

**AVALIAÇÃO DE MODELOS DE REGRESSÃO ESPACIAL PARA
ANÁLISE DE CENÁRIOS DO TRANSPORTE RODOVIÁRIO DE
CARGA**

ALAN RICARDO DA SILVA

**DISSERTAÇÃO DE MESTRADO EM TRANSPORTES
DEPARTAMENTO DE ENGENHARIA CIVIL E AMBIENTAL**

FACULDADE DE TECNOLOGIA

UNIVERSIDADE DE BRASÍLIA

UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA CIVIL E AMBIENTAL

AVALIAÇÃO DE MODELOS DE REGRESSÃO ESPACIAL
PARA ANÁLISE DE CENÁRIOS DO TRANSPORTE
RODOVIÁRIO DE CARGA

ALAN RICARDO DA SILVA

ORIENTADORA: YAEKO YAMASHITA, PHD

DISSERTAÇÃO DE MESTRADO EM TRANSPORTES

PUBLICAÇÃO: T.DM-013A/2006

BRASÍLIA/DF: JULHO/2006

UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA CIVIL E AMBIENTAL

AVALIAÇÃO DE MODELOS DE REGRESSÃO ESPACIAL
PARA ANÁLISE DE CENÁRIOS DO TRANSPORTE
RODOVIÁRIO DE CARGA

ALAN RICARDO DA SILVA

DISSERTAÇÃO DE MESTRADO SUBMETIDA AO DEPARTAMENTO DE ENGENHARIA CIVIL E AMBIENTAL DA FACULDADE DE TECNOLOGIA DA UNIVERSIDADE DE BRASÍLIA COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE.

APROVADA POR:

YAEKO YAMASHITA, PhD (UnB)
(ORIENTADORA)

MARIA ALICE PRUDÊNCIO JACQUES, PhD (UnB)
(EXAMINADORA INTERNA)

CARLOS FELIPE GRANGEIRO LOUREIRO, PhD (UFC)
(EXAMINADOR EXTERNO)

Data: Brasília/DF, 28 de julho de 2006

FICHA CATALOGRÁFICA

SILVA, ALAN RICARDO DA

Avaliação de Modelos de Regressão Espacial para Análise de Cenários do Transporte Rodoviário de Carga. Brasília, 2006.

xii, 125p., 210x297mm (ENC/FT/UnB, Mestre, Transportes, 2006).

Dissertação de Mestrado – Universidade de Brasília. Faculdade de Tecnologia. Departamento de Engenharia Civil e Ambiental.

1 – Transporte Rodoviário de Carga

2 – Regressão Espacial

3 – Análise Espacial

4 – Planejamento de Transportes

I – ENC/FT/UnB

II – Título (série)

REFERÊNCIA BIBLIOGRÁFICA

SILVA, A. R. da (2006). Avaliação de Modelos de Regressão Espacial para Análise de Cenários do Transporte Rodoviário de Carga. Dissertação de Mestrado, Publicação T.DM-013A/2006, Departamento de Engenharia Civil e Ambiental, Faculdade de Tecnologia, Universidade de Brasília, DF, 125p.

CESSÃO DE DIREITOS

NOME DO AUTOR: Alan Ricardo da Silva

TÍTULO DA DISSERTAÇÃO DE MESTRADO: Avaliação de Modelos de Regressão Espacial para Análise de Cenários do Transporte Rodoviário de Carga.

GRAU/ANO: Mestre/2006

É concedida à Universidade de Brasília permissão para reproduzir cópias desta dissertação de mestrado e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte desta dissertação de mestrado pode ser reproduzida sem a autorização por escrito do autor.

Alan Ricardo da Silva
QE 15 Conj. Q casa 38 – Guará II
Brasília, DF – Brasil – CEP: 71050-171

Dedicatória

***A Deus, por tudo que sou, à minha família principalmente a meus pais,
Francisca e Benedito (in memorium), pelos ensinamentos de vida
e à minha namorada, Adriana Mara, pela paciência e dedicação .***

Agradecimentos

Muitas pessoas me ajudaram, direta e indiretamente, na conclusão de mais essa etapa da minha vida, e relacionar todas é muito difícil. Por isso espero não esquecer ninguém nessa lista, mas se esquecer, que a pessoa saiba que a reconheço verdadeiramente.

À professora Yaeko, por toda a orientação e pela paciência por eu perturbá-la quase que semanalmente, na discussão sobre o desenvolvimento da dissertação.

Ao professor Raul, orientador de projeto final em estatística, colega de consultoria e principal incentivador para meu ingresso no mestrado em transportes.

Aos colegas que ingressaram comigo e que tiveram muita, mas muita paciência quando eu tocava hits como “Tremendo Vacilão - Perla” e “Emboladeira – As leoas”, André Nunes (Paraíba), Thania Regina PANTOJA da Vera Cruz, Reverson dos Anjos (Cabelo ou Reversito ou Argentino), Leonardo Vasconcelos (Super Léo @uol), Aline Lang, Mariana Rios, Kátia Gomes (Kátia G), Sandra, Gabrielle, Arthur, Flávio, Félix, Clarisse, Giusepe.

Aos colegas que ingressaram antes e depois de mim, Leandro Rodrigues (Sertanejo), Leandro Bernardes (Careca), João Belém, Marianne, Gildemir, Alessandro, Fred, Aline, Gizelle Netto (Nhanhães), Carlos Eduardo (Pituco), Rafael Mathos, Ricardo Garcia, Francesca, Marcus Porfírio, Andréia, Arley (Saracura), Luciany, Miguel, Fernando.

Aos colegas estatísticos Mírian, Kátia, Hélio, Patrick e Fernando, pelas discussões estatísticas a cerca da utilização dos modelos de regressão espacial.

Ao pessoal do CEFTRU (não vou citar nomes porque são muitos), pela ajuda sobre o Transporte Rodoviário e pelos momentos de descontração.

Ao Júlio pelas conversas de corredor, pela ajuda administrativa e pela arrecadação da caixinha para a compra de biscoitos e copos.

A todos os professores, pelos ensinamentos sobre transporte, assunto antes desconhecido para mim.

E à CAPES, pelo apoio financeiro, fundamental para a conclusão desse trabalho....

RESUMO

AVALIAÇÃO DE MODELOS DE REGRESSÃO ESPACIAL PARA ANÁLISE DE CENÁRIOS DO TRANSPORTE RODOVIÁRIO DE CARGA

O Transporte Rodoviário de Carga (TRC) se apresenta atualmente como um dos modos de transporte mais utilizados no Brasil, representando 61,1% da matriz de transportes, contra 20,7% do ferroviário, 13,6% do aquaviário, 4,2% do dutoviário e 0,4% do aéreo. A possibilidade de entrega na porta do comprador, bem como a facilidade de embarque e desembarque de mercadorias são algumas das características que corroboram sua participação no mercado. Dessa forma, o conhecimento acerca das características do TRC, como também a identificação dos fatores que influenciam sua distribuição espacial auxiliam os interessados nesse modo na tomada de decisão. A distribuição espacial da frota de veículos rodoviários de carga depende geralmente de características regionais, como a produção de grãos, existência de um porto etc, indicando que o “espaço”, no caso o geográfico, influencia diretamente em tal distribuição. Assim, modelos de regressão convencional não são os mais adequados para trabalhar com dados espaciais, onde está presente a dependência espacial, fazendo-se necessário o uso de modelos de regressão espacial. O presente trabalho apresenta, além de uma metodologia para a construção de modelos de regressão espacial, uma avaliação dos mesmos quando comparados aos modelos de regressão convencional, no sentido de identificar e analisar os fatores intervenientes na distribuição espacial da frota de caminhões. Para isso, estudou-se os veículos rodoviários de carga do tipo graneleiro com 2 ou 3 eixos e os caminhões simples, também com 2 ou 3 eixos, onde as variáveis explicativas utilizadas foram densidade demográfica, valor das exportações, valor dos investimentos, quantidade produzida de soja e quantidade de estabelecimentos industriais. Os resultados mostraram que, na existência de dependência espacial, os modelos de regressão espacial apresentaram diagnósticos que permitem compreender melhor a realidade do que os modelos de regressão convencional, e que a forma da matriz de proximidades influencia diretamente em tais diagnósticos. Conclui-se do trabalho que modelos que incorporam o fator aleatório “espaço” fornecem melhores resultados para um planejamento mais eficiente.

ABSTRACT

EVALUATION OF SPATIAL REGRESSION MODELS TO ANALYSIS OF SCENES OF THE ROAD FREIGHT TRANSPORTATION

The Road Freight Transportation (RFT) currently presents itself as one of the most used ways of transportation in Brazil, representing 61.1% of the transportation matrix, against 20.7% of railroad, 13.6% of the river-way, 4.2% of duct-way and 0.4% of the aerial. The possibility of delivery at the door of the purchaser, as well as the easiness of loading and unloading of merchandises are some of the characteristics that corroborate its participation in the market. Therefore, the knowledge concerning the characteristics of the RFT, as well as the identification of the factors that influence its spatial distribution assist the interested parties in decision taking. The spatial distribution of the road freight fleet vehicles generally depends on regional characteristics, such as the production of grains, the existence of a port and so forth, indicating that the "space", in this case, the geographic one, influences directly in such distribution. Thus, conventional regression models are not adjusted to work with spatial data, where the spatial dependence is present, making necessary the use of spatial regression models. The present work shows, through a methodology for the construction of spatial regression models, an evaluation of them when compared with the conventional regression models, aiming at identifying and analyzing the intervening factors in the trucks fleet spatial distribution. For this, the road freight vehicles of the type granary ship with 2 or 3 axles and the simple trucks, also with 2 or 3 axles, was studied where the independent variables used were demographic density, exportations value, investments value, soy produced amount and industrials establishments amount. The results showed that, there being spatial dependence, the spatial regression models presented diagnoses that allow to understand the reality better than the conventional regression models, and that the form of the neighborhoods matrix influences directly in such diagnoses. It is concluded from the work that models that incorporate the random factor "space" supply better results to a more efficient planning.

SUMÁRIO

1. INTRODUÇÃO	1
1.1. FORMULAÇÃO DO PROBLEMA.....	2
1.2. HIPÓTESES	2
1.3. JUSTIFICATIVA	2
1.4. OBJETIVOS	3
1.5. ESTRUTURA DA DISSERTAÇÃO	4
 2. TRANSPORTE RODOVIÁRIO DE CARGA	 5
2.1. APRESENTAÇÃO	5
2.2. DEFINIÇÃO E CARACTERÍSTICAS.....	6
2.2.1.Evolução do Transporte Rodoviário de Carga no Brasil.....	6
2.2.2.Aspectos Positivos e Negativos do Transporte Rodoviário de Carga	8
2.2.2.1.Pontos Positivos	8
2.2.2.2.Pontos Negativos.....	9
2.3. TERMINOLOGIA DOS VEÍCULOS RODOVIÁRIOS DE CARGA	9
2.3.1.Definições	10
2.3.1.1.Veículo Rodoviário de Carga Incompleto	12
2.3.1.2.Veículo Rodoviário de Carga Completo.....	19
2.3.1.3.Caminhão-trator	20
2.3.1.4.Veículo Rodoviário de Carga Combinado.....	20
2.3.2.Tipos de Cargas.....	23
2.4. PARTICIPANTES DO TRANSPORTE RODOVIÁRIO DE CARGA	24
2.5. TÓPICOS CONCLUSIVOS	26
 3. REGRESSÃO ESPACIAL	 28
3.1. APRESENTAÇÃO	28
3.2. MATRIZ DE PROXIMIDADES	29
3.3. TESTES DE AUTOCORRELAÇÃO ESPACIAL.....	31
3.3.1.Índices Globais de Associação Espacial.....	32
3.3.2.Índices Locais de Associação Espacial.....	34
3.4. ANÁLISE GRÁFICA DE DEPENDÊNCIA ESPACIAL	35
3.4.1.Diagrama de Espalhamento de Moran.....	36
3.4.2.Box map, LISA map e Moran map.....	37
3.5. MODELOS DE REGRESSÃO ESPACIAL	39
3.5.1.Modelos com Efeitos Espaciais Globais.....	43
3.5.1.1.Modelo Espacial Autoregressivo de Primeira Ordem.....	44
3.5.1.2.Modelo Espacial Autoregressivo Misto.....	45
3.5.1.3.Modelo com Erro Espacial Autoregressivo	45
3.5.1.4.Modelo Espacial Durbin.....	46

3.5.2.Modelos com Efeitos Espaciais Locais.....	47
3.5.2.1.Modelagem de forma discreta	47
3.5.2.2.Modelagem de forma contínua.....	48
3.6. DEFINIÇÃO DE ESPAÇO NO CONTEXTO DA REGRESSÃO ESPACIAL	49
3.7. TÓPICOS CONCLUSIVOS	52
 4. METODOLOGIA PARA CONSTRUÇÃO DE MODELOS DE REGRESSÃO ESPACIAL	54
4.1. APRESENTAÇÃO	54
4.2. ESTRUTURA.....	55
4.3. ETAPAS	57
4.3.1.Etapa 1: Identificação do Objeto e das Variáveis	57
4.3.2.Etapa 2: Montagem do Banco de Dados	57
4.3.3.Etapa 3: Verificação de Dependência Espacial	59
4.3.4.Etapa 4: Calibração e Validação do Modelo	60
4.4. TÓPICOS CONCLUSIVOS	62
 5. ESTUDO DE CASO: TRANSPORTE RODOVIÁRIO DE CARGA	63
5.1. APRESENTAÇÃO	63
5.2. APLICAÇÃO DA METODOLOGIA I: VEÍCULOS RODOVIÁRIOS DE CARGA DO TIPO GRANELEIRO.....	64
5.2.1.Etapa 1: Identificação do Objeto e das Variáveis	64
5.2.2.Etapa 2: Montagem do Banco de Dados	65
5.2.3.Etapa 3: Verificação de Dependência Espacial	68
5.2.4.Etapa 4: Calibração do Modelo.....	78
5.3. COMPARAÇÃO ENTRE OS MODELOS ESPACIAL E CONVENCIONAL.....	89
5.3.1.Estudo de caso: Carrocerias GRANELEIRAS	91
5.4. APLICAÇÃO DA METODOLOGIA II: VEÍCULOS RODOVIÁRIOS DE CARGA DO TIPO CAMINHÃO SIMPLES	95
5.5. TÓPICOS CONCLUSIVOS	103
 6. CONCLUSÕES E RECOMENDAÇÕES	105
6.1. APRESENTAÇÃO	105
6.2. CONCLUSÕES.....	105
6.3. RECOMENDAÇÕES E SUGESTÕES PARA TRABALHOS FUTUROS.....	107
 BIBLIOGRAFIA	108
 ANEXOS.....	112

LISTA DE TABELAS

Tabela 2.1. Tipos de Veículos Rebocados.....	12
Tabela 2.2. Tipos de Cargas.....	24
Tabela 3.1. Softwares para análise espacial disponíveis no mercado e suas funcionalidades..	49
Tabela 5.1. Estatísticas Descritivas das Variáveis Estudadas.	67
Tabela 5.2. Índices de Dependência Espacial da Variável Frota	72
Tabela 5.3. Índices de Dependência Espacial para Diferentes Distâncias entre Centróides. ...	72
Tabela 5.4. Índices de Dependência Espacial da Variável Frota_{alg}	76
Tabela 5.5. Matriz de Correlação das Variáveis Seleccionadas.	78
Tabela 5.6. Resultados do Modelo de Regressão Convencional e Espacial.....	81
Tabela 5.7. Testes de Normalidade para os Modelos de Regressão Convencional e Espacial.	82
Tabela 5.8. Índices de Dependência Espacial para Diferentes Distâncias entre Centróides. ...	84
Tabela 5.9. Resultados do Modelo de Regressão Convencional e Espacial.....	85
Tabela 5.10. Testes de Normalidade para os Modelos de Regressão Convencional e Espacial.	87
Tabela 5.11. Resultados do Modelo de Regressão Convencional e Espacial.....	87
Tabela 5.12. Testes de Normalidade para os Modelos de Regressão Convencional e Espacial.	88
Tabela 5.13. Estatísticas de Ajuste dos Modelos de Regressão Convencional e Espacial.....	91
Tabela 5.14. Parâmetros dos Modelos de Regressão Convencional e Espacial.	93
Tabela 5.15. Índices de Dependência Espacial para Diferentes Distâncias entre Centróides. .	97
Tabela 5.16. Índices de Dependência Espacial da Variável Frota_B	97
Tabela 5.17. Matriz de Correlação das Variáveis Seleccionadas.	99
Tabela 5.18. Resultados do Modelo de Regressão Convencional e Espacial.....	101
Tabela 5.19. Resultados do Modelo de Regressão Convencional e Espacial.....	101
Tabela 5.20. Estatísticas de Ajuste dos Modelos de Regressão Convencional e Espacial....	102
Tabela 5.21. Parâmetros dos Modelos de Regressão Convencional e Espacial.	103

LISTA DE QUADROS

Quadro 2.1. Tipos de Carrocerias Abertas.	13
Quadro 2.2. Tipos de Carrocerias Fechadas.....	15
Quadro 2.3. Tipos de Mecanismo Operacional.	16
Quadro 2.4. Tipos de Distribuidor de Peso.	18
Quadro 2.5. Tipos de Veículo Rodoviário de Carga Completo.	19
Quadro 2.6. Tipos de caminhões simples classificados segundo o chassi.	20
Quadro 2.7. Configurações de carretas a capacidade de carga.....	21
Quadro 2.8. Combinações de veículos segundo CONTRAN 68/98.	22

LISTA DE FIGURAS

Figura 2.1: Matriz de Transportes Brasileira.	5
Figura 2.2: Classificação dos Veículos Rodoviários de Carga.	11
Figura 2.3: Veículo rodoviário de Carga Incompleto - Chassi.	12
Figura 2.4: Dispositivo de Acoplamento Mecânico – Quinta Roda.	19
Figura 2.5: Caminhão-trator.	20
Figura 2.6: Desenho esquemático de uma carreta.	21
Figura 2.7: Semi-reboque.	21
Figura 2.8: Reboque.	21
Figura 2.9: Interessados no Transporte Rodoviário de Carga.	26
Figura 3.1: Exemplo de Criação da Matriz de Proximidade Espacial.	31
Figura 3.2: Exemplo do Diagrama de Espalhamento de Moran.	37
Figura 3.3: Exemplo de um <i>Box map</i> (ou <i>Moran Scatterplot</i> em forma de mapa).	38
Figura 3.4: Exemplo de um <i>Moran map</i>	39
Figura 3.5: Representação de Autocorrelação Positiva (a) e (b) e Negativa (c) e (d).	40
Figura 3.6: Retas estimadas de um modelo de regressão com uma variável <i>dummy</i>	43
Figura 3.7: Exemplo de Regiões Interligadas por Estradas.	50
Figura 3.8: Tempo necessário (minutos) para ir de uma região <i>i</i> a uma região <i>j</i>	51
Figura 3.9: Matriz <i>W</i> utilizando o tempo necessário para ir de uma região <i>i</i> a uma região <i>j</i>	51
Figura 4.1: Fluxograma para Construção de Modelos de Regressão.	55
Figura 4.2: Fluxograma Metodológico.	56
Figura 4.3: Distribuição dos resíduos na existência de dependência espacial.	60
Figura 5.1: Histograma da Variável Frota	68
Figura 5.2: Distribuição Espacial da Frota de Carrocerias do Tipo GRANELEIRA.	69
Figura 5.3: Distribuição Espacial da Quantidade de Soja Produzida (em Toneladas).	70
Figura 5.4: Distribuição Espacial da Quantidade de Milho Produzido (em Toneladas).	70
Figura 5.5: Distribuição Espacial da Quantidade de Café Produzido (em Toneladas).	71
Figura 5.6: Distribuição Espacial da Quantidade de Grãos Produzido (em Toneladas).	71
Figura 5.7: <i>Moran Scatterplot</i> da Variável Frota Utilizando a Matriz A.	74
Figura 5.8: <i>Moran Scatterplot</i> da Variável Frota Utilizando a Matriz B.	74
Figura 5.9: <i>Moran Scatterplot</i> (Mapa) da Variável Frota Utilizando a Matriz A.	74
Figura 5.10: <i>Moran Scatterplot</i> (Mapa) da Variável Frota Utilizando a Matriz B.	74
Figura 5.11: <i>Moran Map</i> da Variável Frota Utilizando a Matriz A.	75
Figura 5.12: <i>Moran Map</i> da Variável Frota Utilizando a Matriz B.	75
Figura 5.13: <i>Moran Scatterplot</i> da Variável Frota Utilizando a Matriz A.	76
Figura 5.14: <i>Moran Scatterplot</i> da Variável Frota Utilizando a Matriz B.	76
Figura 5.15: <i>Moran Scatterplot</i> (Mapa) da Variável Frota Utilizando a Matriz A.	77
Figura 5.16: <i>Moran Scatterplot</i> (Mapa) da Variável Frota Utilizando a Matriz B.	77
Figura 5.17: <i>Moran Map</i> da Variável Frota Utilizando a Matriz A.	77
Figura 5.18: <i>Moran Map</i> da Variável Frota Utilizando a Matriz B.	77
Figura 5.19: Gráficos de Dispersão das Variáveis Dens, Exp, Inv, Soja e Milho	80
Figura 5.20: Resíduos da Regressão Convencional.	82
Figura 5.21: Resíduos da Regressão Espacial Utilizando a Matriz A.	82
Figura 5.22: Resíduos da Regressão Espacial Utilizando a Matriz B.	82
Figura 5.23: <i>qq-plot</i> dos Resíduos da Regressão Convencional.	82
Figura 5.24: <i>qq-plot</i> dos Resíduos da Regressão Espacial Utilizando a Matriz A.	82
Figura 5.25: <i>qq-plot</i> dos Resíduos da Regressão Espacial Utilizando a Matriz B.	82
Figura 5.26: Gráficos de Dispersão das Variáveis Dens, Exp, Inv, Soja e Milho	83

Figura 5.27: Distribuição Espacial da Frota de Carrocerias do Tipo GRANELEIRA, onde Existe Produção de Soja, Milho e onde há Empresas Exportadoras.	85
Figura 5.28: Resíduos da Regressão Convencional.	86
Figura 5.29: Resíduos da Regressão Espacial Utilizando a Matriz A.	86
Figura 5.30: Resíduos da Regressão Espacial Utilizando a Matriz B.	86
Figura 5.31: <i>qq-plot</i> dos Resíduos da Regressão Convencional.	87
Figura 5.32: <i>qq-plot</i> dos Resíduos da Regressão Espacial Utilizando a Matriz A.	87
Figura 5.33: <i>qq-plot</i> dos Resíduos da Regressão Espacial Utilizando a Matriz B.	87
Figura 5.34: Resíduos da Regressão Convencional.	88
Figura 5.35: Resíduos da Regressão Espacial Utilizando a Matriz A.	88
Figura 5.36: Resíduos da Regressão Espacial Utilizando a Matriz B.	88
Figura 5.37: <i>qq-plot</i> dos Resíduos na Regressão Convencional.	89
Figura 5.38: <i>qq-plot</i> dos Resíduos na Regressão Espacial Utilizando a Matriz A.	89
Figura 5.39: <i>qq-plot</i> dos Resíduos na Regressão Espacial Utilizando a Matriz B.	89
Figura 5.40: Distribuição Espacial da Frota de CAMINHÕES SIMPLES no Espírito Santo.	96
Figura 5.41: <i>Moran Scatterplot</i> da Variável FrotaB Utilizando a Matriz A.	98
Figura 5.42: <i>Moran Scatterplot</i> da Variável FrotaB Utilizando a Matriz B.	98
Figura 5.43: <i>Moran Scatterplot</i> (Mapa) da Variável FrotaB Utilizando a Matriz A.	98
Figura 5.44: <i>Moran Scatterplot</i> (Mapa) da Variável FrotaB Utilizando a Matriz B.	98
Figura 5.45: <i>Moran Map</i> da Variável FrotaB Utilizando a Matriz A.	99
Figura 5.46: <i>Moran Map</i> da Variável FrotaB Utilizando a Matriz B.	99
Figura 5.47: Gráficos de Dispersão das Variáveis Com e Ind	100

1. INTRODUÇÃO

O aumento da produção nacional, seja em *commodities* como a soja ou em produtos manufaturados como o calçado, tem estimulado o desenvolvimento econômico e social do Brasil. O transporte, dentre outros, é um fator que contribui para esse desenvolvimento, pois possibilita a movimentação da carga para distribuição interna ou até um porto para exportação, sem mencionar no transporte das pessoas envolvidas em todo o processo de produção. No tocante à movimentação de carga, o modo rodoviário se destaca dos outros modos por representar cerca de 61% da matriz de transportes brasileira, ressaltando assim sua relevância dentro da economia nacional (ANTT, 2005).

De um modo geral, pode-se dizer que o mercado de transporte rodoviário de carga é composto pelos usuários, pelas montadoras, pelos operadores e pelos transportadores, que podem ser divididos em autônomos, empresas e cooperativas. Assim, o conhecimento acerca das características da frota de caminhões, bem como sua localização espacial, auxilia esse mercado e o poder público na definição da estrutura de preços e na política de regulação do setor, respectivamente.

Aos transportadores e operadores fica a responsabilidade de definir o valor do frete, enquanto as montadoras definem os preços dos caminhões, segundo o tipo e a demanda. Cabe aos usuários, que são os principais atores, movimentar esse mercado. O poder público tem o papel de definir políticas de regulação e fiscalização para garantir a segurança e o transporte da produção nacional. Para tal, tem como responsabilidade manter as rodovias em condições razoáveis de circulação, utilizar balanças para o controle de peso, além da possibilidade de fornecer subsídios aos transportadores para a compra de novos caminhões.

Além disso, identificando as variáveis que influenciam diretamente na quantidade de caminhões, é possível realizar uma análise de cenários, a fim de prever possíveis alterações na frota de caminhões quando ocorrerem variações nas variáveis identificadas. Dessa forma, a análise de cenários constitui uma importante ferramenta no planejamento estratégico.

1.1. FORMULAÇÃO DO PROBLEMA

Para o conhecimento sobre a frota de caminhões é imprescindível a realização de um diagnóstico do sistema de Transporte Rodoviário de Carga (TRC). Acontece que muitos são os fatores que influenciam na distribuição espacial dessa frota, e conhecer a influência de alguns desses fatores nessa distribuição é relevante para o setor de transportes, pois permite um planejamento mais direcionado. Assim, o problema da pesquisa é: “Como identificar e analisar os fatores que influenciam na distribuição espacial da frota de veículos rodoviários de carga, a fim de subsidiar seu diagnóstico?”.

1.2. HIPÓTESES

A hipótese considerada nessa pesquisa é:

- um modelo de regressão espacial permite identificar os fatores intervenientes e avaliar o impacto da variação destes na distribuição espacial da frota de veículos rodoviários de carga.

Como hipóteses secundárias tem-se que:

1. A frota de veículos rodoviários de carga é uma variável espacialmente dependente;
2. Um modelo de regressão espacial permite compreender melhor a realidade do que um modelo de regressão convencional, a fim de uma previsão mais adequada;

1.3. JUSTIFICATIVA

A concentração da frota de caminhões depende principalmente de características como a concorrência e a produção. Além disso, esses dois fatores são interdependentes, ou seja, onde há uma grande produção existe uma grande concorrência e onde há uma grande concorrência existe uma grande produção. Por esse motivo, aglomerações maiores de caminhões em algumas áreas e aglomerações menores em outras configuram a distribuição espacial da oferta. Esse fato sugere que o espaço geográfico influencia diretamente em tal distribuição, o que está de acordo com a Primeira Lei da Geografia enunciada por Tobler (1979): “tudo está relacionado a tudo, mas as coisas mais próximas estão mais relacionadas que as coisas mais distantes”.

Vários fatores são responsáveis por caracterizar essa distribuição espacial da frota de caminhões. O processo de identificação desses fatores engloba as etapas de coleta, tratamento dos dados e análise de correlações. Problemas na coleta de dados ou nos próprios dados, como precisão e atualidade, podem gerar correlações espúrias e levar o planejador a uma tomada de decisão inadequada (Correia, 2004). Para medir a qualidade da informação é necessária a análise do contexto no qual a mesma será utilizada. A mesma informação que possui boa qualidade para um determinado fim pode não apresentar qualidade para outro propósito qualquer.

Além disso, a utilização de um modelo matemático que não incorpore o fator “espaço” em sua estrutura, poderá subestimar ou superestimar as verdadeiras relações entre as variáveis, pois a contribuição do fator aleatório “espaço” estará distribuída entre as variáveis do modelo. Caso isso ocorra, o modelo perderá sua principal função que é tentar representar um fenômeno real. O transporte, em geral, depende de fatores geográficos, físicos, políticos e sócio-econômicos sendo que o fator geográfico é muitas vezes não observado.

Por isso, modelos de regressão convencional não são os mais adequados para se trabalhar com dados de transporte, pois não levam em consideração a espacialidade, ocasionando uma valoração incorreta dos parâmetros e, conseqüentemente, gerando um modelo irreal. Assim, para uma correta definição dos parâmetros do modelo, faz-se necessária a utilização de modelos de regressão espacial que consideram a influência do “espaço” na ocorrência dos eventos, ou seja, tornando as observações dependentes de seus “vizinhos”.

1.4. OBJETIVOS

O objetivo geral dessa pesquisa é o desenvolvimento de um modelo de regressão espacial para identificar os fatores intervenientes e avaliar o impacto da variação destes na distribuição espacial da frota de veículos rodoviários de carga, a fim de subsidiar seu diagnóstico.

Os objetivos específicos são:

1. Definir critérios para verificação da qualidade dos dados;
2. Definir quais variáveis influenciam a frota de veículos rodoviários de carga;

3. Comparar os resultados de um modelo de regressão convencional e um modelo de regressão espacial.

1.5. ESTRUTURA DA DISSERTAÇÃO

A presente dissertação foi dividida em 6 capítulos no intuito de se alcançar os objetivos propostos. Após esse capítulo introdutório, os dois capítulos seguintes apresentam a base teórica, que dá sustentação ao trabalho, e que fornece os conceitos contidos no mesmo. O Capítulo 2 trata dos aspectos referentes ao Transporte Rodoviário de Carga (TRC), como suas características, definições e tipos de veículos rodoviários de carga, como também os participantes e principais interessados no TRC. O Capítulo 3 introduz os principais conceitos da regressão espacial, iniciando pelo diagnóstico da dependência espacial e finalizando com os principais modelos espaciais utilizados. O Capítulo 4 detalha a metodologia utilizada para construção de modelos de regressão espacial, e os resultados dos estudos de caso elaborados a partir do banco de dados do RNTRC (Registro Nacional de Transportadores Rodoviários de Carga) da ANTT (Agência Nacional de Transportes Terrestres) são apresentados no Capítulo 5. Por fim, as principais conclusões bem como as recomendações para estudos futuros fecham o trabalho no Capítulo 6.

2. TRANSPORTE RODOVIÁRIO DE CARGA

2.1. APRESENTAÇÃO

O transporte rodoviário se destaca dos outros modos de transporte por representar cerca de 61,1% da matriz de transportes de carga, contra 20,7% do ferroviário, 13,6% do aquaviário, 4,2% do dutoviário e 0,4% do aéreo, conforme Figura 2.1 (ANTT, 2005). O seu grande mercado, no Brasil, deve-se a inúmeros fatores dentre os quais se destacam a possibilidade de entrega na porta do comprador, rapidez e disponibilidade, além das políticas públicas que favoreceram seu fortalecimento.

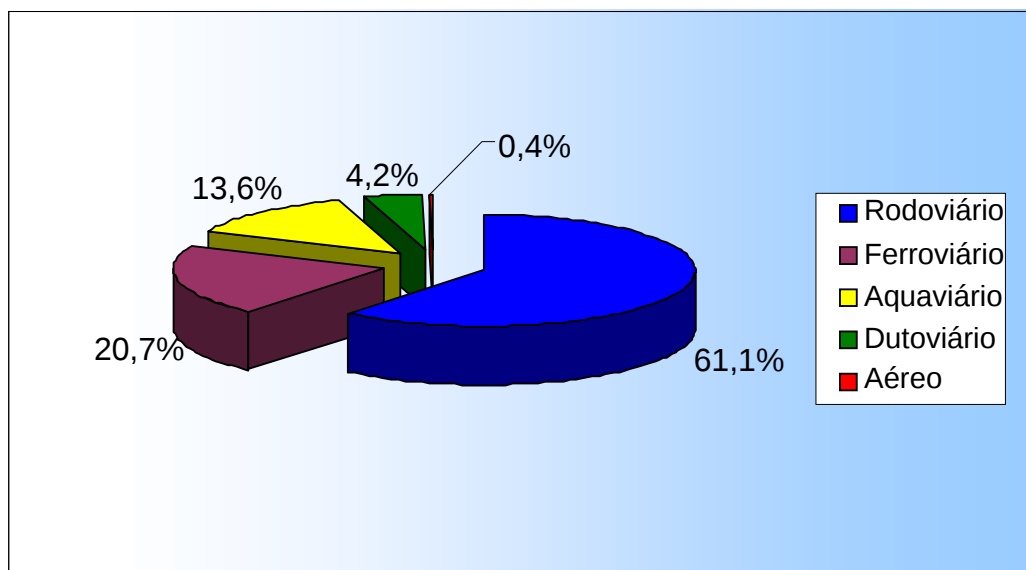


Figura 2.1: Matriz de Transportes Brasileira.
Fonte: ANTT(2005).

O alto custo de manutenção das rodovias, quando comparada com outras infra-estruturas, é um dos principais problemas relacionados a esse modo. No entanto, mesmo com algumas rodovias em condições precárias, o transporte rodoviário é o mais expressivo no transporte de cargas, pois atinge praticamente todo o território nacional.

Ao longo desse capítulo serão apresentadas algumas características desse modo, que servirão de base teórica para o trabalho e para ressaltar a importância do mesmo. Merecem destaque um breve histórico referente à evolução do transporte rodoviário de carga no Brasil e a terminologia dos veículos rodoviários de carga e seus elementos.

2.2. DEFINIÇÃO E CARACTERÍSTICAS

Freitas (2004) conceitua o transporte rodoviário como “... aquele que se realiza em estradas de rodagem, com utilização de veículos como caminhões e carretas”. A opção pelo modo rodoviário, no que diz respeito ao transporte de carga, é um fenômeno quase mundial que se observa desde a década de 50, tendo como base a expansão da indústria automobilística associada aos baixos preços dos combustíveis derivados do petróleo (Schroeder e Castro, 2000). Assim, antes de destacar as principais características desse modo (aspectos positivos e negativos), é importante repassar um pouco de sua história com o objetivo de acompanhar sua evolução.

2.2.1. Evolução do Transporte Rodoviário de Carga no Brasil

O transporte rodoviário está a mais de 100 anos presente no Brasil, acompanhando diretamente o seu desenvolvimento. A predominância desse modo depende da história e do processo de crescimento econômico e social de cada país. China e EUA, por exemplo, que são países com dimensões continentais parecidas com o Brasil, seguiram caminhos diferentes, por isso o modo rodoviário não tem um peso tão grande na matriz de transporte desses países. Nesse sentido, Macohin (2001) apresenta uma retrospectiva do transporte rodoviário de carga no Brasil, indicando os fatos mais relevantes:

- **1902:** registro de uma empresa de transporte rodoviário de cargas (de que se tem notícia), Companhia de União dos Transportes, que se mantém até 1942;
- **1908 – 1909:** registro dos primeiros caminhões utilizados no transporte rodoviário;
- **1914 – 1918:** Primeira Guerra Mundial interrompe a importação de produtos manufaturados para o Brasil; estimula a criação de novas indústrias; inicia-se uma série de ações em favor da rodovia e o setor de transportes cresce e torna-se mais atuante;
- **1918:** final da guerra. A navegação volta a ocupar seu lugar de destaque entre os meios de transporte do país;
- **1919:** cresce o tráfego de automóveis no país. A *Ford Motor Company* decide criar a Ford brasileira;
- **1927:** implanta-se o transporte aéreo; instala-se a *General Motors* do Brasil;

- **Final dos anos 20:** crise na ferrovia e o monopólio ameaçado; cresce o número de veículos automotores fabricados no Brasil; o automóvel e o caminhão são vistos como uma necessidade ao progresso;
- **1929:** 144.776 veículos automotores trafegam pelas ruas e estradas brasileiras;
- **1930:** o governo Vargas incentiva o processo industrial e a economia entra num período de recuperação a partir de 1933;
- **1937:** criação do Departamento Nacional de Estradas de Rodagem (DNER), que em 1944 apresenta o Plano Rodoviário Nacional;
- **1939:** a frota nacional é composta por 190.000 veículos automotores, dos quais um terço eram caminhões; surgem os pioneiros do transporte rodoviário;
- **1943:** Segunda Guerra Mundial interrompe as trocas comerciais entre as nações; o transporte rodoviário chega à beira do colapso, em função das dificuldades de importação de petróleo;
- **1944:** aumento considerável das rodovias que chegam a 277.000 km; o transporte rodoviário de carga inicia sua fase de expansão;
- **Década de 50:** novo modelo de industrialização no governo do Presidente Juscelino Kubitschek; a indústria automobilística assume papel preponderante no processo de industrialização; construção de Brasília; a implantação da indústria automobilística viabiliza a criação do sistema de Transportes Rodoviário de Carga (TRC), em substituição ao sistema ferroviário;
- **Final dos anos 60:** à exceção de Belém e Manaus, todas as capitais brasileiras estavam interligadas por estradas federais;
- **Década de 70:** A estrada virou questão de soberania nacional, sendo que em 1980, o Brasil tinha 47 mil quilômetros de estradas federais pavimentadas;
- **Após segunda metade da década de 90:** com a privatização das ferrovias e a modernização dos portos, efetivamente começou a se esboçar o “fenômeno” da competição ao amplo predomínio exercido pelo transporte rodoviário;

Verifica-se então que a inserção do modo rodoviário no Brasil aconteceu após o declínio do transporte ferroviário e início do processo de industrialização. Mello (2000) comenta que a partir da década de 50 o desenvolvimento acentuou-se, foi implantada a indústria automobilística, a pauta de exportações diversificou-se, com predominância crescente para

os manufaturados, e desde então o PIB cresceu mais de 30 vezes, estimulando também a expansão da malha de transportes.

As indústrias começavam a produzir em alta escala e precisavam rapidamente distribuir sua produção. Nesse ponto, o transporte rodoviário se mostrava o mais atraente, devido à forma de entrega porta a porta e pela malha rodoviária existente que interligava todo o Brasil. Outras características se destacam no transporte rodoviário conforme discutidas a seguir.

2.2.2. Aspectos Positivos e Negativos do Transporte Rodoviário de Carga

Freitas (2004) comenta que, assim como nas demais modalidades, o transporte rodoviário apresenta pontos positivos de elevada consideração e outros negativos que dificultam sua utilização. Para uma melhor visão de seu funcionamento é importante que se faça uma análise destes aspectos positivos e negativos.

2.2.2.1. Pontos Positivos

No caso de países com dimensões continentais como o Brasil, o transporte rodoviário apresenta-se como um dos mais flexíveis e ágeis no acesso às cargas, pois possibilita a interação de diferentes regiões, mesmo as mais remotas. Cabe mencionar que esta praticidade torna-se mais visível no caso de não haver outros modos à disposição (Freitas, 2004).

Outra qualidade de grande valia desta modalidade é a simplicidade de seu funcionamento e a rapidez de sua disponibilidade quando exigida pelo embarcador. A seguir estão enumeradas algumas características que viabilizam a utilização desse modo e que fazem do transporte rodoviário um dos mais utilizados no Brasil:

- agilidade e rapidez na entrega da mercadoria em curtos espaços a percorrer;
- a unidade de carga chega até a mercadoria, enquanto nos outros modos de transporte a mercadoria deve ir ao encontro da unidade de carga;
- vendas que possibilitam a entrega na porta do comprador;
- embalagens com um custo bem menor;

- uma menor movimentação da mercadoria, reduzindo assim os riscos de avarias.

2.2.2.2. Pontos Negativos

Assim como a modalidade em destaque traz vantagens em sua utilização, também possui algumas desvantagens, tais como apresentadas por Freitas (2004):

- o custo de fretamento é mais expressivo que os demais concorrentes com características próximas;
- a capacidade de tração de carga é bastante reduzida;
- os veículos utilizados para tração produzem um elevado grau de poluição ao meio ambiente;
- a malha rodoviária deve estar constantemente em manutenção ou em construção, gerando custos ao erário ou ao contribuinte, visto que, existem estradas privatizadas que cobram pedágio.

Freitas (2004) comenta ainda que o modo rodoviário é o mais adequado para o transporte de mercadorias de alto valor agregado ou perecíveis, em deslocamentos de curta e média distância. Porém, perde muito em competitividade ao transportar produtos agrícolas a granel, devido ao seu baixo valor agregado. Assim, uma forma de reduzir o custo final é utilizar apenas uma fração do caminhão (carga fracionada) e dividir a viagem com outros embarcadores.

2.3. TERMINOLOGIA DOS VEÍCULOS RODOVIÁRIOS DE CARGA

Existem vários tipos de caminhões e carrocerias que são utilizados para diferentes fins, e que também podem ser combinados no intuito de aumentar sua capacidade de carga. O conhecimento das características e da capacidade do veículo auxiliam o embarcador na escolha do mais adequado para o transporte da sua mercadoria, no entanto, isso não é praticado com frequência pelas transportadoras. Dificilmente uma transportadora comprará uma carroceria do tipo “cerealeiro”, por exemplo, para o transporte de cereais porque fora os períodos de colheita, a carroceria ficará parada aguardando a próxima safra. As transportadoras preferem adquirir caminhões e carrocerias que podem ser utilizados para o transporte de diferentes cargas, mesmo sabendo que determinados tipos de veículos não

são os mais adequados para o transporte de alguns produtos. Para o melhor entendimento desse tipo de transporte, seguem as definições e classificações dos veículos rodoviários de carga, bem como suas características e finalidades.

2.3.1. Definições

A ABNT (1997), através da NBR 9762, definiu a classificação dos veículos rodoviários de carga, conforme a Figura 2.2. De uma forma geral, um veículo rodoviário de carga é definido como todo aquele construído especificamente para o transporte geral de cargas, sejam gasosas, líquidas ou sólidas, e que transita em vias de rolamento. Pode ser classificado em quatro tipos: veículo rodoviário de carga incompleto, completo, caminhão-trator ou veículo rodoviário de carga combinado.

Um veículo rodoviário de carga incompleto é todo aquele dependente de complementação para transformá-lo em veículo rodoviário de carga, com ou sem cabina e com permissão para circular em vias públicas. Por sua vez, um veículo rodoviário de carga completo é aquele totalmente construído na fábrica de origem ou acrescido de equipamento veicular, apto a desempenhar os trabalhos de transporte a que se destina. Um caminhão-trator é um veículo rodoviário automotor, com no mínimo quatro rodas, equipado para tracionar veículos rodoviários de carga rebocados. E, finalmente, um veículo rodoviário de carga combinado é aquele constituído de um caminhão automotor tracionando um ou mais rebocados. Todas essas categorias serão detalhadas a seguir.

Veículo Rodoviário De carga	Veículo Rodoviário de Carga Incompleto	Chassi	Implemento Rodoviário	Veículo Rebocado	Equipamento Veicular	Base Equipado Semi-equipado		
						Fixo	Carroceria aberta Carroceria fechada	
							Removível	Contêiner Caçamba estacionária
								Mecanismo Operacional
						Distribuidor De peso	Dolly Eixo veicular auxiliar	
						Dispositivo	Acoplamento Fixação	
						Veículo Rodoviário de Carga Completo	Caminhão Camioneta Motocicleta Motoneta	
							Caminhão-trator	
							Veículo Rodoviário de Carga Combinado	Reboque Múltiplos Semi- reboque

Figura 2.2: Classificação dos Veículos Rodoviários de Carga.
Fonte: ABNT (1997).

2.3.1.1. Veículo Rodoviário de Carga Incompleto

Esse tipo de veículo pode ser classificado em: chassi ou implemento rodoviário. Segundo a ABNT (1997), o chassi é um veículo autopropelido com ou sem cabina, construído de longarinas principais e com equipamento que permita seu deslocamento em vias de rolamento, conforme a Figura 2.3. Já o implemento rodoviário é o veículo rebocado acoplado e/ou equipamento veicular adaptado a chassi de caminhão, permitindo sua funcionalidade. As carrocerias, nesse sentido, nada mais são que implementos rodoviários.



Figura 2.3: Veículo rodoviário de Carga Incompleto - Chassi.

O implemento rodoviário também pode ser de dois tipos: veículo rebocado ou equipamento veicular. O veículo rebocado é aquele que se move tracionado por veículo unidade de tração (caminhão-trator ou veículo rodoviário automotor), através de dispositivos adequados, conforme os tipos descritos pela Tabela 2.1.

Tabela 2.1. Tipos de Veículos Rebocados.

Veículo Rebocado	Definição
Base	necessita da adaptação de um equipamento veicular.
Equipado	totalmente construído na fábrica de origem ou veículo rebocado base com adaptação de um equipamento veicular, especificamente destinado ao transporte de cargas.
Semi-equipado	veículo equipado reforçado que pode receber outro equipamento veicular.

O equipamento veicular é um equipamento específico incorporado a um veículo rodoviário incompleto, construído para complementá-lo, permitindo assim o transporte de cargas. Pode ser de cinco tipos: fixo, removível, mecanismo operacional, distribuidor de peso e

dispositivo. O equipamento veicular fixo é aquele em que não é possível a remoção do chassi ao qual foi adaptado. Pode ser do tipo carroceria aberta ou carroceria fechada.

A carroceria aberta é toda carroçaria sem teto destinada para o transporte de carga que não necessite de cuidados especiais. O Quadro 2.1 apresenta os tipos de carrocerias abertas existentes.

Quadro 2.1. Tipos de Carrocerias Abertas.

	COMUM OU CARGA SECA: utilizada para o transporte de vários tipos de produtos (carga seca), como caixas, enlatados etc. Possui grades laterais, frontal e traseira que facilitam a descarga.
	GRANELEIRA: utilizada para o transporte de grãos a granel. Possui grades laterais, frontal e traseira que facilitam a descarga.
	CANAVIEIRA: carroceria específica para o transporte de cana-de-açúcar picada. Pode ser com laterais fixas ou escamoteáveis, dependendo do sistema de descarga instalado na usina.
	CANAVIEIRA: carroceria específica para o transporte de cana-de-açúcar inteira. Pode ser com laterais fixas ou escamoteáveis, dependendo do sistema de descarga instalado na usina.
	GAIOLA: utilizada para transporte de animais vivos.
	GAIOLA: utilizada para transportar botijões de gás, tambores vazios, etc.

Fonte: IDELT (Instituto de Desenvolvimento, Logística, Transporte e Meio Ambiente) com adaptações.

Quadro 2.1. (Continuação)

	PRANCHA ou CARREGA TUDO: normalmente utilizada no transporte de produtos super dimensionados e de difícil movimentação para o interior da carroceria. Exemplo: tratores, turbinas, etc.
	BUG OU PORTA CONTÊINER: utilizada para transportar contêiner.
	CEREALEIRO: carroceria própria para transporte de cereais.
	CEREALEIRO: carroceria própria para transporte de cereais. Este modelo pode levar dois tipos diferentes de cereais.
	FLORESTAL: utilizada para transportar toras nativas.
	CEGONHA: carroceria própria para o transporte de veículos.
	TRANSPORTE DE BEBIDAS: utilizada para o transporte de engradados específicos para bebidas, normalmente cobertos com lona para proteção.

Fonte: IDELT (Instituto de Desenvolvimento, Logística, Transporte e Meio Ambiente) com adaptações.

A carroceria fechada é toda carroceria com teto rígido, destinada ao transporte de cargas que necessitam ser protegidas contra intempéries, cargas onde a temperatura é influência para sua perecibilidade ou cargas que necessitem de proteção extra para seu transporte. O Quadro 2.2 apresenta os tipos de carrocerias fechadas existentes.

Quadro 2.2. Tipos de Carrocerias Fechadas.

	BAÚ ou FURGÃO SIMPLES: utilizada com as mesmas finalidades da carroceria aberta comum. Porém, o baú oferece maior proteção a carga, por se tratar de uma carroceria fechada.
	BAÚ LONADO ou SIDER: possui as mesmas vantagens da carroceria do tipo baú, além da vantagem de possibilitar o carregamento e descarregamento pelas laterais, com a utilização de empilhadeiras, diminuindo o tempo de carregamento e descarregamento.
	BARRIGUDA OU REBAIXADO: por ter o seu assoalho rebaixado possibilita o transporte de mercadorias que ocupam um grande espaço na carroceria, com um peso menor que os outros. Ex: blocos de isopor, peças de plástico ou de alumínio, etc.
	FRIGORIFICADA ou FURGÃO FRIGORÍFICO: utilizada para transporte de cargas perecíveis. Possui um freezer para manter a temperatura adequada à conservação dos produtos. Também conhecida como câmara fria.
	FURGÃO ISOTÉRMICO: destinada ao transporte de mercadorias que não podem ter alterações de temperatura. Possui revestimento interno total de material isolante, para propiciar manutenção da temperatura.
	FURGÃO BLINDADO: destinada ao transporte de valores, explosivos ou armamentos. É dotada de blindagem metálica, fechada totalmente, suprida de visores e aberturas de ventilação para controle de guarnição interna.
	TRANSPORTE DE BEBIDAS: utilizada para o transporte de bebidas. Especialmente construída para o uso de paletes que propiciam rapidez na carga ou descarga. As laterais são fechadas através de portas de correr para evitar a violação da carga. Normalmente utilizada para distribuição.

Fonte: IDELT (Instituto de Desenvolvimento, Logística, Transporte e Meio Ambiente) com adaptações.

Quadro 2.2. (Continuação)

	TRANSPORTE DE BEBIDAS: utilizada para o transporte de bebidas. Especialmente construída para o uso de paletes que propiciam rapidez na carga ou descarga. As laterais são fechadas através de portas de correr para evitar a violação da carga. Normalmente utilizada para distribuição. Este modelo aproveita melhor o espaço entre as rodas.
	TANQUE: utilizada para o transporte de cargas líquidas ou gasosas a granel.
	SILO: destinada ao transporte de materiais pulverulentos ou grãos, como cimento, cal, farinha de trigo, etc

Fonte: IDELT (Instituto de Desenvolvimento, Logística, Transporte e Meio Ambiente) com adaptações.

O equipamento veicular removível é aquele que pode ser removido do chassi ao qual foi adaptado. Basicamente existem três tipos: contentor ou contêiner, caçamba estacionária para detritos sólidos e caçamba estacionária para fins múltiplos.

O equipamento veicular mecanismo operacional é aquele destinado à prestação de serviços e transportes específicos, operado desde o chassi de um veículo rodoviário automotor até um veículo rebocado base. O Quadro 2.3 apresenta os tipos de mecanismo operacional existentes.

Quadro 2.3. Tipos de Mecanismo Operacional.

	CAÇAMBA BASCULANTE: muito usada no transporte de produtos sólidos a granel, por exemplo, grãos, terra, entulhos, areia, pedras etc. É constituído de compartimento aberto com um sistema para basculamento da caixa no sentido lateral ou traseiro, para rápido escoamento.
	GUINCHO: destinado a içar, puxar ou sustentar uma carga por intermédio de cabo de aço, barra ou dispositivo específico. Possui acionamento hidráulico, elétrico, mecânico ou uma composição destes.

Quadro 2.3. (Continuação)

	POLIGUINDASTE: utilizado para puxar e transportar equipamento veicular removível como caçamba estacionária.
	BETONEIRA: destinado a transportar os componentes de concreto e misturá-los continuamente durante o trajeto do transporte.
	ESPARGIDOR: destinado a transportar e espargir material líquido, a frio ou a quente, em pavimentos.
	COLETOR COMPACTADOR DE RESÍDUOS SÓLIDOS: destinado à coleta, compactação, transporte e descarga de resíduos sólidos.
	GUINDASTE VEICULAR: utilizado pra rebocar veículos através de içamento.
	GUINDASTE VEICULAR: utilizado para rebocar veículos. Neste modelo o veículo é transportado sobre o mecanismo.
	EQUIPAMENTO DE PREVENÇÃO, PROTEÇÃO E EXTINÇÃO DE INCÊNDIOS: destinado a prevenir, proteger e extinguir incêndios.
	EQUIPAMENTO DE RESGATE, SALVAMENTO E OUTRAS EMERGÊNCIAS: destinado às atividades de resgate, salvamento e outras emergências em geral.

Quadro 2.3. (Continuação)

	EQUIPAMENTO DE INSPEÇÃO DE REDES AÉREAS: utilizado especificamente para elevação de uma ou mais pessoas à altura de redes elétricas e de telecomunicações, para execução ou reparos nas linhas destas redes.
	PLATAFORMA ELEVATÓRIA: plataforma aplicada na parte traseira ou, às vezes, na lateral do compartimento de carga, servindo para baixar cargas ao solo ou elevá-las ao nível do assoalho do veículo.

O equipamento veicular distribuidor de peso é aquele destinado a aumentar a capacidade de carga, visando a adequada distribuição de peso sobre a superfície do pavimento. O Quadro 2.4 apresenta os tipos de distribuidor de peso existentes.

Quadro 2.4. Tipos de Distribuidor de Peso.

	EIXO VEICULAR AUXILIAR: eixo veicular adaptado em veículo rodoviário automotor de dois eixos veiculares, mediante reforço no chassi, com a finalidade de propiciar elevação de sua capacidade de carga.
	DOLLY: veículo rebocado especial, apoiado ou não sobre a unidade de tração e destinado a aliviar a carga e distribuí-la sobre o pavimento, facilitando o deslocamento do conjunto tracionado.

O equipamento veicular dispositivo de fixação ou acoplamento mecânico é aquele incorporado a um chassi de um veículo rodoviário automotor para tracionar rebocados ou veículos rodoviários rebocados. A quinta roda (Figura 2.4) é o principal dispositivo de acoplamento mecânico, e serve para unir a unidade de tração a um veículo rebocado ou para combinação de veículos. Outro dispositivo de acoplamento é o engate.

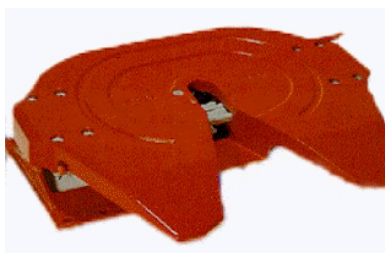


Figura 2.4: Dispositivo de Acoplamento Mecânico – Quinta Roda.

2.3.1.2. Veículo Rodoviário de Carga Completo

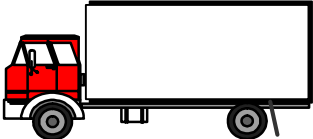
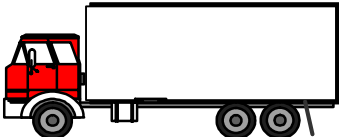
Um veículo rodoviário de carga completo é aquele totalmente construído na fábrica de origem ou acrescido de equipamento veicular, apto a desempenhar os trabalhos de transporte a que se destina. O Quadro 2.5 apresenta os tipos de veículo rodoviário de carga completo.

Quadro 2.5. Tipos de Veículo Rodoviário de Carga Completo.

	CAMINHÃO ou CAMINHÃO SIMPLES: veículo de carga automotor, com no mínimo quatro rodas e com cabina, equipado com equipamento veicular, carroceria e destinado ao transporte de carga com capacidade de carga útil superior a 1.500 kg. A carroceria pode ser fechada ou aberta.
	CAMIONETA DE CARGA: veículo rodoviário automotor, como no mínimo quatro rodas, equipado com carroceria destinada ao transporte de carga, com capacidade de carga útil de até 1.500 kg. A carroceria pode ser aberta ou fechada.
	MOTOCICLETA DE CARGA: veículo rodoviário automotor de duas rodas, dirigido por condutor em posição montada, provido de carro lateral (neste caso de três rodas) ou caçamba para carga. Também existe uma versão com menos potência denominada MOTONETA DE CARGA.

Os caminhões simples também possuem algumas variações quanto ao chassi, sendo classificados como leves ($\frac{3}{4}$ três quartos), médios (toco) ou semipesados (truck) conforme o Quadro 2.6.

Quadro 2.6. Tipos de caminhões simples classificados segundo o chassi.

Desenho Ilustrativo	Descrição
	$\frac{3}{4}$ TRÊS QUARTOS: tem capacidade para 3 toneladas de carga útil.
	TOCO: é o nome dado ao caminhão que possui apenas um eixo traseiro, com rodagem simples, isto é, dois pneus por eixo, ou de rodagem dupla, ou seja, quatro pneus por eixo. Tem capacidade para até 6 toneladas.
	TRUCK: é o nome dado ao caminhão que possui dois eixos traseiros, sendo um deles o de tração motriz, e rodagem dupla. O eixo de tração motriz é aquele que recebe a força do motor e a transmite as rodas. Tem capacidade entre 10 e 14 toneladas.

Fonte: IDELT (Instituto de Desenvolvimento, Logística, Transporte e Meio Ambiente) com adaptações.

2.3.1.3. Caminhão-trator

Um caminhão-trator (Figura 2.5) é um veículo rodoviário automotor, com no mínimo quatro rodas, equipado para tracionar veículos rodoviários de carga rebocados. Dependendo de sua capacidade de tração, pode ser classificado como pesado ou extrapesado.



Figura 2.5: Caminhão-trator.

2.3.1.4. Veículo Rodoviário de Carga Combinado

Um veículo rodoviário de carga combinado é qualquer veículo constituído de dois ou mais outros veículos rodoviários de carga. A Carreta (Figura 2.6), que é um exemplo de veículo rodoviário de carga combinado, é um veículo formado por um caminhão trator (ou usualmente chamado de cavalo mecânico) que puxa um semi-reboque. O número de eixos é variado e sua capacidade de carga depende da configuração dos eixos. O semi-reboque

(Figura 2.7), por sua vez, é um veículo não-tracionado que se movimenta articulado e apoiado no cavalo mecânico. Já o reboque (Figura 2.8) é um veículo que se movimenta puxado por outro veículo, e não apoiado como o semi-reboque.

Caminhão-trator

semi-reboque

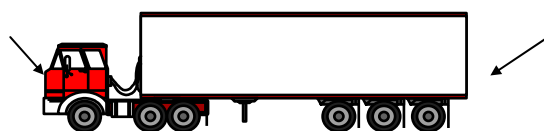


Figura 2.6: Desenho esquemático de uma carreta.



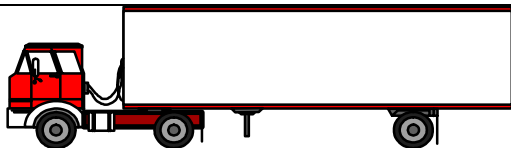
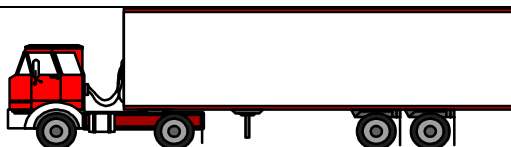
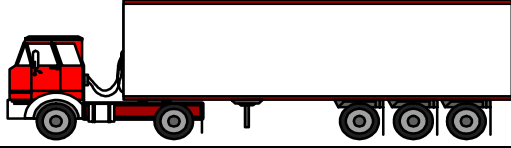
Figura 2.7: Semi-reboque.



Figura 2.8: Reboque.

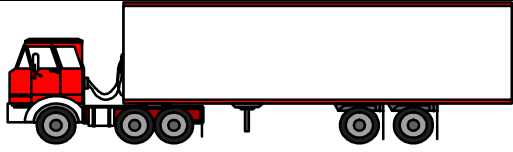
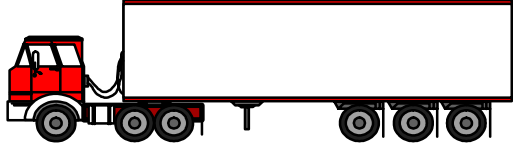
O Quadro 2.7 apresenta as configurações possíveis para carretas, como também o Peso Bruto Total Combinado (PBTC) de cada uma. O PBTC é o peso máximo transmitido ao pavimento pela combinação de um caminhão-trator mais seu semi-reboque ou reboque ou ambos. Mais informações podem ser obtidas em DNIT (2006).

Quadro 2.7. Configurações de carretas a capacidade de carga.

<i>CARRETA</i>	<i>PESO BRUTO MÁXIMO</i>
	Carreta com 3 eixos, sendo 2 para o trator e 1 para o semi-reboque PBTC = 26 T
	Carreta com 4 eixos, sendo 2 para o trator e 2 para o semi-reboque PBTC = 40 T
	Carreta com 5 eixos, sendo 2 para o trator e 3 para o semi-reboque PBTC = 41,5 T

Fonte: IDELT (Instituto de Desenvolvimento, Logística, Transporte e Meio Ambiente) com adaptações.

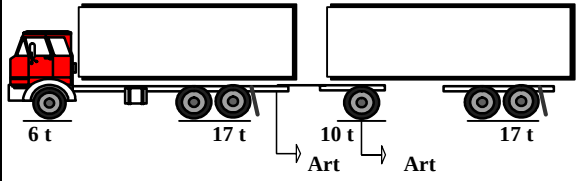
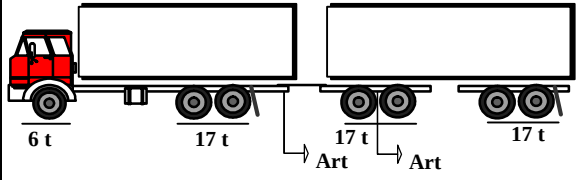
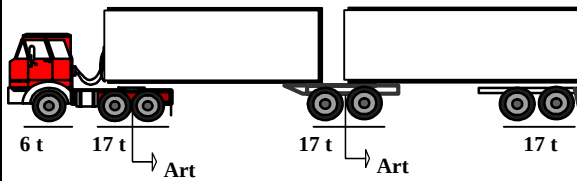
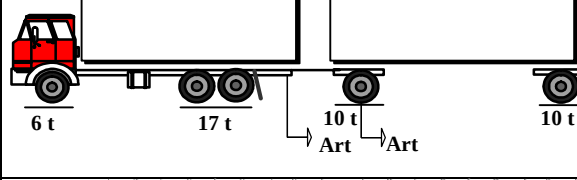
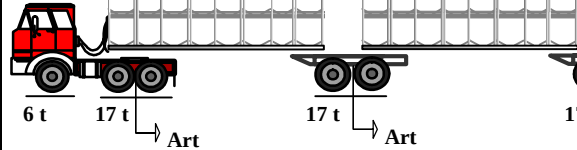
Quadro 2.7. (Continuação)

<i>CARRETA</i>	<i>PESO BRUTO MÁXIMO</i>
	Carreta com 5 eixos, sendo 3 para o trator e 2 para o semi-reboque PBTC = 40 T
	Carreta com 6 eixos, sendo 3 para o trator e 3 para o semi-reboque PBTC = 45 T

Fonte: IDELT (Instituto de Desenvolvimento, Logística, Transporte e Meio Ambiente) com adaptações.

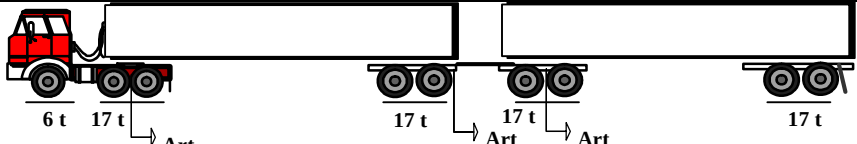
A resolução CONTRAN 68/98 estabelece outras combinações de veículos, que são apresentadas no Quadro 2.8.

Quadro 2.8. Combinações de veículos segundo CONTRAN 68/98.

	Romeu e Julieta comprimento 19,80 m Duas articulações PBTC = 50 T
	Romeu e Julieta comprimento 19,80 m Duas articulações PBTC = 57 T
	Bi Trem Articulado Comprimento 19,80 m Duas articulações PBTC = 57 T
	Treminhão Comprimento 30 m Quatro articulações PBTC = 63 T
	Tri Trem Comprimento 30 m Três Articulações PBTC = 74 T

Fonte: IDELT (Instituto de Desenvolvimento, Logística, Transporte e Meio Ambiente) com adaptações.

Quadro 2.8. (Continuação)

	Rodotrem Comprimento 30 m Três Articulações PBTC = 74 T
--	---

Fonte: IDELT (Instituto de Desenvolvimento, Logística, Transporte e Meio Ambiente) com adaptações.

Essa terminologia apresentada pela ABNT (1997) é muito técnica e por isso não é muito utilizada pelos profissionais da área de transportes. Em geral, os veículos rodoviários de carga são divididos em “autotracionados” e “rebocáveis”. Os autotracionados compreendem os caminhões simples e os caminhões tratores enquanto os rebocáveis compreendem os reboques e semi-reboques. A ANFAVEA (Associação Nacional dos fabricantes de Veículos Automotores), classifica ainda os veículos quanto ao peso máximo que transmitem ao pavimento:

- Semileves → $3,5t < PBT < 6t$
- Leves → $6t < PBT < 10t$
- Médios → $10t < PBT < 15t$
- Semipesados
 - Caminhão-Chassi → $PBT > 15t$ e $CMT < 45t$
 - Caminhão-Trator → $PBT > 15t$ e $PBTC < 40t$
- Pesados
 - Caminhão-Chassi → $PBT > 15t$ e $CMT > 45t$
 - Caminhão-Trator → $PBT > 15t$ e $PBTC > 40t$

Onde PBT significa Peso Bruto Total e CMT significa Capacidade Máxima de Tração.

2.3.2. Tipos de Cargas

Existem basicamente dois tipos de cargas: as cargas a granel e as cargas em geral. Cada um desses tipos possui uma categoria denominada “especial”, que devido suas características, necessitam de caminhões especialmente preparados para seu transporte. A Tabela 2.1 apresenta os tipos de cargas com uma breve descrição sobre cada um.

Tabela 2.2. Tipos de Cargas.

CARGA	TIPO DE CARGA	DESCRIÇÃO
A GRANEL	Granel Sólido	Cargas soltas em pó ou grãos como: cereais, cimento, areia, entulho etc. Transportadas geralmente em carrocerias abertas ou do tipo basculante, devidamente protegidas com lonas e encerados, conforme a necessidade.
	Granel Líquido	Utilizam-se carrocerias tanque específicas, de acordo com o produto a ser transportado.
GERAL	Embaladas	Transportadas em caixas, fardos, feixes, sacarias etc, geralmente em carrocerias abertas, protegidas com lonas e encerados, ou do tipo baú, conforme a necessidade.
	Diversos	São cargas de diferentes tipos e são transportadas sem qualquer tipo de embalagem. Ex. Canos, telhas, pneus, etc.
	Móveis e outros	Materiais que não podem sofrer umidade. Transportados em caminhões do tipo baú.
ESPECIAL	Produtos perigosos	São todos aqueles que por suas características possam trazer riscos à saúde, à vida ou ao meio ambiente. Utilizam-se vários tipos de carrocerias (fechadas, abertas, tanque etc.) de acordo com o tipo do produto a ser transportado, com a devida sinalização de segurança, como placas, símbolos, rótulos, etc.
	Produtos Perecíveis	São utilizados baús frigoríficos para o seu transporte.
	Veículos	Cargas que, por suas características, são transportadas em caminhões especialmente preparados para esse fim.
	Valores	Transportados em carrocerias fechadas e blindadas, de acordo com a segurança exigida.

Fonte: IDELT (Instituto de Desenvolvimento, Logística, Transporte e Meio Ambiente).

2.4. PARTICIPANTES DO TRANSPORTE RODOVIÁRIO DE CARGA

O conhecimento acerca do ambiente de transporte é o primeiro ponto para compreender as tomadas de decisões em transportes. Na maioria das transações comerciais, se não em todas, existem apenas dois participantes: compradores e vendedores. Embora o envolvimento do governo seja necessário em algumas transações, ele não é comum na maioria delas (Bowersox e Closs, 2001).

No entanto, as transações de transporte normalmente são influenciadas por cinco componentes: o embarcador (ponto de origem), o destinatário (ponto de destino ou receptor), a transportadora, o governo e o público.

Embarcadores e Destinatários: Ambos têm como objetivo movimentar as mercadorias da origem até o destino num determinado tempo e com o menor custo possível. Os serviços incluem especificação dos tempos de coleta, de entrega e de trânsito previsível, perda e avaria zero, bem como faturamento e troca de informações precisas e em tempo hábil.

Transportadoras: por serem intermediárias, têm como objetivo aumentar suas receitas brutas mediante a transação e minimizar os custos necessários para concluí-la. A transportadora sempre cobra a taxa mais alta aceitável pelo embarcador (ou destinatário) e sempre procura minimizar os custos de mão-de-obra, combustível e desgaste dos veículos. Para atingir esse objetivo, ela tenta obter flexibilidade nos tempos de entrega e coleta para permitir movimentações mais econômicas.

Governo: tem grande interesse na transação em virtude do impacto do transporte na economia, por isso tem como objetivo manter um ambiente de transporte estável e eficiente a fim de sustentar o crescimento econômico. Uma economia estável e eficiente, por sua vez, leva as transportadoras e/ou operadoras a oferecer serviços competitivos e ao mesmo tempo operar de forma lucrativa.

Público: é quem determina a necessidade de transporte. Preocupa-se com o acesso ao serviço, segurança, meio ambiente e preço. Embora o fator preço seja principal, os padrões de segurança e meio ambiente também merecem consideração, pois os efeitos da poluição do ar e os derramamentos de óleo, decorrentes do transporte, atingem diretamente toda a população.

Assim, toda a cadeia de transportes é influenciada por esses cinco atores, independente do modo de transporte, e, em termos gerais, existem dois interessados em todo esse processo: o poder público (englobando também o público em geral) e o mercado. Tratando mais especificamente do transporte rodoviário de carga, esse mercado é composto pelos usuários (embarcador, destinatário e/ou clientes), operadores e/ou transportadores e pelas montadoras. As montadoras, nesse contexto, compreendem os fabricantes de caminhões, carrocerias ou qualquer outro tipo de equipamento/acessório como pneus, lonas etc. A

Figura 2.9 apresenta um desenho esquemático dos principais interessados no transporte rodoviário de carga.

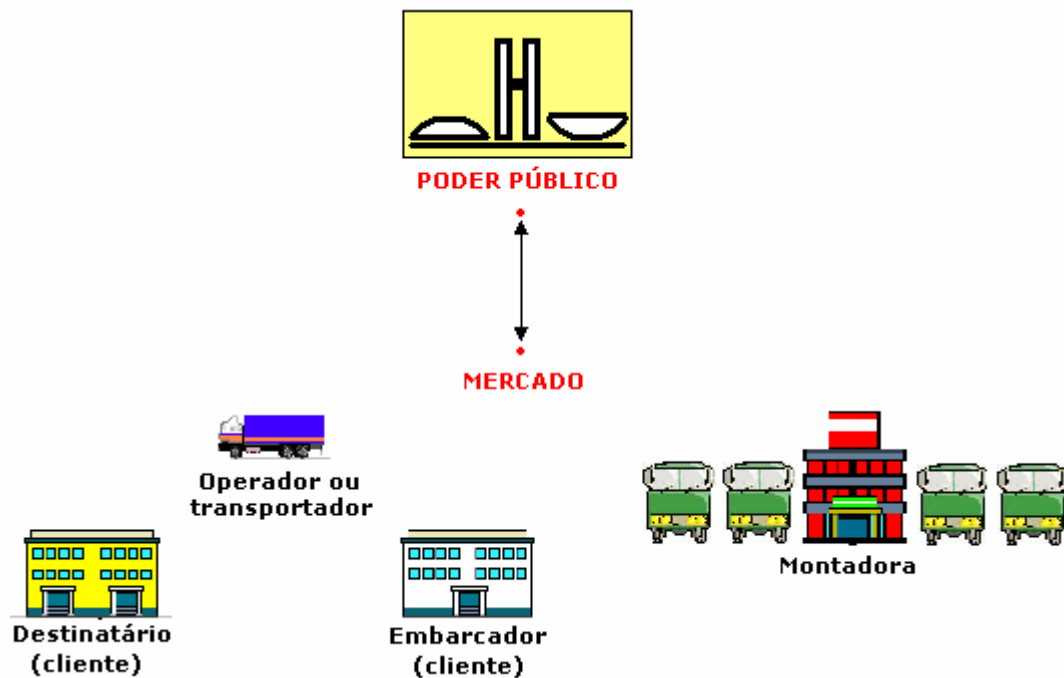


Figura 2.9: Interessados no Transporte Rodoviário de Carga.
Fonte: Adaptado de Alano (2003).

2.5. TÓPICOS CONCLUSIVOS

Este capítulo teve como objetivo ressaltar a importância do transporte rodoviário de carga bem como apresentar os tipos de veículos utilizados no transporte rodoviário de carga. Assim, pelo que foi visto, merecem destaque os seguintes tópicos conclusivos:

- Estudar, entender e propor novas soluções, como o projeto de novas carrocerias destinadas a produtos específicos, para o modo rodoviário é relevante, pois o mesmo representa mais de 60% da matriz de transportes brasileira;
- A história do transporte rodoviário de carga no Brasil seguiu caminhos diferentes dos trilhados por outros países com dimensões continentais equivalentes. Esse modo acompanhou diretamente o desenvolvimento do Brasil e teve sua ascendência após a crise do modo ferroviário;
- As principais características positivas do transporte rodoviário de carga são a agilidade e a possibilidade de entrega porta a porta, além de uma menor

movimentação da mercadoria. Como pontos negativos merecem destaque o alto índice de poluição e a capacidade de tração reduzida, quando comparada a outros modos.

- Existem diversos implementos rodoviários adequados a cada necessidade. No entanto, nem todos esses tipos são utilizados por não produzirem os lucros esperados às transportadoras. A consequência imediata disso é um maior tempo para descarga e a perda de parte da mercadoria nessa operação.
- As transações em transporte são influenciadas normalmente por cinco componentes: o embarcador, o destinatário, a transportadora, o governo e o público. O poder público e o mercado, composto pelos usuários, transportadores e montadoras são os principais interessados nessa cadeia, e o conhecimento acerca desse ambiente auxilia ambos na definição de políticas para o setor.
- Não é possível estudar o transporte rodoviário de carga sem fazer a distinção do tipo de veículo rodoviário, pois como foi visto, existem diferentes tipos desses para diferentes fins. Como exemplo, acredita-se que existe uma maior concentração de carrocerias do tipo *Bug* ou porta contêiner onde estão localizados os portos, assim como uma maior concentração de carrocerias do tipo graneleira onde existe produção de grãos. Dessa forma, verifica-se que o espaço geográfico é fator principal para a distribuição dos veículos rodoviários de carga, o qual merece um tratamento especial no que se refere à modelagem, como será visto no próximo capítulo.

3. REGRESSÃO ESPACIAL

3.1. APRESENTAÇÃO

Modelos de regressão são ferramentas estatísticas que utilizam o relacionamento existente entre duas ou mais variáveis, de maneira que uma delas possa ser explicada pelas demais. No entanto, na situação de dados espaciais, quando está presente a autocorrelação espacial (Seção 3.5), as estimativas do modelo devem incorporar essa estrutura espacial, uma vez que a dependência entre as observações altera o poder explicativo do modelo (Câmara *et al.*, 2002).

Segundo Câmara *et al.* (2002), dois objetivos são pretendidos ao se fazer uma análise de regressão: (a) ter um menor erro entre os valores preditos pelo modelo e os valores observados; b) encontrar variáveis independentes que contribuam significativamente para explicar a variável dependente. Para tanto, supõe-se que as observações são não correlacionadas e os erros i.i.d. (independentes e identicamente distribuídos) seguindo uma distribuição normal com média zero e variância constante.

Entretanto, no caso de dados espaciais, onde está presente a dependência espacial, é pouco provável que o pressuposto padrão de observações não correlacionadas seja verdadeiro. 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 (Câmara *et al.*, 2002).

O presente capítulo tem como objetivo introduzir conceitos que sirvam de base para o estudo da regressão espacial. Antes de introduzir os modelos espaciais é imprescindível que se faça, primeiro, uma análise sobre aspectos espaciais, como a definição da matriz de proximidades e de alguns indicadores de dependência espacial. Pois, caso não haja uma estrutura espacial presente nos dados, os modelos de regressão espacial irão gerar os mesmos resultados de um modelo de regressão convencional.

Assim, esse capítulo foi dividido em 5 partes a fim de criar uma seqüência lógica de raciocínio. A primeira parte se refere à matriz de proximidades, que é a responsável pela estrutura espacial. A segunda introduz os principais testes utilizados para identificação de

dependência espacial. A terceira parte apresenta formas gráficas que auxiliam a análise. Já na quarta parte são estudados os principais modelos de regressão espacial, e por fim a quinta parte define o “espaço” dentro do contexto da regressão espacial.

3.2. MATRIZ DE PROXIMIDADES

A matriz de proximidade espacial ou usualmente denominada “matriz W” de dimensão $n \times n$ é a responsável pela estrutura espacial. Assunção (2003) comenta que cada elemento w_{ij} da matriz W representa o peso ou o grau de conectividade ou de proximidade espacial entre as áreas i e j e, por definição, a diagonal dessa matriz é nula: $w_{ii} = 0$ para todo $i = 1, 2, \dots, n$. Isso porque a área i não sofre influência dela mesma ou porque está distante zero quilômetros (ou qualquer outra unidade de medida) de si própria. A escolha dos elementos w_{ij} é arbitrária e depende do problema específico.

Assunção (2003) apresenta algumas escolhas mais comuns para a matriz W que são apresentadas a seguir:

- 1) $w_{ij} = 1$, se as áreas i e j compartilham fronteiras e $w_{ij} = 0$, caso contrário. Observe que $w_{ij} = w_{ji}$, fazendo com que $W = W^t$, ou seja, W é uma matriz simétrica. Isso quer dizer que se i é vizinho de j , então j é vizinho de i .
- 2) $w_{ij} = 1$, se o centróide da área i está distante a menos de “x” quilômetros da área j e $w_{ij} = 0$, caso contrário. O valor de “x” é escolhido de acordo com a problemática estudada. Neste caso a matriz W também é simétrica.
- 3) $w_{ij} = 1$, se o centróide da área j é um dos k centróides mais próximos daquela área i e $w_{ij} = 0$, caso contrário. O k pode assumir qualquer valor inteiro dependendo do caso. Note que nesse caso a matriz W não é simétrica, pois não necessariamente os k vizinhos mais próximos de i serão os mesmos para j , a não ser para $k = 1$.

Até agora foram tratados apenas os casos binários 0 e 1, indicando proximidade geográfica ou não. Pode ser interessante especificar graus intermediários de vizinhança, ou seja, quantificar quão próxima a área i está de j .

4) Seja d_{ij} a distância entre os centróides de i e j . Fazendo $w_{ij} = 1/(1 + d_{ij})$, as áreas extremamente próximas, com $d_{ij} \approx 0$, teriam assim $w_{ij} \approx 1$, e à medida que estas áreas se afastam, teriam $w_{ij} \approx 0$ (lê-se: tenderia a 0), dando a mesma idéia dos indicadores binários 0 e 1. Note que nesse caso, deve-se forçar o caso $w_{ii} = 0$, pois como a distância entre i e i é zero, faria $w_{ij} = 1$.

5) Seja l_{ij} o comprimento da fronteira comum entre as áreas i e j e seja l_i o perímetro de i . Então $w_{ij} = l_{ij} / l_i$ para $i \neq j$ e $w_{ij} = 0$ para $i = j$. Esse modelo leva em consideração o tamanho da fronteira entre a área i e j . Essa forma é bastante útil quando se deseja discriminar o tamanho das áreas.

Em outras situações, pode ser útil utilizar a matriz W de uma forma que não considere as posições e os limites geográficos das áreas. Assunção (2003) fornece um exemplo:

Um ponto **A** pode estar equidistante dos pontos **B** e **C** mas **A** pode estar ligado a **B** por uma estrada de asfalto, de 2 pistas de cada lado, enquanto que pode estar ligada a **C** por uma estrada de terra em más condições de tráfego. Assim as trocas comerciais e o trânsito de pessoas entre **A** e **B** são mais vizinhas entre si que **A** e **C**.

6) Uma forma de operacionalizar isso é fazer w_{ij} da mesma forma que a escolha (4), apenas trocando d_{ij} pelo tempo necessário para ir de i até j em um determinado meio de transporte.

7) Outra forma seria fazer w_{ij} como o valor ou a quantidade de trocas comerciais entre as áreas i e j . Com isso, áreas distantes geograficamente mas com trocas comerciais significativas teriam mais peso na matriz W do que áreas próximas geograficamente mas com poucas trocas comerciais.

Toda matriz W pode ser transformada em uma matriz W^* simétrica definindo $w_{ij}^* = (w_{ij} + w_{ji}) / 2$. Entretanto, costuma-se padronizar as linhas da matriz W , criando uma nova matriz assimétrica, a fim de facilitar a derivação de fórmulas e as propriedades estatísticas envolvidas. A padronização consiste em fazer com que a soma da linha i seja igual a 1. Isso pode ser feito definindo:

$$w_{ij}^* = w_{ij} / w_{i.} \quad (3.1)$$

onde $w_{i.} = \sum_{j=1}^n w_{ij}$.

Um exemplo de construção da matriz W utilizando os indicadores binários 0 e 1 é apresentado na Figura 3.1.

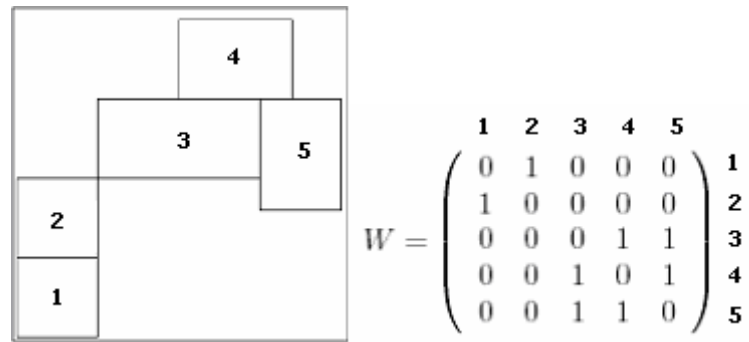


Figura 3.1: Exemplo de Criação da Matriz de Proximidade Espacial.

Fonte: LeSage (1999a).

A matriz W , por si própria, não fornece informações que indiquem a presença de dependência espacial. Serve apenas para indicar a estrutura espacial existente. Para identificar e quantificar tal dependência são necessários testes específicos como o I de Moran e C de Geary que serão estudados a seguir.

3.3. TESTES DE AUTOCORRELAÇÃO ESPACIAL

O índice r conhecido como “coeficiente de correlação de Pearson” mede a correlação entre duas variáveis x e y (Bussab e Morettin, 2004) e é definido por:

$$r = \frac{1}{n} \sum_{i=1}^n \left(\frac{y_i - \bar{y}}{s_y} \right) \left(\frac{x_i - \bar{x}}{s_x} \right) \quad (3.2)$$

onde:

$\bar{y}$ = média de y ;

$\bar{x}$ = média de x ;

s_y = desvio padrão de y ;

s_x = desvio padrão de x ;

O valor de r varia entre -1 e +1, sendo que -1 indica correlação negativa máxima e +1 correlação positiva máxima. Na ausência de correlação o índice r é igual a 0.

Os testes de autocorrelação espacial têm o mesmo objetivo do coeficiente de correlação de Pearson, introduzindo apenas em sua estrutura a informação referente ao fator “espaço”. Segundo Câmara *et al.* (2002), medem o nível de interdependência geográfica entre os valores da variável, bem como a natureza e a força deste relacionamento.

Existem duas formas para medir a dependência espacial: através de indicadores globais que capturam tal dependência em apenas um fator, e através de indicadores locais que verificam a dependência espacial em cada área estudada. Ambos serão discutidos com mais detalhes a seguir.

3.3.1. Índices Globais de Associação Espacial

O índice I de Moran, introduzido pelo estatístico australiano P.A.P. Moran (1950), é o mais antigo e mais utilizado teste para identificação de autocorrelação espacial global. Em transportes, já foi utilizado em diferentes aplicações, conforme Teixeira (2003), Krempi (2004) e Lopes (2005). Foi desenvolvido a partir da mesma idéia do coeficiente de correlação de Pearson. Difere deste por medir a correlação entre os pares y_i e y_j e por ser ponderado pela proximidade espacial.

Seja uma matriz W binária e simétrica, ou seja, w_{ij} assume os valores 0 ou 1 e $w_{ij} = w_{ji}$.

Assim como r , o índice I considera o desvio padronizado $(y_i - \bar{y})/s_y$ mas o relaciona com o desvio padronizado da mesma variável y em uma área j vizinha a i , ou seja,

$$\left(\frac{y_i - \bar{y}}{s_y} \right) \left(\frac{y_j - \bar{y}}{s_y} \right) \quad (3.3)$$

A idéia é verificar se, quando o desvio de uma área é grande, os desvios de suas áreas vizinhas também tendem a ser grandes (a mesma idéia vale para desvios pequenos). É claro que o que interessa é o comportamento médio desse pares. Assim, fazendo a média da Equação (3.3) e, após algumas operações matemáticas chega-se à Equação (3.4):

$$I = \frac{1}{\sum_{i \neq j} w_{ij}} \sum_{i \neq j} w_{ij} \left(\frac{y_i - \bar{y}}{s_y} \right) \left(\frac{y_j - \bar{y}}{s_y} \right) \quad (3.4)$$

onde,

y_i = valor da variável y na região i ;

y_j = valor da variável y na região j ;

$\bar{y}$ = média de y ;

w_{ij} = elemento ij da matriz de proximidade espacial;

s_y = desvio padrão de y ;

n = número de observações;

Como o coeficiente de correlação de Pearson, o índice I varia de -1 a +1, onde valores próximos de 0 indicam ausência de autocorrelação espacial, próximos de -1 autocorrelação negativa e próximos de +1 autocorrelação positiva.

Outro índice global utilizado para medir a dependência espacial é o C de Geary. Sua estrutura é parecida com o índice I de Moran, porém o numerador apresenta a diferença entre um ponto na área i e um ponto na área j . O índice C varia entre 0 e 2, onde valores próximos de 1 indicam ausência de autocorrelação espacial, próximos de 0 indicam autocorrelação positiva e próximos de 2 autocorrelação negativa. O coeficiente de Geary é mais indicado quando existe uma pequena quantidade de vizinhanças (Lembo, 2005).

$$C = \frac{n-1}{2} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - y_j)^2}{\left(\sum_{i=1}^n (y_i - \bar{y})^2 \right) \cdot \left(\sum_{i=1}^n \sum_{j=1}^n w_{ij} \right)} \quad (3.5)$$

onde,

y_i = valor da variável y na região i ;

y_j = valor da variável y na região j ;

$\bar{y}$ = média de y ;

w_{ij} = elemento ij da matriz de proximidade espacial;

n = número de observações;

3.3.2. Índices Locais de Associação Espacial

Os índices locais, ou LISA (*Local Indicators of Spatial Association*) como são conhecidos na literatura, são utilizados quando se deseja analisar a contribuição espacial de cada área individualmente (Anselin, 1995). Servem também para verificar a hipótese de estacionariedade do processo e a identificação de agrupamentos e pontos atípicos, ou *outliers* (Câmara *et al.*, 2002). Um LISA, segundo Anselin (1995), é qualquer estatística que satisfaça as seguintes condições:

- a) O LISA para cada observação fornece uma indicação de aglomerações significativas ao redor daquela observação.
- b) A soma dos LISAS's para todas as observações é proporcional ao indicador global de associação espacial.

O índice local I_i de Moran é definido conforme a Equação (3.6)

$$I_i = \frac{z_i \sum_{j=1}^n w_{ij} z_j}{\sum_{i=1}^n z_i^2 / n} \quad (3.6)$$

onde,

z_i = desvio da região i em relação a média, ou seja, $(y_i - \bar{y})$;

w_{ij} = elemento ij da matriz de proximidade espacial;

Como o índice global de Moran, o índice de Geary também possui um indicador local, assim definido:

$$C_i = \frac{\sum_{j=1}^n w_{ij} (z_i - z_j)^2}{\sum_{i=1}^n z_i^2} \quad (3.7)$$

onde, os parâmetros são como em (3.5).

Mais detalhes sobre a derivação dos índices locais de Moran, de Geary e sobre as propriedades estatísticas, como a esperança matemática e a variância dos índices, necessárias para a definição da significância estatística, podem ser vistos em Anselin (1995).

A variável original y pode ser utilizada para o cálculo dos índices global e local, entretanto os valores na forma de desvios são mais recomendados por facilitarem a derivação das fórmulas e as propriedades estatísticas. O uso da matriz W padronizada (quando as linhas somam 1) também é aconselhável pelas mesmas razões ditas anteriormente, e por atribuir um peso a cada região j vizinha à região i . Por exemplo, no caso de uma área i possuir 4 vizinhos, o peso ou a contribuição de cada vizinho j para explicar i será de 25% do seu valor.

Uma vez calculado o valor dos índices é necessário estabelecer sua validade estatística, ou seja, verificar a significância de tal correlação. Para isso, deve-se atribuir a esses índices uma distribuição de probabilidade, sendo a distribuição normal a mais utilizada como pode ser visto em Moran (1950). Outra forma de estabelecer a significância é utilizar a distribuição empírica do indicador de dependência espacial. Para isso são geradas milhões de permutações que constituirão a distribuição empírica do índice. Caso o valor calculado esteja na cauda dessa distribuição, constituirá um evento significativo. A hipótese nula desses testes é a não existência de dependência espacial.

3.4. ANÁLISE GRÁFICA DE DEPENDÊNCIA ESPACIAL

Outra forma de visualizar a dependência espacial é através de gráficos usualmente denominados “Diagrama de Espalhamento de Moran”. Em casos onde a significância

estatística do índice local já foi determinada, é interessante indicar no mapa as regiões referentes a esses índices. Tal forma de apresentação, além de ser mais amigável, possibilita a localização geográfica das áreas, permitindo assim a identificação de aglomerações.

3.4.1. Diagrama de Espalhamento de Moran

O índice global I de Moran tem sua forma gráfica proposta por Anselin (1996), denominada *Moran Scatterplot* ou simplesmente “Diagrama de Espalhamento de Moran”. Na forma matricial, o índice de Moran é dado por:

$$I = \frac{z^t \cdot Wz}{z^t \cdot z} \quad \text{ou} \quad I = (z^t z)^{-1} z^t Wz \quad (3.8)$$

onde,

z^t = vetor $1 \times n$ dos desvios de y ;

z = vetor $n \times 1$ dos desvios de y ;

Wz = matriz $n \times n$ dos desvios de Wz (matriz de proximidade espacial multiplicada por z);

Neste caso, o índice I é equivalente ao coeficiente angular β de uma regressão linear de Wz em z . Como o β indica a inclinação da reta de regressão, é possível analisar a associação espacial de z_i com a média dos seus vizinhos Wz através da disposição desses pontos ao redor da reta. Assim, o Diagrama de Espalhamento de Moran é dividido em quatro quadrantes (Q1, Q2, Q3 e Q4) conforme a Figura 3.2: os pontos que estão em Q1 são chamados alto-alto (ou *High-High*) por indicarem que para altos valores de z , na média existem altos valores de Wz ; os pontos que estão em Q3 são chamados baixo-baixo (ou *Low-Low*) por indicarem que para baixos valores de z , na média existem baixos valores de Wz ; os pontos em Q2 e Q4 são chamados de baixo-alto (ou *Low-High*) e alto-baixo (ou *High-Low*) respectivamente, indicando que para baixos (ou altos) valores de z , na média existem altos (ou baixos) valores de Wz .

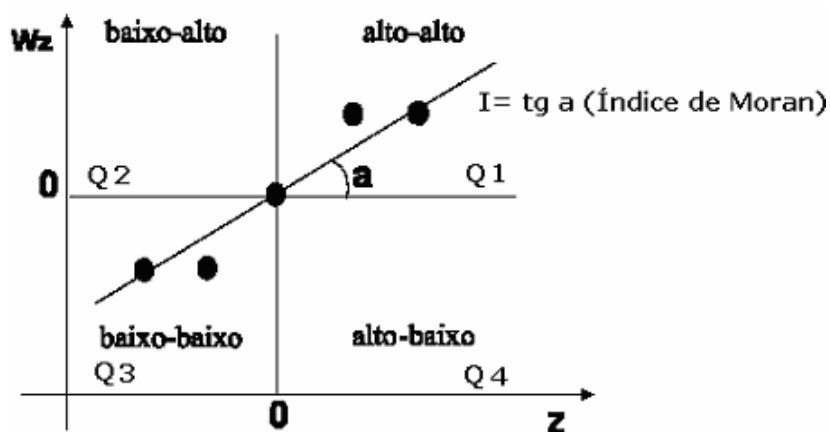


Figura 3.2: Exemplo do Diagrama de Espalhamento de Moran.

Fonte: Kampel e Câmara (2000).

A dependência espacial apresenta maior intensidade à medida que mais pontos se aglomeram no primeiro e no terceiro quadrante, enquanto que a existência de pontos no segundo e quarto quadrante descaracteriza esse fato. Esses últimos sugerem que o fator “espaço” não influencia diretamente na valoração da variável y , caracterizando-se assim, como eventos aleatórios.

3.4.2. *Box map, LISA map e Moran map*

O Diagrama de Espalhamento de Moran também pode ser visto através de um mapa (quando existir a malha geográfica georreferenciada) denominado *Box Map* (Figura 3.3). Os quatro quadrantes são representados por cores diferentes para facilitar a identificação.

Moran Scatterplot

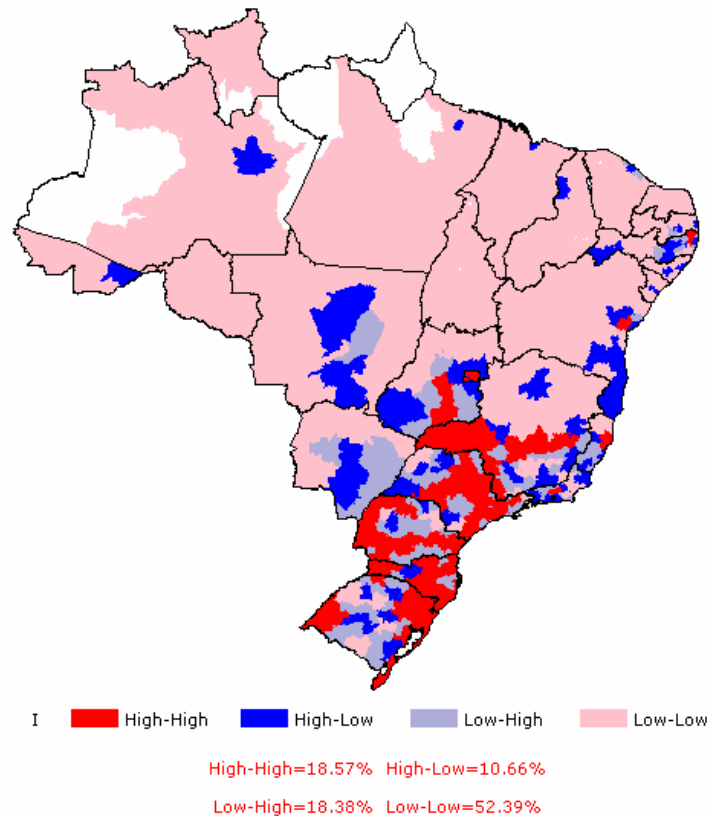


Figura 3.3: Exemplo de um *Box map* (ou *Moran Scatterplot* em forma de mapa).

O *LISA map* é utilizado para localizar, no mapa, os pontos estatisticamente significantes calculados pelos índices locais. Geralmente usam-se os níveis de significância de 99,9%, 99% e 95%. Como o nível 95% é o mais utilizado, a combinação do *Box map* com o *LISA map* fornece um novo tipo de mapa denominado *Moran map* (Figura 3.4), o qual permite identificar quais das classificações definidas no *Moran Scatterplot* (*High-High*, *Low-Low*, *High-Low* e *Low-High*) são verdadeiramente significativas.



Figura 3.4: Exemplo de um *Moran map*.

Constatada a autocorrelação espacial com as ferramentas descritas anteriormente, o próximo passo é utilizar modelos de regressão que incorporem os efeitos espaciais. Estes modelos podem ser globais ou locais. Os modelos globais utilizam um único parâmetro para capturar a estrutura de correlação espacial, enquanto que os locais utilizam k parâmetros que variam continuamente no espaço.

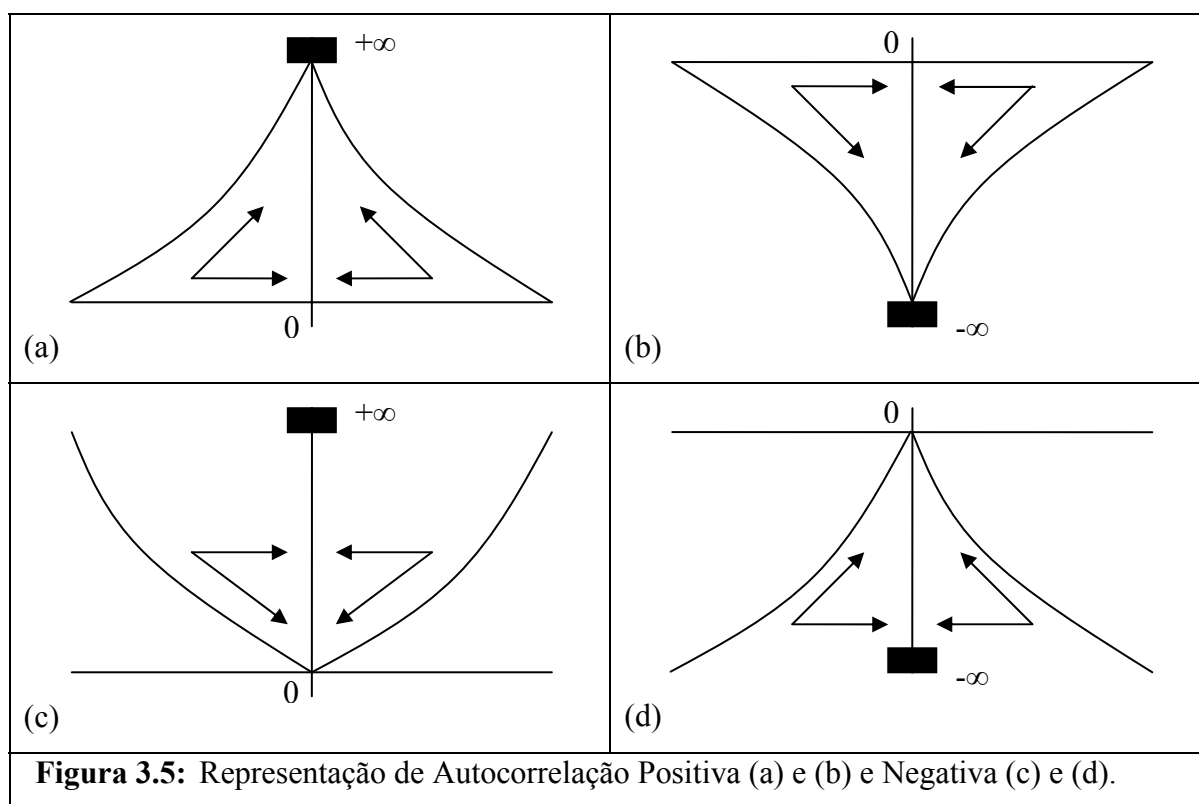
3.5. MODELOS DE REGRESSÃO ESPACIAL

Dados espaciais agregados são caracterizados pela dependência (autocorrelação espacial) e pela heterogeneidade ou estrutura espacial (Anselin, 1988). Esses efeitos espaciais são importantes pois, em alguns casos, são os principais responsáveis pela realização dos eventos. Entretanto, invalidam os resultados dos modelos tradicionais de regressão, por violarem alguns pressupostos como a independência e a homocedasticidade. Assim, pela necessidade de se incorporar tais fenômenos à estrutura de um modelo é que foram

desenvolvidos os modelos de regressão espacial ou *spatial econometric models* como são conhecidos na literatura.

A dependência espacial ou autocorrelação espacial está diretamente relacionada com a Primeira Lei da Geografia enunciada por Tobler (1979): “tudo está relacionado a tudo, mas as coisas mais próximas estão mais relacionadas que as coisas mais distantes”. Câmara *et al.* (2002) comentam que essa dependência é uma característica inerente à representação dos dados através de subdivisões territoriais, ou seja, os dados de uma determinada área tendem a ser mais parecidos com os de seus vizinhos do que com os de áreas mais distantes. Vale ressaltar que o termo “vizinho” está baseado no padrão espacial adotado: geográfico (fronteiras, distancia, etc) ou conectividade (tempo de viagem, trocas comerciais, etc).

Essa dependência relacionada com a Primeira Lei da Geografia é conhecida como a dependência ou autocorrelação positiva (Figura 3.5(a) e (b)). A autocorrelação negativa (Figura 3.5(c) e (d)) é exatamente o contrário, ou seja, as coisas mais distantes entre si estão mais relacionadas que as coisas mais próximas.



De acordo com a Figura 3.5(a) e tomando como referência o retângulo em negrito que representa um alto valor, verifica-se que à medida que os pontos tendem ao centro, maiores ficam seus valores, ou seja, à medida que os vizinhos estão mais próximos do retângulo, mais parecidos com o retângulo eles são. A recíproca também é verdadeira: à medida que os vizinhos estão mais distantes do retângulo, mais diferentes do retângulo eles são. A mesma idéia é apresentada na Figura 3.5(b), apenas tendo como diferença o retângulo que representa um baixo valor.

No caso da Figura 3.5(c), tem-se o mesmo retângulo da Figura 3.5(a) e verifica-se que à medida que os pontos tendem ao centro, menores ficam seus valores, ou seja, à medida que os vizinhos estão mais próximos do retângulo, mais diferentes do retângulo eles são. E à medida que os vizinhos estão mais distantes do retângulo, mais parecidos com o retângulo eles são. A mesma idéia é apresentada na Figura 3.5(d), apenas tendo como diferença o retângulo que representa um baixo valor.

A heterogeneidade espacial diz respeito a aspectos da estrutura socioeconômica do espaço geográfico, ou seja, é o processo em que as respostas variam de lugar pra lugar (Anselin, 1988). Essa característica pode ocasionar instabilidade estrutural (coeficientes variáveis) e variância não constante (heterocedasticidade) que distorcem os resultados do modelo. Mais detalhes sobre dependência e heterogeneidade espacial podem ser encontrados em Câmara *et al.* (2002), Anselin (1988) e LeSage (1999a).

Os modelos de regressão espacial também necessitam dos três principais pressupostos do modelo de regressão convencional, porém ao incorporar em sua estrutura o fator “espaço”, eliminam, na maioria das vezes, os problemas de autocorrelação e heterocedasticidade mencionados anteriormente. Isso porque esse último ocorre, dentre outros motivos, devido à ausência de variáveis, sejam observáveis ou não observáveis como é o caso do “espaço”. As três hipóteses (ou pressupostos) necessárias para validação de um modelo de regressão são:

- a) os erros são normais com média zero;
- b) os erros têm variância constante (homocedasticidade);
- c) os erros são não correlacionados.

Lopes (2005) fez uma primeira tentativa de abordagem matemática espacial ao modelar a demanda por transportes no município de Porto Alegre no Estado do Rio Grande do Sul. Seu objetivo era avaliar o uso de indicadores de dependência espacial em modelos de demanda por transportes. Obteve como conclusão que os modelos que utilizavam variáveis espaciais produziram melhores resultados do que os modelos sem tais variáveis. Isso porque foram introduzidas variáveis *dummies* (sejam globais ou locais) que ajudavam a explicar a demanda por transporte. Acontece que as variáveis *dummies* alteram apenas a média de y (intercepto), não influenciando diretamente no coeficiente angular β , conforme demonstração a seguir. Seja o modelo apresentado na Equação (3.9).

$$y_i = \alpha + \beta_1 x_i + \beta_2 D + \varepsilon_i \quad (3.9)$$

onde:

y_i = valor do atributo i da variável dependente y ;

α = intercepto do modelo;

x_i = valor do atributo i da variável independente x ;

D = variável *dummy* com valor 0 ou 1;

ε = erro aleatório correspondente a observação i ;

β_1 e β_2 = parâmetros do modelo correspondente a x e D , respectivamente.

Fazendo a esperança matemática da Equação (3.9) em relação à y , tem-se que:

$$E(y) = \begin{cases} \alpha + \beta_1 x & , \text{ se } D = 0 \\ (\alpha + \beta_2) + \beta_1 x & , \text{ se } D = 1 \end{cases} \quad (3.10)$$

A Equação (3.9) pode ser estendida para n variáveis *dummies*, gerando resultados similares a Equação (3.10), quando feita a esperança matemática. Verifica-se então que a inclusão de variáveis *dummies* altera apenas o valor inicial de y , mantendo constante a inclinação da reta (Figura 3.6). É importante ressaltar que o valor de β pode sofrer pequenas alterações devido a inclusão de variáveis *dummies*, porque estas podem acrescentar informações que antes estavam no erro aleatório. Outra ressalva que deve ser feita é que o valor de β também é alterado quando uma *dummy* cruzada é adicionada ao modelo (Equação 3.11):

$$y_i = \alpha + \beta_1 x_i + \beta_2 D x_i + \varepsilon_i \quad (3.11)$$

onde os parâmetros são como na Equação (3.9).

Neste caso, a esperança matemática da Equação (3.11) em relação à y é dada por:

$$E(y) = \begin{cases} \alpha + \beta_1 x & , \text{ se } D = 0 \\ \alpha + (\beta_1 + \beta_2)x & , \text{ se } D = 1 \end{cases} \quad (3.12)$$

Porém, os modelos de regressão espacial, por terem o fator “espaço” em sua estrutura, corrigem os valores dos coeficientes angulares β , além de se ajustarem melhor aos dados. O termo “corrige” deve-se ao fato da contribuição do fator “espaço” estar distribuída, em um modelo de regressão convencional, entre as variáveis do modelo. Assim, o valor do β estimado na amostra está super ou subestimando o verdadeiro β populacional, como será visto mais adiante.

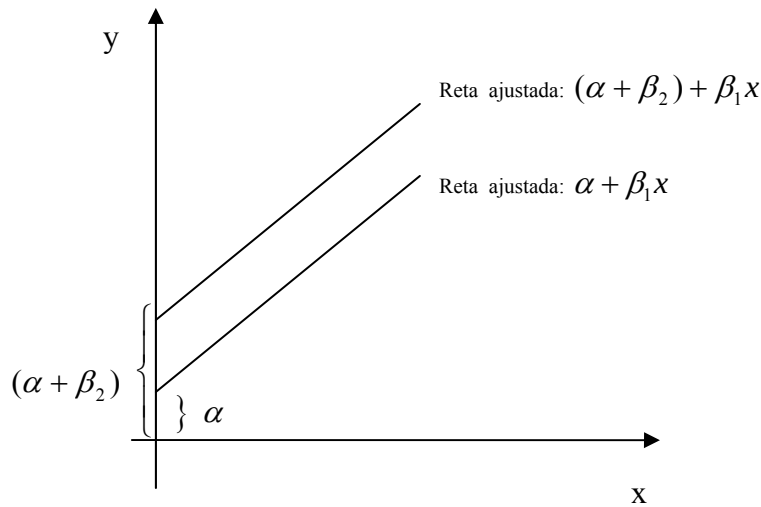


Figura 3.6: Retas estimadas de um modelo de regressão com uma variável *dummy*.

3.5.1. Modelos com Efeitos Espaciais Globais

Segundo Anselin (1988), a forma geral de um modelo espacial autoregressivo global está representado na Equação (3.13). Um processo autoregressivo, por sua vez, é aquele em que uma variável é explicada por si própria, como acontece nos modelos de séries temporais. No caso dos modelos espaciais, a variável y é explicada pelos seus “vizinhos”.

$$\begin{aligned}
\mathbf{y} &= \rho \mathbf{W}_1 \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \mathbf{u} \\
\mathbf{u} &= \lambda \mathbf{W}_2 \mathbf{u} + \boldsymbol{\varepsilon} \\
\boldsymbol{\varepsilon} &\sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)
\end{aligned}
\tag{3.13}$$

onde:

$\mathbf{y}$ = vetor de dimensão $n \times 1$ (variável dependente);

ρ, λ = constantes a serem estimadas (parâmetros espaciais);

$\mathbf{X}$ = matriz de dimensão $n \times k$ (variáveis independentes);

$\mathbf{W}_1, \mathbf{W}_2$ = matrizes de dimensão $n \times n$ (matriz de proximidade espacial);

$\boldsymbol{\beta}$ = vetor de dimensão $k \times 1$ a ser estimado (coeficientes);

$\boldsymbol{\varepsilon}, \mathbf{u}$ = vetores de dimensão $n \times 1$ (erro aleatório);

$\mathbf{I}_n$ = matriz identidade de dimensão $n \times n$;

σ^2 = constante a ser estimada (variância do modelo);

$\mathbf{0}$ = vetor de zeros de dimensão $n \times 1$;

Este modelo considera a dependência espacial na variável dependente y e no erro aleatório u . Note também que não necessariamente as matrizes $\mathbf{W}_1$ e $\mathbf{W}_2$ precisam ser diferentes. Os outros modelos, como FAR, SAR, SEM e DURBIN são derivados a partir deste, como serão apresentados a seguir.

3.5.1.1. Modelo Espacial Autoregressivo de Primeira Ordem

Esse modelo também conhecido como *First-order spatial AR model* (FAR) é utilizado quando se deseja explicar a variável dependente y a partir de seus vizinhos. Note que a dependência espacial está na própria variável y . Fazendo $\mathbf{W}_2 = \mathbf{0}$ e $\mathbf{X} = \mathbf{0}$ chega-se a Equação (3.14).

$$\begin{aligned}
\mathbf{y} &= \rho \mathbf{W}_1 \mathbf{y} + \boldsymbol{\varepsilon} \\
\boldsymbol{\varepsilon} &\sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)
\end{aligned}
\tag{3.14}$$

onde todos os parâmetros são como na Equação (3.13).

3.5.1.2. Modelo Espacial Autoregressivo Misto

Também conhecido como *Spatial Autoregressive Model* (SAR) ou *Spatial Lag Models* é utilizado quando se deseja explicar a variável dependente y a partir dela mesma e de outras variáveis explicativas. É obtido da Equação (3.13) fazendo $\mathbf{W}_2 = 0$.

$$\begin{aligned} \mathbf{y} &= \rho \mathbf{W}_1 \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &\sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n) \end{aligned} \quad (3.15)$$

onde todos os parâmetros são como na Equação (3.13).

Verifique que se o parâmetro espacial ρ for zero, então o modelo resultante é exatamente igual a um modelo de regressão convencional. O vetor de parâmetros do modelo $\boldsymbol{\beta}$, em um modelo de regressão convencional, é obtido a partir da Equação (3.16), enquanto que no modelo SAR é obtido a partir da Equação (3.17).

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} \quad (3.16)$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{I}_n \mathbf{y} - \rho (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{W}_1 \mathbf{y} \quad (3.17)$$

A diferença entre as Equações (3.16) e (3.17) está exatamente na inclusão da estrutura espacial. Quando o valor de ρ estiver próximo de zero (baixa dependência espacial), pouca informação será agregada a $\boldsymbol{\beta}$, enquanto que se estiver próximo de +1 ou -1 (alta dependência espacial), um valor significativo será agregado a $\boldsymbol{\beta}$. Esse fato explica o porquê da regressão espacial “corrigir” os parâmetros do modelo, quando comparada à regressão convencional.

3.5.1.3. Modelo com Erro Espacial Autoregressivo

Esse modelo também conhecido como *Spatial Error Model* (SEM) tem a mesma função do modelo SAR, porém a estrutura espacial está no erro aleatório. É obtido a partir da Equação (3.13) fazendo $\mathbf{W}_1 = 0$.

$$\begin{aligned} \mathbf{y} &= \mathbf{X} \boldsymbol{\beta} + \mathbf{u} \\ \mathbf{u} &= \lambda \mathbf{W}_2 \mathbf{u} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &\sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n) \end{aligned} \quad (3.18)$$

onde todos os parâmetros são como na Equação (3.13).

Na prática, a distinção entre esses dois modelos é difícil. No entanto, o modelo SEM tem mais indícios de ser utilizado quando o resíduo resultante de um modelo de regressão convencional possui dependência espacial, constatada pelo *I* de Moran. Outros fatores que corroboram a utilização desse modelo são os erros de medida, ausência de variáveis explicativas ou variáveis não-observáveis, além da heterocedasticidade.

3.5.1.4. Modelo Espacial Durbin

Esse modelo é similar ao modelo SAR, porém é utilizado quando se deseja incorporar também uma estrutura espacial nas variáveis independentes. Dessa forma, as variáveis explicativas são influenciadas pela dependência espacial da mesma maneira que a variável dependente, refinando assim a caracterização da estrutura espacial.

$$\begin{aligned} \mathbf{y} &= \rho \mathbf{W}_1 \mathbf{y} + \mathbf{X} \boldsymbol{\beta}_1 + \mathbf{W}_1 \mathbf{X} \boldsymbol{\beta}_2 + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &\sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n) \end{aligned} \quad (3.19)$$

onde todos os parâmetros são como na Equação (3.13).

Existem também outros modelos espaciais como *probit*, *tobit* e os baseados na perspectiva Bayesiana. Os modelos *probit* são utilizados quando a variável dependente apresenta uma resposta qualitativa, ou seja, acontece um determinado fenômeno ou não. Para isso, atribui-se os valores 0 e 1 às respectivas respostas. Já os modelos *tobit* são utilizados quando existirem valores censurados na variável dependente e os modelos baseados na perspectiva Bayesiana são aqueles onde é atribuído uma informação a priori para o método de estimação de verossimilhança. Tais modelos não serão aqui detalhados, mas podem ser melhor estudados em Gujarati (1995) e LeSage (1999b).

Os modelos espaciais globais partem do pressuposto que o processo espacial analisado é estacionário, ou seja, possui um único padrão de associação espacial. Greene (1999) comenta que uma série temporal é estacionária (estacionariedade fraca) se suas médias, variâncias e covariâncias permanecem as mesmas, independente do período de tempo em que sejam medidas. Diante desse conceito e fazendo uma analogia ao processo espacial, a estacionariedade é verificada, por exemplo, quando não existir áreas significativas nos

testes locais, indicando dessa forma a ausência de regiões com padrões diferenciados. Tomando esse fato como premissa, os modelos globais capturam a espacialidade em um único parâmetro. Câmara *et al.* (2002) argumentam que a estacionariedade nem sempre ocorre:

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. Esta hipótese, que pode ser verificada, por exemplo, pelos indicadores locais de autocorrelação espacial, está na origem aos modelos cujos parâmetros variam no espaço.

Dessa forma foram desenvolvidos os modelos com efeitos espaciais locais, a fim de se ter um refinamento maior do modelo e um melhor entendimento do processo espacial.

3.5.2. Modelos com Efeitos Espaciais Locais

Esses modelos são utilizados quando o processo é não-estacionário, e por isso, há a necessidade de que os coeficientes da regressão reflitam essa heterogeneidade espacial (Câmara *et al.*, 2002). O primeiro modelo desse tipo foi introduzido por Caseti em 1972 e foi chamado de “modelo espacial de expansão” ou *spatial expansion model* (LeSage, 1999a).

Existem duas formas de se trabalhar com efeitos locais: modelar a variação espacial de forma discreta, dividindo o espaço em sub-regiões estacionárias denominadas regimes espaciais, ou modelar a tendência espacial de forma contínua, com parâmetros variando no espaço.

3.5.2.1. Modelagem de forma discreta

A idéia é identificar primeiramente os regimes espaciais, com a ajuda do diagrama de espalhamento de Moran e/ou dos indicadores locais. Dessa forma é possível dividir os dados em sub-regiões, criando indicadores como por exemplo na Equação (3.20), e assim estimar os parâmetros do modelo conjuntamente.

$$\begin{aligned} Y_1 &= X_1\beta_1 + \varepsilon_1, ind = 1 \\ Y_2 &= X_2\beta_2 + \varepsilon_2, ind = 2 \end{aligned} \quad (3.20)$$

onde:

Y_1 e Y_2 = variáveis dependentes;

X_1 e X_2 = variáveis independentes;

β_1 e β_2 = parâmetros do modelo;

ε_1 e ε_2 = erros aleatórios;

Câmara *et al.* (2002) comentam sobre a utilização prática desses modelos:

Na prática, para os dados sócio-econômicos 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. Isto 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 em nossas grandes cidades.

3.5.2.2. Modelagem de forma contínua

A técnica empregada nesse caso é conhecida como GWR (*Geographically Weighted Regression*) e a idéia é ajustar um modelo de regressão a cada ponto observado, ponderando todas as demais observações como função da distância a esse ponto (Câmara *et al.*, 2002). A forma do modelo é apresentada na Equação (3.21).

$$Y_i = \beta_i X + \varepsilon \quad (3.21)$$

onde:

Y_i = variável dependente representando o processo no ponto i ;

β_i = indica que os parâmetros são estimados no ponto i ;

X = variável independente;

ε = erro aleatório.

Os parâmetros podem ser estimados por

$$\hat{\beta}_i = (\mathbf{X}'\mathbf{W}_i\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}_i\mathbf{y} \quad (3.22)$$

Diversos autores tais como Feitosa (2005) e Câmara *et al.* (2002) comentam que os modelos locais tendem a se ajustar melhor aos dados, em relação à regressão tradicional e aos modelos espaciais globais, devido à maior especificação do processo espacial. No entanto, a interpretação do modelo pode se tornar um pouco complicada devido à quantidade de parâmetros estimados. Mais detalhes sobre os modelos com efeitos espaciais locais podem ser vistos em LeSage (1999a) e Câmara *et al.* (2002).

Muitos softwares especializados em análises espaciais estão disponíveis no mercado. A Tabela 3.1 apresenta uma lista com os principais testes e modelos implementados.

Tabela 3.1. Softwares para análise espacial disponíveis no mercado e suas funcionalidades.

SOFTWARE	TESTES E MODELOS IMPLEMENTADOS
GeoDa	<i>I</i> de Moran, LISA <i>map</i> , Regressão Clássica e Espacial (Spatial Lag & Spatial Error)
SPRING	<i>I</i> de Moran, LISA <i>map</i>
Terraview	<i>I</i> de Moran, LISA <i>map</i> , Regressão Clássica, Espacial (Spatial Lag & Spatial Error) e GWR
SpaceStat	<i>I</i> de Moran, Regressão Clássica e Espacial (Spatial Lag & Spatial Error)
R	Regressão Clássica, Espacial (Spatial Lag & Spatial Error) e GWR
ART	Regressão Clássica, Espacial (Spatial Lag & Spatial Error) e GWR
GWR 3.0	Regressão Clássica e Espacial (GWR)
MATLAB	<i>I</i> de Moran, Regressão Clássica, Espacial (Spatial Lag, Spatial Error, Spatial Durbin, General Spatial Model, Probit, Tobit, Bayesianos) e GWR

Fonte: Feitosa (2005) com adaptações.

3.6. DEFINIÇÃO DE ESPAÇO NO CONTEXTO DA REGRESSÃO ESPACIAL

Definir espaço é uma tarefa complexa, pois depende do contexto a ser considerado. O caso mais comum, e mais usual, é o espaço geográfico, ou seja, aquele onde é possível atribuir uma coordenada (“x”, “y”) – latitude, longitude, UTM etc – a um determinado ponto. Contudo, um “espaço” em termos mais gerais, fornece informações mais amplas como relacionamento e dependência. Pensando nessa última é que foram desenvolvidos os

modelos de regressão espacial, por um modelo de regressão convencional ter como pressuposto a independência entre as observações.

Acontece que alguns fenômenos dependem de outros, que também dependem de terceiros, e assim por diante, formando uma cadeia de recorrência. Ao se modelar tais fenômenos é imprescindível considerar tal estrutura de dependência. Sabidamente, à medida que uma região se desenvolve economicamente, seus vizinhos também tendem a se desenvolver. Então como fazer para que o computador entenda tal informação? Como foi visto, a matriz W é a responsável por indicar a estrutura “espacial” existente. Matematicamente ao trabalhar com um modelo de regressão espacial, considerando o “espaço” como sendo o espaço geográfico e tomando como exemplo as informações da Figura 3.1, tem-se que:

$$Y_1 = \rho(Y_2) + \beta X_1 \quad (3.23)$$

$$Y_2 = \rho(Y_1) + \beta X_2 \quad (3.24)$$

$$Y_3 = \rho(Y_4 + Y_5) + \beta X_3 \quad (3.25)$$

$$Y_4 = \rho(Y_3 + Y_5) + \beta X_4 \quad (3.26)$$

$$Y_5 = \rho(Y_3 + Y_4) + \beta X_5 \quad (3.27)$$

Ou seja, Y_1 é explicada pela variável X e também por seu “vizinho” Y_2 assim como Y_3 é explicada pela variável X e por seus “vizinhos” Y_4 e Y_5 . O mesmo vale para Y_2 , Y_4 e Y_5 devido à simetria das informações (se A está perto de B então B está perto de A).

Agora considere que foram construídas estradas para interligar as regiões da Figura 3.1, conforme a Figura 3.7. Como incorporar tal informação em um modelo de regressão espacial?

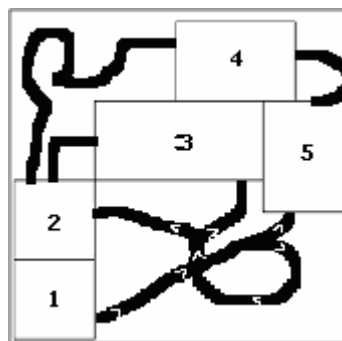


Figura 3.7: Exemplo de Regiões Interligadas por Estradas.

Claramente nesse caso não é possível manter o espaço geográfico como sendo o “espaço” responsável pela realização dos eventos, sob pena de não captar a influência de tal melhoria. Um novo “espaço” poderia ser definido, como por exemplo, o tempo necessário para ir da região i para a região j , conforme a Figura 3.8.

$$W = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{pmatrix} 0 & 25 & 20 & 37 & 18 \end{pmatrix} & \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} \end{matrix}$$

Figura 3.8: Tempo necessário (minutos) para ir de uma região i a uma região j .

Verifica-se na Figura 3.8 que é necessário um menor tempo para ir da região 1 para a região 5 do que da região 1 para a região 2, as quais são geograficamente mais próximas. Refazendo a matriz de acordo com a forma (4) apresentada na seção 3.2 e já padronizando-a, chega-se à Figura 3.9.

$$W = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{pmatrix} 0 & .23 & .29 & .16 & .32 \\ .12 & 0 & .47 & .25 & .16 \\ .17 & .50 & 0 & .18 & .15 \\ .10 & .29 & .20 & 0 & .41 \\ .21 & .19 & .17 & .44 & 0 \end{pmatrix} & \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} \end{matrix}$$

Figura 3.9: Matriz W utilizando o tempo necessário para ir de uma região i a uma região j .

Matematicamente ao trabalhar com um modelo de regressão espacial, considerando o “espaço” como sendo o tempo necessário para ir de uma região i para uma região j , e tomando as informações da Figura 3.9, tem-se que:

$$Y_1 = \rho(0,23Y_2 + 0,29Y_3 + 0,16Y_4 + 0,32Y_5) + \beta X_1 \quad (1)$$

$$Y_2 = \rho(0,12Y_1 + 0,47Y_3 + 0,25Y_4 + 0,16Y_5) + \beta X_2 \quad (2)$$

$$Y_3 = \rho(0,17Y_1 + 0,50Y_2 + 0,18Y_4 + 0,15Y_5) + \beta X_3 \quad (3)$$

$$Y_4 = \rho(0,10Y_1 + 0,29Y_2 + 0,20Y_3 + 0,41Y_5) + \beta X_4 \quad (4)$$

$$Y_5 = \rho(0,21Y_1 + 0,19Y_2 + 0,17Y_3 + 0,44Y_4) + \beta X_5 \quad (5)$$

Ou seja, Y_1 é explicada pela variável X e também por 23% de Y_2 , 29% de Y_3 , 16% de Y_4 e 32% de Y_5 . Já Y_2 é explicada pela variável X e por 12% de Y_1 , 47% de Y_3 , 25% de Y_4 e 16% de Y_5 . A interpretação é a mesma para Y_3 , Y_4 e Y_5 . Verifica-se então que os “vizinhos” passaram a ser aqueles onde é necessário um menor tempo de chegada, e por isso o modelo usa a maior parte da informação destes “vizinhos” mais próximos para explicar o Y_i correspondente.

Outro tipo de “espaço” que pode ser definido é a partir da matriz de insumo-produto, para explicar o valor da produção das atividades econômicas (setor têxtil, confecções, automobilística etc). Tal matriz fornece a quantidade de mercadorias negociadas entre um setor e o outro. É clara a dependência entre os setores, como por exemplo o setor têxtil e de confecções (pois um movimenta o outro) e tentar modelar tal fenômeno sem considerar uma estrutura de dependência (ou em termos mais gerais, uma estrutura “espacial”) ocasionaria uma valoração incorreta dos parâmetros do modelo.

3.7. TÓPICOS CONCLUSIVOS

Este capítulo buscou levantar os principais aspectos referentes à regressão espacial, bem como as ferramentas que auxiliam na identificação da dependência espacial. Assim, pelo que foi visto, merecem destaque os seguintes tópicos conclusivos:

- Uma boa definição da matriz de proximidade espacial garante resultados mais fidedignos ao problema estudado;
- Os testes para identificação de dependência espacial, sejam globais ou locais, são capazes de quantificar tal dependência, além de terem validade estatística;
- As ferramentas gráficas *Box map*, *LISA map* e *Moran map* auxiliam o planejador na identificação de padrões espaciais, além de possibilitarem uma identificação visual de possíveis aglomerações e tendências;
- A dependência e a heterogeneidade espacial são características intrínsecas de dados espaciais agregados. Assim, ao mesmo tempo que são importantes para realização de um evento, invalidam os resultados dos modelos de regressão convencional. Essa foi a principal motivação para o estudo dos modelos de regressão espacial;

- Os modelos de regressão espacial, por incorporarem em sua estrutura o fator “espaço”, tendem a apresentar um melhor ajuste aos dados do que um modelo de regressão tradicional, quando identificada a existência de dependência espacial. Por consequência, os coeficientes β são “corrigidos”, representando assim sua verdadeira contribuição à variável dependente;
- Os modelos espaciais globais, que captam a espacialidade em um único parâmetro, são utilizados quando os dados forem estacionários. Constatada a não-estacionariedade dos dados, a saída é utilizar os modelos espaciais locais, que captam a espacialidade em k parâmetros;
- A definição de espaço é muito complexa, porém no contexto da regressão espacial, o “espaço” pode ser definido como o ambiente responsável pela realização de um evento o qual considera uma estrutura de relacionamento e dependência. Nesse sentido, pode-se dizer que o transporte é um problema espacial por interferir permanentemente na dinâmica urbana e regional, gerando automaticamente, dependência das áreas com menos investimentos às áreas com mais investimentos.

4. METODOLOGIA PARA CONSTRUÇÃO DE MODELOS DE REGRESSÃO ESPACIAL

4.1. APRESENTAÇÃO

A busca de relação entre variáveis é um processo demasiadamente complexo, porque necessita primeiramente de um conhecimento prévio acerca da causa/efeito das variáveis envolvidas ou de uma boa base em estatística. Este último pode criar uma armadilha para o planejador devido à existência de correlações espúrias entre as variáveis. Tome como exemplo, a quantidade de igrejas e o número de bêbados: constatada uma correlação de 70%, é condição suficiente para afirmar que quanto mais igrejas existirem, maior é a quantidade de bêbados?

Aliado a isso está também um bom planejamento para a coleta de dados (ou a aquisição a partir de fontes confiáveis), porque as ferramentas estatísticas apenas refletirão o que está presente nos dados. Outro ponto que merece atenção é quanto à temporalidade. Uma determinada variável, medida um ano depois que a outra variável analisada, pode apresentar um acréscimo, ou decréscimo, devido a uma melhoria nesse período decorrido, e não apenas por estar correlacionada com a outra variável. Assim, essa não adequabilidade temporal resultará em uma correlação não totalmente verdadeira.

O avanço computacional também facilitou a utilização de ferramentas de modelagem mais robustas, que buscam retratar a aleatoriedade do mundo real com mais fidedignidade. As redes neurais e a regressão espacial são alguns dos exemplos dessas ferramentas. A regressão espacial, conforme apresentado no Capítulo 3, é a mais indicada para se trabalhar com dados regionais (ou em um sentido mais amplo, dados “espaciais”) do que a regressão convencional, por incorporar em sua estrutura o fator “espaço”.

Dessa forma, este capítulo apresenta o método proposto para a identificação, construção e análise de um modelo de regressão espacial. A metodologia padrão para construção de modelos de regressão e a estrutura geral da metodologia estão na primeira seção. A seção posterior descreve com detalhes cada etapa necessária para atingir os objetivos propostos.

4.2. ESTRUTURA

A metodologia padrão para a construção de modelos de regressão, esquematizada por Ogliari (2004a), é apresentada na Figura 4.1. Verifica-se a necessidade de 4 etapas até a consolidação do modelo final: coleta e preparação dos dados, redução do número de variáveis regressoras, refinamento e seleção de modelos e validação do modelo.

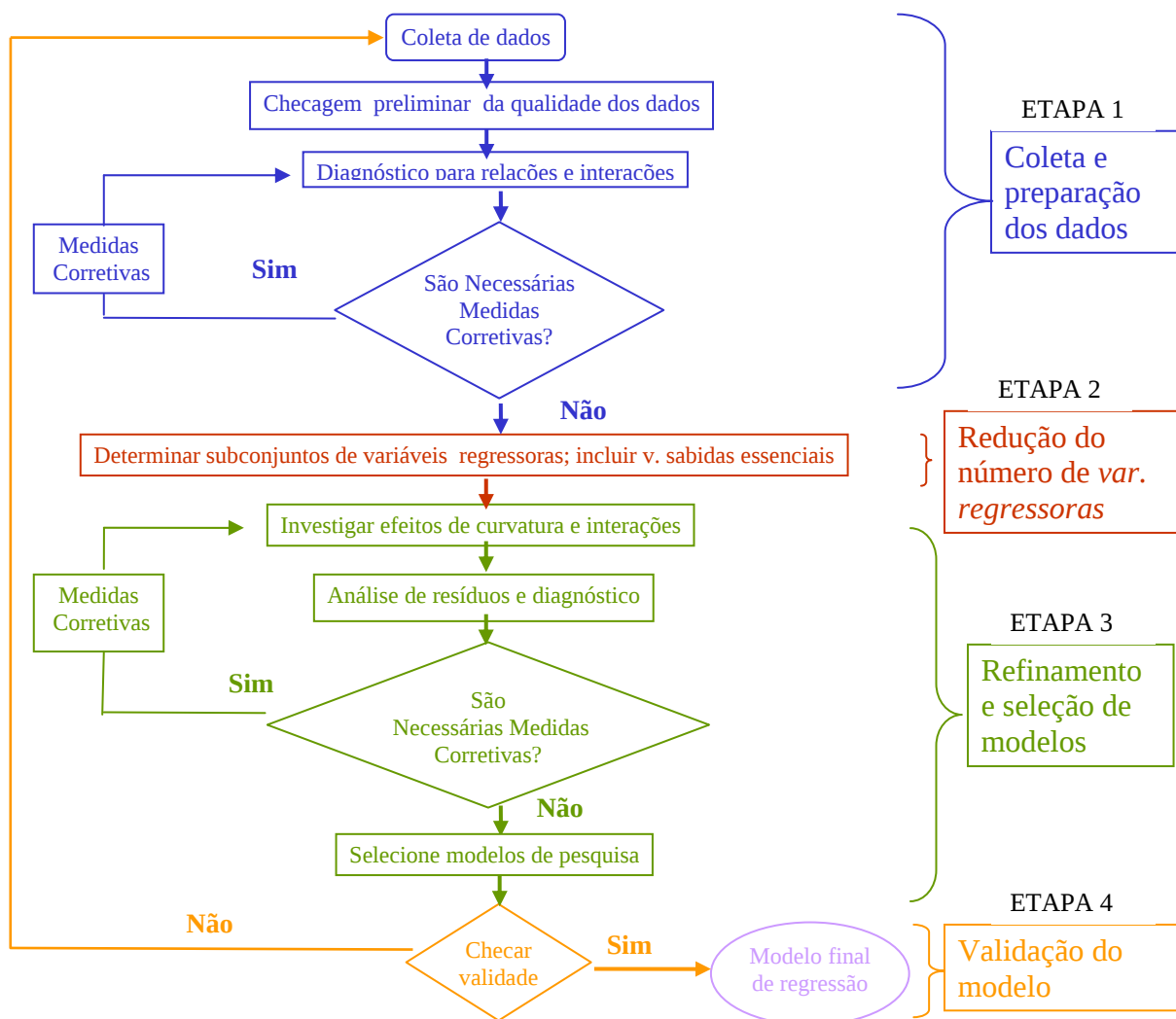


Figura 4.1: Fluxograma para Construção de Modelos de Regressão.
Fonte: Ogliari (2004a).

A Figura 4.2, por sua vez, apresenta as 4 etapas necessárias para desenvolver um modelo de regressão espacial, tendo como base a metodologia descrita por Ogliari (2004a) com algumas considerações para o caso da regressão espacial. Tais considerações são referentes à definição da unidade espacial de análise e a inserção da etapa da verificação de dependência espacial. A descrição de cada uma das etapas será apresentada a seguir.

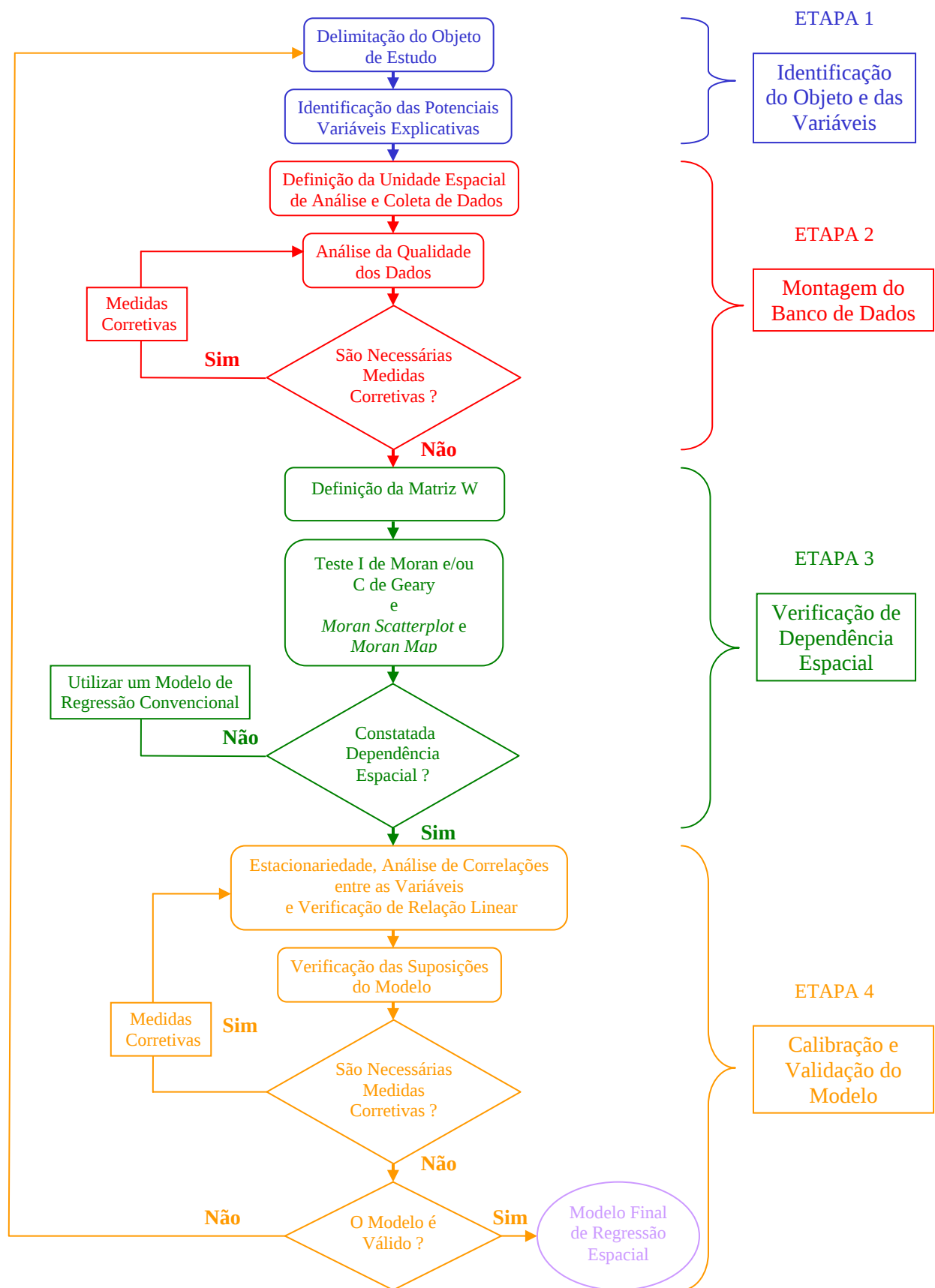


Figura 4.2: Fluxograma Metodológico.

4.3. ETAPAS

4.3.1. Etapa 1: Identificação do Objeto e das Variáveis

A delimitação do objeto de estudo (fenômeno ou população da variável) é o ponto de partida para elaboração de qualquer modelo, pois permite que o planejador tenha um foco mais direcionado à resolução do problema. Seja por exemplo um estudo sobre os estudantes brasileiros. Como todos os estudantes de nível básico, médio e superior estão sendo analisados, fica muito difícil decidir o que estudar e como estudar, pois cada um desses grupos possui uma particularidade diferente. Assim, o correto seria estudar apenas um desses grupos, devido à interpretação e aplicabilidade dos resultados.

A identificação das variáveis explicativas se torna mais fácil após a delimitação do objeto de estudo, e pode ou deve estar constantemente em revisão. O primeiro contato com essas variáveis pode ser obtido através da aplicação de questionários destinados a especialistas da área. Outra forma é através de um conhecimento prévio, ou seja, utilizando variáveis que de forma direta ou indireta podem influenciar na ocorrência do evento em estudo, porém sem nenhuma base teórica para sustentá-las. Quando a identificação das variáveis se utilizar dessa última forma, muito cuidado deve ser tomado a fim de não cair em correlações espúrias.

4.3.2. Etapa 2: Montagem do Banco de Dados

Antes de iniciar a coleta dos dados, deve-se definir qual a unidade espacial de análise que será utilizada. Os casos mais comuns de unidades espaciais de análise (geográficas) são zona de tráfego, setor censitário, município, microrregião ou mesorregião. A UF não é muito utilizada devido a existência de apenas 27 pontos, entretanto, nada impede sua utilização.

O próximo passo é a aquisição dos dados propriamente ditos, que podem ser obtidos através de fontes primárias (coleta de dados) ou secundárias (pesquisa documental). Como a regressão espacial utiliza dados agregados, ou seja, os dados na forma *cross-sectional* são agrupados segundo uma unidade maior, uma coleta de dados pode se tornar bastante cara dependendo da definição do “espaço”. Por isso é recomendável utilizar as bases de dados do IBGE ou de outra fonte de pesquisa conceituada, devido ao baixo custo de aquisição.

Para exemplificar, seja a base de dados da RAIS (Relação Anual de Informações Sociais) do MTE (Ministério do Trabalho e Emprego), composta pelas empresas brasileiras. Através de uma variável indicadora de localização municipal é possível agrupar as empresas segundo o município, criando uma nova variável referente à quantidade de empresas em cada um desses municípios (a mesma idéia pode ser feita para microrregião, macrorregião e UF). Dessa forma, é possível conseguir a quantidade de empresas sem ir a campo e a um custo menor.

Nesta etapa deve ser feita também uma análise da qualidade dos dados, a fim de que a informação gerada seja consistente. Esta análise, não aprofundada, é composta basicamente por 3 itens: identificação do domínio; média, mediana, máximos e mínimos; e informações perdidas ou *missing*.

A primeira avaliação que deve ser feita em qualquer banco de dados quantitativo é a identificação do domínio de cada campo. A partir dessa identificação facilmente são detectadas as observações problemáticas e que devem ser removidas. Caso todas as observações estejam dentro dos limites do domínio, o passo seguinte é a caracterização dos campos.

O objetivo dessa caracterização é a identificação de *outliers* (valores discrepantes) e pode ser feita por meio de quatro medidas: média, mediana, máximos e mínimos. A média e a mediana, caso estejam muito discordantes, fornecem uma primeira indicação da presença de *outliers*, que podem ser confirmados através do valor máximo e mínimo da variável. Uma forma de visualizar graficamente todas essas medidas é através do *Boxplot*, que é uma ferramenta estatística bastante empregada para identificação de valores extremos.

Os *outliers* devem ser tratados de acordo com a realidade e o contexto. Confirmados erros de digitação, as observações erradas devem ser retiradas do banco de dados. Caso um valor discrepante esteja dentro de um contexto real, como por exemplo, qualquer estatística da cidade de São Paulo, a qual é muito diferente em relação a outra cidade brasileira, o mesmo deve ser mantido a fim de caracterizar tal observação. Nesse caso, o termo mais adequado para tal tipo de “*outlier*” é valor extremo.

A última checagem no banco de dados diz respeito à quantidade de observações faltantes. Como os *outliers*, as observações perdidas também devem ser tratadas caso a caso. Cabe aqui uma distinção entre valor zero e valor *missing*. Seja por exemplo um banco de dados contendo todas as empresas exportadoras, e seus respectivos faturamentos. Se existir algum município sem o valor das exportações, após agregação das empresas em municípios, então esse município deve receber valor zero, pois no mesmo não existem empresas que exportam. Agora no caso do índice de alfabetização, por exemplo, se existir algum município sem o respectivo valor, então este deve ser mantido em branco, porque não se sabe qual o verdadeiro valor. Sabe-se apenas que com certeza o índice de alfabetização é diferente de zero.

Os bancos de dados que apresentam grandes quantidades de observações *missing* não são recomendados para análises estatísticas, devido à falta de dados e ao viés que gerariam nas informações, caso fossem utilizados.

Não é obrigatório que a base de dados esteja georreferenciada. No entanto, análises gráficas como *Moran Scatterplot* e *Moran Map* ficam impossibilitadas de serem construídas, o que reduz a caracterização da dependência espacial.

4.3.3. Etapa 3: Verificação de Dependência Espacial

Essa etapa é a responsável por sinalizar a necessidade ou não de um modelo de regressão espacial para a modelagem dos dados. Caso não seja constatada a existência de dependência espacial, o esforço de se utilizar um modelo de regressão espacial não valerá a pena, pois além do maior tempo de processamento, apresentará os mesmos resultados de um modelo de regressão convencional.

Após a montagem do banco de dados, o próximo passo é definir as matrizes de proximidade espacial que serão utilizadas, segundo as formas apresentadas na seção 3.2. É importante definir mais de uma matriz W a fim de refinar e melhor caracterizar a dependência espacial, que pode estar presente tanto na variável dependente quanto no erro aleatório, ou em ambos.

Nesse sentido, pode-se partir de duas maneiras diferentes para verificar a dependência espacial: gerar um mapa coroplético ou temático (mapa onde o valor da variável é diferenciada por cores) da variável dependente, observar a existência de tendências espaciais e depois quantificar tal dependência através dos índices I de Moran e/ou C de Geary, ou verificar se os resíduos de um modelo de regressão convencional possuem tendência crescente ou decrescente, conforme a Figura 4.3. Constatada essa forma de cone (ou V) dos resíduos, quantificar a dependência através dos índices I de Moran e/ou C de Geary. Podem ser utilizadas também as formas gráficas *Moran Scatterplot* e *Moran Map* para auxiliar na constatação da dependência espacial de ambas as maneiras.

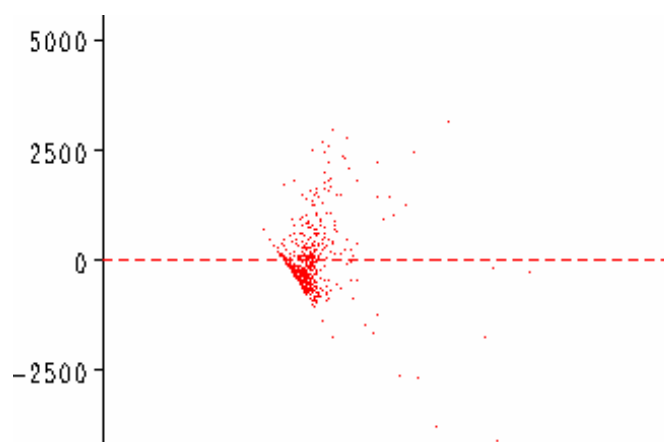


Figura 4.3: Distribuição dos resíduos na existência de dependência espacial.

Caso seja identificado que a dependência espacial está presente na variável dependente, e que os resíduos também possuem tal dependência, pode ser interessante especificar um modelo em que a dependência espacial esteja presente em ambos os termos. A escolha do melhor modelo poderá ser feita por estatísticas de ajuste discutidas na seção 4.3.4, ou por meio de critérios como o de AIC (*Akaike Information Criterion*) e/ou SBC (*Schwarz Bayesian Criterion*) discutidos na seção 5.3.

4.3.4. Etapa 4: Calibração e Validação do Modelo

Depois de verificada a existência de dependência espacial, o passo seguinte é a construção do modelo propriamente dito. O melhor tipo de modelo a ser utilizado (globais ou locais) depende primeiramente de uma análise a respeito da estacionariedade, onde os indicadores locais podem ser utilizados para tal verificação. Após definido qual o melhor modelo a ser utilizado, deve ser construída uma matriz de correlações, (utilizando o índice r de Pearson

conforme a Equação 3.2) a fim de identificar as variáveis explicativas mais relevantes e as possíveis correlações entre as mesmas. Alta correlação entre as variáveis independentes pode levar a um problema denominado multicolinearidade, o que invalida o modelo. Este problema pode ser identificado através da inversão de sinais dos parâmetros em relação ao sinal da matriz de correlação.

Outro ponto que deve ser analisado é a relação linear entre as variáveis, o que pode ser facilmente visto pelos gráficos de dispersão. Na falta de linearidade, medidas corretivas devem ser aplicadas para garantir tal pressuposto. Também é fundamental verificar os pressupostos básicos da análise de regressão (seção 3.5), pois são eles que garantem a validade do modelo. Recapitulando, os pressupostos são normalidade dos resíduos, erros com variância constante e erros não correlacionados.

A normalidade dos resíduos pode ser testada através de testes específicos como Shapiro-Wilk, Kolmogorov-Smirnov, Cramer-von Mises, Anderson-Darling e Jarque-Bera. A hipótese nula para todos esses testes é que a distribuição dos resíduos segue uma distribuição normal. Outra forma de verificar esse pressuposto é através do gráfico *qq-plot*, que compara os quantis (ou percentil) da distribuição normal com os quantis da distribuição empírica dos dados: se os dados se ajustam a uma distribuição normal, então o gráfico tem a forma de uma reta. Mais detalhes sobre esses testes podem ser vistos em Gujarati (1995) e SAS (1999).

Existem diferentes testes para verificar heterocedasticidade (variância não constante), tais como Breusch-Pagan, Koenker-Bassett e White, sendo este último o mais utilizado. A hipótese nula considerada nesses testes é a variância constante dos dados (homocedasticidade). Mais detalhes podem ser vistos em Gujarati (1995) e White (1980).

As estatísticas R^2 e R^2_{ajustado} indicam a qualidade do ajuste. Quanto mais próximo os valores de R^2 e R^2_{ajustado} estiverem de 1, mais ajustado está o modelo; enquanto que mais próximo de 0, pior é esse ajuste. O valor do R^2_{ajustado} é mais aconselhável do que R^2 por levar em consideração a quantidade de parâmetros do modelo, pois à medida que novas variáveis são inseridas, mais próximo de 1 tende a ser o valor de R^2 (quando existem poucas variáveis essas estatísticas tendem a ser muito próximas). Os modelos sem o intercepto, ou seja, com a reta passando pela origem apresentam maiores valores de R^2 e

R^2_{ajustado} , no entanto essas medidas perdem o sentido por poderem assumir valores negativos, conforme Ogliari (2004b).

A estatística F verifica se todos os parâmetros do modelo são não significativos, ou seja, verifica se os parâmetros são iguais a zero (conjuntamente). Em outras palavras, ao se rejeitar a hipótese nula pode-se dizer que o modelo existe porque pelo menos um parâmetro é diferente de zero. Por outro lado, a estatística t indica a significância de cada variável individualmente. Ambas estatísticas são utilizadas para verificar a influência ou não de cada variável independente na variável dependente em estudo.

A validade do modelo fica então condicionada à verificação desses valores. Geralmente valores de R^2 abaixo de 0,20 (se o foco for previsão) indicam que não houve uma grande explicação por parte das variáveis explicativas, necessitando, portanto, utilizar outras variáveis. Caso o teste F seja significativo e algum teste t não seja significativo, também pode ser um indício de não validade do modelo.

4.4. TÓPICOS CONCLUSIVOS

Este capítulo teve como objetivo apresentar a metodologia para a construção de modelos de regressão espacial, necessária para atingir os objetivos propostos. Dessa forma, merecem destaque os seguintes tópicos conclusivos:

- A metodologia geral da pesquisa foi baseada na metodologia de construção de modelos, proposta por Ogliari (2004a), fazendo as alterações pertinentes ao modelo de regressão espacial, a saber: definição da unidade espacial de análise e a inserção da etapa da verificação de dependência espacial;
- Foram especificadas 4 etapas na metodologia, compreendendo: identificação do objeto e das variáveis; montagem do banco de dados; verificação de dependência espacial e calibração e validação do modelo;
- É importante verificar no fluxograma metodológico que, na ausência de dependência espacial, é recomendando a utilização de um modelo de regressão convencional devido sua simplicidade.

5. ESTUDO DE CASO: TRANSPORTE RODOVIÁRIO DE CARGA

5.1. APRESENTAÇÃO

O presente capítulo tem por objetivo verificar a aplicabilidade da metodologia proposta no Capítulo 4. Para tal, foi realizado um estudo de caso para o Transporte Rodoviário de Carga (TRC). O objetivo específico do estudo é identificar os fatores intervenientes e analisar o impacto da variação destes na distribuição espacial da oferta do TRC.

De acordo com o Capítulo 2, o conhecimento sobre a oferta do transporte rodoviário nas diversas regiões do Brasil, por meio de veículos disponíveis, permite conhecer a capacidade de carga transportada. Além disso, permite subsidiar tomadas de decisão em termos de mercado, órgão gestor e da indústria de veículos.

O modelo de regressão espacial foi aplicado para analisar a distribuição da oferta de veículos rodoviários de carga do tipo GRANELEIRO, para o transporte de grãos, e CAMINHÃO SIMPLES com 2 ou 3 eixos, para o transporte de mercadorias em geral. A base de dados utilizada para identificar tais quantidades foi obtida do banco de dados do RNTRC (Registro Nacional de Transportadores Rodoviários de Carga) da ANTT (Agência Nacional de Transportes Terrestres).

Sendo assim, este capítulo apresenta-se dividido em três partes: a primeira seção diz respeito à aplicação da metodologia aos veículos rodoviários de carga do tipo GRANELEIRO, e a segunda descreve as etapas para a comparação entre os modelos de regressão espacial e convencional. Por fim, a terceira seção faz uma segunda aplicação da metodologia aos veículos rodoviários de carga do tipo CAMINHÃO SIMPLES no Estado do Espírito Santo.

5.2. APLICAÇÃO DA METODOLOGIA I: VEÍCULOS RODOVIÁRIOS DE CARGA DO TIPO GRANELEIRO

Esta seção pretende desenvolver um modelo de regressão espacial, seguindo a metodologia proposta, para o veículo rodoviário de carga carroceria aberta do tipo GRANELEIRA. Além disso, pretende mostrar o ganho de adequabilidade, na existência de dependência espacial, do modelo de regressão espacial quando comparado ao modelo de regressão tradicional (seção 5.3.1). Para isso, a calibração dos modelos foi feita utilizando dados do ano 2000 e testados com os dados do ano de 2003.

Como apresentado no Capítulo 3, existem diversos softwares que já trabalham com os modelos de regressão espacial, no entanto, o programa utilizado nessa dissertação foi totalmente desenvolvido no software SAS 8.2, conforme Silva (2004). No Anexo II estão as demonstrações dos parâmetros do modelo e da matriz de informação de Fisher, utilizada para estimar o desvio padrão dos mesmos. Ambos fazem parte do algoritmo desenvolvido.

5.2.1. Etapa 1: Identificação do Objeto e das Variáveis

Foram aplicados questionários às empresas do setor de transportes, com objetivo de auxiliar na delimitação do objeto de estudo e na identificação de variáveis que influenciam na quantidade de veículos rodoviários de carga. O questionário aplicado encontra-se no Anexo I. As empresas respondentes disseram que acreditam na influência de características regionais na quantidade de veículos rodoviários de carga. Esse fato revela a importância de se incorporar o “espaço”, no caso o espaço geográfico, nos modelos matemáticos utilizados para previsão. Uma das empresas escreveu o seguinte:

Poderíamos citar diversas características, mas sugerimos um foco maior na questão do PIB, o desempenho regional / concentração de áreas agrícolas, crescimento da malha viária. Só como exemplo, o transporte rodoviário pesado está hoje fortemente concentrado em carga geral, o qual sofre forte influência da agricultura e da safra. Por outro lado, o desempenho industrial está fortemente ligado às frotas mais especializadas (logística). A questão é bastante ampla e de difícil resposta em poucas palavras. A sugestão é analisar a composição da frota rodoviária por segmentos, tomando por base a venda de implementos (Anfir), e associá-la às indústrias servidas e seus principais indicadores de crescimento.

Conforme informação do especialista, não faz sentido estudar a oferta do TRC sem distinguir o tipo de veículo rodoviário de carga, pois este se distribui espacialmente conforme uma demanda específica. Assim, a partir dos questionários e devido à sua relevância para o Brasil, optou-se estudar primeiramente os produtos em grãos como a soja, milho, trigo, arroz, feijão, amendoim etc. Segundo o MDICE (2006), os mesmos correspondem a aproximadamente 29% das exportações brasileiras em 2000 e a 28,12% das exportações em 2003, sendo que a soja representa 58,34% desse percentual em 2000 e 71,66% em 2003.

O RNTRC é composto por diversas variáveis, dentre as quais merecem destaque: placa do veículo, quantidade de eixo, capacidade de carga e tração, ano de fabricação, tipo do veículo (caminhão simples, caminhão trator, reboque, semi-reboque, veículo operacional de apoio e caminhonete furgão), tipo de carroceria (porta contêiner, granel sólido, carga seca, cegonha, baú simples, baú frigorífico, silo, tanque, sider e outros), localidade do veículo e localidade da transportadora (empresa, cooperativa, autônomo).

Nesse sentido, o objeto de estudo ou o veículo rodoviário de carga utilizado para estudar esses produtos foi do tipo carroceria aberta GRANELEIRA com 2 ou 3 eixos. A quantidade de caminhões tratores não foi levada em consideração, pois o que determina o produto transportado, no caso os produtos em grãos, é o tipo de carroceria. Porém, a quantidade de caminhões tratores está diretamente relacionada com a quantidade de carrocerias, por ser este último, um veículo rebocado.

Foram selecionadas também variáveis que, de alguma maneira, representam a necessidade do transporte de carga e que caracterizam a região (além de estarem disponíveis) como o PIB (Produto Interno Bruto), Receita Advinda do IPTU, Investimentos, Impostos, Quantidade de Estabelecimentos do Ramo do Comércio, Indústria e Construção, Densidade Populacional, Quantidade Produzida da Lavoura Temporária (Soja em Grão e Milho), Quantidade Produzida da Lavoura Permanente (Café) e Valor das Exportações.

5.2.2. Etapa 2: Montagem do Banco de Dados

A unidade espacial de análise considerada nessa pesquisa foi a divisão territorial brasileira. Como a pesquisa abrange todo o Brasil, trabalhar com todos os 5507 municípios (em

2000), levaria um maior tempo de processamento, além do viés de não espacialidade ocasionada pela grande heterogeneidade espacial. Esse último porque a grande quantidade de unidades espaciais ocasionaria uma dependência não contínua, ou seja, regiões com altos valores estariam cercadas por regiões com baixos valores, que por sua vez estariam cercadas por regiões de altos valores e assim por diante. Por esse motivo, optou-se por agrupar os dados, que originalmente foram obtidos segundo os municípios, em microrregiões (no total são 556). Todos os dados foram obtidos do IBGE e da SECEX (Secretaria de Comércio Exterior), além da quantidade de veículos rodoviários de carga que foi obtido no RNTRC.

Conforme a metodologia descrita no Capítulo 4, é importante verificar após a coleta dos dados, a qualidade do dado que está sendo utilizado. Na etapa anterior foram selecionadas 13 variáveis, segundo as respostas dos questionários e outras consideradas importantes. São elas:

1. **Frota** = Quantidade de carrocerias do tipo GRANELEIRA;
2. **Com** = Quantidade de estabelecimentos comerciais (geral, excluindo construção civil);
3. **Const** = Quantidade de estabelecimentos comerciais do ramo da construção civil;
4. **Dens** = Densidade Demográfica;
5. **Exp** = Valor das Exportações (US\$ - mil);
6. **Imp** = Valor arrecadado de impostos (R\$);
7. **Ind** = Quantidade de Estabelecimentos Industriais;
8. **Inv** = Valor gasto em Investimentos em geral (R\$);
9. **IPTU** = Valor arrecadado do IPTU (R\$);
10. **PIB** = Produto Interno Bruto (R\$);
11. **Soja** = Quantidade produzida de soja em grão (Tonelada);
12. **Milho** = Quantidade produzida de milho (Tonelada);
13. **Café** = Quantidade produzida de café (Tonelada);

É válido ressaltar que um problema aparente do RNTRC para um estudo espacial é a localização do veículo, pois a variável “localidade do veículo” é informada apenas para a Unidade da Federação (UF). Uma forma encontrada para corrigir esse problema foi utilizar a localização municipal da transportadora. Acredita-se que a maioria dos veículos se

encontra na sede da empresa, entretanto, devido à existência de filiais ou por outro motivo, alguns destes podem estar em localidades diferentes das matrizes.

A Tabela 5.1 apresenta os resultados do domínio, média, mediana, máximos e mínimos das variáveis selecionadas a fim de caracterizá-las.

Tabela 5.1. Estatísticas Descritivas das Variáveis Estudadas.

Variável	Domínio	Média	Mediana	Mínimo	Máximo
Frota	0 a ∞	55	15	1	911
Com	0 a ∞	5.108	2.395	127	220.599
Const	0 a ∞	276	65	4	14.898
Dens	0 a ∞	113,16	31,91	0,60	5.471
Exp	0 a ∞	130.771.632	10.953.097	0	5.752.106.580
Imp	0 a ∞	31.148.612	3.136.271	9.892	3.427.546.486
Ind	0 a ∞	1.066	407	17	53.911
Inv	0 a ∞	18.868.633	8.222.937	169.550	667.436.067
IPTU	0 a ∞	13.046.086	1.120.890	0	1.436.232.418
PIB	0 a ∞	29.847.953.974	25.637.157.334	8.881.506	125.963.966.801
Soja	0 a ∞	81.323	6	0	2.697.413
Milho	0 a ∞	76.465,66	28.630	0	1.538.168
Cafe	0 a ∞	8.516,31	45	0	206.882

De acordo com a Tabela 5.1, constata-se que todas as variáveis estão com seus valores dentro dos respectivos domínios. No entanto, existe uma discrepância muito grande entre a média e a mediana. Como o valor da mediana é menor que o da média, a distribuição das variáveis assume uma forma assimétrica à direita, conforme a Figura 5.1, que apresenta o histograma da variável **Frota**. Isso indica a presença de valores extremos positivos, o que é comprovado pelos valores de máximo e mínimo. Uma consequência imediata dessa assimetria é a não normalidade dos dados, o que pode interferir nos pressupostos de normalidade dos resíduos e de variância homogênea.

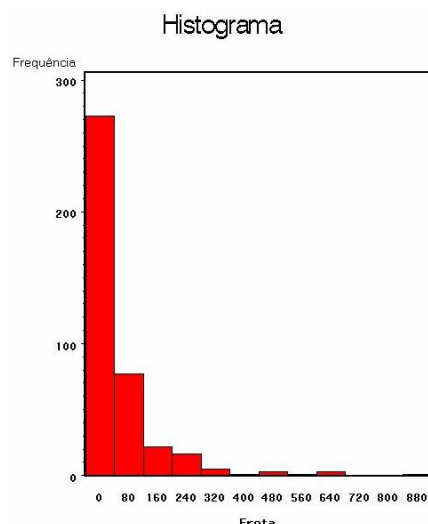


Figura 5.1: Histograma da Variável **Frota**.

As variáveis **Const**, **Exp**, **Imp**, **IPTU**, **Inv**, **Soja**, **Milho** e **Cafe** apresentaram valores faltantes ou *missing*. No caso das variáveis **Const**, **Exp**, **Soja**, **Milho** e **Cafe**, os valores faltantes podem ser substituídos por zero, porque a ausência desses indica que as microrregiões não possuem estabelecimentos de construção, empresas exportadoras e produção de soja, milho ou café, respectivamente. No caso das variáveis **Imp**, **IPTU** e **Inv**, os valores faltantes devem permanecer desconhecidos, pois todo município arrecada impostos, IPTU e recebe investimento, mesmo que esse último seja mínimo. A perda dessa informação, no entanto, não é significativa, pois existem apenas 3 observações (entre 402) nessa situação.

5.2.3. Etapa 3: Verificação de Dependência Espacial

No caso do transporte rodoviário de carga, a melhor forma que a matriz de proximidades poderia assumir é no que se refere às trocas comerciais (ou de produtos entre as regiões) ou o tempo necessário para ir de uma localidade a outra, de caminhão via rodovia, pois essas formas melhor evidenciariam a influência desse modo. Devido à dificuldade de aquisição dessas informações, as matrizes *W* utilizadas foram do tipo binária 1 e 0, indicando proximidade geográfica ou não entre as microrregiões, e do tipo distância euclidiana entre os centróides, conforme as formas 1 e 4 descritas na seção 3.2. Os centróides nesse caso foram os centros geométricos dos polígonos.

A Figura 5.2 apresenta a distribuição espacial da frota de carrocerias do tipo GRANELEIRA, sendo as quatro classificações utilizadas retiradas a partir dos quartis. Verifica-se uma grande aglomeração de carrocerias no centr-sul brasileiro, principalmente nos municípios próximos às rodovias de acesso aos portos de Santos e Paranaguá, como a BR-374 e a BR-116. No Mato Grosso e no Mato Grosso do Sul também existe uma forte tendência de proximidade à BR-163 que cruza esses dois estados. Essa distribuição se configura dessa forma por estarem ali concentrados os maiores produtores de soja no Brasil.

Distribuição Espacial da Frota de Carrocerias do Tipo Graneleira

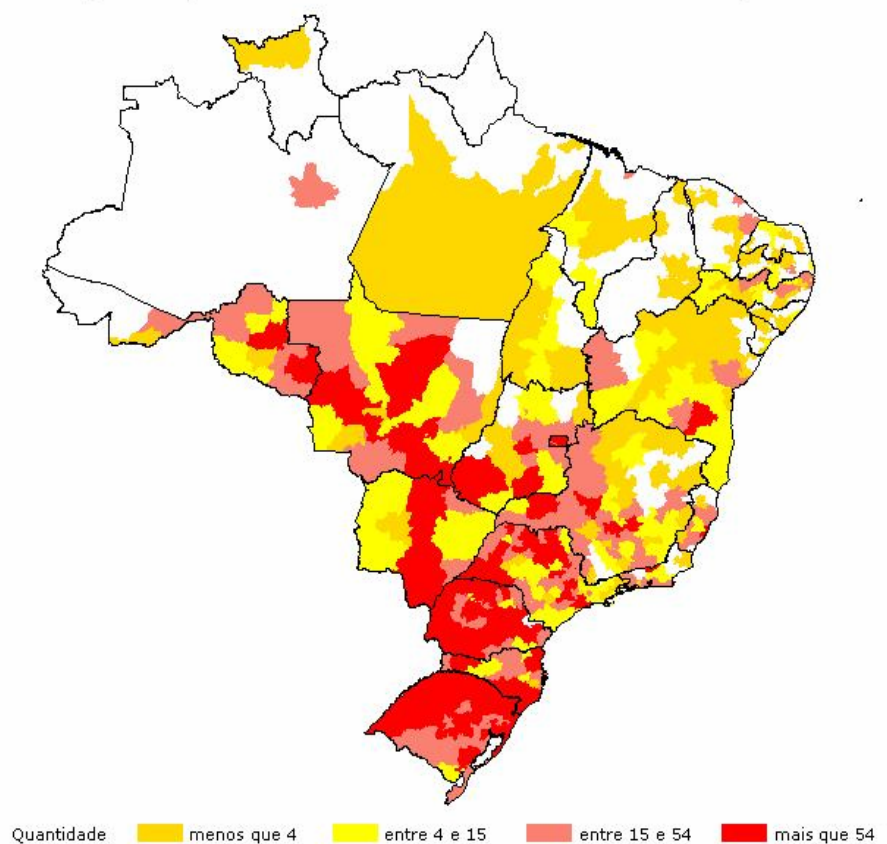


Figura 5.2: Distribuição Espacial da Frota de Carrocerias do Tipo GRANELEIRA.

As Figuras 5.3 a 5.6 mostram a distribuição espacial da quantidade produzida de soja, milho, café e dos três juntos (denominado “grãos”), respectivamente. Aparentemente a configuração espacial é muito parecida com a de carrocerias GRANELEIRAS, a menos da distribuição do café.

Distribuição Espacial da Quantidade de Soja

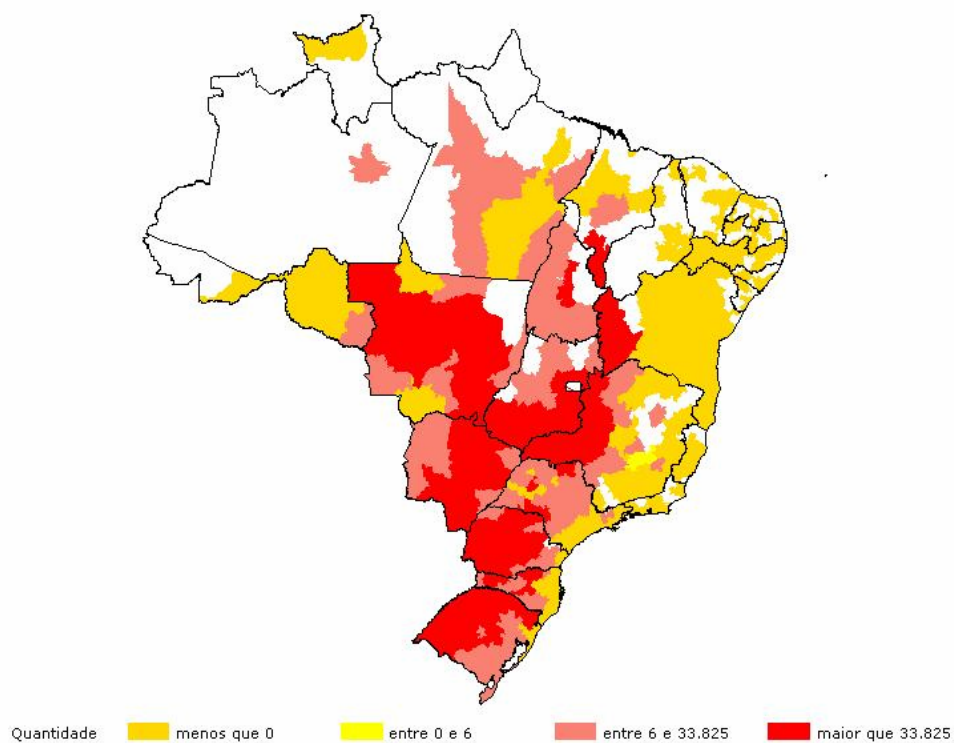


Figura 5.3: Distribuição Espacial da Quantidade de Soja Produzida (em Toneladas).

Distribuição Espacial da Quantidade de Milho

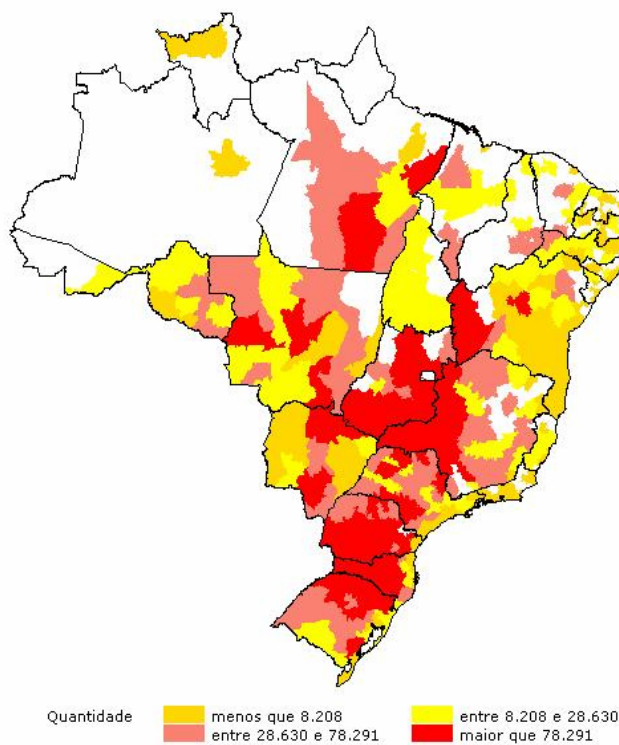


Figura 5.4: Distribuição Espacial da Quantidade de Milho Produzido (em Toneladas).

Distribuição Espacial da Quantidade de Café

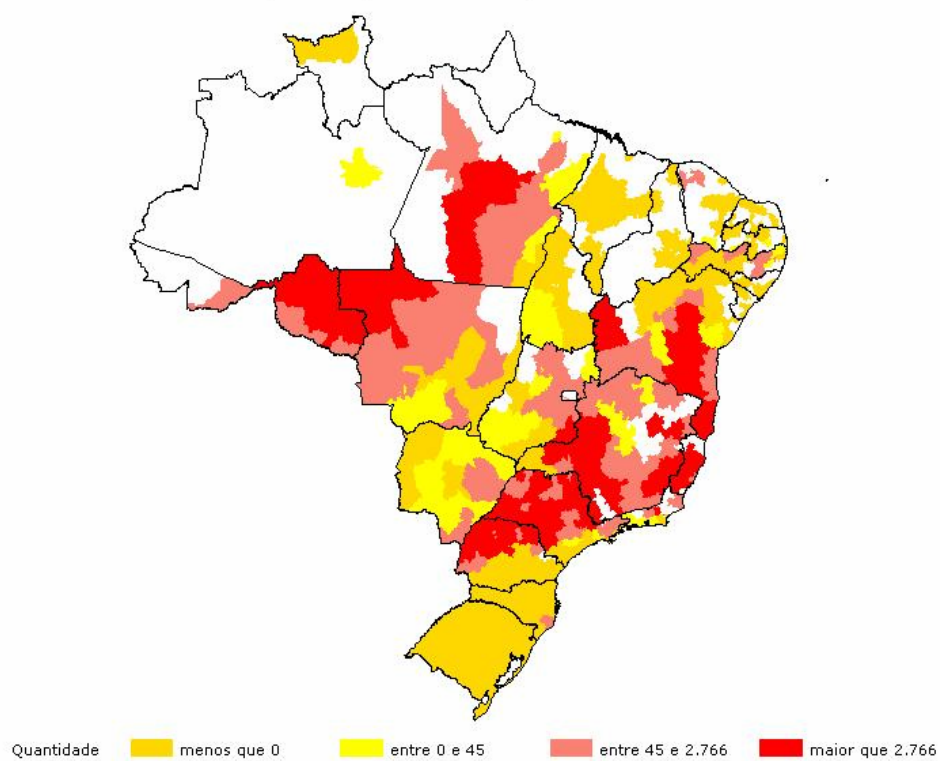


Figura 5.5: Distribuição Espacial da Quantidade de Café Produzido (em Toneladas).

Distribuição Espacial da Quantidade de Grãos

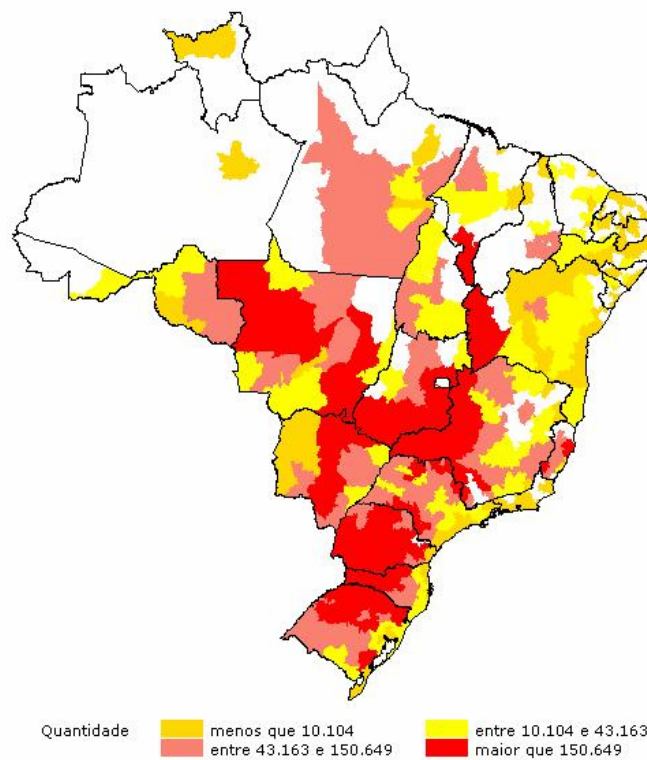


Figura 5.6: Distribuição Espacial da Quantidade de Grãos Produzido (em Toneladas).

Verifica-se, portanto, um indício de dependência espacial na quantidade de carrocerias do tipo GRANELEIRA (a partir daqui esta variável será chamada de **Frota**). Utilizando os 2 tipos de matriz de proximidades (Matriz A = binária e Matriz B = distância entre centróides), chega-se aos valores dos índices I de Moran e C de Geary apresentados na Tabela 5.2.

Tabela 5.2. Índices de Dependência Espacial da Variável **Frota**.

Matriz A	Matriz B
I = 0,22 (<0.001)	I = 0,09 (<0.001)
C = 0,79 (<0.001)	C = 0,88 (<0.001)

p-valor entre parênteses (Regra de decisão: aceita-se a hip. nula se p-valor >0.05)

A variável **Frota**, utilizando a Matriz A, apresentou uma dependência espacial significativa de 0,22 (I de Moran) e de 0,79 (C de Geary), no entanto esses valores indicam baixa dependência. Utilizando a Matriz B esses valores foram 0,09 e 0,88, respectivamente, indicando uma dependência espacial ainda menor. Os baixos valores dessa última se justificam pelo fato da matriz apresentar uma relação de dependência quase total entre todas as microrregiões, ou seja, a microrregião *i* utiliza um percentual da informação de todas as outras microrregiões (mesmo que seja mínimo). Isso faz com que o percentual utilizado dos seus “verdadeiros vizinhos” seja menor.

Uma maneira de corrigir esse problema é definir a Matriz B, segundo a forma 4 apresentada na seção 3.2, fazendo $w_{ij} = 1/(1 + d_{ij})$ apenas quando a distância entre os centróides de *i* e *j* for menor que uma certa distância *d*. Dessa forma, os “vizinhos” mais próximos terão um maior peso dentre aqueles que são os “verdadeiros vizinhos”. A Tabela 5.3 apresenta os índices I de Moran e C de Geary para diferentes distâncias.

Tabela 5.3. Índices de Dependência Espacial para Diferentes Distâncias entre Centróides.

Índice	d= 20 km	d= 25 km	d= 30 km	d= 35 km	d= 40 km	d= 45 km
I de Moran	0,09 (0.307)	0,18 (0.072)	0,18 (0.025)	0,15 (0.029)	0,23 (<0.001)	0,25 (<0.001)
C de Geary	0,38 (0.008)	0,23 (<0.001)	0,23 (<0.001)	0,38 (<0.001)	0,58 (<0.001)	0,66 (<0.001)

p-valor entre parênteses (Regra de decisão: aceita-se a hip. nula se p-valor >0.05)

Tabela 5.3. (Continuação)

Índice	d= 50 km	d= 55 km	d= 60 km	d= 65 km	d= 70 km	d= 80 km
I de Moran	0,35 (<0.001)	0,35 (<0.001)	0,31 (<0.001)	0,27 (<0.001)	0,26 (<0.001)	0,20 (<0.001)
C de Geary	0,59 (<0.001)	0,65 (<0.001)	0,69 (<0.001)	0,67 (<0.001)	0,65 (<0.001)	0,72 (<0.001)
Índice	d= 90 km	d= 100 km	d= 150 km	d= 200 km	d= 250 km	d= 300 km
I de Moran	0,20 (<0.001)	0,18 (<0.001)	0,18 (<0.001)	0,19 (<0.001)	0,18 (<0.001)	0,17 (<0.001)
C de Geary	0,82 (<0.001)	0,80 (<0.001)	0,79 (<0.001)	0,78 (<0.001)	0,79 (<0.001)	0,80 (<0.001)
Índice	d= 350 km	d= 400 km	d= 450 km	d= 500 km	d= 550 km	d= 600 km
I de Moran	0,17 (<0.001)	0,17 (<0.001)	0,16 (<0.001)	0,16 (<0.001)	0,16 (<0.001)	0,16 (<0.001)
C de Geary	0,80 (<0.001)	0,80 (<0.001)	0,81 (<0.001)	0,81 (<0.001)	0,82 (<0.001)	0,82 (<0.001)

p-valor entre parênteses (Regra de decisão: aceita-se a hip. nula se p-valor >0.05)

Verifica-se na Tabela 5.3 que o maior valor significativo para o índice de Moran ($I=0,35$) aconteceu para as distâncias iguais a 50 e 55 km, e à medida que as distâncias se afastavam desses valores, menor ficava o índice. A distância igual a 55 km foi escolhida devido ao fato do índice C de Geary estar mais estável a partir desse valor, ou seja, o que se espera é que quanto mais próximo de 1 está o índice I de Moran, mais próximo de 0 está o índice C de Geary (observe a discordância entre os valores de C para as distâncias menores que 55 km). Essa discordância pode ter acontecido devido à sensibilidade do índice C à quantidade de zeros da matriz W, por este ser mais sensível a poucos vizinhos. Vale advertir que quanto menor a distância, maior a quantidade de zeros na matriz W, o que pode descaracterizar a estrutura espacial.

Ressalta-se ainda que a matriz binária apresenta 98,8% dos seus valores iguais a zero (desconsiderando a diagonal que por definição é zero), enquanto que a matriz de distâncias entre centróides, calculada para distâncias menores que 55 km, apresenta 99,25% dos seus valores iguais a zero. A proximidade do número de zeros em relação à matriz binária acontece a partir de distâncias menores que 65 km (com distância igual a 65 km existem 98,92% de zeros).

Com o objetivo de auxiliar na caracterização da dependência espacial, as Figuras 5.7 a 5.12 apresentam as análises gráficas para as matrizes A e B.

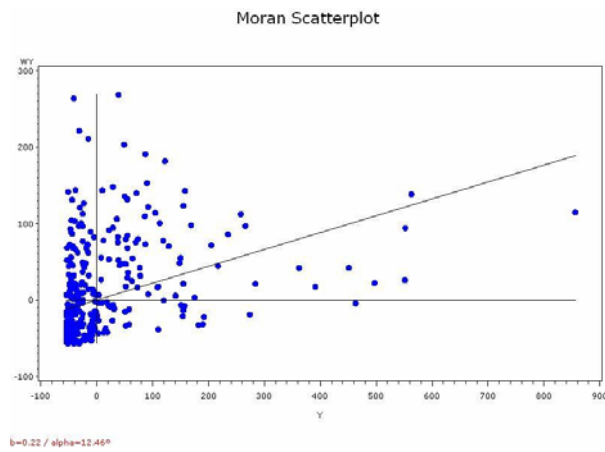


Figura 5.7: *Moran Scatterplot* da Variável **Frota** Utilizando a Matriz A.

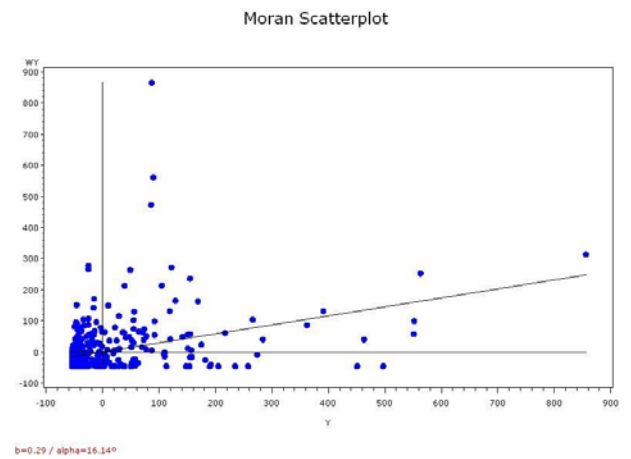


Figura 5.8: *Moran Scatterplot* da Variável **Frota** Utilizando a Matriz B.

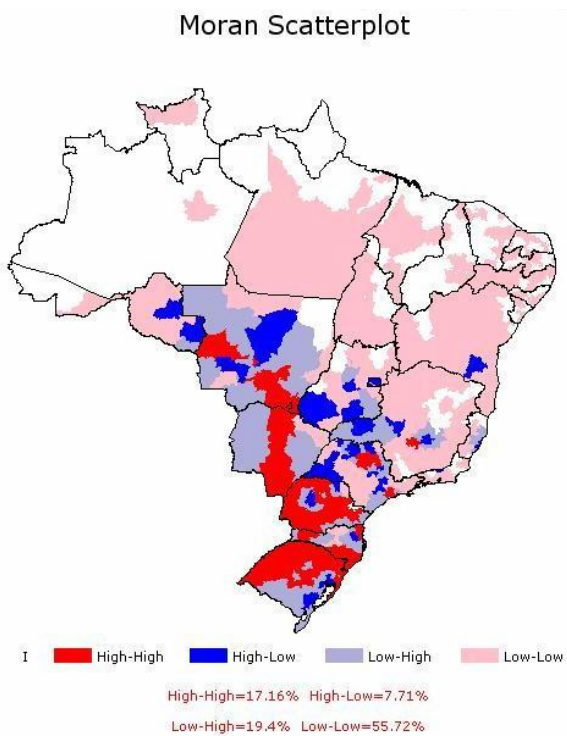


Figura 5.9: *Moran Scatterplot (Mapa)* da Variável **Frota** Utilizando a Matriz A.

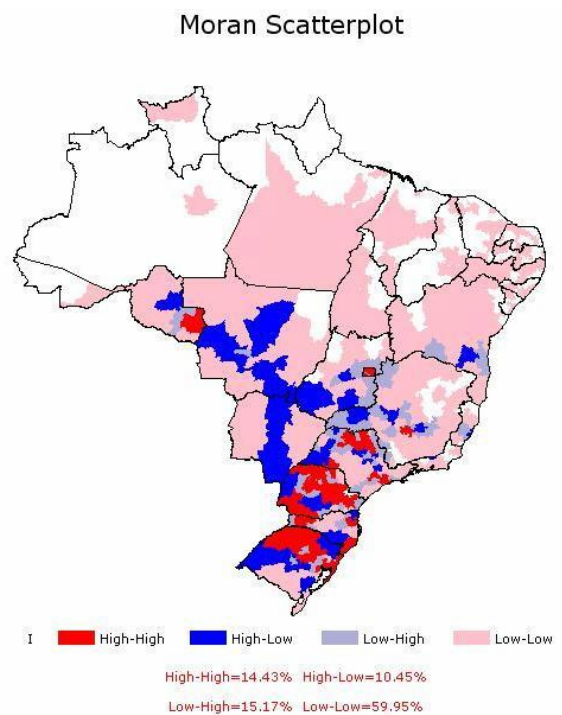


Figura 5.10: *Moran Scatterplot (Mapa)* da Variável **Frota** Utilizando a Matriz B.

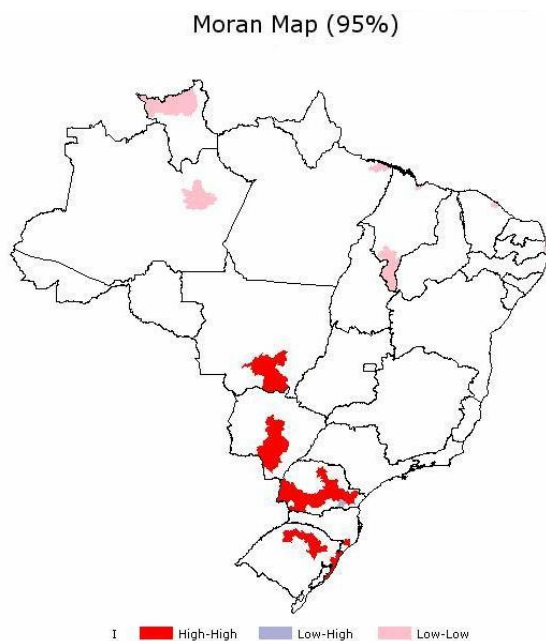


Figura 5.11: *Moran Map da Variável Frota Utilizando a Matriz A.*

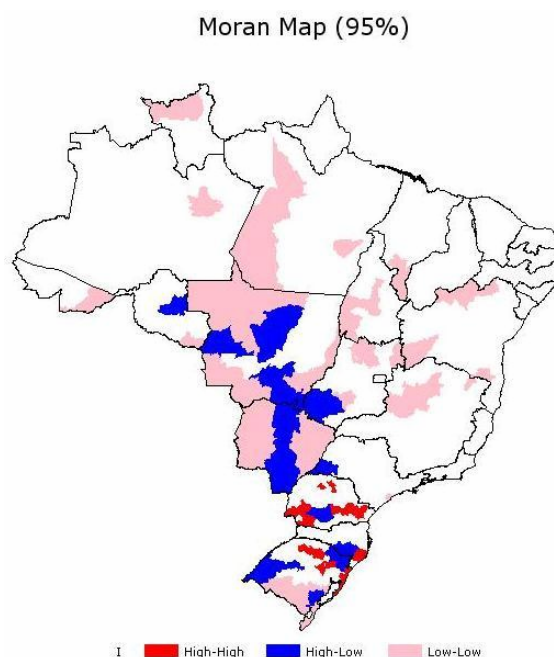


Figura 5.12: *Moran Map da Variável Frota Utilizando a Matriz B.*

Verifica-se na Figura 5.9 (a versão georreferenciada da Figura 5.7) que a quantidade de microrregiões classificadas como *High-High* e *Low-Low* – definidoras da dependência espacial – somam 72,88%, enquanto que na Figura 5.10 somam 74,38%, ou seja, conforme já indicado pelo índice I de Moran, a Matriz B melhor caracterizou a dependência espacial. É interessante observar também a mudança de classificação das microrregiões dos Estados do Mato Grosso e do Mato Grosso do Sul: *High-High* utilizando a Matriz A e *High-Low* utilizando a Matriz B. Por fim, a Matriz B apresentou uma maior heterogeneidade espacial, segundo a Figura 5.12 que destaca as microrregiões significativas, sugerindo um estudo diferenciado para essas regiões.

Devido à baixa dependência espacial apresentada pela variável **Frota** e também pela presença de valores extremos, decidiu-se aplicar o logaritmo na mesma variável, a fim de se verificar uma melhor caracterização da dependência espacial e a homogeneização dos dados. Tal transformação também se justifica pelo fato dos coeficientes terem uma interpretação econômica, ou seja, a elasticidade (Gujarati, 1995). Outro motivo para essa transformação é que segundo Assunção (2003), dados contínuos apresentam melhores resultados do que dados discretos (ou de contagem), como é o caso da quantidade de

carrocerias do tipo GRANELEIRA. Assim, a Tabela 5.4 apresenta os valores dos índices I de Moran e C de Geary para o logaritmo da variável **Frota**, agora denominada **Frota_{lg}**.

Tabela 5.4. Índices de Dependência Espacial da Variável **Frota_{lg}**.

Matriz A	Matriz B
I = 0,53 (<0.001)	I = 0,57 (<0.001)
C = 0,49 (<0.001)	C = 0,46 (<0.001)

p-valor entre parênteses (Regra de decisão: aceita-se a hip. nula se p-valor >0.05)

Verifica-se que a variável **Frota_{lg}**, utilizando a Matriz A, apresenta agora uma forte dependência espacial significativa de 0,53 (I de Moran) e de 0,49 (C de Geary). Utilizando a Matriz B, esses valores aumentaram para 0,57 e 0,46 respectivamente. Esse aumento se justifica porque o logaritmo homogeniza os dados, ou seja, agrupa os valores altos e separa os valores baixos. No entanto, vale lembrar que tal transformação é utilizada apenas para satisfazer os pressupostos da análise de regressão, necessitando posteriormente a volta aos dados originais. As Figuras 5.13 a 5.18 apresentam as análises gráficas para as matrizes A e B.

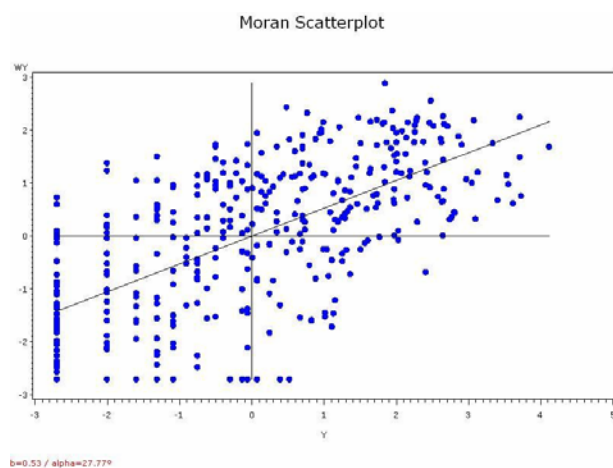


Figura 5.13: Moran Scatterplot da Variável **Frota_{lg}** Utilizando a Matriz A.

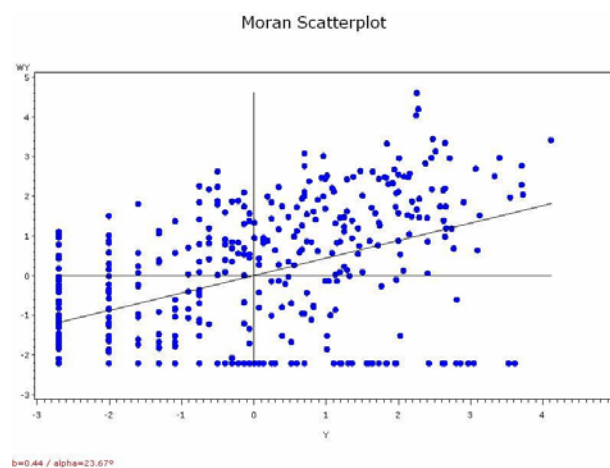


Figura 5.14: Moran Scatterplot da Variável **Frota_{lg}** Utilizando a Matriz B.

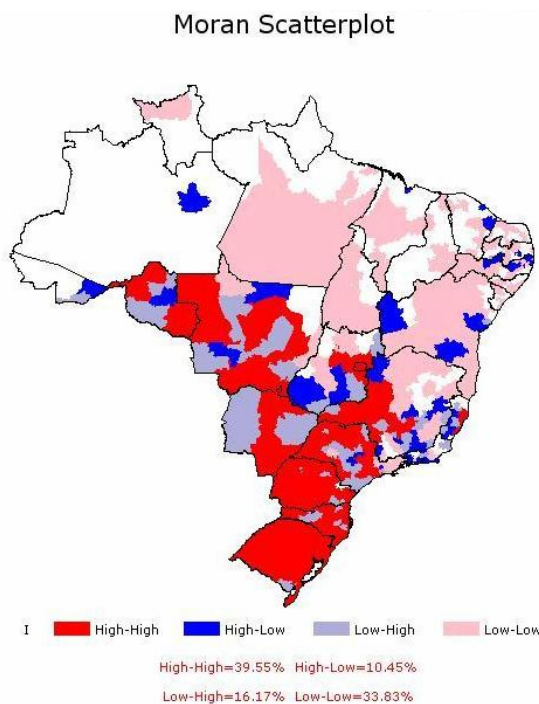


Figura 5.15: *Moran Scatterplot* (Mapa) da Variável **Frotalg** Utilizando a Matriz A.

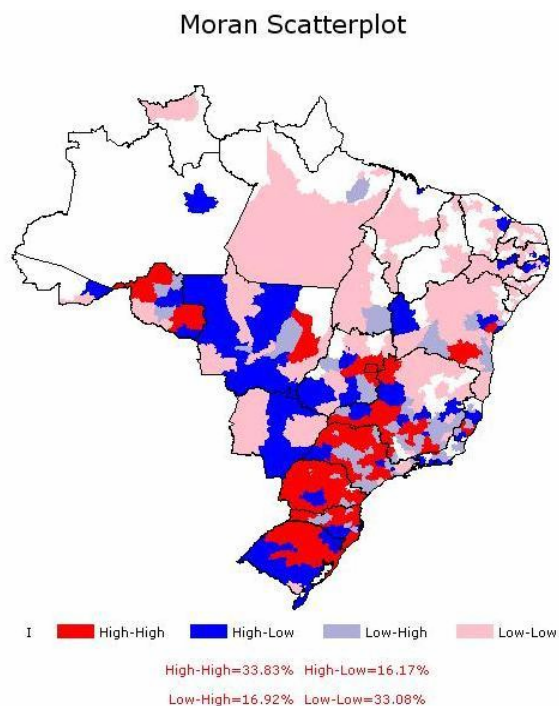


Figura 5.16: *Moran Scatterplot* (Mapa) da Variável **Frotalg** Utilizando a Matriz B.

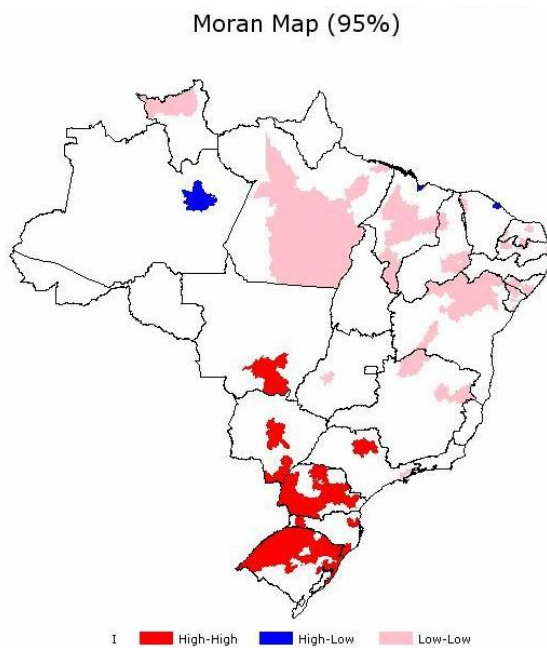


Figura 5.17: *Moran Map* da Variável **Frotalg** Utilizando a Matriz A.

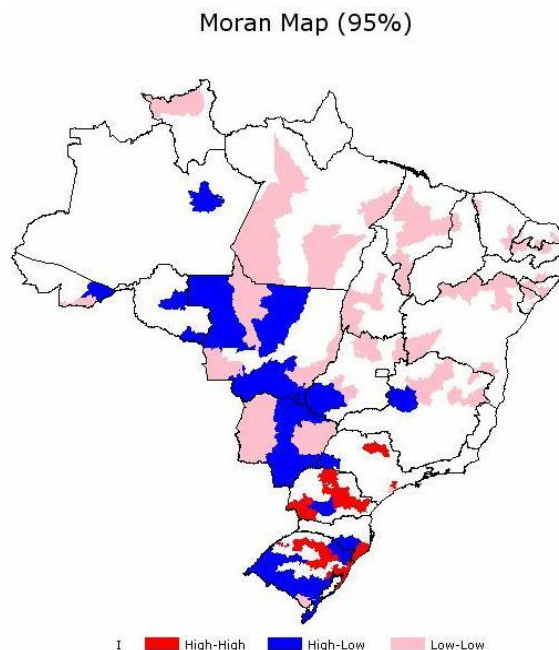


Figura 5.18: *Moran Map* da Variável **Frotalg** Utilizando a Matriz B.

Observa-se na Figura 5.13 e 5.14 a homogeneização dita anteriormente, ocasionada pela transformação dos dados. A Figura 5.15 mostra que a quantidade de microrregiões classificadas como *High-High* e *Low-Low* – definidoras da dependência espacial – somam

73,38%, enquanto que na Figura 5.16 somam 66,91%. Houve uma queda acentuada desse percentual quando utilizada a Matriz B, mostrando que existe uma certa instabilidade em tal matriz, quando comparada à Matriz A, que manteve seu percentual muito próximo ao anterior. Sobre a heterogeneidade espacial, ambas as matrizes apresentaram muitas regiões significativas, indicando padrões diferenciados.

Como foi constatada uma forte dependência espacial na frota de carrocerias do tipo GRANELEIRA (após a transformação logaritma), pode-se partir diretamente para calibração do modelo espacial, não necessitando nesse momento a verificação de dependência espacial no erro aleatório. Tal verificação será feita na etapa de calibração dos modelos.

5.2.4. Etapa 4: Calibração do Modelo

A etapa anterior mostrou que a variável **Frota_{lg}** apresentou um maior índice de dependência espacial. Dessa forma, o modelo a ser desenvolvido nesta etapa será do tipo log-log, ou seja, tanto a variável dependente quanto as independentes sofrerão a transformação logaritma. Como pretende-se comparar os modelos de regressão espacial e convencional, o modelo convencional sempre será calibrado juntamente com o espacial. Além disso, será utilizado um modelo espacial global, por este ser mais simples e pelo trabalho ter um caráter investigativo entre um modelo convencional e espacial. A Tabela 5.5 apresenta a matriz de correlação das variáveis selecionadas (já aplicada a função logaritma, inclusive na variável **Frota**).

Tabela 5.5. Matriz de Correlação das Variáveis Selecionadas.

	Frota	Com	Const	Dens	Exp	Imp	Ind	Inv	IPTU	PIB	Soja	Milho	Cafe
Frota	1.00	0.55	0.48	0.21	0.43	0.54	0.27	0.41	0.15	0.04	0.50	0.25	0.05
Com		1.00	0.92	0.64	0.48	0.91	0.54	0.85	0.41	0.12	0.06	0.00	0.12
Const			1.00	0.66	0.49	0.89	0.53	0.86	0.40	0.11	-0.01	-0.13	-0.03
Dens				1.00	0.28	0.65	0.43	0.61	0.36	-0.07	-0.31	-0.32	-0.07
Exp					1.00	0.50	0.18	0.42	0.11	-0.05	0.16	0.01	0.10
Imp						1.00	0.50	0.87	0.40	-0.05	0.06	-0.13	0.09

Tabela 5.5. (Continuação)

	Frota	Com	Const	Dens	Exp	Imp	Ind	Inv	IPTU	PIB	Soja	Milho	Cafe
Ind							1.00	0.50	0.91	0.03	-0.02	-0.04	-0.05
Inv								1.00	0.89	-0.01	-0.09	-0.13	-0.05
Iptu									1.00	-0.06	-0.08	-0.15	-0.05
Pib										1.00	0.04	0.21	-0.14
Soja											1.00	0.43	0.03
Milho												1.00	0.17
Cafe													1.00

Verifica-se que as variáveis **Com**, **Const**, **Dens**, **Exp**, **Imp**, **Ind**, **Inv**, **Soja** e **Milho** apresentam uma razoável correlação com a variável dependente **Frota**. No entanto, a variável **Com** apresenta forte correlação com as variáveis **Const**, **Dens** e **Imp**. O mesmo acontece com as variáveis **Dens** e **Ind** que apresentam alta correlação com as variáveis **Imp** e **Inv**, respectivamente. Tais correlações indicam um possível problema de multicolinearidade. Dessa forma, as variáveis candidatas a entrar no modelo são **Dens**, **Exp**, **Inv**, **Soja** e **Milho**.

Apesar da alta correlação das variáveis **Com** e **Const**, não existe uma relação positiva lógica (ou uma explicação) para a quantidade de carrocerias do tipo GRANELEIRA e a quantidade de estabelecimentos do ramo do comércio e da construção. Essas correlações indicam claramente um problema de correlação espúria, o que também pode ser visto na variável **Ind**.

Como visto na seção 5.2.2, as variáveis **Exp**, **Soja** e **Milho** apresentam valores nulos, o que indica a não existência de empresas exportadoras e a não produção de soja e milho na referida microrregião. Acontece que o logaritmo de zero não existe, e para que não houvesse perda de informação, o logaritmo dessas observações foi calculado para um valor muito próximo de zero (à medida que um número tende a 0, seu logaritmo tende para menos infinito ($-\infty$)).

O estudo foi feito dessa maneira, pois um dos objetivos dessa dissertação é a verificação de possíveis problemas nos dados, que ocasionariam resultados incorretos. Dessa forma, foram testadas as variáveis **Dens**, **Exp**, **Inv**, **Soja** e **Milho** em um modelo de regressão

convencional e espacial, onde os resultados dos modelos se encontram na Tabela 5.6. Antes disso, a Figura 5.19 mostra os gráficos de dispersão de todas as variáveis. Verifica-se que, com exceção dos valores repetidos à esquerda (devido ao cálculo do “logaritmo de 0”), existe uma tendência linear positiva entre a variável **Frota** e todas as outras variáveis.

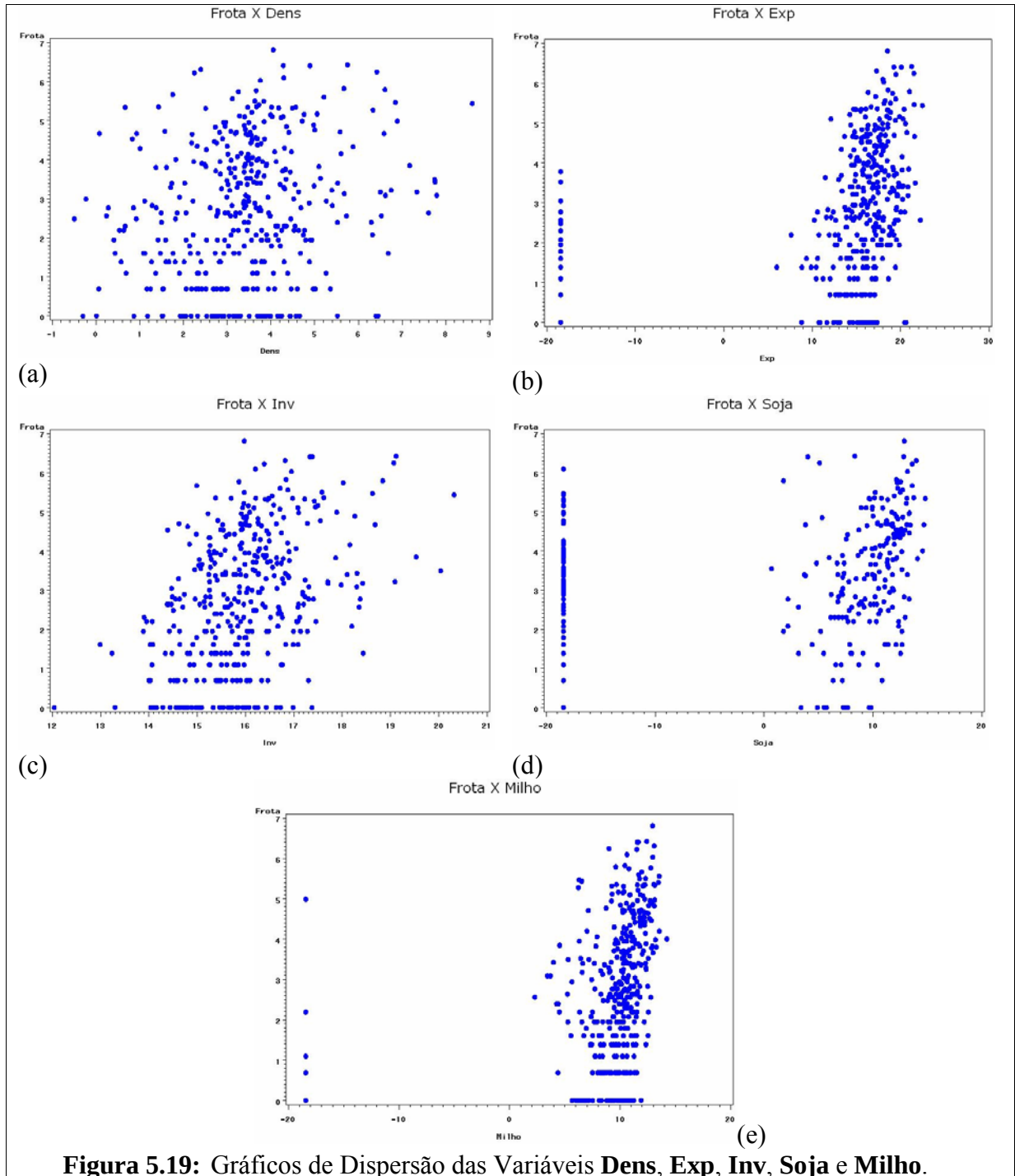


Tabela 5.6. Resultados do Modelo de Regressão Convencional e Espacial.

Variável	Regressão Convencional		Regressão Espacial MA		Regressão Espacial MB	
	Parâmetro	Teste t	Parâmetro	Teste t	Parâmetro	Teste t
Intercepto	-4,631* (<0,001)	-4,20	-6,926* (<0,001)	-7,27	-5,664* (<0,001)	-5,39
Dens	0,247* (<0,001)	4,04	0,143* (0,008)	2,65	0,101 ^{NS} (0,106)	1,62
Exp	0,029* (<0,001)	4,73	0,020* (<0,001)	3,64	0,027* (<0,001)	4,57
Inv	0,366* (<0,001)	4,73	0,463* (<0,001)	6,95	0,443* (<0,001)	6,00
Soja	0,060* (<0,001)	11,66	0,034* (<0,001)	6,87	0,052* (<0,001)	10,29
Milho	0,054* (<0,005)	2,83	0,043* (0,009)	2,60	0,037** (0,044)	2,01
ρ	-	-	0,451* (<0,001)	10,61	0,212* (<0,001)	5,80
R^2	0,493		0,620		0,539	
$R^2_{ajustado}$	0,486		0,615		0,533	
F	76,41*		128,48*		92,089*	
AIC	180,568		67,006		144,12	
White	22,32 ^{NS} (0,323)		51,53* (<0,001)		73,29* (<0,001)	
n	399		399		399	

(p-valor), (*) = muito significativo ($p < 0,01$), (**) = pouco significativo ($p < 0,05$), NS = não significativo

De acordo com a Tabela 5.6, verifica-se um ajuste razoável dos modelos (49% para o modelo convencional e 62% para o modelo espacial) e que todas as variáveis selecionadas foram significativas, exceto a variável **Dens** do modelo espacial que utilizou a Matriz B. Verifica-se também que a hipótese de homocedasticidade foi rejeitada nos modelos espaciais (conforme o teste de White), e que também pode ser comprovada no gráfico dos resíduos (Figuras 5.20 a 5.22), o mesmo não acontecendo para o modelo convencional. Fatalmente essa heterocedasticidade ocorreu em decorrência da transformação das variáveis **Exp**, **Soja** e **Milho** com valores nulos. Apesar disso, os três modelos apresentaram os resíduos normalmente distribuídos, segundo os testes de normalidade da Tabela 5.7 e dos testes gráficos *qq-plot* que estão nas Figuras 5.23 a 5.25.

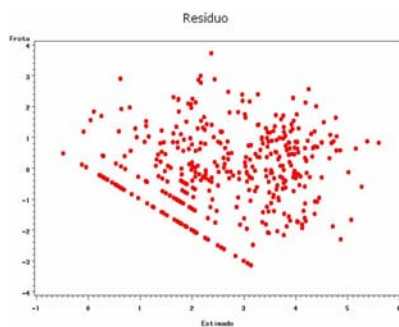


Figura 5.20: Resíduos da Regressão Convencional.

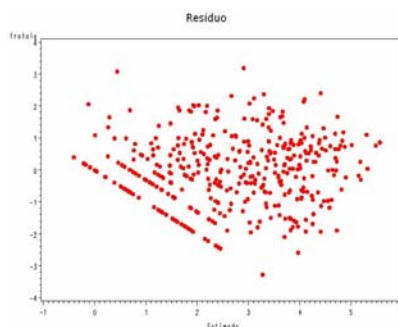


Figura 5.21: Resíduos da Regressão Espacial Utilizando a Matriz A.

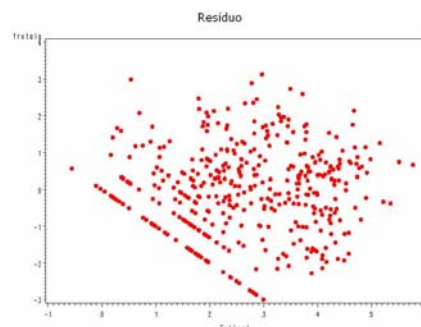


Figura 5.22: Resíduos da Regressão Espacial Utilizando a Matriz B.

Tabela 5.7. Testes de Normalidade para os Modelos de Regressão Convencional e Espacial.

Teste	p-valor (Convencional)	p-valor (Espacial – Matriz A)	p-valor (Espacial – Matriz B)
Shapiro-Wilk	0,469	0,542	0,295
Kolmogorov-Smirnov	0,150	0,150	0,150
Cramer-von Mises	0,250	0,250	0,240
Anderson-Darling	0,250	0,250	0,250

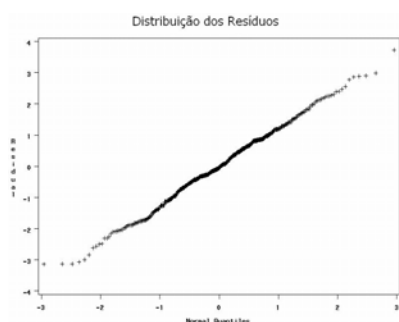


Figura 5.23: qq-plot dos Resíduos da Regressão Convencional.

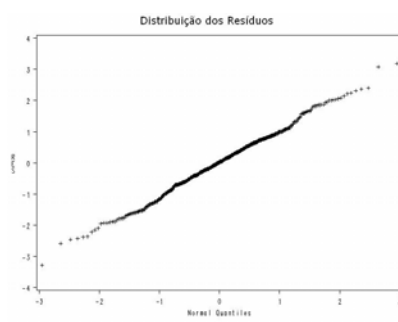


Figura 5.24: qq-plot dos Resíduos da Regressão Espacial Utilizando a Matriz A.

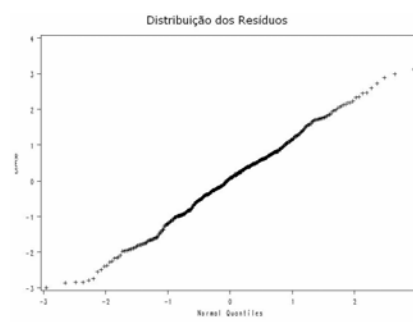
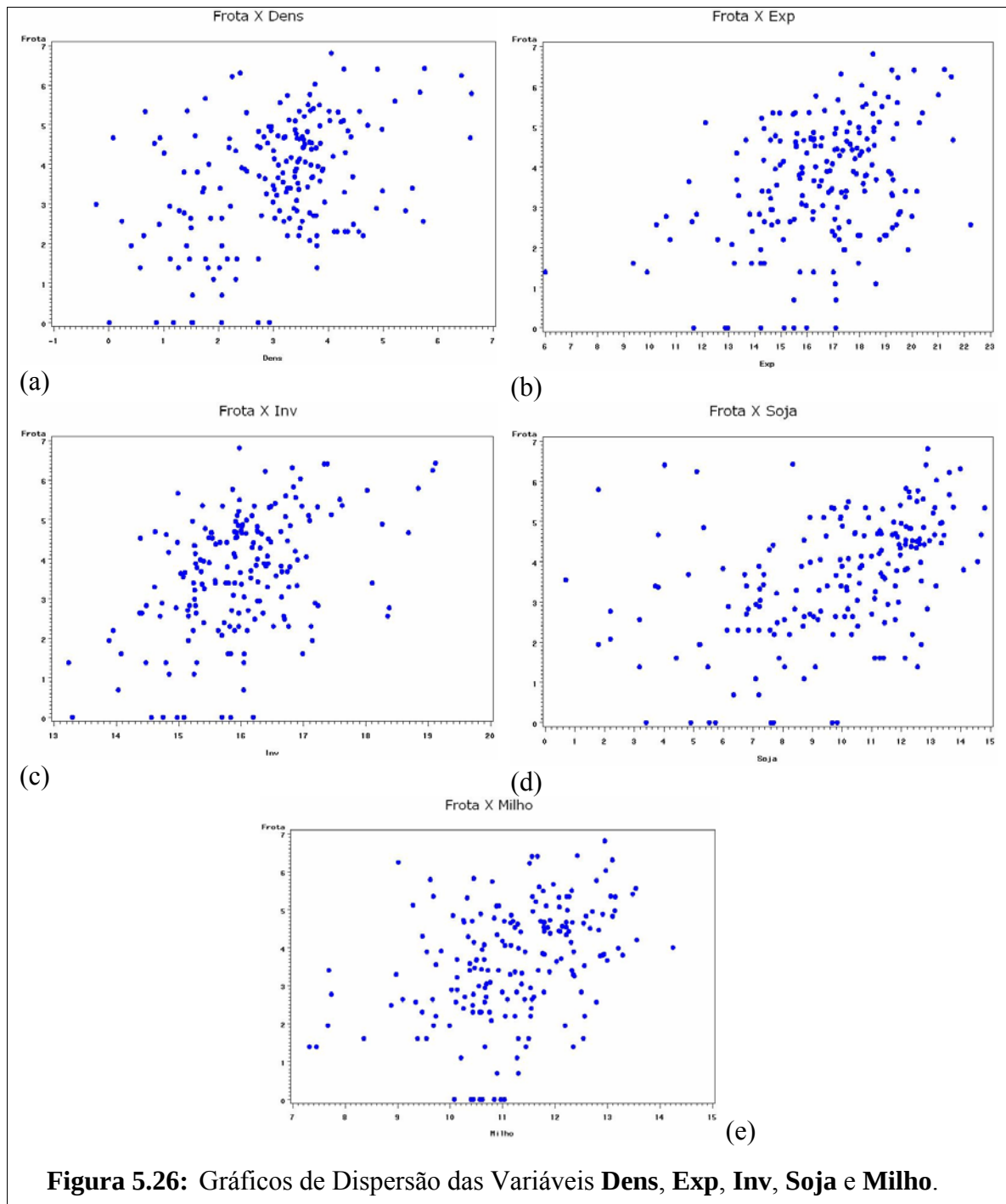


Figura 5.25: qq-plot dos Resíduos da Regressão Espacial Utilizando a Matriz B.

Retirando as localidades que não produzem soja, milho e que não exportam, a análise passa de 399 microrregiões analisadas para 189, ou seja, uma perda de informação de quase 50%. Porém, acredita-se que o modelo gerado será mais representativo. A Figura 5.26 apresenta os novos gráficos de dispersão, em que a linearidade entre a variável **Frota** e as variáveis **Dens**, **Exp**, **Inv**, **Soja** e **Milho** é mantida.



Como a quantidade de observações do modelo foi alterada, é necessário definir uma nova configuração para a Matriz B. Por isso foram calculadas e testadas novas distâncias, através dos testes I de Moran e C de Geary, onde os resultados se encontram na Tabela 5.8. Foi definida a distância inicial de 200 km porque as distâncias menores apresentaram estatísticas I de Moran não significativas.

Tabela 5.8. Índices de Dependência Espacial para Diferentes Distâncias entre Centróides.

Índice	d= 200 km	d= 210 km	d= 220 km	d= 230 km	d= 240 km	d= 250 km
I de Moran	0,27 (<0.001)	0,26 (<0.001)	0,25 (<0.001)	0,24 (0.029)	0,24 (<0.001)	0,24 (<0.001)
C de Geary	0,64 (<0.001)	0,66 (<0.001)	0,66 (<0.001)	0,66 (<0.001)	0,66 (<0.001)	0,67 (<0.001)
Índice	d= 260 km	d= 270 km	d= 280 km	d= 290 km	d= 300 km	d= 350 km
I de Moran	0,24 (<0.001)	0,25 (<0.001)	0,24 (<0.001)	0,24 (<0.001)	0,24 (<0.001)	0,24 (<0.001)
C de Geary	0,67 (<0.001)	0,68 (<0.001)	0,68 (<0.001)	0,68 (<0.001)	0,67 (<0.001)	0,68 (<0.001)
Índice	d= 400 km	d= 450 km	d= 500 km	d= 550 km	d= 600 km	d= 650 km
I de Moran	0,23 (<0.001)	0,28 (<0.001)	0,27 (<0.001)	0,26 (<0.001)	0,26 (<0.001)	0,25 (<0.001)
C de Geary	0,69 (<0.001)	0,69 (<0.001)	0,70 (<0.001)	0,70 (<0.001)	0,71 (<0.001)	0,72 (<0.001)

p-valor entre parênteses (Regra de decisão: aceita-se a hip. nula se p-valor >0.05)

Segundo a Tabela 5.8, a distância de 450 km apresentou o melhor percentual de dependência espacial (28% segundo o índice de Moran), e por essa razão será utilizada como a distância para a nova Matriz B. A dependência espacial utilizando a Matriz A, constatada pelo índice I de Moran, foi de 34%. Assim, a Tabela 5.9 apresenta os resultados dos modelos de regressão convencional e espacial para os novos dados e para a nova Matriz B. Antes disso, a Figura 5.27 apresenta a nova distribuição espacial das carrocerias do tipo GRANELEIRA.

Verifica-se na Figura 5.27 uma distribuição similar à apresentada pela Figura 5.2, diferenciando desta última por uma aglomeração mais forte no centro-sul do Brasil, e pela ausência de carrocerias GRANELEIRAS na Região Nordeste e no litoral brasileiro.

Distribuição Espacial da Frota de Carrocerias do Tipo Graneleira

onde Existe Produção de Soja, Milho e onde há Empresas Exportadoras

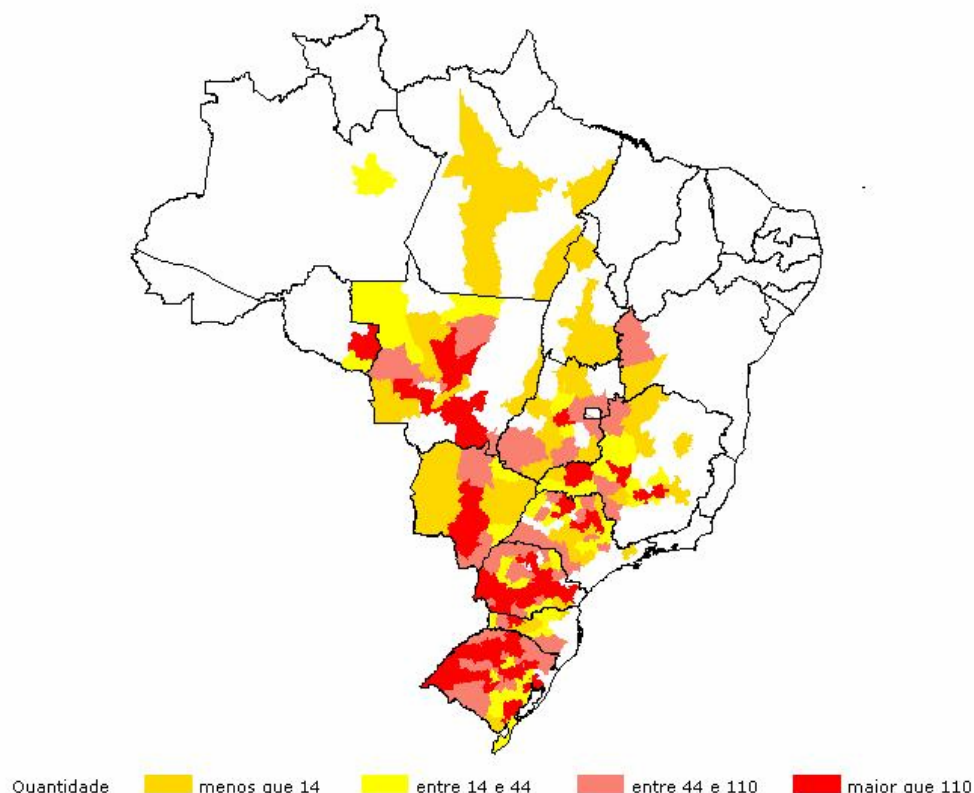


Figura 5.27: Distribuição Espacial da Frota de Carrocerias do Tipo GRANELEIRA, onde Existe Produção de Soja, Milho e onde há Empresas Exportadoras.

Tabela 5.9. Resultados do Modelo de Regressão Convencional e Espacial.

Variável	Regressão Convencional		Regressão Espacial MA		Regressão Espacial MB	
	Parâmetro	Teste t	Parâmetro	Teste t	Parâmetro	Teste t
Intercepto	-5,892* ($<0,001$)	-3,66	-8,541* ($<0,001$)	-5,88	-9,314* ($<0,001$)	-6,06
Dens	0,388* ($<0,001$)	4,59	0,216* (0,009)	2,61	0,182** (0,047)	1,99
Exp	0,099** (0,017)	2,40	0,093** (0,011)	2,55	0,091** (0,015)	2,43
Inv	0,274** (0,015)	2,44	0,414* ($<0,001$)	4,10	0,431* ($<0,001$)	4,10
Soja	0,283* ($<0,001$)	7,99	0,202* ($<0,001$)	5,94	0,208* ($<0,001$)	6,07
Milho	-0,042 ^{NS} (0,606)	-0,52	-0,004 ^{NS} (0,948)	-0,06	0,001 ^{NS} (0,989)	0,01
ρ	-	-	0,389* ($<0,001$)	5,99	0,528* ($<0,001$)	5,73

(p-valor), (*) = muito significativo ($p < 0,01$), (**) = pouco significativo ($p < 0,05$), NS = não significativo

Tabela 5.9. (Continuação)

	<i>Regressão Convencional</i>	<i>Regressão Espacial MA</i>	<i>Regressão Espacial MB</i>
R^2	0,505	0,600	0,584
$R^2_{ajustado}$	0,492	0,589	0,573
F	37,46*	55,01*	51,429*
AIC	33,608	-4,595	2,952
White	27,70 ^{NS}	27,39 ^{NS}	29,98 ^{NS}
	(0,116)	(0,158)	(0,092)
n	189	189	189

(p-valor), (*) = muito significativo ($p < 0,01$), (**) = pouco significativo ($p < 0,05$), NS = não significativo

De acordo com a Tabela 5.9, verifica-se um aumento considerável do parâmetro da variável **Soja**, quando comparada ao modelo anterior. Isso porque as localidades não produtoras de soja fizeram tal valor diminuir, a fim de caracterizá-las. Outro ponto a observar é que a variável **Milho** deixou de ser significativa nos três modelos e que pelo teste de White, os modelos passaram a ter os erros com variância constante (homocedasticidade), o que também pode ser comprovado pelo gráfico dos resíduos que estão nas Figuras 5.28 a 5.30. É interessante observar também que os resíduos dos modelos deixaram de seguir a distribuição normal (considerando um nível de significância de 5%), conforme a Tabela 5.10 e os testes gráficos *qq-plot* (Figura 5.31 a 5.33). Tal violação pode invalidar os resultados dos modelos (no que diz respeito ao desvio padrão dos parâmetros), no entanto, é importante recalibrá-los retirando a variável **Milho**.

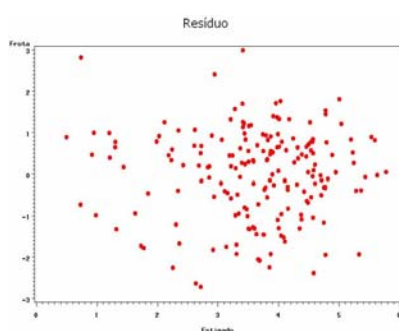


Figura 5.28: Resíduos da Regressão Convencional.

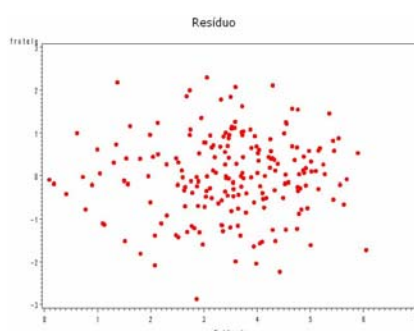


Figura 5.29: Resíduos da Regressão Espacial Utilizando a Matriz A.

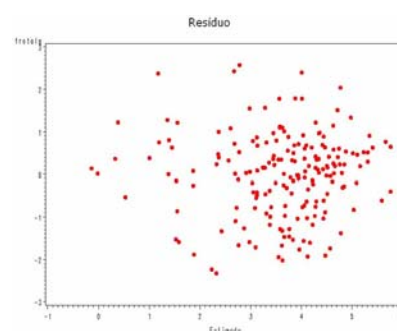


Figura 5.30: Resíduos da Regressão Espacial Utilizando a Matriz B.

Tabela 5.10. Testes de Normalidade para os Modelos de Regressão Convencional e Espacial.

Teste	p-valor (Convencional)	p-valor (Espacial – Matriz A)	p-valor (Espacial – Matriz B)
Shapiro-Wilk	0,013	0,038	0,043
Kolmogorov-Smirnov	0,012	0,019	0,023
Cramer-von Mises	0,005	0,016	0,010
Anderson-Darling	0,005	0,013	0,010

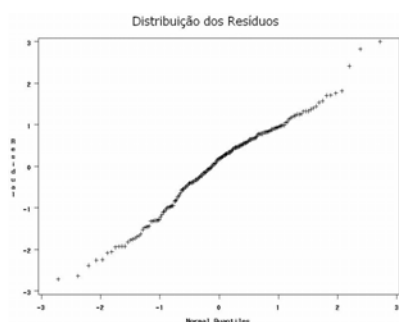


Figura 5.31: qq-plot dos Resíduos da Regressão Convencional.

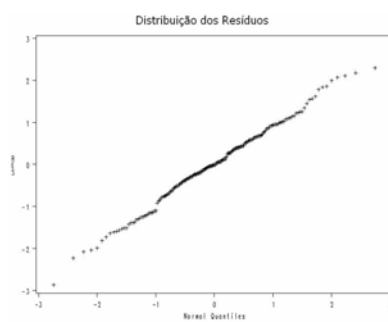


Figura 5.32: qq-plot dos Resíduos da Regressão Espacial Utilizando a Matriz A.

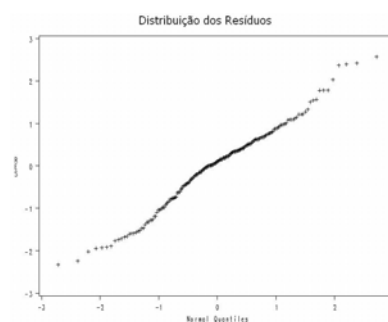


Figura 5.33: qq-plot dos Resíduos da Regressão Espacial Utilizando a Matriz B.

Os resultados dos modelos, convencional e espacial, com as variáveis **Dens**, **Exp**, **Inv** e **Soja**, encontram-se na Tabela 5.11.

Tabela 5.11. Resultados do Modelo de Regressão Convencional e Espacial.

Variável	Regressão Convencional		Regressão Espacial MA		Regressão Espacial MB	
	Parâmetro	Teste t	Parâmetro	Teste t	Parâmetro	Teste t
Intercepto	-6,200* (<0,001)	-4,16	-8,577* (<0,001)	-6,34	-9,307* (<0,001)	-6,44
Dens	0,379* (<0,001)	4,59	0,215* (0,008)	2,64	0,182** (0,044)	2,02
Exp	0,097** (0,018)	2,37	0,093** (0,011)	2,55	0,091** (0,015)	2,44
Inv	0,274** (0,015)	2,45	0,414* (<0,001)	4,10	0,431* (<0,001)	4,10
Soja	0,271* (<0,001)	9,99	0,201* (<0,001)	7,27	0,208* (<0,001)	7,52
ρ	-	-	0,390* (<0,001)	6,00	0,527* (<0,001)	5,73
R^2	0,505		0,600		0,584	
$R^2_{ajustado}$	0,494		0,591		0,575	

(p-valor), (*) = muito significativo ($p < 0,01$), (**) = pouco significativo ($p < 0,05$), NS = não significativo

Tabela 5.11. (Continuação)

	Regressão Convencional	Regressão Espacial MA	Regressão Espacial MB
F	46,952*	69,150*	64,637*
AIC	31,883	-6,603	0,954
White	22,96 ^{NS} (0,061)	22,97 ^{NS} (0,084)	27,87 ^{**} (0,022)
n	189	189	189

(p-valor), (*) = muito significativo ($p < 0,01$), (**) = pouco significativo ($p < 0,05$), NS = não significativo

De acordo com a Tabela 5.11, verifica-se que os modelos continuaram com os erros homocedásticos, à exceção do modelo espacial que utilizou a Matriz B, segundo o teste de White e o gráfico dos resíduos que estão nas Figuras 5.34 a 5.36. O critério de AIC reduziu para os três modelos e os resíduos, a um nível de significância de 2%, apresentam distribuição normal, com exceção do modelo convencional. Tais resultados podem ser vistos na Tabela 5.12 e nos testes gráficos *qq-plot* (Figura 5.37 a 5.39).

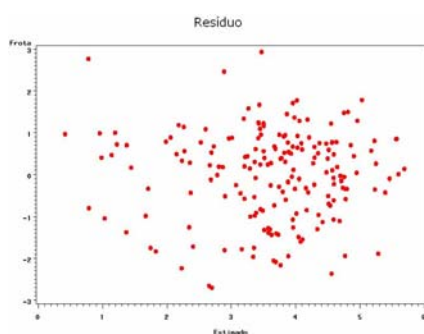


Figura 5.34: Resíduos da Regressão Convencional.

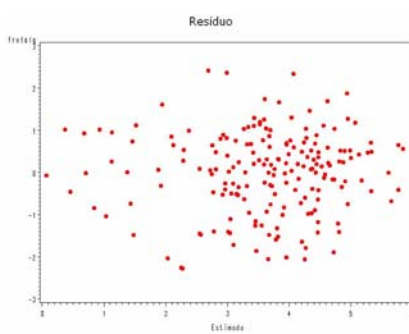


Figura 5.35: Resíduos da Regressão Espacial Utilizando a Matriz A.

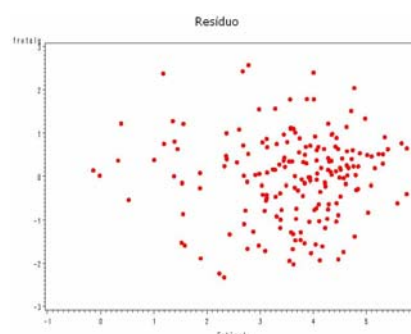


Figura 5.36: Resíduos da Regressão Espacial Utilizando a Matriz B.

Tabela 5.12. Testes de Normalidade para os Modelos de Regressão Convencional e Espacial.

Teste	p-valor (Convencional)	p-valor (Espacial – Matriz A)	p-valor (Espacial – Matriz B)
Shapiro-Wilk	0,010	0,039	0,042
Kolmogorov-Smirnov	0,010	0,027	0,023
Cramer-von Mises	0,005	0,017	0,010
Anderson-Darling	0,005	0,013	0,010

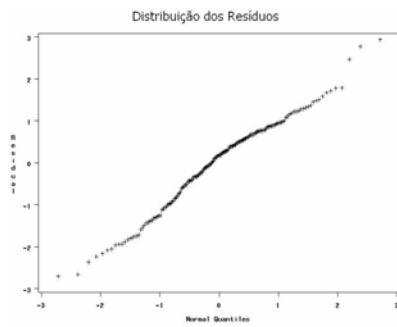


Figura 5.37: *qq-plot* dos Resíduos na Regressão Convencional.

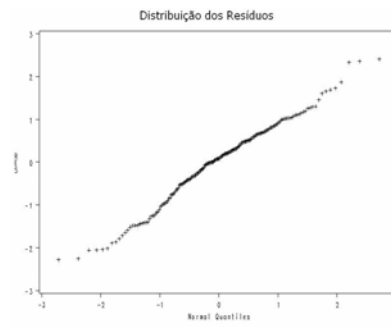


Figura 5.38: *qq-plot* dos Resíduos na Regressão Espacial Utilizando a Matriz A.

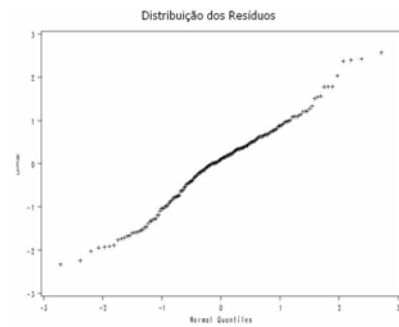


Figura 5.39: *qq-plot* dos Resíduos na Regressão Espacial Utilizando a Matriz B.

O modelo com autocorrelação espacial no erro também foi testado, e seus resultados foram semelhantes ao modelo espacial misto, porém o critério de AIC apresentou um maior valor (AIC igual a 36,907). O modelo geral apresentou um valor de AIC igual a -8,367, mas o parâmetro espacial λ (coeficiente espacial do erro) foi não significativo. Por esse motivo decidiu-se não utilizar tais modelos. De posse dos resultados dos modelos de regressão convencional e espacial, o próximo passo é a comparação entre os mesmos, a fim de verificar a diferença entre eles.

5.3. COMPARAÇÃO ENTRE OS MODELOS ESPACIAL E CONVENCIONAL

Esta seção tem por objetivo indicar algumas formas para comparar os resultados dos modelos de regressão convencional e espacial, na existência de dependência espacial, a fim de verificar diferenças significativas na utilização dos mesmos. A primeira comparação pode ser feita a partir do valor do coeficiente de determinação (R^2). Se as suposições dos modelos foram cumpridas e não existirem problemas aparentes, então o melhor modelo apresentará um maior valor de R^2 (a mesma interpretação vale para o $R^2_{ajustado}$).

Aliado ao coeficiente de determinação (R^2) estão os testes F e t. Quanto mais explicativo for o modelo, ou seja, maior o R^2 , mais significativos, mas não necessariamente apenas por esse motivo, serão esses testes. Os resíduos também devem ser avaliados para verificação de uma maior aleatoriedade em torno do zero.

A comparação mais robusta entre modelos é feita a partir dos critérios de AIC (*Akaike Information Criterion*) e/ou SBC (*Schwarz Bayesian Criterion*) cujas fórmulas são dadas por:

$$AIC = T \ln (\text{soma do quadrado dos resíduos}) + 2n$$

$$SBC = T \ln (\text{soma do quadrado dos resíduos}) + n \ln(T)$$

Onde,

n = número de parâmetros estimados;

T = número de observações utilizadas.

O melhor modelo é aquele que apresentar o menor valor de AIC e/ou SBC. Mais detalhes podem ser vistos em Akaike (1974).

Outra comparação que pode ser feita é através do ERM (Erro Relativo Médio) que avalia os resultados das estimativas. O melhor modelo para previsão é aquele que apresentar o menor valor ERM. A fórmula é dada pela Equação 5.1:

$$ERM = \frac{\sum_{i=1}^n \frac{|ER_i - Yref_i|}{Yref_i}}{n} \quad (5.1)$$

onde:

ERM = erro relativo médio;

ER_i = valor estimado pelo modelo;

$Yref_i$ = valor real;

n = número de observações;

Por fim, a última comparação entre os modelos de regressão espacial e tradicional é a análise de cenários. Segundo Coates (2000), um cenário é um estudo prospectivo acerca do futuro aliado à organização das informações obtidas, de modo a oferecer um conjunto de informações coerente, sistemático, plausível e compreensível, com o objetivo de descrever um determinado evento e oferecer instrução e suporte à tomada de decisões. Se a base para criação de cenários não está correta, então de nada serve essa previsão para o futuro. Dessa forma, a análise de cenários é feita a partir da comparação entre os parâmetros estimados dos modelos, espacial e tradicional, e os dados reais, tendo como objetivo a busca pelo modelo que apresente um menor erro de previsão.

Os cenários, na técnica de análise de regressão, são construídos a partir dos parâmetros estimados (Gujarati, 1995). A interpretação é feita da seguinte forma: aumentando 1 unidade na variável “x” (independente), a variável “y” (dependente) é acrescida de β unidades (parâmetro estimado correspondente à variável x). No caso de um modelo log-log (é aplicada a função logarítima na base “e” tanto na variável dependente quanto nas independentes), a interpretação toma a forma percentual: aumentando 1% na variável “x”, a variável “y” é acrescida de $\beta\%$. A transformação logarítima é amplamente utilizada por fornecer a elasticidade diretamente no parâmetro β , além de adequar os dados à distribuição normal.

5.3.1. Estudo de caso: Carrocerias GRANELEIRAS

A comparação foi feita utilizando os resultados do último modelo apresentado na seção 5.2.5, ou seja, onde existe produção de soja. As estatísticas R^2 , F e AIC servem para comparar os modelos quanto à adequabilidade aos dados, enquanto que a estatística ERM (erro relativo médio) serve para comparar os mesmos quanto ao poder de previsão. Tais estatísticas estão reproduzidas na Tabela 5.13.

Tabela 5.13. Estatísticas de Ajuste dos Modelos de Regressão Convencional e Espacial.

<i>Estatísticas</i>	<i>Modelo Convencional</i>	<i>Modelo Espacial Utilizando a Matriz A</i>	<i>Modelo Espacial Utilizando a Matriz B</i>
R^2	0,505	0,600	0,584
$R^2_{ajustado}$	0,494	0,591	0,575
F	46,95	69,15	64,64
AIC	31,883	-6,603	0,954
ERM (2000)	0,286	0,286	0,290
ERM (2003)	0,268	0,293	0,284

Verifica-se na Tabela 5.13 um ganho de adequabilidade do ajuste por parte do modelo espacial. O R^2 passou de 0,505 no modelo convencional para 0,600 no modelo espacial que utilizou a Matriz A, ou seja, um aumento de 18%. O valor do R^2 e do $R^2_{ajustado}$ não foram muito discordantes devido à pouca quantidade de variáveis explicativas. Uma maior diferença entre os mesmos pode ser vista quando existir uma maior quantidade de variáveis independentes.

A estatística F também teve um aumento de 47%, passando de 46,95 no modelo convencional para 69,15 no modelo espacial que utilizou a Matriz A. Esse aumento está relacionado diretamente com a maior explicação por parte do modelo espacial, ou seja, o “espaço” – no caso o geográfico – que antes estava no erro aleatório agora passa a ter um papel fundamental na explicação da quantidade de veículos rodoviários de carga do tipo GRANELEIRO.

A comparação mais robusta entre os dois modelos está no critério de AIC, que passou de 31,883 no modelo convencional para -6,603 no modelo espacial que utilizou a Matriz A. Verifica-se no modelo espacial um menor AIC, o que confirma a melhor adequabilidade. Em relação aos modelos espaciais, verifica-se que o modelo que utilizou a matriz binária (Matriz A) produziu sempre melhores resultados do que o que utilizou a matriz de distâncias entre centróides (Matriz B). Esse fato revela que o simples compartilhamento de fronteiras entre as microrregiões já é o suficiente para ambas se desenvolverem, não importando muito quão distante seus centróides estão.

Comparando os modelos através do ERM (erros relativos médios) para o ano 2000, verifica-se que o modelo espacial que utilizou a Matriz A apresentou o mesmo valor que o modelo convencional, e que o modelo espacial que utilizou a Matriz B apresentou o maior valor. Em 2003, o modelo convencional apresentou o menor valor e o modelo espacial que utilizou a Matriz A, o maior valor. Acontece que de 2000 a 2003 o Brasil diversificou muito sua pauta de exportações, o que pode ter descalibrado o modelo no que se refere aos produtos exportados. Para se ter uma idéia, se a frota for estimada sem o valor das exportações e da quantidade de soja produzida, os ERM passam para 0,904 no modelo espacial e 1,192 no modelo convencional, mostrando a influência do “espaço” na realização dos eventos.

Terminada a análise de adequabilidade dos modelos e o poder de previsão, o próximo passo é estudar cada parâmetro individualmente. Dessa forma, encontram-se na Tabela 5.14 os parâmetros estimados dos modelos convencional e espacial, necessários para a análise de cenários.

Tabela 5.14. Parâmetros dos Modelos de Regressão Convencional e Espacial.

<i>Variável</i>	<i>Regressão Convencional</i>		<i>Regressão Espacial MA</i>		<i>Regressão Espacial MB</i>	
	<i>Parâmetro</i>	<i>Teste t</i>	<i>Parâmetro</i>	<i>Teste t</i>	<i>Parâmetro</i>	<i>Teste t</i>
Intercepto	-6,200*	-4,16	-8,577*	-6,34	-9,307*	-6,44
Dens	0,379*	4,59	0,215*	2,64	0,182**	2,02
Exp	0,097**	2,37	0,093**	2,55	0,091**	2,44
Inv	0,274**	2,45	0,414*	4,10	0,431*	4,10
Soja	0,271*	9,99	0,201*	7,27	0,208*	7,52
ρ	-	-	0,390*	6,00	0,527*	5,73

(*) = muito significativo ($p < 0,01$), (**) = pouco significativo ($p < 0,05$)

Verifica-se que o parâmetro da variável **Inv** mudou de 0,274 do modelo convencional para 0,414 do modelo espacial que utilizou a Matriz A. Isso quer dizer que para um aumento de 1% no valor dos investimentos, a quantidade de carrocerias do tipo GRANELEIRA aumentaria em 0,274% no modelo convencional e 0,414% no modelo espacial. Fazendo o intervalo de confiança dos parâmetros (0,054; 0,494) no convencional e (0,215; 0,613) no espacial, verifica-se que estatisticamente os dois valores são iguais, o que também é confirmado pelo teste t que apresentou estatística igual a -0,135 e p-valor igual 0,892. Mesmo assim é válido avaliar o valor das estimativas pontuais, conforme apresentado a seguir.

Tome por exemplo o investimento recebido em Vilhena/RO: em 2000 Vilhena recebeu R\$ 4.791.246,31 de investimentos, e em 2003 recebeu R\$ 18.071.459,90, ou seja, um aumento de R\$13.280.213,59 ou 277,17%. Como em 2000 a microrregião possuía 212 veículos rodoviários de carga do tipo GRANELEIRO, fazendo $277,17 \times 0,274$ tem-se um aumento de 75,94% da frota, ou seja, 161 carrocerias do tipo GRANELEIRA. No caso do modelo espacial, esse aumento seria de 114,75% ou 243 veículos. Não se sabe de fato qual o aumento real, mas acredita-se que devido ao melhor ajuste do modelo espacial, mais correto está o resultado deste último. Isso porque, caso o município invista em infraestrutura, melhores condições irá fornecer para os produtores transportarem seus produtos, ou seja, quanto maior a produção maior a quantidade de veículos necessários para o transporte. É importante ressaltar que a validade dos resultados está condicionada ao pressuposto da localização da frota no município produtor.

Analisando agora a variável **Soja**, verifica-se o valor de 0,271 no modelo convencional e 0,201 no modelo espacial, ou seja, aumentando 1% na quantidade produzida de soja

(tonelada), a quantidade de carrocerias do tipo GRANELEIRA aumentaria em 0,271% no modelo convencional e 0,201% no modelo espacial. Esse resultado parece estar de acordo com a realidade, pois a quantidade de carrocerias GRANELEIRA não está totalmente voltada para o transporte de soja, daí um menor parâmetro por parte do modelo espacial. Pode-se dizer que o fator “espaço” – geográfico – nesse caso corrigiu o parâmetro, pois a produção de uma localidade pode ser transportada por uma frota alocada em uma microrregião vizinha ou próxima, não necessitando, portanto, um número maior de veículos rodoviários de carga para o transporte.

Fazendo o intervalo de confiança dos parâmetros (0,218; 0,324) no convencional e (0,148; 0,254) no espacial, verifica-se que estatisticamente os dois valores também são iguais, o que novamente é confirmado pelo teste t que apresentou estatística igual a 0,266 e p-valor igual 0,789. Como no caso da variável **Inv**, é importante avaliar o valor das estimativas pontuais, a fim de se verificar diferenças.

Para exemplificar, seja a quantidade de soja produzida em Vilhena/RO: em 2000 Vilhena produziu 25.500 Toneladas de soja, e em 2003 produziu 95.640, ou seja, um aumento de 70.140 ou 275,06%. Como em 2000 a microrregião possuía 212 veículos rodoviários de carga do tipo GRANELEIRO, fazendo $275,06 \times 0,271$ tem-se um aumento de 74,54% da frota, ou seja, 158 carrocerias do tipo GRANELEIRA. No caso do modelo espacial, esse aumento seria de 55,29% ou 117 veículos. Partindo da premissa que o modelo espacial produz resultados mais realistas, teria uma subutilização de 41 veículos, ou seja, 35% a mais de carrocerias do tipo GRANELEIRA.

Verifica-se, portanto, que na existência de dependência espacial em um modelo de regressão convencional, o fator “espaço” é distribuído entre as variáveis, gerando uma valoração incorreta dos parâmetros: subestimando no caso do investimento municipal e superestimando no caso da produção de soja. Esse foi um tipo de cenário construído para o transporte rodoviário de carga, e para que não sejam feitas conclusões precipitadas em apenas 1 estudo de caso, decidiu-se estudar também os veículos rodoviários de carga do tipo CAMINHÃO SIMPLES.

5.4. APLICAÇÃO DA METODOLOGIA II: VEÍCULOS RODOVIÁRIOS DE CARGA DO TIPO CAMINHÃO SIMPLES

Os CAMINHÕES SIMPLES são amplamente utilizados nas cidades para transportarem mercadorias em curtas distâncias, devido ao seu menor porte. Normalmente servem para fazer entregas de produtos em supermercados, materiais de construção e distribuição interna de produtos advindo das indústrias. Na maioria das vezes, o transporte desse último para outras cidades é feito pelas carretas, por causa da maior capacidade de carga.

O estudo seguiu a mesma lógica das carrocerias do tipo GRANELEIRA, no entanto, não serão apresentados todos os detalhes e as etapas da metodologia. Verificou-se a dependência espacial dos CAMINHÕES SIMPLES com 2 ou 3 eixos em todo o Brasil, porém o Estado do Espírito Santo apresentou os melhores resultados.

Segundo o MRE (2006), o Espírito Santo é um dos Estados que mais se desenvolveu economicamente nas duas últimas décadas, tendo uma estrutura industrial bastante diversificada. Entre as empresas de expressiva participação na indústria nacional, destacam-se a Vale do Rio Doce, a Companhia Siderúrgica de Tubarão, a Aracruz Celulose, a Samarco Mineração, a Chocolates Garoto, a Xerox do Brasil e a Braspérola. A adaptação do eucalipto, sobretudo ao norte, vem permitindo também o desenvolvimento da indústria de papel e celulose.

Para este mesmo Estado, a unidade espacial utilizada foi a divisão municipal, que é composta por 77 municípios. A Figura 5.40 apresenta a distribuição espacial da frota de CAMINHÕES SIMPLES com 2 ou 3 eixos no Estado do Espírito Santo. Neste caso também foram utilizados os quartis para definir as classificações.

Distribuição Espacial da Frota de Caminhões Simples com 2 ou 3 Eixos

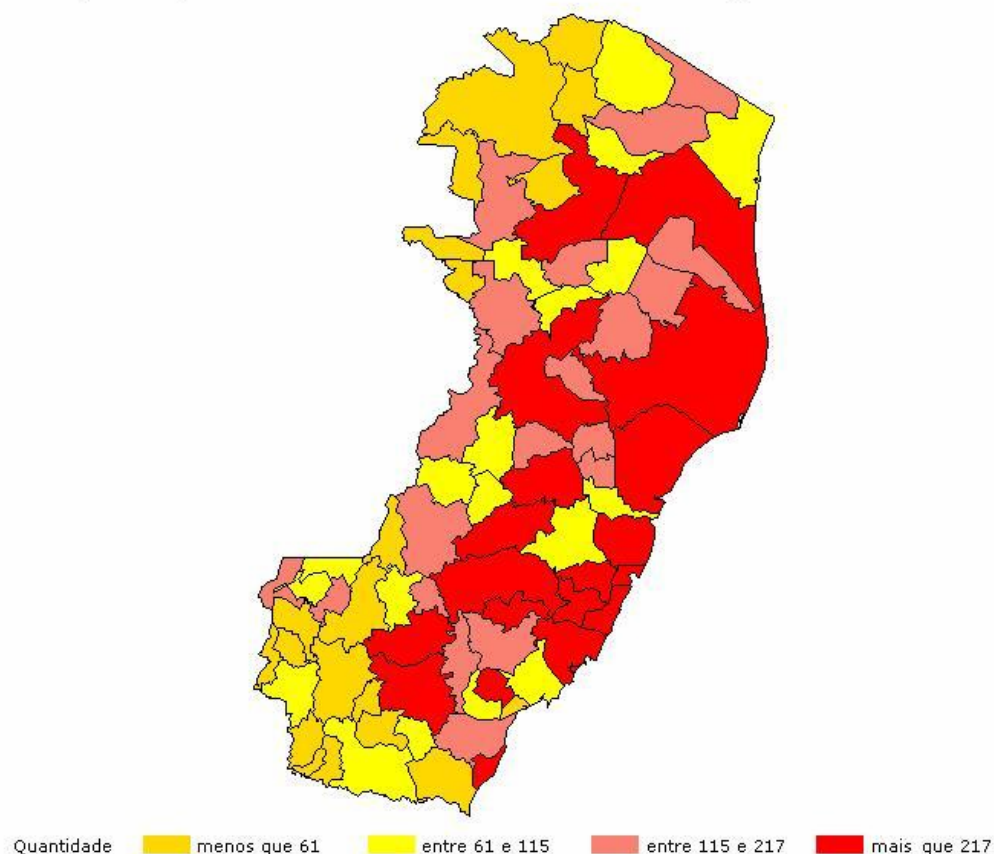


Figura 5.40: Distribuição Espacial da Frota de CAMINHÕES SIMPLES no Espírito Santo.

Verifica-se que existe uma concentração maior de caminhões no litoral do Estado e uma concentração menor à medida que as cidades vão se afastando do litoral. Essa configuração é normalmente encontrada em cidades litorâneas, devido ao maior desenvolvimento e a grande densidade populacional.

Com isso, constata-se também um indicio de dependência espacial na quantidade de CAMINHÕES SIMPLES (a partir daqui esta variável será chamada de **FrotaB**). Da mesma forma feita anteriormente e utilizando os 2 tipos de matriz de proximidades (Matriz A = binária e Matriz B = distância entre centróides), chega-se aos valores dos índices I de Moran e C de Geary apresentados na Tabela 5.16. A distância que melhor caracterizou a Matriz B foi de 21 km, conforme a Tabela 5.15.

Tabela 5.15. Índices de Dependência Espacial para Diferentes Distâncias entre Centróides.

Índice	d= 10 km	d= 20 km	d= 21 km	d= 22 km	d= 23 km	d= 24 km
I de Moran	0,56 (0.05)	0,48 (0.001)	0,48 (0.001)	0,32 (0.009)	0,32 (0.006)	0,28 (0.01)
C de Geary	0,19 (0.04)	0,70 (0.04)	0,67 (0.03)	0,83 (0.14)	0,86 (0.17)	0,87 (0.17)
Índice	d= 25 km	d= 26 km	d= 27 km	d= 28 km	d= 29 km	d=30 km
I de Moran	0,26 (0.01)	0,26 (0.01)	0,25 (0.01)	0,25 (0.008)	0,25 (0.007)	0,25 (0.006)
C de Geary	0,76 (0.02)	0,71 (0.006)	0,74 (0.01)	0,72 (0.005)	0,77 (0.01)	0,78 (0.002)
Índice	d= 40 km	d= 50 km	d= 60 km	d= 70 km	d= 80 km	d= 90 km
I de Moran	0,21 (0.003)	0,18 (0.002)	0,16 (<0.001)	0,15 (<0.001)	0,14 (<0.001)	0,13 (<0.001)
C de Geary	0,76 (<0.001)	0,80 (<0.001)	0,82 (<0.001)	0,85 (<0.001)	0,85 (<0.001)	0,86 (<0.001)

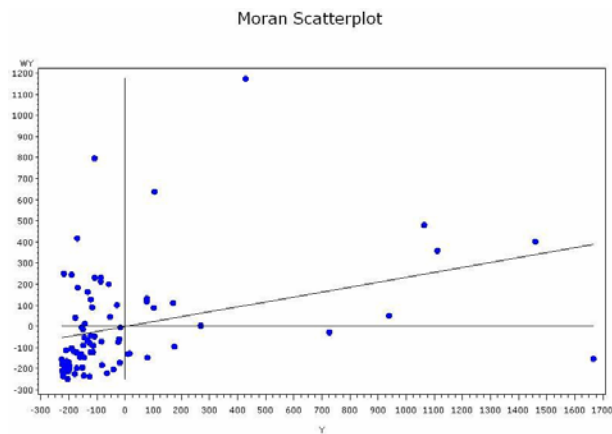
p-valor entre parênteses (Regra de decisão: aceita-se a hip. nula se p-valor >0.05)

Tabela 5.16. Índices de Dependência Espacial da Variável **FrotaB**.

Matriz A	Matriz B
I = 0,23 (0.001)	I = 0,48 (0.001)
C = 0,87 (0.048)	C = 0,67 (0.03)

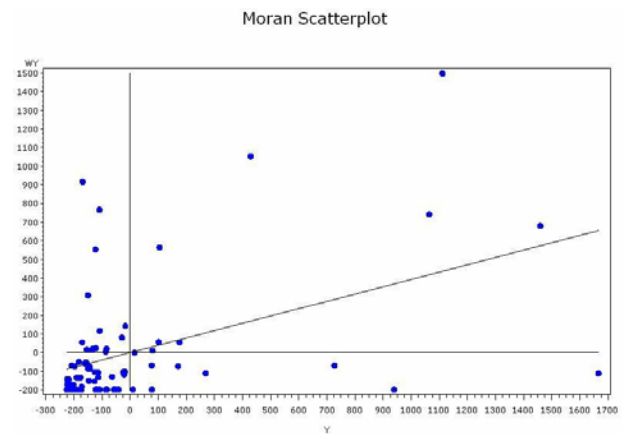
p-valor entre parênteses (Regra de decisão: aceita-se a hip. nula se p-valor >0.05)

Como as carrocerias do tipo GRANELEIRA, verifica-se uma baixa dependência espacial da frota de CAMINHÕES SIMPLES no Estado do Espírito Santo. Note também que por muito pouco tais índices não se tornam significativos (no caso do índice C de Geary). Com o objetivo de auxiliar na caracterização da dependência espacial, as Figuras de 5.41 a 5.46 apresentam as análises gráficas para as matrizes A e B.



b=0.23 / alpha=13.1°

Figura 5.41: *Moran Scatterplot* da Variável **FrotaB** Utilizando a Matriz A.



b=0.39 / alpha=21.45°

Figura 5.42: *Moran Scatterplot* da Variável **FrotaB** Utilizando a Matriz B.

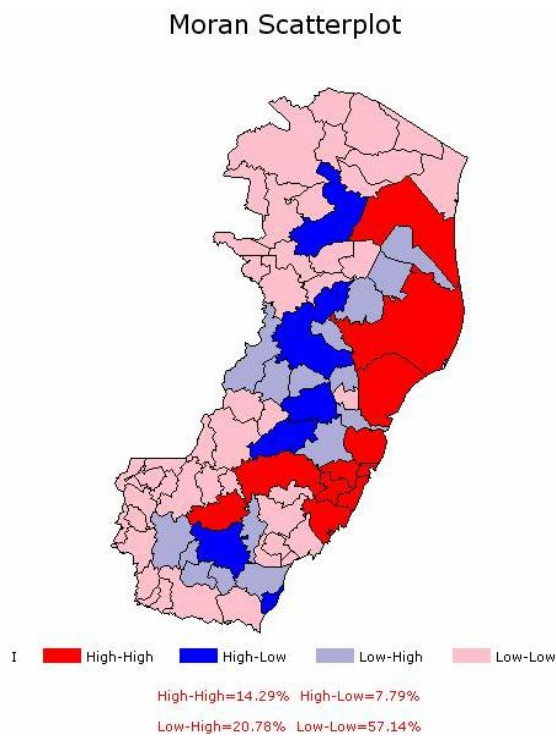


Figura 5.43: *Moran Scatterplot (Mapa)* da Variável **FrotaB** Utilizando a Matriz A.

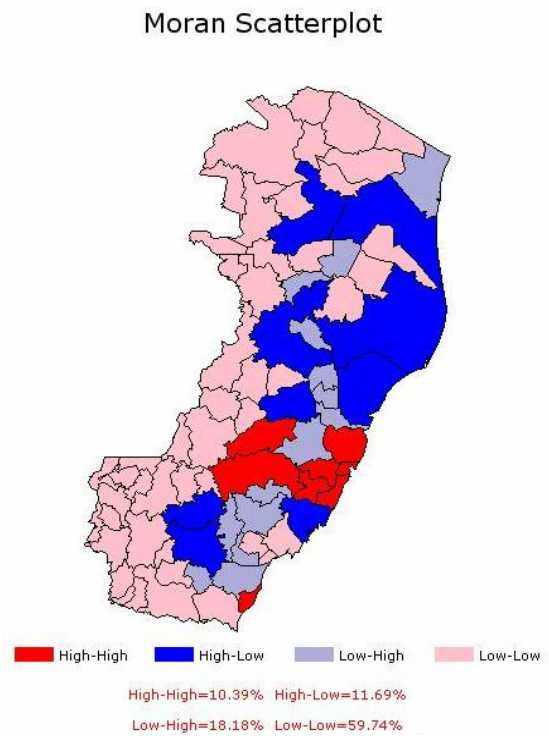


Figura 5.44: *Moran Scatterplot (Mapa)* da Variável **FrotaB** Utilizando a Matriz B.

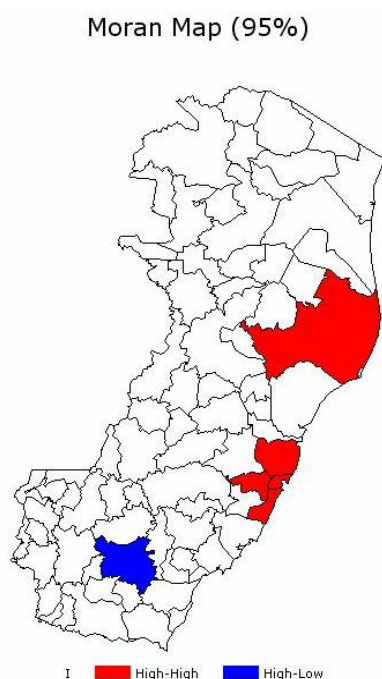


Figura 5.45: *Moran Map* da Variável **FrotaB** Utilizando a Matriz A.

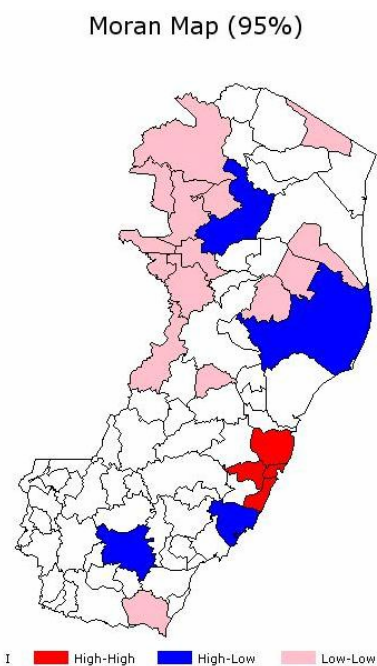


Figura 5.46: *Moran Map* da Variável **FrotaB** Utilizando a Matriz B.

Verifica-se na Figura 5.43 que a quantidade de microrregiões classificadas como *High-High* e *Low-Low* – definidoras da dependência espacial – somam 71,43%, enquanto que na Figura 5.44 somam 70,13%. Esses resultados discordam do índice I de Moran para a Matriz B, que apresentou um índice maior. Verifica-se também a mesma mudança de *High-High* para *High-Low* que ocorreu na variável **FrotaA** quando utilizada a Matriz B, como também a maior heterogeneidade espacial, segundo a Figura 5.46.

A transformação logaritma também foi aplicada nas variáveis pelos mesmos motivos apresentados anteriormente, e para a etapa de calibração dos modelos, a Tabela 5.17 apresenta a matriz de correlações das variáveis analisadas (já aplicado o logaritmo).

Tabela 5.17. Matriz de Correlação das Variáveis Seleccionadas.

	FrotaB	Com	Const	Exp	Imp	Ind	Inv	IPTU	PIB	Dens
FrotaB	1.00	0.83	0.73	0.46	0.79	0.88	0.74	0.72	-0.16	0.59
Com		1.00	0.90	0.40	0.91	0.92	0.75	0.86	-0.24	0.73
Const			1.00	0.47	0.90	0.83	0.76	0.84	-0.27	0.77
Exp				1.00	0.52	0.49	0.59	0.44	-0.20	0.38
Imp					1.00	0.89	0.80	0.93	-0.19	0.73

Tabela 5.17. (Continuação)

	FrotaB	Com	Const	Exp	Imp	Ind	Inv	IPTU	PIB	Dens
Ind						1.00	0.77	0.85	-0.18	0.65
Inv							1.00	0.68	-0.30	0.60
Iptu								1.00	-0.15	0.69
Pib									1.00	-0.26
dens										1.00

Verifica-se que as variáveis **Com**, **Const**, **Imp**, **Ind**, **Inv**, e **IPTU** apresentam uma alta correlação com a variável dependente **FrotaB**. No entanto, a variável **Com** apresenta forte correlação com as variáveis **Const**, **Imp**, **Ind**, **Inv**, **IPTU** e **Dens**. Tais correlações indicam um possível problema de multicolinearidade. Dessa forma, por todas as variáveis estarem correlacionadas entre si e a variável **Ind** apresentar a maior correlação, como foi verificado pelo MRE (2006), esta última se tornou candidata a entrar no modelo. Foi testada também a variável **Com**, mas os resultados foram inferiores aos da variável **Ind**. Mesmo assim, a Figura 5.47 apresenta o gráfico de dispersão de ambas as variáveis, mostrando a tendência linear existente com a variável **FrotaB**.

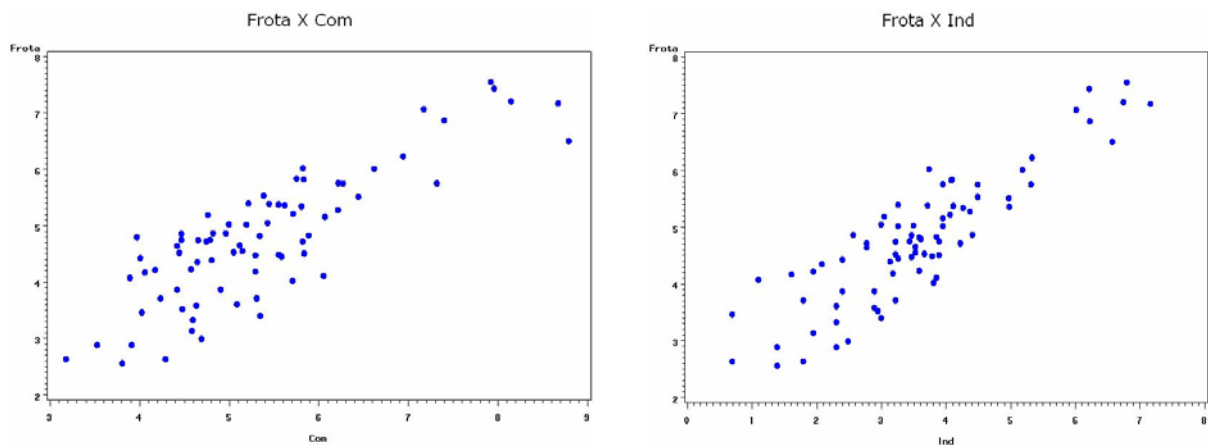


Figura 5.47: Gráficos de Dispersão das Variáveis **Com** e **Ind**.

Tabela 5.18. Resultados do Modelo de Regressão Convencional e Espacial.

	Regressão Convencional		Regressão Espacial MA		Regressão Espacial MB	
<i>Variável</i>	<i>Parâmetro</i>	<i>Teste t</i>	<i>Parâmetro</i>	<i>Teste t</i>	<i>Parâmetro</i>	<i>Teste t</i>
Intercepto	2,190* (<0,001)	12,88	1,365* (<0,001)	3,59	2,176* (<0,001)	12,24
Ind	0,726* (<0,001)	16,28	0,670* (<0,001)	14,16	0,721* (<0,001)	15,77
ρ	-	-	0,208** (0,018)	2,41	0,008 ^{NS} (0,776)	0,28
R^2	0,779		0,799		0,779	
$R^2_{ajustado}$	0,776		0,797		0,776	
F	265,05*		299,85*		265,41*	
AIC	-93,74		-99,24		-91,82	
White	8,32** (0,015)		5,35 ^{NS} (0,148)		8,48** (0,037)	
n	77		77		77	

(p-valor), (*) = muito significativo ($p < 0,01$), (**) = pouco significativo ($p < 0,05$), NS = não significativo

De acordo com a Tabela 5.18, constata-se um bom ajuste dos modelos (77% para o modelo convencional e 79% para o modelo espacial que utilizou a Matriz A). É interessante observar que o parâmetro espacial “ ρ ” do modelo espacial que utilizou a matriz B não foi significativo. Isso se reflete diretamente nos resultados que são praticamente os mesmos do modelo convencional. Para que se tivesse duas matrizes de proximidade espacial, decidiu-se utilizar a distância de 40 km para definir uma nova Matriz B, pois as distâncias menores que esta apresentaram os mesmos resultados (ou seja, no modelo ρ é não significativo) da distância de 21 km. Os novos resultados encontram-se na Tabela 5.19.

Tabela 5.19. Resultados do Modelo de Regressão Convencional e Espacial.

	Regressão Convencional		Regressão Espacial MA		Regressão Espacial MB	
<i>Variável</i>	<i>Parâmetro</i>	<i>Teste t</i>	<i>Parâmetro</i>	<i>Teste t</i>	<i>Parâmetro</i>	<i>Teste t</i>
Intercepto	2,190* (<0,001)	12,88	1,365* (<0,001)	3,59	1,186* (<0,001)	3,56
Ind	0,726* (<0,001)	16,28	0,670* (<0,001)	14,16	0,653* (<0,001)	14,20
ρ	-	-	0,208** (<0,018)	2,41	0,267* (0,001)	3,40
R^2	0,779		0,799		0,813	
$R^2_{ajustado}$	0,776		0,797		0,811	
F	265,05*		299,85*		327,10*	

(p-valor), (*) = muito significativo ($p < 0,01$), (**) = pouco significativo ($p < 0,05$), NS = não significativo

Tabela 5.19. (Continuação)

	<i>Regressão Convencional</i>	<i>Regressão Espacial MA</i>	<i>Regressão Espacial MB</i>
AIC	-93,74	-99,24	-104,64
White	8,32 ^{**} (0,015)	5,35 ^{NS} (0,148)	5,56 ^{NS} (0,135)
n	77	77	77

(p-valor), (*) = muito significativo ($p < 0,01$), (**) = pouco significativo ($p < 0,05$), NS = não significativo

Com a mudança da distância utilizada na Matriz B, os resultados desse modelo espacial melhoraram. A hipótese de heterocedasticidade foi rejeitada e o critério de AIC ficou menor, indicando um melhor ajuste. Todos os modelos apresentaram os erros normalmente distribuídos, mas o modelo convencional possui problema de heterocedasticidade, conforme o teste de White.

A Tabela 5.20 apresenta as estatísticas necessárias para a comparação entre os modelos. Verifica-se uma pequena melhora dos modelos espaciais em relação ao modelo convencional, e os ERM (erros relativos médios) para os anos de 2000 e 2003 são praticamente os mesmos.

Tabela 5.20. Estatísticas de Ajuste dos Modelos de Regressão Convencional e Espacial.

<i>Estatísticas</i>	<i>Modelo Convencional</i>	<i>Modelo Espacial Utilizando a Matriz A</i>	<i>Modelo Espacial Utilizando a Matriz B</i>
R^2	0,779	0,799	0,813
$R^2_{ajustado}$	0,776	0,797	0,811
F	265,05	299,85	327,102
AIC	-93,74	-99,24	-104,64
ERM (2000)	0,103	0,101	0,104
ERM (2003)	0,087	0,093	0,092

Após a análise de adequabilidade dos modelos e o poder de previsão, o próximo passo é estudar cada parâmetro individualmente. A Tabela 5.21 apresenta os parâmetros estimados dos modelos convencional e espacial, necessários para a análise de cenários.

Tabela 5.21. Parâmetros dos Modelos de Regressão Convencional e Espacial.

<i>Variável</i>	<i>Regressão Convencional</i>		<i>Regressão Espacial MA</i>		<i>Regressão Espacial MB</i>	
	<i>Parâmetro</i>	<i>Teste t</i>	<i>Parâmetro</i>	<i>Teste t</i>	<i>Parâmetro</i>	<i>Teste t</i>
Intercepto	2,190*	12,88	1,365*	3,59	1,186*	3,56
Ind	0,726*	16,28	0,670*	14,16	0,653 *	14,20
ρ	-	-	0,208**	2,41	0,267 *	3,40

(*) = muito significativo ($p < 0,01$), (**) = pouco significativo ($p < 0,05$)

Verifica-se que o parâmetro da variável **Ind** mudou de 0,726 do modelo convencional para 0,670 do modelo espacial que utilizou a Matriz A. Isso quer dizer que para um aumento de 1% na quantidade de estabelecimentos industriais, a quantidade de CAMINHÕES SIMPLES aumentaria em 0,726% no modelo convencional e 0,670% no modelo espacial. Fazendo o intervalo de confiança dos parâmetros (0,639; 0,813) no convencional e (0,577; 0,763) no espacial, verifica-se que estatisticamente os dois valores são iguais, o que também é confirmado pelo teste t que apresentou estatística igual a 0,198 e p-valor igual 0,843. Mesmo assim é válido avaliar o valor das estimativas pontuais, conforme apresentado a seguir.

No município de Vila Velha, por exemplo, em 2000 existiam 1228 estabelecimentos industriais e em 2003 já eram 1347, ou seja, um aumento de 119 ou 9,69%. Como em 2000 o município possuía 1302 veículos rodoviários de carga do tipo CAMINHÃO SIMPLES, fazendo $9,69 \times 0,726$ tem-se um aumento de 7,03% da frota, ou seja, 91 CAMINHÕES SIMPLES. No caso do modelo espacial, esse aumento seria de 6,49% ou 84 veículos. Verifica-se mais uma vez que há uma superestimação por parte do modelo convencional.

5.5. TÓPICOS CONCLUSIVOS

Este capítulo teve como objetivo verificar a aplicabilidade da metodologia e a utilização da regressão espacial. Para tal, serviu-se do banco de dados RNTRC juntamente com algumas variáveis do IBGE e da SECEX. Diante do que foi exposto, merecem destaque os seguintes tópicos conclusivos:

- O RNTRC se constitui uma importante ferramenta da ANTT, pois através dele a agência tem mais suporte à tomada de decisões e pode melhor auxiliar o Ministério dos Transportes;

- Através do banco de dados do RNTRC não foi possível determinar a localização exata do veículo rodoviário de carga. Tal localização ficou condicionada ao lugar onde se encontrava o transportador;
- As carrocerias do tipo GRANELEIRA apresentaram dependência espacial de 22%, segundo o teste I de Moran, enquanto que para os CAMINHÕES SIMPLES, tal dependência foi de 23%;
- Os modelos de regressão espacial apresentaram melhores ajustes aos dados do que os modelos de regressão convencional, no entanto, no que se refere à previsão, os erros relativos médios foram praticamente os mesmos.
- As estimativas pontuais de ambos os modelos, convencional e espacial, foram estatisticamente iguais, entretanto, o modelo convencional apresentou na maioria das vezes, intervalos de confiança maiores do que os do modelo espacial. Uma explicação plausível para essa equivalência estatística pode ser atribuída à baixa dependência espacial constatada.
- As estimativas dos parâmetros, em um modelo de regressão convencional que utiliza o método de Mínimos Quadrados Ordinários, independem da distribuição normal. No entanto, os desvios padrões são calculados a partir dessa distribuição. Como se sabe, a técnica de análise de regressão faz uma estimativa do verdadeiro parâmetro populacional, através da amostra estudada. Por isso, problemas na falta de normalidade dos erros ou de heterocedasticidade, podem fornecer estimativas enganosas do intervalo de confiança dos parâmetros e conseqüentemente do intervalo de previsão. Assim, os testes realizados para verificação da igualdade entre os parâmetros podem ser enganosos, pois tais pressupostos não foram satisfeitos pelos modelos convencionais em ambos os estudos de caso.

6. CONCLUSÕES E RECOMENDAÇÕES

6.1. APRESENTAÇÃO

Este capítulo busca apresentar as principais conclusões obtidas dos resultados do capítulo anterior, bem como fazer sugestões para trabalhos futuros no mesmo tema. A pesquisa não tinha a pretensão de encontrar todos os fatores que influenciam na oferta do transporte rodoviário de carga, mas sim identificar alguns desses fatores e iniciar uma discussão sobre a importância de se utilizar ferramentas de modelagem mais robustas. Ao mesmo tempo em que a regressão espacial é uma ferramenta robusta, por incorporar o fator “espaço” em sua estrutura, sua implementação é simples e seus resultados são interpretados como na regressão tradicional.

6.2. CONCLUSÕES

Os resultados de ambos os tipos de veículos rodoviários de carga, carroceria GRANELEIRA e CAMINHÃO SIMPLES, mesmo sendo estudos distintos mostraram que, na presença de dependência espacial, existe diferença entre os modelos de regressão espacial e convencional. De fato, nos exemplos não houve uma diferença tão acentuada, mas isso se deve à baixa dependência espacial constatada na variável original. No caso de alta dependência, uma grande diferença entre os modelos é verificada, como pode ser visto em Gasques *et al* (2005). Apesar disso, é válida a reflexão de que ao se ignorar a influência do “espaço”, e considerar que as unidades espaciais se comportam de forma independente, nenhum modelo conseguirá descrever a relação de desenvolvimento existente entre uma região pólo e seus vizinhos.

Pode-se concluir também que a forma da matriz de proximidades influencia diretamente na constatação da dependência espacial. No caso do transporte rodoviário de carga, a matriz binária se mostrou bastante útil, gerando resultados melhores ou próximos de uma matriz mais refinada (nos exemplos, a distância entre centróides). Contudo, é válido dizer mais uma vez que tais formas não são as melhores para se trabalhar com dados de transporte, pois melhorias na malha rodoviária ou qualquer outro agente externo, não seriam capturados pelas matrizes que consideram apenas a proximidade geográfica.

Além disso, averiguou-se que a mudança da estrutura ou alteração da dependência espacial, ocasionada por políticas públicas ou fatores geográficos, refletem diretamente nas previsões feitas pelos modelos de regressão espacial. Um exemplo disso foi a diversificação feita pelo Brasil em sua pauta de exportações. A soja teve um salto de 58,34% de participação dos produtos em grãos em 2000 para 71,66% em 2003. Esse aumento pode ter ocorrido em decorrência do incentivo dado aos produtores de aumentar a produção de soja, o que pode ter despertado o interesse de agricultores de outras culturas migrarem para soja, descaracterizando assim a estrutura espacial existente em 2000.

Outra conclusão de extrema importância é a correção dos parâmetros feita pelos modelos espaciais. Nos dois estudos de casos, verificou-se praticamente os mesmos erros relativos médios (ERM), porém, os parâmetros das variáveis analisadas apresentaram diferenças. No caso da previsão, os resultados são bastante próximos ao se utilizar todas as variáveis incluídas no modelo, mas são bem diferentes ao se utilizar algumas ou cada variável individualmente (análise de cenários). Esse fato mostra claramente a influência do fator “espaço” nos parâmetros estimados.

Diante dos resultados apresentados, constata-se que as hipóteses da pesquisa foram validadas, como também todos os objetivos foram alcançados. Partindo da premissa que a oferta do transporte rodoviário de carga é uma variável espacialmente dependente, o modelo de regressão espacial se mostrou capaz de identificar os fatores que influenciam na frota de veículos rodoviários de carga. Esse fato corrobora diretamente a hipótese de que um modelo de regressão espacial gera resultados mais realistas do que um modelo de regressão convencional. A expressão “resultados mais realistas” diz respeito à correção feita nos parâmetros devido a retirada da influência do fator “espaço”, o que não acontece no modelo de regressão convencional.

Os dados originais não apresentaram problemas, conforme os critérios definidos para verificação da qualidade dos mesmos. Mas nem por isso, a hipótese de que a avaliação da qualidade da informação elimina possíveis inconsistências nos dados não foi validada. Ao se trabalhar com o “logaritmo de 0”, verificou-se que os parâmetros do modelo mudaram radicalmente, indicando um problema grave de interpretação. A retirada desse “problema” corrigiu os parâmetros, o que confirma a validade da referida hipótese.

6.3. RECOMENDAÇÕES E SUGESTÕES PARA TRABALHOS FUTUROS

O presente trabalho buscou mostrar o ganho de adequabilidade, na existência de dependência espacial, do modelo de regressão espacial em relação ao modelo convencional. A comparação persistiu desde as suposições básicas do modelo até a interpretação de cada parâmetro individualmente. Para que o trabalho não ficasse muito extenso e por impossibilidades, alguns procedimentos deixaram de ser executados, o que de maneira alguma invalida a pesquisa. Dessa forma, seguem algumas sugestões e recomendações para trabalhos futuros:

- Verificar se o veículo se localiza efetivamente onde a empresa, cooperativa ou pessoa física se localiza, a fim de se obter uma melhor caracterização da distribuição espacial da frota de veículos rodoviários de carga.
- Utilizar um modelo de Poisson Espacial para modelar a frota de veículos rodoviários de carga, por ser este tipo de distribuição mais adequada para se trabalhar com dados de contagem (discretos). Vale ressaltar que nenhum software possui esse tipo de algoritmo implementado, necessitando, portanto, estudos mais aprofundados nessa área. A maioria dos softwares comerciais trabalham apenas com a distribuição normal.
- A matriz de distância entre centróides, considerando o centro geométrico do polígono, apresentou resultados interessantes, mas acredita-se que definindo o centróide como o centro populacional do polígono, melhores e mais realistas seriam os resultados. Recomenda-se também utilizar alguma variável não geográfica para definir a matriz de proximidades. Tais variáveis podem ser o tempo necessário para ir de uma localidade a outra, quantidade de rodovias de ligação, quantidade ou valor das trocas comerciais entre as unidades espaciais etc.
- O trabalho utilizou um modelo de regressão espacial global, ou seja, aquele que captura a espacialidade em apenas um parâmetro. Como foram identificadas, pelos testes espaciais locais, algumas áreas significativas é interessante utilizar um modelo que considere padrões diferenciados de espacialidade, o que melhor caracterizaria a heterogeneidade espacial.

BIBLIOGRAFIA

- ABNT (1997), Associação Brasileira de Normas Técnicas. *Veículo Rodoviário de Carga – Terminologia*. NBR 9762, Rio de Janeiro, RJ.
- AKAIKE, H. A. (1974). *New Look at Statistical Model Identification*. IEEE Trans, on Automatic Control, v.19.
- ALANO, F. (2003), *A Decisão de Compra de Frete Rodoviário Internacional – O Caso da Carga Seca, Modalidade Fechada, Porta a Porta*. Dissertação de Mestrado – Porto Alegre, RS.
- ANFAVEA (2006), Associação Nacional dos fabricantes de Veículos Automotores. Disponível em: <<http://www.anfavea.com.br/Index.html>>. Acesso em 04 mai. 2006.
- ANSELIN, L. (1988), *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers, Santa Barbara, EUA.
- ANSELIN, L. (1995), *Local Indicators of Spatial Association — LISA*. Geographical Analysis.
- ANSELIN, L.(1996), *The Moran Scatterplot as ESDA Tool to Assess Local Instability in Spatial Association*. Spatial Analytical Perspectives on GIS,Londres, UK.
- ANTT (2005), Agência Nacional de Transportes Terrestres. Disponível em: < <http://www.antt.gov.br/destaques/ANTTemNumeros20051226.pdf> >. Acesso em 3 abr. 2006.
- ASSUNÇÃO, R. M. (2003), *Índices de Auto-Correlação Espacial*. Notas de Aula. Departamento de Estatística – UFMG.
- BENITE, M. (2003), *Aplicação de Modelos de Redes Neurais na Elaboração e Análise de Cenários Macroeconômicos*. Dissertação de Mestrado – São Carlos, SP.
- BOWERSOX, D.J. e CLOSS, D.J (2001), *Logística Empresarial: O Processo de Integração da Cadeia de Suprimento*. São Paulo: Atlas.
- BUSSAB, W. O. e MORETTIN, P. A. (2004), *Estatística Básica*. 5. ed. – São Paulo: Saraiva.

- CÂMARA, G; CARVALHO, M. S.; CRUZ, O. G.; CORREA, V. (2002), *Análise Espacial de Dados Geográficos*. Instituto Nacional de Pesquisas Espaciais, São José dos Campos.
- COATES, J. F. (2000). *Scenario Planning*. Technological Forescating and Social Change, v. 5.
- CORREIA, D. E. R. (2004), *Metodologia para Identificação da Qualidade da Informação: Uma Aplicação para o Planejamento de Transportes*. Dissertação de Mestrado – Brasília, DF.
- CRESSIE, N. A. C. (1993), *Statistics for Spatial Data – Revised Edition* – John Wiley & Sons, Inc – Nova York, EUA.
- DNIT (2006), Departamento Nacional de Infra-Estrutura de Transportes. Disponível em: <http://www.dnit.gov.br/rodovias/pesagem/peso_maximo.htm>. Acesso em 04 mai. 2006.
- FREITAS, M. B. de (2004), *Transporte rodoviário de cargas e sua respectiva responsabilidade civil*. Disponível em: <<http://www1.jus.com.br/doutrina/texto.asp?id=5231>>. Acesso em: 28 mai. 2005.
- FEITOSA, F. (2005), *Análise Espacial de Áreas: Regressão*. INPE. Disponível em: <www.dpi.inpe.br/cursos/ser301/Aula_Regressao_2005.ppt>. Acesso em 12 mar. 2006.
- GASQUES, J.G.; FREITAS, R. E.; BASTOS, E. T.; SILVA, H. D. P. da; SILVA, A. R. da (2005), *Agricultura Familiar – PRONAF – Análise de Alguns indicadores*. XLIII Congresso da Sociedade Brasileira de Economia e Sociologia Rural, Ribeirão Preto, SP.
- GEIPOT (2000), Empresa Brasileira de Planejamento de Transportes. Disponível em : < <http://www.geipot.gov.br> >. Acesso em 17 fev. 2006.
- GREENE, W. H. (1999), *Econometric Analysis*, 3. ed. New Jersey: Prentice Hall.
- GUJARATI, D. N. (1995), *Econometria Básica*, Makron Books.
- KAMPEL, S. M. e CÂMARA, G. (2000), *Análise Exploratória das Relações Espaciais do Desflorestamento na Amazônia Legal Brasileira*. INPE.
- KREMPI, A. P. (2004), *Explorando Recursos de Estatística Espacial para Análise da Acessibilidade na Cidade de Bauru*. Dissertação de Mestrado – São Carlos, SP.

- LAKATOS, E. M.e MARCONI, M. A. (2001), *Metodologia do Trabalho Científico*. Atlas – São Paulo, SP.
- LESAGE, J. P. (1999a), *The Theory and Practice of Spatial Econometrics*, Department of Economics, University of Toledo, EUA.
- LESAGE, J. P. (1999b), *Applied Econometrics Using MATLAB*, Department of Economics, University of Toledo, EUA.
- LEMBO, A. J. (2005), *Spatial Autocorrelation*. Cornell University. Department of Crop and Soil Sciences. Disponível em: < <http://www.css.cornell.edu/courses/620/lecture9.ppt> >. Acesso em 16 mar. 2006.
- LOPES, S.B. (2005), *Efeitos da Dependência Espacial em Modelos de Previsão de Demanda por Transporte*. Dissertação de Mestrado – São Carlos, SP.
- MACOHIN, G. A (2001), *De Transportador a Operador Logístico – a lacuna a ser preenchida: um estudo de caso*. Dissertação de Mestrado – Centro Tecnológico, UFSC, SC.
- MELLO, J.C. (2000), *Transporte Rodoviário*, Ministério das relações exteriores. Disponível em: <<http://www.mre.gov.br/cdbrasil/itamaraty/web/port/economia/transp/rodo/apresent.htm>>. Acesso em 28 abr. 2005. Brasília, DF.
- MAPA (2006), Ministério da Agricultura, Pecuária e Abastecimento. Disponível em : <www.agricultura.gov.br>. Acesso em 17 mar. 2006.
- MDICE (2006), Ministério do Desenvolvimento, Indústria e Comércio Exterior. Disponível em <www.desenvolvimento.gov.br>. Acesso 17 mar. 2006.
- MORAN, P.A.P. (1950). *Notes on continuous stochastic phenomena*, Biometrika.
- MRE (2006), Ministério das Relações Exteriores. Disponível em: <<http://www.mre.gov.br/cdbrasil/itamaraty/web/port/divpol/sudeste/es/apresent/apresent.htm>>. Acesso em: 21 jun. 2006.
- NEVES, M.A.O. (2004), *No futuro, será possível viver única e exclusivamente do transporte rodoviário de carga no Brasil?*. Disponível em: < <http://www.crowmatec.com.br/noticias.asp?id=86> >. Acesso em 20 dez. 2005.
- NOVAES, A. G. (1981), *Modelos em Planejamento Urbano, Regional e de Transportes*. Editora Edgard Blücher Ltda. São Paulo, SP.

- OGLIARI, P. J. (2004a), *Construção de Modelos: Seleção de Variáveis Regressoras*. Notas de Aula, UFSC. Disponível em : <[www.inf.ufsc.br/~ogliari/arquivos/construcao_de modelos de regressao.ppt](http://www.inf.ufsc.br/~ogliari/arquivos/construcao_de_modelos_de_regressao.ppt)>. Acesso em 21 mar. 2006.
- OGLIARI, P. J. (2004b), *Outros Tópicos na Análise de Regressão*. Notas de Aula, UFSC. Disponível em:
<[www.inf.ufsc.br/~ogliari/arquivos/ outros topicos na analise de regressao.ppt](http://www.inf.ufsc.br/~ogliari/arquivos/outros_topicos_na_analise_de_regressao.ppt)> Acesso em 12 fev. 2006.
- ROSA, A. V. A. (2004), *Análise do Arranjo Produtivo Local como Estratégia Competitiva de Pequenas Empresas de Transporte Rodoviário de Cargas: O Caso do Grupo Oeste Transportes de Osvaldo Cruz – SP*. Dissertação de Mestrado – PUC, PR.
- SAS (1999), SAS Online Doc V8. Disponível em: <<http://v8doc.sas.com>>. Acesso em 10 jan. 2006.
- SCHROEDER, E. M. e CASTRO, J. C. de (2000), *Transporte Rodoviário de Carga: Situação Atual e Perspectivas*. Disponível em:
< www.bndes.gov.br/conhecimento/revista/carga.pdf >. Acesso em 12 fev. 2006.
- SILVA, A.R. da (2004), *Modelos Espaciais Autoregressivos (Spatial Autoregressive Models)*. 13º GUSAS. São Paulo, SP.
- TEIXEIRA, G.L. (2003), *Uso de Dados Censitários para Identificação de Zonas Homogêneas para Planejamento de Transportes Utilizando Estatística Espacial*. Dissertação de Mestrado – Brasília, DF.
- TOBLER, W. R. (1979), *Cellular Geography, Philosophy in Geography*. Edited by S. Gale and G. Olsson. Eds, Amsterdam.
- WHITE, H. (1980), *A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity*. *Econometrica*, Volume 48, Number 4.
- ____ RESOLUÇÃO N° 68/98. Disponível em:
< <http://www.denatran.gov.br/resolucoes.htm> >. Acesso em 13 abr. 2006.

ANEXO I – QUESTIONÁRIO



Universidade de Brasília – UNB
Faculdade de Tecnologia
Departamento de Engenharia Civil e
Ambiental
Programa de Pós-Graduação em Transportes



O presente questionário tem como finalidade colher opiniões das empresas montadoras de veículos comerciais, a fim de auxiliar um estudo sobre fatores que influenciam a oferta do transporte rodoviário de carga. As informações aqui registradas servirão apenas para fins acadêmicos.

Questionário

1) Para um estudo de previsão da frota do transporte rodoviário de carga, estamos adotando as seguintes classificações. Ao seu ver, quais dessas são relevantes estudar.

- a) Caminhão Simples por quantidade de Eixos
- b) Caminhão Trator por quantidade de Eixos
- c) Carrocerias por tipo:
 - a. Porta Contêiner
 - b. Granel Sólido
 - c. Carga Seca
 - d. Cegonha
 - e. Baú Simples
 - f. Baú Frigorífico
 - g. Silo
 - h. Tanque
 - i. Sider

2) Na sua opinião, o que pode impulsionar a venda de caminhões?

3) Você acha que características regionais influenciam no aumento / diminuição da frota de caminhões? (Por exemplo: densidade populacional, pib, quantidade de indústrias...)

☐ SIM

☐ NÃO

Se SIM, quais:

ANEXO II – DEMONSTRAÇÕES DOS PARÂMETROS REFERENTES AOS MODELOS ESPACIAIS AUTOREGRESSIVOS

Modelo Geral

$$\begin{aligned} \mathbf{y} &= \rho \mathbf{W}_1 \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\mu} \\ \boldsymbol{\mu} &= \lambda \mathbf{W}_2 \boldsymbol{\mu} + \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n) \end{aligned} \quad (1)$$

onde: $\mathbf{y}_{n \times 1}$; $\mathbf{W}_1_{n \times n}$; $\mathbf{X}_{n \times k}$; $\boldsymbol{\beta}_{k \times 1}$; $\boldsymbol{\mu}_{n \times 1}$; $\mathbf{W}_2_{n \times n}$; $\boldsymbol{\varepsilon}_{n \times 1}$; $\mathbf{I}_n_{n \times n}$; ρ e λ - constantes

Do modelo (1), pode-se derivar outros modelos impondo restrições. Se $\mathbf{W}_2 = 0$ e $\mathbf{X} = 0$, tem-se o modelo espacial autoregressivo de 1º ordem:

$$\mathbf{y} = \rho \mathbf{W}_1 \mathbf{y} + \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n) \quad (2)$$

Fazendo $\mathbf{W}_2 = 0$ tem-se o modelo espacial autoregressivo misto :

$$\mathbf{y} = \rho \mathbf{W}_1 \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n) \quad (3)$$

Fazendo $\mathbf{W}_1 = 0$ tem-se o modelo com autocorrelação espacial no erro:

$$\begin{aligned} \mathbf{y} &= \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\mu} \\ \boldsymbol{\mu} &= \lambda \mathbf{W}_2 \boldsymbol{\mu} + \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n) \end{aligned} \quad (4)$$

Existe ainda o modelo espacial Durbin que considera também a dependência espacial nas variáveis independentes:

$$\mathbf{y} = \rho \mathbf{W}_1 \mathbf{y} + \mathbf{X} \boldsymbol{\beta}_1 + \mathbf{W}_1 \mathbf{X} \boldsymbol{\beta}_2 + \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n) \quad (5)$$

MODELO ESPACIAL AUTOREGRESSIVO DE 1º ORDEM

$$\mathbf{y} = \rho \mathbf{W}_1 \mathbf{y} + \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$$

Se a matriz $\mathbf{W}_1$ for padronizada, ou seja, a soma das linhas é igual a 1, então o vetor $\mathbf{y}$ deve ser expresso como desvios da média, $(\mathbf{y} - \bar{\mathbf{y}})$, para eliminar o termo constante do modelo.

Utilizando mínimos quadrados ordinários (OLS) para estimar ρ tem-se:

$$\mathbf{y} = \rho \mathbf{W}_1 \mathbf{y} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{\varepsilon} = \mathbf{y} - \rho \mathbf{W}_1 \mathbf{y}$$

$$\frac{\partial \boldsymbol{\varepsilon}' \boldsymbol{\varepsilon}}{\partial \rho} = (-\mathbf{W}_1 \mathbf{y})' (\mathbf{y} - \rho \mathbf{W}_1 \mathbf{y})$$

$$\rho \mathbf{y}' \mathbf{W}_1' \mathbf{W}_1 \mathbf{y} = \mathbf{y}' \mathbf{W}_1' \mathbf{y}$$

$$\hat{\rho} = (\mathbf{y}' \mathbf{W}_1' \mathbf{W}_1 \mathbf{y})^{-1} \mathbf{y}' \mathbf{W}_1' \mathbf{y}$$

Para verificar se o estimador é viesado, deve-se verificar se $E\left(\hat{\rho}\right) = \rho$.

$$\begin{aligned} E\left(\hat{\rho}\right) &= E\left[(\mathbf{y}' \mathbf{W}_1' \mathbf{W}_1 \mathbf{y})^{-1} \mathbf{y}' \mathbf{W}_1' (\rho \mathbf{W}_1 \mathbf{y} + \boldsymbol{\varepsilon})\right] \\ &= E\left[(\mathbf{y}' \mathbf{W}_1' \mathbf{W}_1 \mathbf{y})^{-1} \mathbf{y}' \mathbf{W}_1' \mathbf{W}_1 \mathbf{y} \rho + (\mathbf{y}' \mathbf{W}_1' \mathbf{W}_1 \mathbf{y})^{-1} \mathbf{y}' \mathbf{W}_1' \boldsymbol{\varepsilon}\right] \\ &= \rho + E\left[(\mathbf{y}' \mathbf{W}_1' \mathbf{W}_1 \mathbf{y})^{-1} \mathbf{y}' \mathbf{W}_1' \boldsymbol{\varepsilon}\right] \end{aligned}$$

ou seja, o estimador $\hat{\rho}$ é viesado, pois $E\left(\hat{\rho}\right) \neq \rho$. Por causa da dependência espacial, não pode-se passar o operador da esperança matemática para dentro da expressão e então ter $E(\boldsymbol{\varepsilon}) = 0$, retirando assim o viés. Porém a probabilidade do limite (plim) do termo $\mathbf{y}' \mathbf{W}_1' \boldsymbol{\varepsilon}$ é diferente de 0.

$$p \lim N^{-1} (\mathbf{y}' \mathbf{W}_1' \boldsymbol{\varepsilon}) = p \lim N^{-1} \boldsymbol{\varepsilon}' \mathbf{W}_1' (\mathbf{I}_n - \rho \mathbf{W}_1) \boldsymbol{\varepsilon}$$

Pode-se então estimar o valor de ρ , procurando um $\hat{\rho}$ que maximize a função de verossimilhança :

$$L(\mathbf{y} | \rho, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} |\mathbf{I}_n - \rho \mathbf{W}_1| e^{-\frac{1}{2\sigma^2} (\mathbf{y} - \rho \mathbf{W}_1 \mathbf{y})' (\mathbf{y} - \rho \mathbf{W}_1 \mathbf{y})}$$

Para simplificar a maximização, pode-se substituir σ^2 por $\frac{1}{n} (\mathbf{y} - \rho \mathbf{W}_1 \mathbf{y})' (\mathbf{y} - \rho \mathbf{W}_1 \mathbf{y})$

Prova:

$$\begin{aligned} \ln(L) &= \ln |\mathbf{I}_n - \rho \mathbf{W}_1| - \frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} (\mathbf{y} - \rho \mathbf{W}_1 \mathbf{y})' (\mathbf{y} - \rho \mathbf{W}_1 \mathbf{y}) \\ \frac{\partial \ln(L)}{\partial \sigma^2} &= -\frac{n}{2} \frac{2\pi}{2\pi\sigma^2} + \frac{1}{2(\sigma^2)^2} (\mathbf{y} - \rho \mathbf{W}_1 \mathbf{y})' (\mathbf{y} - \rho \mathbf{W}_1 \mathbf{y}) \end{aligned}$$

$$\frac{2(\sigma^2)^2}{2\sigma^2}n = (\mathbf{y} - \rho\mathbf{W}_1\mathbf{y})'(\mathbf{y} - \rho\mathbf{W}_1\mathbf{y}) \Rightarrow \sigma^2 = \frac{1}{n}(\mathbf{y} - \rho\mathbf{W}_1\mathbf{y})'(\mathbf{y} - \rho\mathbf{W}_1\mathbf{y})$$

Depois de substituir σ^2 por $\frac{1}{n}(\mathbf{y} - \rho\mathbf{W}_1\mathbf{y})'(\mathbf{y} - \rho\mathbf{W}_1\mathbf{y})$ e fazendo o logaritmo natural tem-se:

$$Ln(L) = -\frac{n}{2}\ln\left(2\pi\frac{1}{n}(\mathbf{y} - \rho\mathbf{W}_1\mathbf{y})'(\mathbf{y} - \rho\mathbf{W}_1\mathbf{y})\right) - \frac{(\mathbf{y} - \rho\mathbf{W}_1\mathbf{y})'(\mathbf{y} - \rho\mathbf{W}_1\mathbf{y})}{2\left(\frac{1}{n}(\mathbf{y} - \rho\mathbf{W}_1\mathbf{y})'(\mathbf{y} - \rho\mathbf{W}_1\mathbf{y})\right)} + \ln|\mathbf{I}_n - \rho\mathbf{W}_1|$$

$$Ln(L) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\left(\frac{1}{n}(\mathbf{y} - \rho\mathbf{W}_1\mathbf{y})'(\mathbf{y} - \rho\mathbf{W}_1\mathbf{y})\right) - \frac{n}{2} + \ln|\mathbf{I}_n - \rho\mathbf{W}_1|$$

Para maximizar $Ln(L)$ em ρ é necessário somente maximizar em ρ

$$Ln(L) = -\frac{n}{2}\ln\left(\frac{1}{n}(\mathbf{y} - \hat{\rho}\mathbf{W}_1\mathbf{y})'(\mathbf{y} - \hat{\rho}\mathbf{W}_1\mathbf{y})\right) + \ln|\mathbf{I}_n - \hat{\rho}\mathbf{W}_1|$$

Seja $\hat{\rho}$ este valor. Então $\hat{\sigma}^2 = \frac{1}{n}(\mathbf{y} - \hat{\rho}\mathbf{W}_1\mathbf{y})'(\mathbf{y} - \hat{\rho}\mathbf{W}_1\mathbf{y})$.

Depois de estimado os valores para ρ e σ^2 , é necessário conhecer a variância dos parâmetros para o cálculo do teste t e do intervalo de confiança. Para isso pode-se usar a inversa da matriz de informação de Fisher, que considera as derivadas segundas.

$$[I(\theta)]^{-1} = -E\left(\frac{\partial^2 L}{\partial\theta\partial\theta'}\right)^{-1}, \text{ onde } \theta = (\rho, \sigma^2)$$

Sabendo que:

$$E(\mathbf{y}) = 0 \quad E(\mathbf{y}\mathbf{y}') = (\mathbf{I}_n - \rho\mathbf{W}_1)^{-1}\sigma^2\mathbf{I}_n(\mathbf{I}_n - \rho\mathbf{W}_1)^{-1}$$

$$Ln(L) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}(\mathbf{y} - \rho\mathbf{W}_1\mathbf{y})'(\mathbf{y} - \rho\mathbf{W}_1\mathbf{y}) + \ln|\mathbf{I}_n - \rho\mathbf{W}_1|$$

$$\frac{\partial Ln(L)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}(\mathbf{y} - \rho\mathbf{W}_1\mathbf{y})'(\mathbf{y} - \rho\mathbf{W}_1\mathbf{y})$$

$$\frac{\partial Ln(L)}{\partial \rho} = -tr\left((\mathbf{I}_n - \rho\mathbf{W}_1)^{-1}\mathbf{W}_1\right) + \frac{2}{2\sigma^2}(\mathbf{W}_1\mathbf{y})'(\mathbf{y} - \rho\mathbf{W}_1\mathbf{y})$$

$$\frac{\partial^2 Ln(L)}{\partial(\sigma^2)^2} = \frac{n}{2(\sigma^2)^2} - \frac{1}{(\sigma^2)^3}(\mathbf{y} - \rho\mathbf{W}_1\mathbf{y})'(\mathbf{y} - \rho\mathbf{W}_1\mathbf{y})$$

$$\begin{aligned}
\frac{\partial^2 \text{Ln}(L)}{\partial \rho^2} &= -\text{tr}((\mathbf{I}_n - \rho \mathbf{W}_1)^{-1} \mathbf{W}_1 (\mathbf{I}_n - \rho \mathbf{W}_1)^{-1} \mathbf{W}_1) - \frac{\mathbf{y}' \mathbf{W}_1' \mathbf{W}_1 \mathbf{y}}{\sigma^2} \\
\frac{\partial^2 \text{Ln}(L)}{\partial \rho \partial \sigma^2} &= -\frac{(\mathbf{W}_1 \mathbf{y})' (\mathbf{y} - \rho \mathbf{W}_1 \mathbf{y})}{(\sigma^2)^2} = \frac{\rho \mathbf{y}' \mathbf{W}_1' \mathbf{W}_1 \mathbf{y} - \mathbf{y}' \mathbf{W}_1' \mathbf{y}}{(\sigma^2)^2} \\
-E\left(\frac{\partial^2 \text{Ln}(L)}{\partial (\sigma^2)^2}\right) &= -\frac{n}{2(\sigma^2)^2} + \frac{1}{(\sigma^2)^3} E(\boldsymbol{\varepsilon}' \boldsymbol{\varepsilon}) = -\frac{n}{2(\sigma^2)^2} + \frac{n\sigma^2}{(\sigma^2)^3} = \frac{n}{2(\sigma^2)^2} \\
-E\left(\frac{\partial^2 \text{Ln}(L)}{\partial \rho^2}\right) &= \text{tr}((\mathbf{I}_n - \rho \mathbf{W}_1)^{-1} \mathbf{W}_1 (\mathbf{I}_n - \rho \mathbf{W}_1)^{-1} \mathbf{W}_1) + \text{tr}(\mathbf{W}_1' \mathbf{W}_1 (\mathbf{I}_n - \rho \mathbf{W}_1)^{-1} (\mathbf{I}_n - \rho \mathbf{W}_1)^{-1}) \\
&= \text{tr}(\mathbf{W}_1 (\mathbf{I}_n - \rho \mathbf{W}_1)^{-1} \mathbf{W}_1 (\mathbf{I}_n - \rho \mathbf{W}_1)^{-1}) + \text{tr}(((\mathbf{I}_n - \rho \mathbf{W}_1)' (\mathbf{I}_n - \rho \mathbf{W}_1))^{-1} \mathbf{W}_1' \mathbf{W}_1) \\
-E\left(\frac{\partial^2 \text{Ln}(L)}{\partial \rho \partial \sigma^2}\right) &= -\frac{1}{\sigma^2} \left[\rho \text{tr}(((\mathbf{I}_n - \rho \mathbf{W}_1)' (\mathbf{I}_n - \rho \mathbf{W}_1))^{-1} \mathbf{W}_1' \mathbf{W}_1) \right] \\
&\quad + \frac{1}{\sigma^2} \left[\text{tr}(((\mathbf{I}_n - \rho \mathbf{W}_1)' (\mathbf{I}_n - \rho \mathbf{W}_1))^{-1} \mathbf{W}_1') \right]
\end{aligned}$$

MODELO ESPACIAL AUTOREGRESSIVO MISTO

$$\mathbf{y} = \rho \mathbf{W}_1 \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$$

Neste modelo, para estimar ρ deve-se seguir alguns passos:

- 1) Fazer mínimos quadrados ordinários (OLS) no modelo : $\mathbf{y} = \mathbf{X} \boldsymbol{\beta}_0 + \boldsymbol{\varepsilon}_0$;
- 2) Fazer mínimos quadrados ordinários (OLS) no modelo :
 $\mathbf{W}_1 \mathbf{y} = \mathbf{X} \boldsymbol{\beta}_L + \boldsymbol{\varepsilon}_L$;
- 3) Calcular os resíduos $\mathbf{e}_0 = \mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}_0$ e $\mathbf{e}_L = \mathbf{W}_1 \mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}_L$;
- 4) Com $\mathbf{e}_0$ e $\mathbf{e}_L$ calculados, procurar ρ que maximize

$$\text{Ln}(L) = -\frac{n}{2} \ln \left(\frac{1}{n} (\mathbf{e}_0 - \rho \mathbf{e}_L)' (\mathbf{e}_0 - \rho \mathbf{e}_L) \right) + \ln |\mathbf{I}_n - \rho \mathbf{W}_1|$$

- 5) Com ρ estimado, calcular $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_0 - \hat{\rho} \hat{\boldsymbol{\beta}}_L$

Prova:

$$\mathbf{e}_0 = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}_0$$

$$\frac{\partial \mathbf{e}_0' \mathbf{e}_0}{\partial \boldsymbol{\beta}_0} = (-\mathbf{X})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_0)$$

$$\mathbf{X}'\mathbf{X}\boldsymbol{\beta}_0 = \mathbf{X}'\mathbf{y} \Rightarrow \hat{\boldsymbol{\beta}}_0 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Analogamente, tem-se que:

$$\mathbf{e}_L = \mathbf{W}_1\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_L \quad \hat{\boldsymbol{\beta}}_L = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}_1\mathbf{y}$$

$$(\mathbf{I}_n - \rho\mathbf{W}_1)\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{\varepsilon} = (\mathbf{I}_n - \rho\mathbf{W}_1)\mathbf{y} - \mathbf{X}\boldsymbol{\beta}$$

$$\frac{\partial \boldsymbol{\varepsilon}' \boldsymbol{\varepsilon}}{\partial \boldsymbol{\beta}} = (-\mathbf{X})'[(\mathbf{I}_n - \rho\mathbf{W}_1)\mathbf{y} - \mathbf{X}\boldsymbol{\beta}]$$

$$\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'(\mathbf{I}_n - \rho\mathbf{W}_1)\mathbf{y}$$

$$\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{I}_n\mathbf{y} - \rho\mathbf{X}'\mathbf{W}_1\mathbf{y}$$

$$\boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{I}_n\mathbf{y} - \rho(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}_1\mathbf{y}$$

$$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_0 - \rho\hat{\boldsymbol{\beta}}_L$$

$$\boldsymbol{\varepsilon} = \mathbf{y} - \rho\mathbf{W}_1\mathbf{y} - \mathbf{X}(\hat{\boldsymbol{\beta}}_0 - \rho\hat{\boldsymbol{\beta}}_L)$$

$$\boldsymbol{\varepsilon} = \mathbf{y} - \rho\mathbf{W}_1\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_0 + \rho\mathbf{X}\hat{\boldsymbol{\beta}}_L$$

$$\boldsymbol{\varepsilon} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_0 - \rho(\mathbf{W}_1\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_L) \Rightarrow \boldsymbol{\varepsilon} = \mathbf{e}_0 - \rho\mathbf{e}_L \quad \text{e} \quad \sigma^2 = \frac{1}{n}(\mathbf{e}_0 - \rho\mathbf{e}_L)'(\mathbf{e}_0 - \rho\mathbf{e}_L)$$

Calculando a variância dos parâmetros:

Sabendo que:

$$E(\mathbf{y}) = (\mathbf{I}_n - \rho\mathbf{W}_1)^{-1}\mathbf{X}\boldsymbol{\beta}$$

$$E(\mathbf{y}\mathbf{y}') = (\mathbf{I}_n - \rho\mathbf{W}_1)^{-1}\mathbf{X}\boldsymbol{\beta}(\mathbf{X}\boldsymbol{\beta})'(\mathbf{I}_n - \rho\mathbf{W}_1)^{-1} + (\mathbf{I}_n - \rho\mathbf{W}_1)^{-1}\sigma^2\mathbf{I}_n(\mathbf{I}_n - \rho\mathbf{W}_1)^{-1}$$

$$\ln(L) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}(\mathbf{e}_0 - \rho\mathbf{e}_L)'(\mathbf{e}_0 - \rho\mathbf{e}_L) + \ln |\mathbf{I}_n - \rho\mathbf{W}_1|$$

$$\frac{\partial \ln(L)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}(\mathbf{e}_0 - \rho\mathbf{e}_L)'(\mathbf{e}_0 - \rho\mathbf{e}_L)$$

$$\begin{aligned}
\frac{\partial \ln(L)}{\partial \rho} &= -\text{tr}((\mathbf{I}_n - \rho \mathbf{W}_1)^{-1} \mathbf{W}_1) + \frac{2}{2\sigma^2} (\mathbf{W}_1 \mathbf{y})' (\mathbf{y} - \rho \mathbf{W}_1 \mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\
\frac{\partial \ln(L)}{\partial \boldsymbol{\beta}} &= \frac{2}{2\sigma^2} \mathbf{X}' ((\mathbf{I}_n - \rho \mathbf{W}_1) \mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\
\frac{\partial^2 \ln(L)}{\partial (\sigma^2)^2} &= \frac{n}{2(\sigma^2)^2} - \frac{1}{(\sigma^2)^3} (\mathbf{e}_0 - \rho \mathbf{e}_L)' (\mathbf{e}_0 - \rho \mathbf{e}_L) \\
\frac{\partial^2 \ln(L)}{\partial \rho^2} &= -\text{tr}((\mathbf{I}_n - \rho \mathbf{W}_1)^{-1} \mathbf{W}_1 (\mathbf{I}_n - \rho \mathbf{W}_1)^{-1} \mathbf{W}_1) - \frac{\mathbf{y}' \mathbf{W}_1' \mathbf{W}_1 \mathbf{y}}{\sigma^2} \\
\frac{\partial^2 \ln(L)}{\partial \boldsymbol{\beta}^2} &= -\frac{\mathbf{X}' \mathbf{X}}{\sigma^2} \\
\frac{\partial^2 \ln(L)}{\partial \sigma^2 \partial \rho} &= -\frac{1}{(\sigma^2)^2} (\mathbf{y}' \mathbf{W}_1' \mathbf{y} - \rho \mathbf{y}' \mathbf{W}_1' \mathbf{W}_1 \mathbf{y} - \mathbf{y}' \mathbf{W}_1' \mathbf{X} \boldsymbol{\beta}) \\
\frac{\partial^2 \ln(L)}{\partial \boldsymbol{\beta} \partial \sigma^2} &= \frac{1}{(\sigma^2)^2} \mathbf{X}' ((\mathbf{I}_n - \rho \mathbf{W}_1) \mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\
\frac{\partial^2 \ln(L)}{\partial \boldsymbol{\beta} \partial \rho} &= -\frac{1}{\sigma^2} \mathbf{y}' \mathbf{W}_1 \mathbf{X} \\
-E\left(\frac{\partial^2 \ln(L)}{\partial (\sigma^2)^2}\right) &= -\frac{n}{2(\sigma^2)^2} + \frac{1}{(\sigma^2)^3} E(\boldsymbol{\varepsilon}' \boldsymbol{\varepsilon}) = -\frac{n}{2(\sigma^2)^2} + \frac{n\sigma^2}{(\sigma^2)^3} = \frac{n}{2(\sigma^2)^2} \\
-E\left(\frac{\partial^2 \ln(L)}{\partial \rho^2}\right) &= \text{tr}((\mathbf{I}_n - \rho \mathbf{W}_1)^{-1} \mathbf{W}_1 (\mathbf{I}_n - \rho \mathbf{W}_1)^{-1} \mathbf{W}_1) \\
&\quad + \frac{1}{\sigma^2} (\mathbf{X}\boldsymbol{\beta})' (\mathbf{I}_n - \rho \mathbf{W}_1)^{-1} \mathbf{W}_1' \mathbf{W}_1 (\mathbf{I}_n - \rho \mathbf{W}_1)^{-1} (\mathbf{X}\boldsymbol{\beta}) \\
&\quad + \text{tr}(((\mathbf{I}_n - \rho \mathbf{W}_1)' (\mathbf{I}_n - \rho \mathbf{W}_1))^{-1} \mathbf{W}_1' \mathbf{W}_1) \\
-E\left(\frac{\partial^2 \ln(L)}{\partial \boldsymbol{\beta}^2}\right) &= \frac{\mathbf{X}' \mathbf{X}}{\sigma^2} \\
-E\left(\frac{\partial^2 \ln(L)}{\partial \sigma^2 \partial \rho}\right) &= \frac{1}{(\sigma^2)^2} (\mathbf{X}\boldsymbol{\beta})' (\mathbf{I}_n - \rho \mathbf{W}_1)^{-1} \mathbf{W}_1' (\mathbf{I}_n - \rho \mathbf{W}_1)^{-1} (\mathbf{X}\boldsymbol{\beta}) \\
&\quad + \frac{1}{\sigma^2} \text{tr}(\mathbf{W}_1' ((\mathbf{I}_n - \rho \mathbf{W}_1)' (\mathbf{I}_n - \rho \mathbf{W}_1))^{-1}) \\
&\quad - \frac{1}{(\sigma^2)^2} \rho (\mathbf{X}\boldsymbol{\beta})' (\mathbf{I}_n - \rho \mathbf{W}_1)^{-1} \mathbf{W}_1' \mathbf{W}_1 (\mathbf{I}_n - \rho \mathbf{W}_1)^{-1} (\mathbf{X}\boldsymbol{\beta}) \\
&\quad - \frac{1}{\sigma^2} \rho \text{tr}(((\mathbf{I}_n - \rho \mathbf{W}_1)' (\mathbf{I}_n - \rho \mathbf{W}_1))^{-1} \mathbf{W}_1' \mathbf{W}_1) \\
&\quad - \frac{1}{(\sigma^2)^2} (\mathbf{X}\boldsymbol{\beta})' \mathbf{W}_1 (\mathbf{I}_n - \rho \mathbf{W}_1)^{-1} (\mathbf{X}\boldsymbol{\beta}) \\
-E\left(\frac{\partial^2 \ln(L)}{\partial \boldsymbol{\beta} \partial \sigma^2}\right) &= -\frac{1}{(\sigma^2)^2} [\mathbf{X}' \mathbf{X} \boldsymbol{\beta} - \mathbf{X}' \mathbf{X} \boldsymbol{\beta}] = 0 \\
-E\left(\frac{\partial^2 \ln(L)}{\partial \boldsymbol{\beta} \partial \rho}\right) &= \frac{1}{\sigma^2} \mathbf{X}' \mathbf{W}_1 (\mathbf{I}_n - \rho \mathbf{W}_1)^{-1} \mathbf{X} \boldsymbol{\beta}
\end{aligned}$$

MODELO COM AUTOCORRELAÇÃO ESPACIAL NO ERRO

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu}$$

$$\boldsymbol{\mu} = \lambda \mathbf{W}_2 \boldsymbol{\mu} + \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$$

Como no modelo anterior, deve-se seguir alguns passos:

- 1) Fazer mínimos quadrados ordinários (OLS) no modelo : $\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \boldsymbol{\varepsilon}$;
- 2) Calcular os resíduos : $\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$;
- 3) Procurar o valor de λ que maximiza a função de verossimilhança condicionada aos valores dos $\hat{\boldsymbol{\beta}}$ calculados

$$Ln(L) = -\frac{n}{2} \ln \left(\frac{1}{n} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})' (\mathbf{I}_n - \lambda \mathbf{W}_2)' (\mathbf{I}_n - \lambda \mathbf{W}_2) (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \right) + \ln |\mathbf{I}_n - \lambda \mathbf{W}_2|$$

- 4) Atualizar os valores dos $\hat{\boldsymbol{\beta}}$ usando o valor de $\hat{\lambda}$ calculado. Para obter o novo valor de $\hat{\boldsymbol{\beta}}$ pode-se usar mínimos quadrados generalizados (GLS): $\hat{\boldsymbol{\beta}} = \left[((\mathbf{I}_n - \hat{\lambda} \mathbf{W}_2) \mathbf{X})' ((\mathbf{I}_n - \hat{\lambda} \mathbf{W}_2) \mathbf{X}) \right]^{-1} ((\mathbf{I}_n - \hat{\lambda} \mathbf{W}_2) \mathbf{X})' (\mathbf{I}_n - \hat{\lambda} \mathbf{W}_2) \mathbf{y}$
- 5) Voltar para o passo 3) até obter a convergência dos resíduos;

Prova:

$$\boldsymbol{\mu} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$$

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \lambda \mathbf{W}_2 (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \boldsymbol{\varepsilon}$$

$$\boldsymbol{\varepsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) - \lambda \mathbf{W}_2 (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

$$\boldsymbol{\varepsilon} = (\mathbf{I}_n - \lambda \mathbf{W}_2) (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

Depois de estimar o valor de λ :

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \lambda \mathbf{W}_2 (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \boldsymbol{\varepsilon}$$

$$\boldsymbol{\varepsilon} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \lambda \mathbf{W}_2 \mathbf{y} + \lambda \mathbf{W}_2 \mathbf{X}\boldsymbol{\beta}$$

$$\boldsymbol{\varepsilon} = (\mathbf{I}_n - \lambda \mathbf{W}_2) \mathbf{y} - (\mathbf{I}_n - \lambda \mathbf{W}_2) \mathbf{X}\boldsymbol{\beta}$$

$$\frac{\partial \boldsymbol{\varepsilon}' \boldsymbol{\varepsilon}}{\partial \boldsymbol{\beta}} = -(\mathbf{I}_n - \lambda \mathbf{W}_2) \mathbf{X}' [(\mathbf{I}_n - \lambda \mathbf{W}_2) \mathbf{y} - (\mathbf{I}_n - \lambda \mathbf{W}_2) \mathbf{X}\boldsymbol{\beta}]$$

$$= -\mathbf{X}' (\mathbf{I}_n - \lambda \mathbf{W}_2)' (\mathbf{I}_n - \lambda \mathbf{W}_2) \mathbf{y} + \mathbf{X}' (\mathbf{I}_n - \lambda \mathbf{W}_2)' (\mathbf{I}_n - \lambda \mathbf{W}_2) \mathbf{X}\boldsymbol{\beta} = 0$$

$$\boldsymbol{\beta} = (\mathbf{X}' (\mathbf{I}_n - \lambda \mathbf{W}_2)' (\mathbf{I}_n - \lambda \mathbf{W}_2) \mathbf{X})^{-1} \mathbf{X}' (\mathbf{I}_n - \lambda \mathbf{W}_2)' (\mathbf{I}_n - \lambda \mathbf{W}_2) \mathbf{y}$$

$$\hat{\boldsymbol{\beta}} = \left[((\mathbf{I}_n - \hat{\lambda} \mathbf{W}_2) \mathbf{X})' ((\mathbf{I}_n - \hat{\lambda} \mathbf{W}_2) \mathbf{X}) \right]^{-1} ((\mathbf{I}_n - \hat{\lambda} \mathbf{W}_2) \mathbf{X})' (\mathbf{I}_n - \hat{\lambda} \mathbf{W}_2) \mathbf{y}$$

$$\hat{\beta}_D = (X_S' X_S)^{-1} X_S' Y_S$$

Para a convergência dos resíduos, usar os seguintes valores iniciais :

$$e_c = y - X\hat{\beta}$$

$$critério = 10^{-4}$$

$$convergência = 1$$

Depois de calculado $\hat{\beta}_D$:

$$e_\varepsilon = y - X\hat{\beta}_D$$

$$convergência = \sum_{i=1}^n |e_{\varepsilon_i} - e_{c_i}|$$

Enquanto convergência for maior que critério, continue a calcular e_ε ;

Calculando a variância dos parâmetros:

Sabendo que:

$$E(y) = X\beta$$

$$E((y - X\beta)(y - X\beta)') = (I_n - \lambda W_2)^{-1} \sigma^2 I_n (I_n - \lambda W_2)^{-1}$$

$$Ln(L) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} (y - X\beta)' (I_n - \lambda W_2)^{-1} (I_n - \lambda W_2) (y - X\beta) + \ln |I_n - \lambda W_2|$$

$$\frac{\partial Ln(L)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} (y - X\beta)' (I_n - \lambda W_2)^{-1} (I_n - \lambda W_2) (y - X\beta)$$

$$\frac{\partial Ln(L)}{\partial \lambda} = -tr((I_n - \lambda W_2)^{-1} W_2) + \frac{2}{2\sigma^2} (y - X\beta)' W_2 [(y - X\beta) - \lambda W_2 (y - X\beta)]$$

$$\frac{\partial Ln(L)}{\partial \beta} = \frac{2}{2\sigma^2} X' (I_n - \lambda W_2)^{-1} [(I_n - \lambda W_2) y - (I_n - \lambda W_2) X\beta]$$

$$\frac{\partial^2 Ln(L)}{\partial (\sigma^2)^2} = \frac{n}{2(\sigma^2)^2} - \frac{1}{(\sigma^2)^3} (y - X\beta)' (I_n - \lambda W_2)^{-1} (I_n - \lambda W_2) (y - X\beta)$$

$$\frac{\partial^2 Ln(L)}{\partial \lambda^2} = -tr((I_n - \lambda W_2)^{-1} W_2 (I_n - \lambda W_2)^{-1} W_2) - \frac{(y - X\beta)' W_2' W_2 (y - X\beta)}{\sigma^2}$$

$$\frac{\partial^2 Ln(L)}{\partial \beta^2} = -\frac{X' (I_n - \lambda W_2)^{-1} (I_n - \lambda W_2) X}{\sigma^2}$$

$$\frac{\partial^2 Ln(L)}{\partial \lambda \partial \sigma^2} = -\frac{1}{(\sigma^2)^2} (W_2 (y - X\beta)) [(y - X\beta) - \lambda W_2 (y - X\beta)]$$

$$\frac{\partial^2 Ln(L)}{\partial \beta \partial \sigma^2} = \frac{1}{(\sigma^2)^2} X' (I_n - \lambda W_2)^{-1} [(I_n - \lambda W_2) y - (I_n - \lambda W_2) X\beta]$$

$$\begin{aligned}
\frac{\partial^2 \text{Ln}(L)}{\partial \boldsymbol{\beta} \partial \lambda} &= \frac{1}{\sigma^2} [2\mathbf{X}'\mathbf{W}_2'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) - 2\lambda\mathbf{X}'\mathbf{W}_2'\mathbf{W}_2(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})] \\
-E\left(\frac{\partial^2 \text{Ln}(L)}{\partial (\sigma^2)^2}\right) &= -\frac{n}{2(\sigma^2)^2} + \frac{1}{(\sigma^2)^3} E(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}) = -\frac{n}{2(\sigma^2)^2} + \frac{n\sigma^2}{(\sigma^2)^3} = \frac{n}{2(\sigma^2)^2} \\
-E\left(\frac{\partial^2 \text{Ln}(L)}{\partial \lambda^2}\right) &= \text{tr}((\mathbf{I}_n - \lambda\mathbf{W}_2)^{-1}\mathbf{W}_2(\mathbf{I}_n - \lambda\mathbf{W}_2)^{-1}\mathbf{W}_2) + \text{tr}(\mathbf{W}_2'\mathbf{W}_2(\mathbf{I}_n - \lambda\mathbf{W}_2)^{-1}(\mathbf{I}_n - \lambda\mathbf{W}_2)^{-1}) \\
&= \text{tr}(\mathbf{W}_2(\mathbf{I}_n - \lambda\mathbf{W}_2)^{-1}\mathbf{W}_2(\mathbf{I}_n - \lambda\mathbf{W}_2)^{-1}) + \text{tr}(((\mathbf{I}_n - \lambda\mathbf{W}_2)'(\mathbf{I}_n - \lambda\mathbf{W}_2))^{-1}\mathbf{W}_2'\mathbf{W}_2) \\
-E\left(\frac{\partial^2 \text{Ln}(L)}{\partial \boldsymbol{\beta}^2}\right) &= \frac{\mathbf{X}'(\mathbf{I}_n - \lambda\mathbf{W}_2)'(\mathbf{I}_n - \lambda\mathbf{W}_2)\mathbf{X}}{\sigma^2} \\
-E\left(\frac{\partial^2 \text{Ln}(L)}{\partial \lambda \partial \sigma^2}\right) &= \frac{1}{\sigma^2} [\text{tr}(((\mathbf{I}_n - \lambda\mathbf{W}_2)'(\mathbf{I}_n - \lambda\mathbf{W}_2))^{-1}\mathbf{W}_2') - \lambda \text{tr}(((\mathbf{I}_n - \lambda\mathbf{W}_2)'(\mathbf{I}_n - \lambda\mathbf{W}_2))^{-1}\mathbf{W}_2'\mathbf{W}_2)] \\
-E\left(\frac{\partial^2 \text{Ln}(L)}{\partial \boldsymbol{\beta} \partial \sigma^2}\right) &= -\frac{1}{(\sigma^2)^2} [\mathbf{X}'(\mathbf{I}_n - \lambda\mathbf{W}_2)^{-1}(\mathbf{I}_n - \lambda\mathbf{W}_2)^{-1}\mathbf{X}\boldsymbol{\beta} - \mathbf{X}'(\mathbf{I}_n - \lambda\mathbf{W}_2)^{-1}(\mathbf{I}_n - \lambda\mathbf{W}_2)^{-1}\mathbf{X}\boldsymbol{\beta}] = 0 \\
-E\left(\frac{\partial^2 \text{Ln}(L)}{\partial \boldsymbol{\beta} \partial \lambda}\right) &= \frac{2\mathbf{X}'\mathbf{W}_2'\mathbf{X}\boldsymbol{\beta} - 2\mathbf{X}'\mathbf{W}_2'\mathbf{X}\boldsymbol{\beta}}{\sigma^2} + \frac{2\lambda\mathbf{X}'\mathbf{W}_2'\mathbf{W}_2\mathbf{X}\boldsymbol{\beta} - 2\lambda\mathbf{X}'\mathbf{W}_2'\mathbf{W}_2\mathbf{X}\boldsymbol{\beta}}{\sigma^2} = 0
\end{aligned}$$

MODELO ESPACIAL GERAL

$$\mathbf{y} = \rho\mathbf{W}_1\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu}$$

$$\boldsymbol{\mu} = \lambda\mathbf{W}_2\boldsymbol{\mu} + \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_n)$$

Neste modelo é possível ter $\mathbf{W}_1 = \mathbf{W}_2$, mas podem ocorrer resultados inesperados. Os resultados para esse modelo são apresentados abaixo:

$$\text{Ln}(L) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}(\mathbf{e}'\mathbf{B}'\mathbf{B}\mathbf{e}) + \ln|\mathbf{A}| + \ln|\mathbf{B}|$$

onde :

$$\mathbf{e} = (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

$$\mathbf{A} = (\mathbf{I}_n - \rho\mathbf{W}_1)$$

$$\mathbf{B} = (\mathbf{I}_n - \lambda\mathbf{W}_2)$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{B}'\mathbf{B}\mathbf{X})^{-1}\mathbf{X}'\mathbf{B}'\mathbf{B}\mathbf{A}\mathbf{y}$$

Prova :

$$\begin{aligned}
(\mathbf{I}_n - \rho \mathbf{W}_1) \mathbf{y} &= \mathbf{X} \boldsymbol{\beta} + \lambda \mathbf{W}_2 [(\mathbf{I}_n - \rho \mathbf{W}_1) \mathbf{y} - \mathbf{X} \boldsymbol{\beta}] + \boldsymbol{\varepsilon} \\
\boldsymbol{\varepsilon} &= (\mathbf{I}_n - \rho \mathbf{W}_1) \mathbf{y} - \mathbf{X} \boldsymbol{\beta} - \lambda \mathbf{W}_2 (\mathbf{I}_n - \rho \mathbf{W}_1) \mathbf{y} + \lambda \mathbf{W}_2 \mathbf{X} \boldsymbol{\beta} \\
\boldsymbol{\varepsilon} &= (\mathbf{I}_n - \lambda \mathbf{W}_2) (\mathbf{I}_n - \rho \mathbf{W}_1) \mathbf{y} - (\mathbf{I}_n - \lambda \mathbf{W}_2) \mathbf{X} \boldsymbol{\beta} \\
\boldsymbol{\varepsilon} &= \mathbf{B} \mathbf{A} \mathbf{y} - \mathbf{B} \mathbf{X} \boldsymbol{\beta} \\
\boldsymbol{\varepsilon} &= \mathbf{B} (\mathbf{A} \mathbf{y} - \mathbf{X} \boldsymbol{\beta}) \\
\frac{\partial \boldsymbol{\varepsilon}' \boldsymbol{\varepsilon}}{\partial \boldsymbol{\beta}} &= (-\mathbf{B} \mathbf{X})' (\mathbf{B} \mathbf{A} \mathbf{y} - \mathbf{B} \mathbf{X} \boldsymbol{\beta}) \\
\mathbf{X}' \mathbf{B}' \mathbf{B} \mathbf{X} \boldsymbol{\beta} &= \mathbf{X}' \mathbf{B}' \mathbf{B} \mathbf{A} \mathbf{y} \quad \Rightarrow \quad \hat{\boldsymbol{\beta}} = (\mathbf{X}' \mathbf{B}' \mathbf{B} \mathbf{X})^{-1} \mathbf{X}' \mathbf{B}' \mathbf{B} \mathbf{A} \mathbf{y} \\
\sigma^2 &= \frac{(\mathbf{A} \mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}})' \mathbf{B}' \mathbf{B} (\mathbf{A} \mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}})}{n}
\end{aligned}$$

Calculando a variância dos parâmetros:

Sabendo que:

$$\begin{aligned}
E(\mathbf{y}) &= (\mathbf{I}_n - \rho \mathbf{W}_1)^{-1} \mathbf{X} \boldsymbol{\beta} \\
E(\mathbf{y} \mathbf{y}') &= (\mathbf{I}_n - \rho \mathbf{W}_1)^{-1} \mathbf{X} \boldsymbol{\beta} (\mathbf{X} \boldsymbol{\beta})' (\mathbf{I}_n - \rho \mathbf{W}_1)^{-1} \\
&\quad + [(\mathbf{I}_n - \lambda \mathbf{W}_2) (\mathbf{I}_n - \rho \mathbf{W}_1)]^{-1} \sigma^2 \mathbf{I}_n [(\mathbf{I}_n - \lambda \mathbf{W}_2) (\mathbf{I}_n - \rho \mathbf{W}_1)]^{-1}
\end{aligned}$$

$$\begin{aligned}
Ln(L) &= -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) + \ln |\mathbf{I}_n - \lambda \mathbf{W}_2| + \ln |\mathbf{I}_n - \rho \mathbf{W}_1| \\
&\quad - \frac{1}{2\sigma^2} ((\mathbf{I}_n - \lambda \mathbf{W}_2) (\mathbf{I}_n - \rho \mathbf{W}_1) \mathbf{y} - (\mathbf{I}_n - \lambda \mathbf{W}_2) \mathbf{X} \boldsymbol{\beta})' ((\mathbf{I}_n - \lambda \mathbf{W}_2) (\mathbf{I}_n - \rho \mathbf{W}_1) \mathbf{y} - (\mathbf{I}_n - \lambda \mathbf{W}_2) \mathbf{X} \boldsymbol{\beta})
\end{aligned}$$

$$\begin{aligned}
\frac{\partial Ln(L)}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} (\mathbf{A} \mathbf{y} - \mathbf{X} \boldsymbol{\beta})' \mathbf{B}' \mathbf{B} (\mathbf{A} \mathbf{y} - \mathbf{X} \boldsymbol{\beta}) \\
\frac{\partial Ln(L)}{\partial \rho} &= -tr((\mathbf{I}_n - \rho \mathbf{W}_1)^{-1} \mathbf{W}_1) + \frac{2}{2\sigma^2} ((\mathbf{I}_n - \lambda \mathbf{W}_2) \mathbf{W}_1 \mathbf{y})' [(\mathbf{I}_n - \lambda \mathbf{W}_2) (\mathbf{I}_n - \rho \mathbf{W}_1) \mathbf{y} - (\mathbf{I}_n - \lambda \mathbf{W}_2) \mathbf{X} \boldsymbol{\beta}] \\
\frac{\partial Ln(L)}{\partial \lambda} &= -tr((\mathbf{I}_n - \lambda \mathbf{W}_2)^{-1} \mathbf{W}_2) - \frac{2}{2\sigma^2} (-\mathbf{W}_2 (\mathbf{I}_n - \rho \mathbf{W}_1) \mathbf{y} + \mathbf{W}_2 \mathbf{X} \boldsymbol{\beta})' [(\mathbf{I}_n - \lambda \mathbf{W}_2) (\mathbf{I}_n - \rho \mathbf{W}_1) \mathbf{y} - (\mathbf{I}_n - \lambda \mathbf{W}_2) \mathbf{X} \boldsymbol{\beta}] \\
\frac{\partial Ln(L)}{\partial \boldsymbol{\beta}} &= \frac{2}{2\sigma^2} \mathbf{X}' (\mathbf{I}_n - \lambda \mathbf{W}_2)' [(\mathbf{I}_n - \lambda \mathbf{W}_2) (\mathbf{I}_n - \rho \mathbf{W}_1) \mathbf{y} - (\mathbf{I}_n - \lambda \mathbf{W}_2) \mathbf{X} \boldsymbol{\beta}] \\
\frac{\partial^2 Ln(L)}{\partial (\sigma^2)^2} &= \frac{n}{2(\sigma^2)^2} - \frac{1}{(\sigma^2)^3} (\mathbf{A} \mathbf{y} - \mathbf{X} \boldsymbol{\beta})' \mathbf{B}' \mathbf{B} (\mathbf{A} \mathbf{y} - \mathbf{X} \boldsymbol{\beta}) \\
\frac{\partial^2 Ln(L)}{\partial \rho^2} &= -tr((\mathbf{I}_n - \rho \mathbf{W}_1)^{-1} \mathbf{W}_1 (\mathbf{I}_n - \rho \mathbf{W}_1)^{-1} \mathbf{W}_1) - \frac{1}{\sigma^2} [\mathbf{y}' \mathbf{W}_1' (\mathbf{I}_n - \lambda \mathbf{W}_2)' (\mathbf{I}_n - \lambda \mathbf{W}_2) \mathbf{W}_1 \mathbf{y}]
\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 Ln(L)}{\partial \lambda^2} = & -tr((\mathbf{I}_n - \lambda \mathbf{W}_2)^{-1} \mathbf{W}_2 (\mathbf{I}_n - \lambda \mathbf{W}_2)^{-1} \mathbf{W}_2) - \frac{1}{\sigma^2} [y'(\mathbf{I}_n - \rho \mathbf{W}_1)' \mathbf{W}_2' \mathbf{W}_2 (\mathbf{I}_n - \rho \mathbf{W}_1) y \\ & - y'(\mathbf{I}_n - \rho \mathbf{W}_1)' \mathbf{W}_2' \mathbf{W}_2 \mathbf{X} \beta \\ & - (\mathbf{X} \beta)' \mathbf{W}_2' \mathbf{W}_2 (\mathbf{I}_n - \rho \mathbf{W}_1) y + (\mathbf{X} \beta)' \mathbf{W}_2' \mathbf{W}_2 (\mathbf{X} \beta)]\end{aligned}$$

$$\frac{\partial^2 Ln(L)}{\partial \beta^2} = - \frac{\mathbf{X}'(\mathbf{I}_n - \lambda \mathbf{W}_2)'(\mathbf{I}_n - \lambda \mathbf{W}_2) \mathbf{X}}{\sigma^2}$$

$$\frac{\partial^2 Ln(L)}{\partial \rho \partial \sigma^2} = - \frac{1}{(\sigma^2)^2} [((\mathbf{I}_n - \lambda \mathbf{W}_2) \mathbf{W}_1 y)'((\mathbf{I}_n - \lambda \mathbf{W}_2)(\mathbf{I}_n - \rho \mathbf{W}_1) y - (\mathbf{I}_n - \lambda \mathbf{W}_2) \mathbf{X} \beta)]$$

$$\frac{\partial^2 Ln(L)}{\partial \lambda \partial \sigma^2} = \frac{1}{(\sigma^2)^2} [(-\mathbf{W}_2 (\mathbf{I}_n - \rho \mathbf{W}_1) y + \mathbf{W}_2 \mathbf{X} \beta)'((\mathbf{I}_n - \lambda \mathbf{W}_2)(\mathbf{I}_n - \rho \mathbf{W}_1) y - (\mathbf{I}_n - \lambda \mathbf{W}_2) \mathbf{X} \beta)]$$

$$\frac{\partial^2 Ln(L)}{\partial \beta \partial \sigma^2} = \frac{1}{(\sigma^2)^2} [((\mathbf{I}_n - \lambda \mathbf{W}_2) \mathbf{X})'((\mathbf{I}_n - \lambda \mathbf{W}_2)(\mathbf{I}_n - \rho \mathbf{W}_1) y - (\mathbf{I}_n - \lambda \mathbf{W}_2) \mathbf{X} \beta)]$$

$$\begin{aligned}\frac{\partial^2 Ln(L)}{\partial \rho \partial \lambda} = & - \frac{1}{\sigma^2} [y' \mathbf{W}_1' \mathbf{W}_2 (\mathbf{I}_n - \lambda \mathbf{W}_2)(\mathbf{I}_n - \rho \mathbf{W}_1) y + y'(\mathbf{I}_n - \rho \mathbf{W}_1)' \mathbf{W}_2' (\mathbf{I}_n - \lambda \mathbf{W}_2) \mathbf{W}_1 y] \\ & - \frac{1}{\sigma^2} [y' \mathbf{W}_1' \mathbf{W}_2' (\mathbf{I}_n - \lambda \mathbf{W}_2) \mathbf{X} \beta - (\mathbf{X} \beta)' \mathbf{W}_2' (\mathbf{I}_n - \lambda \mathbf{W}_2) \mathbf{W}_1 y]\end{aligned}$$

$$\frac{\partial^2 Ln(L)}{\partial \rho \partial \beta} = - \frac{\mathbf{X}'(\mathbf{I}_n - \lambda \mathbf{W}_2)'(\mathbf{I}_n - \lambda \mathbf{W}_2) \mathbf{W}_1 y}{\sigma^2}$$

$$\begin{aligned}\frac{\partial^2 Ln(L)}{\partial \lambda \partial \beta} = & \frac{1}{\sigma^2} [-\mathbf{X}' \mathbf{W}_2' (\mathbf{I}_n - \lambda \mathbf{W}_2)(\mathbf{I}_n - \rho \mathbf{W}_1) y - \mathbf{X}'(\mathbf{I}_n - \lambda \mathbf{W}_2)' \mathbf{W}_2 (\mathbf{I}_n - \rho \mathbf{W}_1) y + \mathbf{X}' \mathbf{W}_2' (\mathbf{I}_n - \lambda \mathbf{W}_2) \mathbf{X} \beta \\ & + \mathbf{X}'(\mathbf{I}_n - \lambda \mathbf{W}_2)' \mathbf{W}_2 \mathbf{X} \beta]\end{aligned}$$

$$-E\left(\frac{\partial^2 Ln(L)}{\partial (\sigma^2)^2}\right) = -\frac{n}{2(\sigma^2)^2} + \frac{1}{(\sigma^2)^3} E(\boldsymbol{\varepsilon}' \boldsymbol{\varepsilon}) = -\frac{n}{2(\sigma^2)^2} + \frac{n\sigma^2}{(\sigma^2)^3} = \frac{n}{2(\sigma^2)^2}$$

$$\begin{aligned}-E\left(\frac{\partial^2 Ln(L)}{\partial \rho^2}\right) = & tr((\mathbf{I}_n - \rho \mathbf{W}_1)^{-1} \mathbf{W}_1 (\mathbf{I}_n - \rho \mathbf{W}_1)^{-1} \mathbf{W}_1) \\ & + \frac{1}{\sigma^2} tr(\mathbf{W}_1' (\mathbf{I}_n - \lambda \mathbf{W}_2)' (\mathbf{I}_n - \lambda \mathbf{W}_2) \mathbf{W}_1 (\mathbf{I}_n - \rho \mathbf{W}_1)^{-1} (\mathbf{X} \beta)(\mathbf{X} \beta)' (\mathbf{I}_n - \rho \mathbf{W}_1)^{-1}) \\ & + tr(\mathbf{W}_1' (\mathbf{I}_n - \lambda \mathbf{W}_2)' (\mathbf{I}_n - \lambda \mathbf{W}_2) \mathbf{W}_1 [(\mathbf{I}_n - \lambda \mathbf{W}_2)(\mathbf{I}_n - \rho \mathbf{W}_1)]^{-1})\end{aligned}$$

$$-E\left(\frac{\partial^2 Ln(L)}{\partial \lambda^2}\right) = tr((\mathbf{I}_n - \lambda \mathbf{W}_2)^{-1} \mathbf{W}_2 (\mathbf{I}_n - \lambda \mathbf{W}_2)^{-1} \mathbf{W}_2) + tr(\mathbf{W}_2' \mathbf{W}_2 ((\mathbf{I}_n - \lambda \mathbf{W}_2)' (\mathbf{I}_n - \lambda \mathbf{W}_2))^{-1})$$

$$-E\left(\frac{\partial^2 Ln(L)}{\partial \beta^2}\right) = \frac{\mathbf{X}'(\mathbf{I}_n - \lambda \mathbf{W}_2)'(\mathbf{I}_n - \lambda \mathbf{W}_2) \mathbf{X}}{\sigma^2}$$

$$\begin{aligned}
-E\left(\frac{\partial^2 \text{Ln}(L)}{\partial \rho \partial \sigma^2}\right) &= \frac{1}{(\sigma^2)^2} \text{tr}(\mathbf{W}_1'(\mathbf{I}_n - \lambda \mathbf{W}_2)'(\mathbf{I}_n - \lambda \mathbf{W}_2)(\mathbf{X}\boldsymbol{\beta})(\mathbf{X}\boldsymbol{\beta})'(\mathbf{I}_n - \rho \mathbf{W}_1)^{-1}) \\
&\quad + \frac{1}{\sigma^2} \text{tr}(\mathbf{W}_1'(\mathbf{I}_n - \lambda \mathbf{W}_2)[(\mathbf{I}_n - \lambda \mathbf{W}_2)(\mathbf{I}_n - \rho \mathbf{W}_1)]^{-1}) \\
&\quad - \frac{1}{(\sigma^2)^2} \text{tr}((\mathbf{X}\boldsymbol{\beta})'(\mathbf{I}_n - \lambda \mathbf{W}_2)'(\mathbf{I}_n - \lambda \mathbf{W}_2)\mathbf{W}_1(\mathbf{I}_n - \rho \mathbf{W}_1)^{-1}(\mathbf{X}\boldsymbol{\beta})) \\
&= + \frac{1}{\sigma^2} \text{tr}(\mathbf{W}_1'(\mathbf{I}_n - \rho \mathbf{W}_1)^{-1}) \\
-E\left(\frac{\partial^2 \text{Ln}(L)}{\partial \lambda \partial \sigma^2}\right) &= \frac{1}{(\sigma^2)^2} \text{tr}(\mathbf{W}_2'(\mathbf{I}_n - \lambda \mathbf{W}_2)(\mathbf{X}\boldsymbol{\beta})(\mathbf{X}\boldsymbol{\beta})') + \frac{1}{\sigma^2} \text{tr}(\mathbf{W}_2'(\mathbf{I}_n - \lambda \mathbf{W}_2)^{-1}) \\
&\quad - \frac{1}{(\sigma^2)^2} \text{tr}((\mathbf{X}\boldsymbol{\beta})'(\mathbf{I}_n - \lambda \mathbf{W}_2)' \mathbf{W}_2(\mathbf{X}\boldsymbol{\beta})) - \frac{1}{(\sigma^2)^2} \text{tr}((\mathbf{X}\boldsymbol{\beta})' \mathbf{W}_2'(\mathbf{I}_n - \lambda \mathbf{W}_2)(\mathbf{X}\boldsymbol{\beta})) \\
&\quad + \frac{1}{(\sigma^2)^2} \text{tr}(\mathbf{W}_2'(\mathbf{I}_n - \lambda \mathbf{W}_2)(\mathbf{X}\boldsymbol{\beta})(\mathbf{X}\boldsymbol{\beta})') \\
&= \frac{1}{\sigma^2} \text{tr}(\mathbf{W}_2'(\mathbf{I}_n - \lambda \mathbf{W}_2)^{-1}) \\
-E\left(\frac{\partial^2 \text{Ln}(L)}{\partial \boldsymbol{\beta} \partial \sigma^2}\right) &= \frac{1}{(\sigma^2)^2} (\mathbf{X}'(\mathbf{I}_n - \lambda \mathbf{W}_2)'(\mathbf{I}_n - \lambda \mathbf{W}_2)((\mathbf{I}_n - \rho \mathbf{W}_1)(\mathbf{I}_n - \rho \mathbf{W}_1)^{-1} \mathbf{X}\boldsymbol{\beta} - \mathbf{X}\boldsymbol{\beta})) = 0 \\
-E\left(\frac{\partial^2 \text{Ln}(L)}{\partial \rho \partial \lambda}\right) &= \frac{1}{\sigma^2} \text{tr}(\mathbf{W}_1' \mathbf{W}_2(\mathbf{I}_n - \lambda \mathbf{W}_2)(\mathbf{X}\boldsymbol{\beta})(\mathbf{X}\boldsymbol{\beta})'(\mathbf{I}_n - \rho \mathbf{W}_1)^{-1}) \\
&\quad + \text{tr}(\mathbf{W}_1' \mathbf{W}_2[(\mathbf{I}_n - \lambda \mathbf{W}_2)(\mathbf{I}_n - \rho \mathbf{W}_1)]^{-1}) \\
&\quad + \text{tr}(\mathbf{W}_2'(\mathbf{I}_n - \lambda \mathbf{W}_2)\mathbf{W}_1(\mathbf{I}_n - \rho \mathbf{W}_1)^{-1}[(\mathbf{I}_n - \lambda \mathbf{W}_2)'(\mathbf{I}_n - \lambda \mathbf{W}_2)]^{-1}) \\
&\quad - \frac{1}{\sigma^2} \text{tr}((\mathbf{X}\boldsymbol{\beta})'(\mathbf{I}_n - \lambda \mathbf{W}_2)' \mathbf{W}_2 \mathbf{W}_1(\mathbf{I}_n - \rho \mathbf{W}_1)^{-1}(\mathbf{X}\boldsymbol{\beta})) \\
&= \text{tr}(\mathbf{W}_1' \mathbf{W}_2[(\mathbf{I}_n - \lambda \mathbf{W}_2)(\mathbf{I}_n - \rho \mathbf{W}_1)]^{-1}) \\
&\quad + \text{tr}(\mathbf{W}_2'(\mathbf{I}_n - \lambda \mathbf{W}_2)\mathbf{W}_1(\mathbf{I}_n - \rho \mathbf{W}_1)^{-1}[(\mathbf{I}_n - \lambda \mathbf{W}_2)'(\mathbf{I}_n - \lambda \mathbf{W}_2)]^{-1}) \\
-E\left(\frac{\partial^2 \text{Ln}(L)}{\partial \rho \partial \boldsymbol{\beta}}\right) &= \frac{1}{\sigma^2} (\mathbf{X}'(\mathbf{I}_n - \lambda \mathbf{W}_2)'(\mathbf{I}_n - \lambda \mathbf{W}_2)\mathbf{W}_1(\mathbf{I}_n - \rho \mathbf{W}_1)^{-1} \mathbf{X}\boldsymbol{\beta}) \\
-E\left(\frac{\partial^2 \text{Ln}(L)}{\partial \lambda \partial \boldsymbol{\beta}}\right) &= \frac{1}{\sigma^2} \left(\mathbf{X}'(\mathbf{I}_n - \lambda \mathbf{W}_2)' \mathbf{W}_2((\mathbf{I}_n - \rho \mathbf{W}_1)(\mathbf{I}_n - \rho \mathbf{W}_1)^{-1} \mathbf{X}\boldsymbol{\beta} - \mathbf{X}\boldsymbol{\beta}) \right. \\
&\quad \left. - \mathbf{X}' \mathbf{W}_2'(\mathbf{I}_n - \lambda \mathbf{W}_2)((\mathbf{I}_n - \rho \mathbf{W}_1)(\mathbf{I}_n - \rho \mathbf{W}_1)^{-1} \mathbf{X}\boldsymbol{\beta} - \mathbf{X}\boldsymbol{\beta}) \right) = 0
\end{aligned}$$