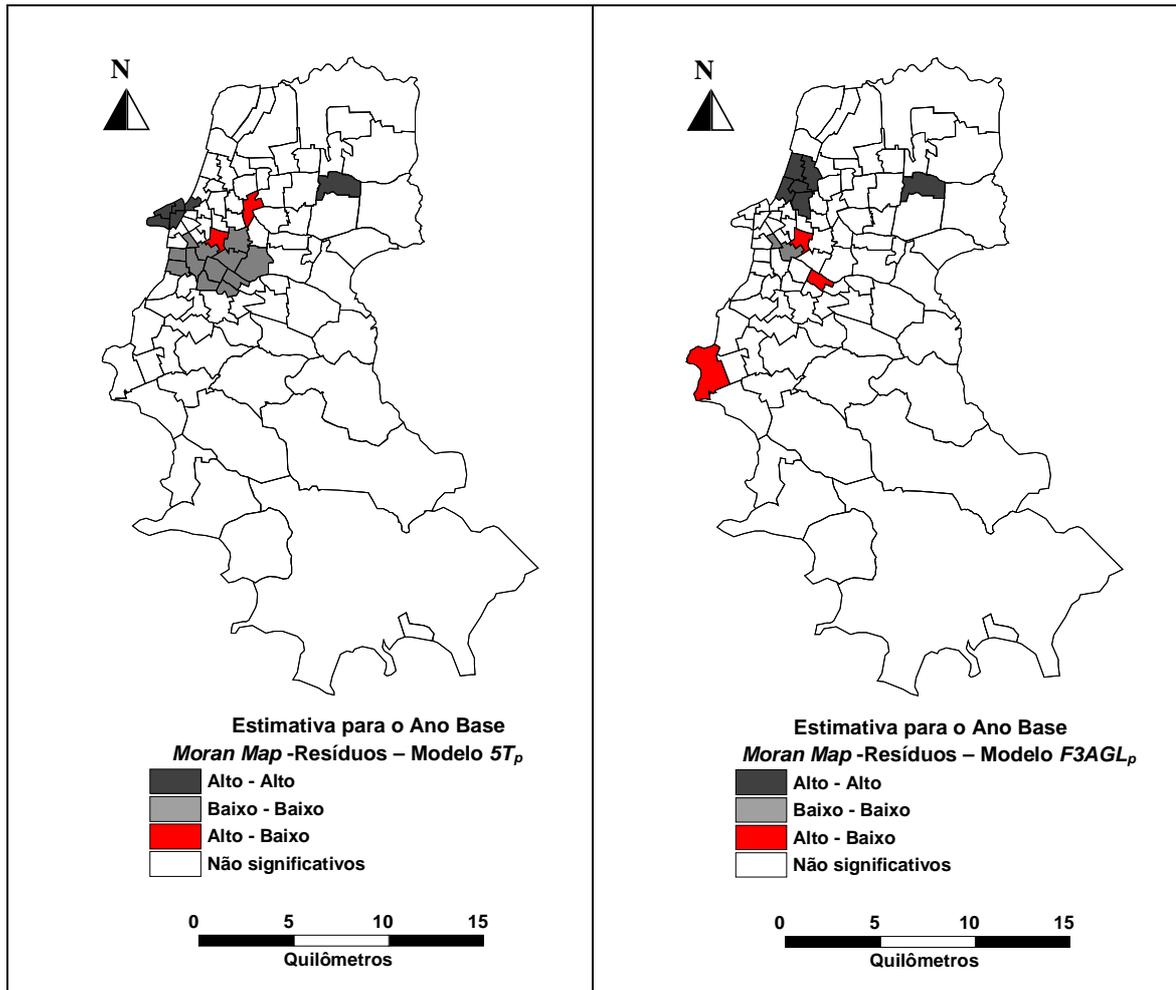


Figura 5.20: Moran Map para os resíduos das estimativas para o ano base, com o modelo tradicional  $5T_p$  e com o modelo alternativo  $F3AGL_p$

## **6. CONCLUSÕES E SUGESTÕES PARA TRABALHOS FUTUROS**

---

Este capítulo apresenta as principais conclusões obtidas nas etapas do desenvolvimento deste trabalho, destacando-se a importância do método desenvolvido e dos resultados encontrados após sua implementação, bem como sugestões para trabalhos futuros nesta área de pesquisa.

### **6.1. CONCLUSÕES**

O foco deste trabalho concentrou-se na análise dos efeitos do fenômeno da dependência espacial (também referida como autocorrelação espacial) na análise de dados de demanda por transportes. A hipótese básica era de que a introdução de indicadores de dependência espacial na modelagem de demanda por transportes poderia produzir resultados mais confiáveis dos que os obtidos com modelos tradicionais.

O objetivo foi avaliar os efeitos da consideração dos indicadores de dependência espacial nos modelos de previsão de demanda por transportes, mediante o uso de ferramentas de planejamento de transportes e de análise espacial em ambiente SIG. Fez-se uma análise comparativa, baseada em pesquisas O-D de dois períodos distintos (ano base e ano meta), na cidade de Porto Alegre, dos resultados obtidos com modelos tradicionais e os obtidos com modelos que consideram variáveis espaciais.

Na análise preliminar das pesquisas O-D disponíveis para a cidade de Porto Alegre foram enfrentados problemas que, provavelmente, serão encontrados em outras localidades. Destacam-se a não compatibilidade entre as pesquisas O-D realizadas em 1974, 1986 e 1997, com diferenças nos limites das ZTs, nas variáveis contempladas e na abrangência das pesquisas.

A pesquisa realizada em 1997 tinha um escopo reduzido, com dados muito agregados, e não foi devidamente concluída, enquanto a EDOM 2003 não estava disponível durante o desenvolvimento deste trabalho. Foram consideradas, portanto, as pesquisas realizadas em 1974 (“ano base”) e 1986 (“ano meta”). A etapa de preparação da base de dados foi trabalhosa e lenta, pois os dados da pesquisa do ano base não estavam georreferenciados e nem em arquivos eletrônicos.

As etapas de análise das pesquisas O-D e preparação da base de dados chamaram a atenção para o fato de que, apesar da realização de uma pesquisa deste porte ser trabalhosa e de custo elevado, nem sempre são tomados cuidados, na sua execução, de manter a compatibilidade com pesquisas anteriores, o que poderia permitir que as informações obtidas servissem melhor aos órgãos de planejamento municipais e metropolitanos. Um zoneamento compatível, assim como a forma de coleta e disponibilização dos dados, são fundamentais para tornar possível a análise comparativa entre as diferentes pesquisas e possibilitar o estudo da evolução espaço-temporal das demandas de viagens.

Foram utilizados, basicamente, dois *software* de SIG, SPRING e TransCAD, uma vez que nenhum deles contemplava todas as ferramentas de planejamento de transporte e de análise espacial necessárias para o desenvolvimento do estudo. Houve a necessidade, portanto, do desenvolvimento de rotinas de exportação e importação de dados, destacando-se, também, as dificuldades decorrentes da não total compatibilidade entre os programas computacionais. Verificou-se, desta forma, que se fosse possível o uso de um SIG que contemplasse todas as ferramentas de planejamento de transportes, assim como as de análise espacial, isso facilitaria consideravelmente a aplicação do método.

Para a análise comparativa dos modelos tradicionais com os modelos que introduzem variáveis espaciais foi considerada a variável dependente Viagens Produzidas de Base Domiciliar (VPBD), agregada em todos os modos e todos os motivos. Isto se deve ao fato que a pesquisa de 1974 (ano-base) não apresentava em seus relatórios resultados de viagens por ZT mais desagregados. Por este motivo, este trabalho se caracterizou mais como um exemplo de aplicação do método.

Inicialmente, foram definidos os modelos tradicionais, partindo-se das variáveis citadas na Pesquisa de 1974 (população residente total, densidade populacional e renda média mensal). Como os resultados não foram satisfatórios, aplicaram-se outros critérios, inclusive com padronização das variáveis e consideração de taxas derivadas, que resultaram na escolha das variáveis explicativas população residente total e número de veículos particulares, obtendo-se, então, um modelo tradicional com elevado coeficiente de determinação ( $R^2 = 0,92$ ). Destaca-se, particularmente, o efeito da padronização de variáveis para a obtenção de melhores modelos tradicionais, enquanto a introdução de variáveis em forma de taxas derivadas não teve efeito significativo. Acredita-se, também, que uma análise de variáveis dependentes mais desagregadas (por modo e por motivo de viagens) possa revelar outras variáveis correlacionadas e resultar em modelos com mais variáveis explicativas. Isto poderia conduzir, conseqüentemente, à obtenção de melhores estimativas.

A obtenção de variáveis espaciais foi através da análise de autocorrelação espacial, cujos maiores índices foram verificados para as variáveis taxas, destacando-se a densidade populacional, o número de domicílios por área e o número de veículos particulares por área. A análise de autocorrelação espacial mostrou, também, que a variável VPBD é significativamente influenciada pela dependência espacial e altamente relacionada com a dependência espacial das mesmas variáveis socioeconômicas.

A introdução de variáveis espaciais no estudo de modelos de demanda por transporte apresentou resultados positivos, em relação aos modelos tradicionais, tanto na etapa de diagnóstico, quando aplicados para os dados do ano base, como na etapa de validação, com estimativas para o ano meta. No entanto, requer, da mesma forma que com os modelos tradicionais, rigor na análise de significância das variáveis. Dentre os métodos analisados, verificou-se que a combinação dos dois tipos de variáveis espaciais

(globais e locais) foi mais eficiente do que a consideração dessas variáveis isoladamente.

O fato que o modelo alternativo mais ajustado para os dados do ano base não foi o mesmo que apresentou os melhores resultados para as estimativas futuras permite a conclusão complementar de que a dinâmica apresentada pelo desenvolvimento urbano, como é o caso de Porto Alegre, acarreta alterações nas relações entre as diferentes variáveis com o fenômeno estudado, modificando, inclusive, os padrões espaciais. Considera-se, portanto, que a análise desta dinâmica e o estudo de formas de introduzi-la nos modelos de demanda por transportes podem produzir resultados ainda melhores.

## **6.2. SUGESTÕES PARA TRABALHOS FUTUROS**

Os estudos realizados e os resultados obtidos representam uma etapa inicial da análise da influência das variáveis espaciais nos modelos de demanda por transportes. Muitas pesquisas complementares podem e precisam ser desenvolvidas, destacando-se:

- Estudo de outras formas de introdução de indicadores de dependência espacial, como por exemplo, a regressão espacial, que necessita de um mesmo zoneamento para os dados do ano base e do ano meta;
- Estender o estudo para as estimativas de viagens atraídas e também as de base não domiciliar, completando a etapa de geração de viagens. Posteriormente, estender para as outras três etapas: distribuição, divisão modal e alocação de viagens;
- Investigar a hipótese de que a dinâmica do desenvolvimento urbano e o estudo de formas de considerá-la nos modelos de demanda por transportes pode produzir resultados ainda melhores, da mesma forma que a análise de variáveis dependentes mais desagregadas. Sugere-se, portanto, a aplicação das análises realizadas neste trabalho a outras cidades e, no caso de Porto Alegre, o estudo dos dados da EDOM de 2003, uma vez que são os mais recentes e estão todos georreferenciados.

- Estudar uma forma de introdução de indicadores de dependência espacial em modelos de viagens encadeadas.

## REFERÊNCIAS BIBLIOGRÁFICAS

---

- Anselin, L. (1992) *Spatial data analysis with GIS: An introduction to application in the Social Sciences National Center for Geographic Information and Analysis* – University of California – Califórnia, EUA – Disponível em: <[http://www.ncgia.ucsb.edu/Publications/Tech\\_Reports/92/92-10.PDF](http://www.ncgia.ucsb.edu/Publications/Tech_Reports/92/92-10.PDF)> Acesso em 25/04/2004
- Anselin, L. (1995) Local indicators of spatial association - LISA. *Geographical Analysis*, 27:91-114
- Anselin, L. (2002) *Under the hood: issues in the specification and interpretation of spatial regression models (forthcoming: Agricultural Economics)*, obtido em: <<http://agec221.agecon.uiuc.edu/users/anselin/papers/hood.pdf>> Acesso em 03/03/2004
- Anselin, L. e Griffith, D. A. (1988) *Do spatial effects really matter in regression analysis?* Regional Science Association, 65, 11-34.
- Assunção, R. M. (2001) *Estatística Espacial com aplicações em Epidemiologia, Economia, Sociologia*. 7ª Escola de Modelos de Regressão, São Carlos, SP.
- Batty, M. (1976) Entropy in spatial aggregation. *Geographical Analysis*. 8, 1-31.
- Batty, M. e Sikdar, P.K. (1982) Spatial aggregation in spatial interaction models – Two dimensional trip distribution and location models. *Environment and Planning A*. 14, 629-658.

- Bivand, R. (1998) *A review of spatial statistical techniques for location studies*. Department of Geography - Norwegian School of Economics and Business Administration Disponível em: <http://www.nhh.no/geo/gib/gib1998/gib98-3/lund.pdf>
- CALIPER (1996) *TransCAD User's Guide*. Newton, Massachusetts, Caliper Corporation.
- Câmara, G.; Carvalho, M.S.; Cruz, O.G.; Correa, V. (2002a). *Análise Espacial de Áreas*, Em: *Análise Espacial de Dados Geográficos*, eds. Fuks, S.D.; Carvalho, M.S.; Câmara, G.; Monteiro, A.M.V. – Divisão de Processamento de Imagens – Instituto Nacional de Pesquisas Espaciais – São José dos Campos, Brasil – Disponível em: <http://www.dpi.inpe.br/gilberto/livro/analise/> Acesso em 23 de março de 2004
- Câmara, G.; Monteiro, A.M.; Carvalho, M.S. (2002b). *Análise Espacial e Geoprocessamento*, In: *Análise Espacial de Dados Geográficos*, eds. Fuks, S.D.; Carvalho, M.S.; Câmara, G.; Monteiro, A.M.V. – Divisão de Processamento de Imagens – Instituto Nacional de Pesquisas Espaciais – São José dos Campos, Brasil – Disponível em: <http://www.dpi.inpe.br/gilberto/livro/analise/> Acesso em 23 de março de 2004
- Carvalho, M. S. (1997) *Aplicação de Métodos de Análise Espacial na Caracterização de Áreas de Risco à Saúde*. Tese de Doutorado em Engenharia Biomédica, COPPE/UFRJ. Disponível em: [http://www.dpi.inpe.br/cursos/ser301/referencias/marilia\\_tese.pdf](http://www.dpi.inpe.br/cursos/ser301/referencias/marilia_tese.pdf) >. Acesso em: 21 de abril de 2004.
- Case, A. (1992) Neighborhood influence and technological change - *Regional Science and Urban Econometrics*, 22, 491-508.
- Cressie, N. (1995) Bayesian smoothing of rates in small geographic areas. *Journal of Regional Science*. 35, 659-673.
- Cressie, N (1996) Change of support and the modifiable areal unit problem. *Geographical Systems*. 3, 159-180.
- Durrett, R. (1994) *Stochastic spatial models, forefronts (newsletter of the Cornell theory center)*, 9 (#4, Spring), 4-6.
- Fotheringham, A. S. e Brunsdon, C. e Charlton, M. (2000) *Quantitative Geography - Perspectives on Spatial Data Analysis*. Ed. Sage, London.
- Geary, R. (1954) The contiguity ratio and statistical mapping. *The Incorporated Statistician*. v. 5, p. 115-145.

- Griffith, D. A. (1996) *Introduction: the need for spatial statistics*, In: Arlinghaus, S.L. and Griffith, D.A. (eds.) *Practical Handbook of Spatial Statistics*. CRC Press, Boca Raton.
- Henrique, C. S. (2004). *Diagnóstico Espacial da Mobilidade e da Acessibilidade dos Usuários do Sistema Integrado de Transporte de Fortaleza*. Dissertação de Mestrado, Programa de Mestrado em Engenharia de Transportes, Universidade Federal do Ceará, Fortaleza, CE, 165 fl.
- Hutchinson, B.G. (1979) *Princípios de Planejamento dos Sistemas de Transporte Urbano* - Editora Guanabara Dois S.A. - RJ
- IBGE (2000) *Censo Demográfico 2000 - Resultados do universo - Porto Alegre*. Acesso e, 10 de Maio de 2004 (<http://www.ibge.gov.br/english/estatistica/populacao/censo2000/universo.php?tipo=31&uf=43>).
- INPE (2004) *SPRING - Sistema de Processamento de Informações Georreferenciadas* - Instituto Nacional de Pesquisas Espaciais – Divisão de Processamento de Imagens, São José dos Campos, São Paulo. Disponível em: <<http://www.dpi.inpe.br/spring/portugues/download.php>>. Acesso em: 20 de março de 2004
- Kawamoto, E. (1993) *Análise de Sistemas de Transporte*. Relatório Técnico, EESC, Escola de Engenharia de São Carlos.
- Krempi, A. P. (2004) *Explorando Recursos de Estatística Espacial para Análise da Acessibilidade de Cidade de Bauru*. Dissertação de Mestrado, Universidade de São Paulo, São Carlos, SP, Brasil.
- Krempi, A.P., Brondino, N.C.M., Silva, A.N.R. (2002) *Evaluating transportation accessibility with spatial statistics tools in a GIS environment*. In: International Conference on Design and Decision Support Systems in Architecture and Urban Planning, 6, Ellecom – Holanda, 2002. Proceedings... Eindhoven, Eindhoven University of Technology, p. 132-144.
- Landim, P.M.B. (1998) *Análise Estatística de Dados Geológicos* – Editora UNESP – São Paulo, Brasil
- Lopes, S. B. e A. N. R. Silva (2004) *An Assessment Study of The Spatial Dependence In Transportation Demand Models*. Proceedings XIII Congresso Panamericano de Engenharia de Transito e Transportes, Anbany, NY, EUA, 26 a 29 de Setembro.

- Messner, S. e Anselin, L. (2002) *Spatial analyses of homicide with areal data*, obtido em: <http://agec221.agecon.uiuc.edu/users/anselin/papers/smla.pdf> acesso em 25/04/2004
- Miller, H. J. (1999) Potential contribution of spatial analysis to Geographic Information Systems for Transportation (GIS-T), *Geographical Analysis*, 31, 373-399.
- Moellering, H. e Tobler, W. (1972) Geographical variances. *Geographical Analysis*. 4, 34-50.
- Moran, P. (1948) The interpretation of statistical maps. *Journal of the Royal Statistic Society B*. v. 10, p. 243-251.
- Openshaw, S. (1978) An optimal zoning approach to the study of spatially aggregated data. *Spatial Representation and Spatial Interaction* – Editores: I. Masser e P.J.B. Brown. Leiden – Martinus Nijhoff, 93-113
- Ortúzar, J. de D. e Willumsen, L. G. (1994) *Modelling Transport*, 2nd edition, New York: John Wiley and Sons.
- Peixoto, N.M.O. (2002) *A Evolução Temporal da Mobilidade da População da Região Metropolitana de Porto Alegre, entre 1986 e 1997* – Dissertação de Mestrado – UFRGS – Porto Alegre-RS – 136p.
- Queiroz, M.P. (2003) *Análise Espacial dos Acidentes de Trânsito do município de Fortaleza* -Dissertação de Mestrado – UFC- Fortaleza – CE – 124p
- Saraiva, M. (2000) *A Cidade e o Tráfego – Uma Abordagem Estratégica* – Editora Universitária – UFPE – Recife
- Serrano, R. M.; Valcarce, E. V (2000) *Técnicas econométricas para el tratamiento de datos espaciales: la econometría espacial*. Barcelona, Edicions Universita de Barcelona.
- Stock, J.H. e Watson, M.W. (2004) *Econometria* - Pearson Education do Brasil, SP
- Taco, P.W.G. (1997) *Modelo de Geração de Viagens com aplicação dos Sistemas de Informação Geográfica e Sensoriamento Remoto* Dissertação de Mestrado, Publicação TU.DM-02A/97, Departamento de Engenharia Civil e Ambiental, Universidade de Brasília, Brasília, 155p.
- Teixeira, G.L. (2003) *Uso de Dados Censitários para Identificação de Zonas Homogêneas para Planejamento de Transportes Utilizando Estatística Espacial* – Dissertação de Mestrado – Publicação T.DM-010A/03- Departamento de Engenharia Civil e Ambiental- Faculdade de Tecnologia- Universidade de Brasília- DF- 155p.

- Tobler, W.R. (1970) *A computer movie simulating urban growth in the Detroit region*, *Economic Geograpy*, 46, 234-240.
- Vasconcellos, E.A. (1996) *Transporte Urbano nos Países em Desenvolvimento: Reflexões e Propostas*, 1<sup>a</sup>. Edição – Annablume – São Paulo, Brasil
- Wise, S.; R. Haining e C. Signoretta (1998) *The Role of Visualization in the Exploratory Spatial Data Analysis of Area-based Data – Proceedings of the 3rd International Conference on GeoComputation* – University of Bristol, Reino Unido. Disponível em: <<http://www.geocomputation.org/1998/>> Acesso em: 20 de setembro 2003.