



**UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA DE TRANSPORTES
CURSO DE ENGENHARIA CIVIL**

LUCAS TITO PEREIRA SOBREIRA

**ANÁLISE EXPLORATÓRIA DE ACIDENTES COM PEDETTRES EM ZONAS DE
TRÁFEGO DE FORTALEZA**

FORTALEZA

2016

LUCAS TITO PEREIRA SOBREIRA

ANÁLISE EXPLORATÓRIA DE ACIDENTES COM PEDESTRES EM ZONAS DE
TRÁFEGO DE FORTALEZA

Monografia apresentada ao Departamento de Engenharia de Transportes do Curso de Engenharia Civil da Universidade Federal do Ceará, como requisito parcial para a obtenção do título de Bacharel em Engenharia Civil.

Orientador: Prof. Dr. Flávio José Craveiro
Cunto

FORTALEZA

2016

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

- S659a Sobreira, Lucas Tito Pereira.
Análise Exploratória de Acidentes com Pedestres em Zonas de Tráfego de Fortaleza / Lucas Tito Pereira Sobreira. – 2016.
77 f. : il. color.
- Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Centro de Tecnologia, Curso de Engenharia Civil, Fortaleza, 2016.
Orientação: Prof. Dr. Flávio José Craveiro Cunto.
1. Segurança Viária. 2. Análise Exploratória. 3. Análise Exploratória Espacial. 4. Modelos de Previsão de Acidentes. I. Título.

CDD 620

AGRADECIMENTOS

À minha família, Clarice, Glauco, Davi, Rafinha e Deusito, por todo o carinho e incentivo durante minha caminhada acadêmica.

À minha namorada, Maiara, por todo o amor e companheirismo durante todos esses anos.

Ao professor Flávio pela excelente orientação, pelo incentivo durante toda a minha graduação e por conselhos durante minha caminhada.

Aos demais professores que contribuíram para meu crescimento como estudante, em especial os professores Bruno Bertoncini e Felipe Loureiro.

Às empresas Porto Freire Engenharia, Aval Engenharia e Napax Engenharia por terem me proporcionado um ambiente de aprendizado prático na profissão.

Ao CNPq pela bolsa de iniciação científica concedida ao longo da graduação.

À CAPES pela oportunidade de participar do programa Ciência sem Fronteiras.

Aos meus amigos de faculdade, Thomás, João Victor e Caio, pelas divertidas e longas noites de estudo compartilhadas.

Aos avaliadores do presente trabalho, professores Francisco Moraes e Marcos Timbó, pela atenção e prontidão que aceitaram participar da minha banca.

RESUMO

O desenvolvimento de estudos em longo prazo de segurança viária para melhoria das condições aos usuários de trânsito vem se tornando uma prática comum no Brasil e fora dele. Este Projeto de Graduação apresenta uma análise exploratória das variáveis relacionadas à ocorrência de atropelamentos em áreas de Fortaleza. A princípio, um banco de dados de 283 zonas de tráfego foi consolidado com variáveis representando o uso do solo, a geometria e operação das vias e características socioeconômicas. Estudos anteriores indicam que análises realizadas em área apresentam melhores resultados quando incorporado o fator espacial da amostra. Objetivando o levantamento de hipóteses sobre essas variáveis, a análise foi realizada em três etapas: i) análise exploratória não espacial dos dados, ii) análise exploratória espacial dos dados, e iii) comparação de Modelos de Previsão de Acidentes (MPA) binomiais negativos e espaciais. Os resultados indicaram que variáveis relacionadas à geração de viagens e ao constante movimento de pedestres estão ligadas à ocorrência de atropelamentos. A dependência espacial dos acidentes foi analisada e confirmada. A comparação dos parâmetros e da distribuição de resíduos entre MPA binomiais negativos e espaciais confirmou que a utilização de fatores espaciais em análises em área é de extrema significância para a obtenção de modelos mais ajustados.

Palavras-chave: Segurança viária. Análise exploratória. Análise exploratória espacial. Modelos de previsão de acidentes.

ABSTRACT

The development of long term studies related to road safety improvement has become a common practice in Brazil and abroad. This Under Graduation Project presents an exploratory analysis of variables related to pedestrian accidents in areas of the city of Fortaleza-Brazil. Firstly, a 283 traffic zones database was consolidated with variables that represent land use, road geometry and operation and socioeconomics characteristics. Previous studies indicate that analysis made in traffic zones presents better results when the spatial factor is considered. To achieve the main goal, which was to investigate the relationship between the collected variables and the pedestrian accidents, three steps were followed: i) exploratory non spatial data analysis, ii) exploratory spatial data analysis, and iii) comparison of Accidents Prediction Models (APM) using binomial negative distribution and spatial dependence. The results showed that the variables related to trip generation and constant pedestrian flow are linked to the frequency of pedestrian accidents. The spatial dependence of the accidents was analyzed and proven. The comparison of parameters and residuals distribution between APM binomial negative and spatial confirmed that the use of spatial factors in area APM is highly significant to increase the model fitness.

Keywords: Road safety. Exploratory analysis. Exploratory spatial analysis. Accidents prediction models.

LISTA DE FIGURAS

Figura 1 - Tipos usuais de correlação encontrados em gráficos de dispersão. Sem correlação (a), correlação linear forte (b), correlação moderada e variância constante (c) e correlação moderada e variância considerável (d).	18
Figura 2 - Identificação dos componentes do <i>boxplot</i>	19
Figura 3 - Representação da contiguidade. Critério de Rook (a) e critério de Queen (b)	20
Figura 4 - Diagrama de espalhamento de Moran	22
Figura 5 - Mapa de <i>LISA</i> da variável acidentes com ciclistas em 2009 a 2011 em Fortaleza..	24
Figura 6 - Mapa de Moran da variável acidentes com ciclistas em 2009 a 2011 em Fortaleza	24
Figura 7 - Exemplo de <i>CURE plot</i> : resíduos acumulados X fluxo veicular.....	29
Figura 8 - Resumo da metodologia do trabalho	31
Figura 9 - 283 zonas de tráfego do Fortaleza 2040	32
Figura 10 - Base de logradouros georreferenciados da AMC sobre as zonas de tráfego do Fortaleza 2040	33
Figura 11 - Metodologia de georreferenciamento dos acidentes.....	34
Figura 12 - Mapa com <i>shapes</i> de zonas de tráfego e fiscalização eletrônica	35
Figura 13 - <i>Boxplot</i> dos atropelamentos	42
Figura 14 - <i>Boxplot</i> das interseções semaforizadas	46
Figura 15 - <i>CURE plots</i> dos modelos binomiais negativos	52
Figura 16 - Mapa de Moran dos atropelamentos	55
Figura 17 - <i>CURE plots</i> dos modelos espaciais.....	57
Figura 18 - Mapa de Moran dos resíduos do modelo 8.....	58
Figura 19 - Mapa de Moran dos resíduos do modelo <i>SAR</i>	58

LISTA DE TABELAS

Tabela 1 - Resumo das variáveis dependentes utilizadas em estudos de MPA em área	14
Tabela 2 - Diretrizes para avaliação do coeficiente de Pearson segundo Dancey e Reidy (2005)	16
Tabela 3 - Estatística descritiva dos atropelamentos	39
Tabela 4 - Estatística descritiva das variáveis de operação do tráfego.....	40
Tabela 5 - Estatística descritiva das variáveis de características de infraestrutura das vias.....	40
Tabela 6 - Estatística descritiva das variáveis de uso do solo	41
Tabela 7 - Estatística descritiva dos dados de características socioeconômicas	41
Tabela 8 - Correlação de Pearson entre as variáveis que apresentaram correlação igual ou superior a 0,40 e a variável de atropelamentos.....	44
Tabela 9 - Resumo da interpretação dos gráficos de dispersão	45
Tabela 10 - Variáveis selecionadas na análise exploratória não espacial dos dados.....	49
Tabela 11 - Coeficientes dos modelos desenvolvidos com a aplicação da técnica inclusão sequencial de variáveis	50
Tabela 12 - Indicadores do ajuste global dos modelos binomiais negativos.....	51
Tabela 13 - Influência das variáveis na predição dos atropelamentos utilizando o modelo 8 ..	53
Tabela 14 - Índice global de Moran e significância estatística do índice de Moran das variáveis.....	54
Tabela 15 - Coeficientes e parâmetros dos modelos espaciais	56
Tabela 16 - Indicadores do ajuste global do modelo 8 binomial negativo e do modelo SAR ..	57
Tabela 17 - Índice global de Moran dos resíduos dos modelos 8 e SAR.....	57
Tabela 18 - Influência de cada variável na predição de atropelamentos utilizando o modelo SAR	60

SUMÁRIO

1 INTRODUÇÃO	10
1.1 Considerações iniciais e problemas motivadores	10
1.2 Questões motivadoras	11
1.3 Objetivos	12
<i>1.3.1 Objetivo geral</i>	12
<i>1.3.2 Objetivos específicos</i>	12
2 REVISÃO BIBLIOGRÁFICA	13
2.1 Estudos anteriores de desempenho de segurança viária em área	13
2.2 Análise exploratória dos dados	15
<i>2.2.1 Análise exploratória não espacial dos dados</i>	15
<i>2.2.2 Análise exploratória espacial dos dados</i>	19
2.3 Modelos de previsão de acidentes	25
<i>2.3.1 Modelos de previsão de acidentes com distribuição binomial negativa</i>	25
<i>2.3.2 Modelos de previsão de acidentes espaciais</i>	26
<i>2.3.3 Avaliação do ajuste do modelo e comparação entre modelos</i>	27
3 METODOLOGIA	31
3.1 Coleta de dados	32
3.2 Consolidação do banco de dados	34
3.3 Análise exploratória não espacial dos dados	35
3.4 Análise exploratória complementar dos dados	36
3.5 Análise exploratória espacial dos dados	37
3.6 MPA binomial negativo Vs MPA espacial	38
4. RESULTADOS E ANÁLISES DOS RESULTADOS	39
4.1 Apresentação dos dados	39
4.2 Análise exploratória não espacial dos dados	41
4.3 Análise exploratória complementar dos dados	49

4.4 Análise exploratória espacial dos dados	53
4.5 MPA binomial negativo Vs MPA espacial	55
5 CONCLUSÕES E SUGESTÕES PARA FUTUROS TRABALHOS.....	61
6 REFERÊNCIAS BIBLIOGRÁFICAS	63
ANEXOS A – CORRELAÇÃO DE PEARSON DE TODAS AS VARIÁVEIS	69
ANEXOS B – GRÁFICOS DE DISPERSÃO	70
ANEXOS C – MAPAS DE MORAN	76

1 INTRODUÇÃO

1.1 Considerações iniciais e problemas motivadores

Acidentes de trânsito representam uma considerável parcela no número de mortalidade da população brasileira. De acordo com o Ministério da Saúde, em 2012, acidentes de transporte terrestre representaram a 9ª colocação no índice de mortalidade com 8.058 mortes. Além de fatalidades, os acidentes de trânsito correspondem, também, a um elevado valor de custo anual para a sociedade. No ano de 2014, foram gastos aproximadamente 40 bilhões de reais com acidentes em rodovias federais, estaduais e municipais (IPEA 2015), valor que representa 0,72% do PIB brasileiro de 2014.

Com relação a acidentes fatais com pedestres, em 2012, 19,7% do total de vítimas no trânsito brasileiro foram pedestres, segundo a Organização Mundial da Saúde (OMS). Já em Fortaleza, essa porcentagem representa 45% do total (165 fatalidades no total de 367) conforme dados apresentados pela Autarquia Municipal de Trânsitos, Serviços Públicos e de Cidadania de Fortaleza (AMC), para o ano de 2011. Comparando com outros países, nos Estados Unidos, em 2013, foram registradas 32.719 mortes no trânsito, sendo 4.735 (14,47%) de pedestres. Dessas fatalidades, 73% ocorreram em zonas urbanas (NHTSA, 2015).

A utilização de indicadores de desempenho de segurança viária (DSV) em planejamentos de transporte em longo prazo vem se tornando uma prática comum em estudos fora do Brasil, como mostrado por Washington *et al* (2006) e Herbel *et al* (2009). De acordo com Torres (2016), o entendimento de indicadores do DSV e a tomada de decisão baseada na análise dos mesmos podem diminuir aspectos negativos em uma rede de transportes, ocasionando uma importante redução nos gastos econômicos e sociais, refletidos numa queda no número de acidentes e na necessidade de intervenções para melhoria da segurança viária. A prefeitura de Fortaleza está desenvolvendo um planejamento de transporte estratégico para a cidade. O programa chamado Fortaleza 2040 divide a cidade em 283 zonas de tráfego e busca incorporar o fator da segurança em seu desenvolvimento. Ferreira (2010) cita que a inclusão de parâmetros de DSV no planejamento de transportes tem potencial reconhecido e deveriam ser mais utilizados em planos estratégicos.

A grande maioria das pesquisas relacionadas à segurança viária foca no desenvolvimento de modelos de previsão de acidentes (MPA) que consistem na relação entre o número de acidentes esperado para cada zona e variáveis explicativas de exposição, de uso do solo, entre outros (Hadayeghi *et al*, 2003; Ukkusuri *et al*, 2011; Dolatsara, 2014). O foco

principal deste Projeto de Graduação será a análise exploratória das variáveis que influenciam a ocorrência de atropelamentos em zonas de tráfego, visando obter-se um maior entendimento da relação entre eles. MPA serão utilizados como parte de uma análise exploratória complementar, para uma maior investigação da ordem de grandeza de influência de cada variável.

A quantidade e os tipos de variáveis que podem contribuir para que um acidente ocorra são diversas: desde características geométricas da via ao tipo de agressividade dos motoristas em determinada região da cidade. Em análises feitas com atropelamentos em interseções e segmentos desenvolvidas em outros países, constatou-se que as mais significativas variáveis são a exposição de veículos e pedestres (Persaud e Lyon, 2012 e Dolatsara, 2014). Para análises em nível de zonas de tráfego, a estimação dos fluxos de veículos e pedestres é feita de forma indireta, a partir de fatores de uso do solo (área comercial, residencial, escolar), sociais (densidade da população) e econômicos (renda de acordo com domicílios) (Ukkusuri *et al*, 2011). Além das variáveis de exposição, os fatores geométricos (número de faixas e quilometro de vias) e operacionais (número de interseções semaforizadas e de paradas de ônibus) das vias de cada zona também serão considerados (Quddus, 2008; Siddiqui *et al*, 2012).

1.2 Questões motivadoras

A partir da contextualização da problemática apresentada torna-se possível o estabelecimento das seguintes questões motivadoras:

- I. Quais variáveis relacionadas ao uso do solo, à geometria e à operação da via e a características socioeconômicas podem ser utilizadas para representar a ocorrência de atropelamentos em um nível de agregação espacial de zonas de tráfego?
- II. Como efetivar uma análise exploratória não espacial a fim de levantar hipóteses sobre a relação entre os atropelamentos e as variáveis coletadas?
- III. Como efetivar uma análise exploratória espacial a fim de investigar a autocorrelação espacial dos atropelamentos e a correlação espacial entre os atropelamentos e as variáveis coletadas?
- IV. Quais as principais diferenças nos parâmetros das variáveis significativas e nos resíduos de modelos não espaciais e espaciais globais para a estimação da frequência de atropelamentos?

1.3 Objetivos

1.3.1 Objetivo geral

O objetivo geral desse Projeto de Graduação é avaliar através de uma análise exploratória não espacial e espacial de dados a relação entre os atropelamentos em zonas de tráfego de fortaleza com variáveis que representem características relativas ao uso do solo, aos aspectos sociais e econômicos, além de atributos operacionais e geométricos das vias.

1.3.2 Objetivos específicos

Os seguintes objetivos específicos foram considerados no desenvolvimento deste Projeto de Graduação:

- I. Consolidar um banco de dados com variáveis relacionadas ao uso do solo, à geometria e à operação da via e a características socioeconômicas que possam ser utilizadas para representar a ocorrência de atropelamentos em um nível de agregação espacial de zonas de tráfego;
- II. Realizar uma análise exploratória não espacial a fim de levantar hipóteses sobre a relação entre os atropelamentos e as variáveis coletadas;
- III. Realizar uma análise exploratória espacial das variáveis para investigação da possível autocorrelação espacial dos atropelamentos e da correlação espacial entre os atropelamentos e as variáveis coletadas;
- IV. Desenvolver MPA espaciais e não espaciais a fim de avaliar as principais diferenças nos parâmetros das variáveis significativas e nos resíduos gerados na estimação da frequência dos atropelamentos.

2 REVISÃO BIBLIOGRÁFICA

A revisão bibliográfica tem como objetivo verificar as metodologias e os conceitos aplicados em estudos de temas relacionados a análises exploratórias não espaciais e espaciais de atropelamentos, além de verificar técnicas aplicadas em MPA com pedestres desenvolvidos em área.

2.1 Estudos anteriores de desempenho de segurança viária em área

Segundo Gomes *et al*, 2015, o nível de agregação utilizado em relação às possíveis variáveis explicativas está diretamente relacionado ao tipo de análise dos sistemas de transportes que se deseja efetivar. Técnicas macroscópicas desenvolvidas para zonas de tráfego são utilizadas para planejamentos estratégicos, buscando uma melhor compreensão da relação entre as variáveis e a ocorrência dos acidentes. Já análises microscópicas feitas com interseções e segmentos viários apresentam melhores resultados quando aplicados a planejamentos de nível operacional.

Em geral, as análises realizadas em estudos internacionais de atropelamentos em zonas de tráfego utilizam variáveis de caráter sociais e econômicos e, também, de uso do solo para representar os níveis de exposição de cada entidade. Além disso, características operacionais e geométricas também são amplamente encontradas nesses modelos.

Modelos de previsão de acidentes representam formulações matemáticas em que as variáveis independentes são associadas a parâmetros para representar a frequência de acidentes. Os MPA são reconhecidos como potentes ferramentas na determinação de locais críticos e no entendimento da relação causa-efeito entre as variáveis e os acidentes. Além disso, a eficácia de estratégias de segurança viária pode ser verificada de maneira mais coerente (Persaud e Lyon, 2007; Geedipally e Lord, 2010). A formação e os diversos tipos de MPA serão apresentados com mais detalhes nas próximas seções.

Ukkusuri *et al* (2011) desenvolveram dois tipos de MPA para atropelamentos em zonas da cidade de Nova Iorque. O primeiro deles considerou a divisão das zonas de acordo com o CEP, o outro utilizou regiões segmentadas pelo censo da cidade. Utilizando diversos tipos de variáveis socioeconômicas, de uso do solo e de características das vias e agregando cinco anos de dados de atropelamentos, os autores verificaram que as variáveis *população total*, *população com menos de 17 anos*, *uso do solo comercial* e *uso do solo industrial*

estavam relacionadas ao aumento de atropelamentos nas zonas. Já as variáveis *uso do solo residencial e população com mais de 65 anos* foram associadas a menos atropelamentos.

Hashimoto (2005) realizou uma análise exploratória espacial de atropelamentos em 732 zonas da cidade de Hillsborough-EUA. Mapas de densidade dos acidentes foram usados para análise da espacialidade desses e MPA de distribuição binomial negativa foram aplicados para verificação da significância estatística de cada variável. Foi utilizada uma série histórica de três anos de acidentes e os resultados mostraram que as variáveis de *uso do solo de comércio e de serviços, uso do solo residencial e densidade de rodovias estaduais* foram diretamente proporcionais à frequência de acidentes. Já as variáveis de *renda média por domicílio e a porcentagem de moradores idosos* foram inversamente proporcionais aos acidentes. A ocorrência de espacialidade dos atropelamentos foi verificada nos mapas de densidade.

Para 463 zonas de tráfego de Toronto-CAN, Hadayeghi *et al* (2003) desenvolveram MPA binomiais negativos e espaciais para a frequência de acidentes totais e severos (com feridos e fatalidades) para o ano de 1996. Aplicando MPA binomiais negativos as seguintes variáveis foram consideradas significativas: *exposição de veículos, número de domicílios, quilômetros de via principal e densidade de interseções*. Os pesquisadores também checaram que em locais com elevados limites de velocidade, o número de acidentes tendeu a diminuir. A inclusão da espacialidade aos modelos melhorou o valor do R^2 em 17% em relação aos binomiais negativos. De acordo com os autores, isso se deu devido ao aumento de parâmetros no modelo (um parâmetro por zona para cada variável), pois apenas a variável *densidade de interseções* apresentou dependência espacial. A Tabela 1 apresenta um resumo das variáveis dependentes utilizadas em estudos de MPA em área.

Tabela 1 - Resumo das variáveis dependentes utilizadas em estudos de MPA em área

Autores	Local	Tamanho amostral	Variáveis dependentes
Ukkusuri <i>et al</i> , 2011	NY, USA	180	Pop. Total, pop. Com menos de 17 anos, pop. Com mais de 65 anos e uso do solo.
Huang <i>et al</i> , 2010	FL, USA	67	Pop. Entre 15 e 24 anos, intensidade de tráfego.
Hashimoto, 2005	Hillsborough, USA	732	Uso do solo de comércio e de serviços, uso do solo residencial e densidade de rodovias estaduais.
Hadayeghi <i>et al</i> , 2003	Toronto, CAN	463	Fluxo de veículos, número de domicílios, quilômetros de via principal e densidade de interseções.

Fonte: Próprio autor (2016)

Além das variáveis citadas acima, o índice de entropia vem se destacando como uma variável importante para a caracterização de atropelamentos e do fluxo de pedestres. Ouyang e Bejleri (2014) desenvolveram um estudo sobre a ocorrência de acidentes de trânsito em zonas em que uma das variáveis era o índice de entropia, que apresentou proporcionalidade direta com todos os tipos de acidentes analisados: total de acidentes, atropelamentos, acidentes fatais, acidentes com danos materiais e acidentes com bicicletas.

O índice de entropia (Equação 2.1) é uma variável que representa o grau de diversidade de uso do solo de certa zona. Esse índice tem valores de 0 a 1, em que 0 corresponde a locais de uso do solo único e 1 é relativo a zonas de diversidade igualmente distribuída. De acordo com Ewing e Cervero (2010), o índice de entropia é altamente utilizado em estudos de geração de viagens.

$$I_{ent} = \sum_{i=1}^n - p_i * \ln p_i / \ln n \quad (2.1)$$

Em que I_{ent} é o índice de entropia, p_i é a área do tipo de uso do solo i e n representa o total de tipos de uso do solo.

Chen e Shen (2016) desenvolveram modelos para analisar a severidade dos acidentes com ciclistas na cidade de Seattle-EUA e utilizaram o índice de entropia como uma das variáveis do estudo. Os resultados mostraram que o aumento da variedade do uso do solo estava relacionado com a diminuição da severidade dos acidentes.

2.2 Análise exploratória dos dados

Nessa seção serão discutidas as principais atividades propostas para a realização de análises exploratórias de dados espaciais e não espaciais com foco na identificação das principais variáveis que influenciam a frequência e a severidade dos atropelamentos em zonas de tráfego.

2.2.1 Análise exploratória não espacial dos dados

Nessa seção serão apresentadas algumas das técnicas mais utilizadas na análise exploratória não espacial dos dados. Análises de duas variáveis (bivariadas) são realizadas geralmente com a aplicação do coeficiente de Pearson e de gráficos de dispersão. Já

investigação do comportamento de uma variável se dá com a utilização de histogramas, *boxplots* e mapas de distribuição (Guimarães, 2008; Medri, 2011).

2.2.1.1 Correlação de Pearson

O coeficiente de correlação de Pearson (Equação 2.2) é uma importante ferramenta estatística para a análise do comportamento de duas variáveis. Garson (2009) define esse coeficiente como “uma medida de associação bivariada do grau de relacionamento entre duas variáveis”.

$$R = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2.2)$$

Em que: R representa o coeficiente de correlação de Pearson, x_i e y_i são os valores para as variáveis x e y na posição i , \bar{x} e \bar{y} são as médias globais das variáveis x e y , n representa o número total de observações de cada variável.

Os valores obtidos com o uso dessa equação variam de -1 a 1, em que coeficientes de valor igual a 1 representam proporcionalidade linear direta perfeita, ou seja, com o crescimento de uma variável, a outra crescerá de igual modo. Já o resultado de -1 indicaria uma proporcionalidade inversa às variáveis. Caso o resultado seja nulo, a relação entre as variáveis não apresenta nenhuma linearidade.

Resultados de valores extremos (-1, 0 ou 1) raramente são encontrados nos estudos práticos. Dancey e Reidy (2005) sugerem diretrizes, apresentadas na Tabela 2 com valores em módulo, para a avaliação dos resultados encontrados para o coeficiente.

Tabela 2 - Diretrizes para avaliação do coeficiente de Pearson segundo Dancey e Reidy (2005)

Valor de R	Grau de correlação
0,10 a 0,39	Fraco
0,40 a 0,69	Moderado
0,70 a 1,00	Forte

Fonte: Dancey e Reidy (2005)

Huang *et al* (2010) utilizaram a correlação de Pearson como uma das ferramentas para investigar a relação entre variáveis explicativas em MPA. Nesse estudo, os pesquisadores avaliaram a inclusão de três variáveis no modelo: *intensidade de tráfego*, *densidade da*

população e nível de urbanização. Essas três variáveis apresentaram correlação forte entre si e a apenas *intensidade de tráfego* foi incorporada ao modelo por apresentar maior correlação com a frequência de acidentes.

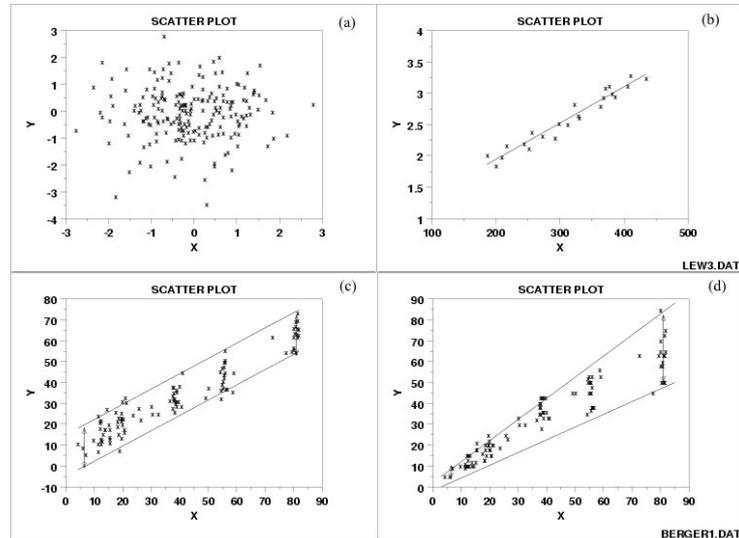
Gomes *et al* (2015) analisaram em nível estratégico e tático a correlação entre as variáveis explicativas e os acidentes totais e com vítimas, com o intuito de agregar ou excluir variáveis da análise. Para o nível estratégico, as variáveis *consumo de energia na indústria e consumo de energia no comércio e serviço* foram agregadas por apresentarem forte correlação entre si e com os acidentes. Já em âmbito tático, a variável *extensão de vias locais* foi excluída da análise por apresentar alta correlação com *interseções não semaforizadas* e possuir menor correlação com a frequência dos acidentes.

2.2.1.2 Gráficos de dispersão

O coeficiente de correlação de Pearson é uma boa maneira de se averiguar a intensidade da relação entre duas variáveis, entretanto ela não é uma boa ferramenta para avaliar como se dá essa relação. Os gráficos de dispersão bivariados apresentam de forma visual uma ideia do comportamento das variáveis, visando responder às seguintes questões: 1) as variáveis são relacionadas? 2) As variáveis são linearmente relacionadas? 3) As variáveis são não-linearmente relacionadas? 4) A variação do eixo y depende do eixo x? (Chambers, 1983).

O gráfico de dispersão consiste na plotagem bidimensional de pontos das variáveis dependente e independente. NIST/SEMATECH (2013) apresenta alguns tipos de correlação usualmente encontradas nos dados: sem correlação, correlação linear forte, correlação moderada e variância relativamente constante e correlação moderada e variância considerável. A Figura 1 apresenta os tipos de correlação supracitados.

Figura 1 - Tipos usuais de correlação encontrados em gráficos de dispersão. Sem correlação (a), correlação linear forte (b), correlação moderada e variância constante (c) e correlação moderada e variância considerável (d).



Fonte: <http://www.itl.nist.gov/div898/handbook/eda/section3/scatter1.htm>

Gomes *et al* (2015) utilizaram gráficos de dispersão para analisar a relação e a dispersão entre as variáveis *consumo de energia* e *número de semáforos* e os acidentes totais e com vítimas. Foi verificada relação positiva entre os acidentes e as variáveis dependentes, além de maior dispersão dos dados para maiores valores da variável dependente.

Kim *et al* (2016) aplicaram gráficos de dispersão na análise da relação das variáveis *população*, *emprego* e *renda* com os acidentes com ciclistas, automóveis e pedestres para 131 zonas de tráfego de Honolulu-Hawaii. Os autores verificaram que as três variáveis apresentaram correlação positiva com todos os tipos de acidentes.

2.2.1.3 Boxplot

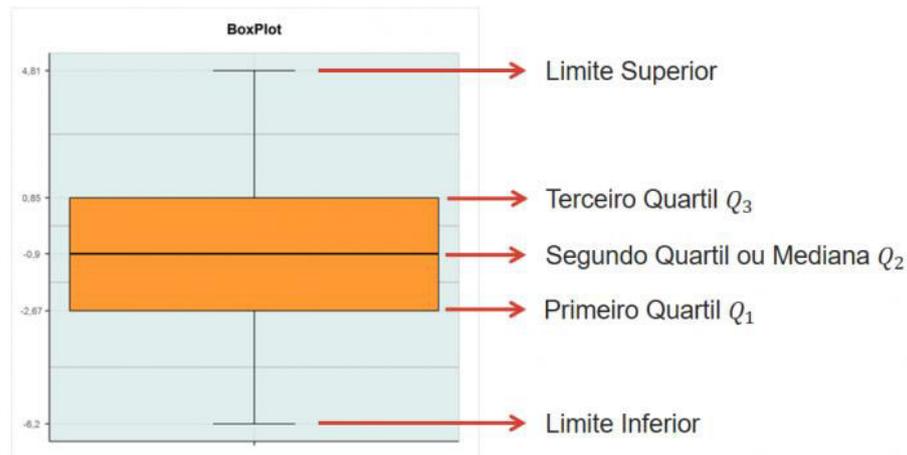
O *boxplot* é uma ferramenta que representa graficamente a distribuição de uma variável quantitativa. Suas principais funções são a verificação da variabilidade dos dados e identificação de *outliers* presentes na amostra. O *boxplot* (Figura 2) é formado pelo primeiro e terceiro quartil e pela mediana dos dados, além de exibir limites superiores e inferiores (Equações 2.3 e 2.4) para a identificação visual de valores extremos (Medri, 2011).

$$\text{Limite Superior} = Q_1 + 1,5 (Q_3 - Q_1) \quad (2.3)$$

$$\text{Limite Inferior} = Q_3 - 1,5 (Q_3 - Q_1) \quad (2.4)$$

Em que: Q_1 representa a observação referente ao primeiro quartil, Q_3 é a observação referente ao terceiro quartil.

Figura 2 - Identificação dos componentes do *boxplot*



Fonte: <http://www.portalection.com.br/estatistica-basica/31-boxplot>

Simões *et al* (2015) e Gopaul *et al* (2016) utilizaram o *boxplot* para caracterizar os dados de acidentes e para identificar a presença de *outliers* na amostra.

2.2.2 Análise exploratória espacial dos dados

De acordo com Anselin (1998), a análise exploratória espacial de dados é constituída por um conjunto de técnicas que busca facilitar o entendimento e a visualização da distribuição espacial dos dados, a identificação de locais atípicos e *outliers* e a verificação de padrões de associação espacial. O conceito central dessa ferramenta passa pelo princípio da autocorrelação espacial entre zonas, em que as análises são feitas levando em consideração as características de zonas com proximidades à estudada.

Um estudo realizado por Anselin (1992) demonstrou que para dados com dependência espacial, a utilização de procedimentos estatísticos convencionais, que não incorporam a espacialidade dos dados, afeta a validade dos resultados. Nos últimos anos, estudos de frequência de acidentes em zonas de tráfego vêm sendo realizados comparando técnicas com e sem a incorporação da espacialidade, os resultados apontam que a aplicação de métodos espaciais gera melhores resultados (Hadayeghi *et al*, 2003; Huang *et al*, 2010; Hadayeghi *et al*, 2010; Siddiqui *et al*, 2012; Pirdavani *et al*, 2012).

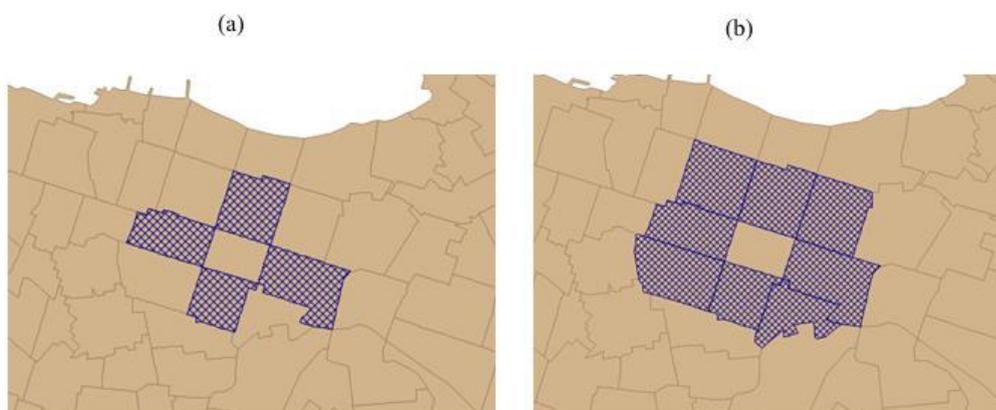
2.2.2.1 Índices de autocorrelação e matrizes de adjacência

De acordo com Queiróz (2003), na estatística espacial, a correlação entre uma ou mais variáveis representa a tendência delas, quando associadas a uma determinada localização, apresentarem maior semelhança a zonas vizinhas do que a outras zonas da amostra. A autocorrelação é utilizada para a investigação do comportamento de uma única variável. Autocorrelações positivas indicam que a ocorrência da variável em uma zona está associada com a presença dessa variável em zonas vizinhas, gerando um efeito de aglomeração. Já uma autocorrelação negativa representa casos em que a ocorrência da variável em uma zona repulsa a presença da variável nas zonas de proximidade.

Os índices de autocorrelação têm a função de representar quantitativamente a relação da zona estudada com as zonas vizinhas. Já a matriz de adjacência, geralmente representada pela letra W , apresenta a proximidade das zonas vizinhas. Para um estudo que contenha n zonas, a matriz W apresentaria tamanho $(n \times n)$, em que cada elemento w_{ij} quantificaria a ponderação espacial entre a zona i e a zona j .

Anselin (1992) apresenta dois modos diferentes para o cálculo dos w_{ij} : contiguidade e distância. A contiguidade possibilita a análise de três formas diferentes das proximidades das zonas: apenas fronteira em comum (critério de *Rook*), apenas vértices em comum (critério de *Bishop*) e fronteiras e vértices em comum (critério de *Queen*). Para o caso de contiguidade, a matriz de adjacência seria formada por binários (0,1), em que valores de 1 representam zonas próximas e de 0, zonas não próximas. A Figura 3 ilustra os critérios de *Rook* e *Queen*.

Figura 3 - Representação da contiguidade. Critério de *Rook* (a) e critério de *Queen* (b)



Fonte: Próprio autor (2016)

Já w_{ij} baseados na distância podem ser calculados de duas formas baseadas nos centroides das zonas. Uma das formas é a definição de raios de distância que limitariam a proximidade ou não das zonas, gerando uma matriz W binária (0,1). O outro método utiliza o inverso da distância entre os centroides das zonas para gerar a matriz W .

2.2.2.2 Índice global de Moran

O índice global de Moran é um indicador de autocorrelação que tem como objetivo analisar a dependência espacial de um conjunto de dados. A formulação desse índice será exposta na Equação 2.5.

$$I = \left[\frac{n}{\sum_{i=1}^n z_i^2} \right] * \left[\frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} * z_i * z_j}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \right] \quad (2.5)$$

Em que: I corresponde ao índice global de Moran, n é o número total de zonas, w_{ij} é o elemento correspondente da matriz de adjacência W , z_i é a diferença entre o valor do atributo na zona i e a média de todos os atributos, z_j é a diferença entre o valor dos atributos vizinhos da zona i e a média de todos os atributos.

Valores positivos para o índice global de Moran indicam uma autocorrelação positiva, em que a maior parte dos pares de zonas vizinhas analisados tem seu valor do mesmo lado quando comparados à média, ou seja, os dois são superiores ou inferiores a ela, representando locais de concentração de altos ou baixos valores. No caso de valores negativos ocorre o inverso, em que a maior parte dos pares de zonas vizinhas possui resultados de sinal diferente quando comparados à média, representando aleatoriedade à amostra de dados. Valores de I superiores a 0,3 são indicações de forte autocorrelação espacial (O'Sullivan e Unwin, 2010).

A verificação da validade estatística do índice de Moran é de suma importância para a sua utilização. Para realizá-la, é necessária a associação dos dados com uma distribuição estatística, em geral as abordagens de normal padrão e técnica de permutação são aplicadas (Queiróz, 2003). A abordagem normal padrão consiste na consideração da distribuição da variável com normal padrão, média igual a zero e variância igual a um, assim a significância é obtida a partir da comparação do escore Z obtido e valores de tabelas dessa distribuição. Já a técnica de permutação consiste na obtenção de índices de Moran simulados a partir da substituição aleatória dos n valores correspondentes às zonas, com isso, compara-se o I observado com os diversos I obtidos empiricamente. Se o observado estiver situado em um

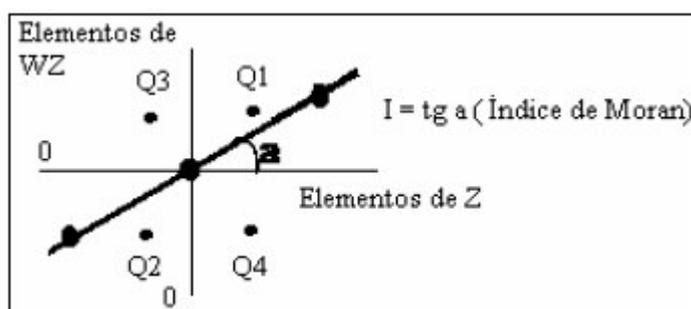
extremo da distribuição simulada, existe a significância estatística (Anselin, 1992). A existência de significância estatística representa a rejeição da hipótese nula de completa aleatoriedade espacial na amostra, que representaria uma variável que se distribui sem qualquer relação de espacialidade (Carvalho, 1997).

Quddus (2008), Siddiqui *et al* (2012), Pirdavani *et al* (2012) e Simões *et al* (2015) utilizaram o índice global de Moran em suas análises para a verificação de espacialidade nas variáveis utilizadas.

2.2.2.3 Diagramas de espalhamento de Moran

Desenvolvido por Anselin (1992), o diagrama de espalhamento de Moran (Figura 4) consiste na comparação do valor normalizado (Z) de uma variável da zona i com a média normalizada (WZ) da mesma variável de seus vizinhos. O diagrama é dividido em quatro quadrantes, em que Q1 e Q2 representam áreas de autocorrelação espacial positiva (Q1: valores positivos na zona i e na média das vizinhas, Q2: valores negativos na zona i e na média das vizinhas) e Q3 e Q4 apresentam áreas de autocorrelação espacial negativa (Q3: valor negativo na zona i e média positiva nas vizinhas, Q4: valor positivo na zona i e média negativa nas vizinhas). A inclinação da reta indica o valor do índice global de Moran.

Figura 4 - Diagrama de espalhamento de Moran



Fonte: Câmara *et al*, 2000

O diagrama permite a identificação de valores extremos representados por pontos que estejam distantes da tendência central. De acordo com Queiróz (2003), a presença desses *outliers* deve-se a problemas na especificação da matriz de adjacência ou na escala espacial de observação dos dados. Além disso, os *outliers* podem representar regiões de transição espacial, geralmente presentes nos quadrantes Q3 e Q4.

2.2.2.4 Índice local de Moran, Mapa de Moran e Mapa de LISA

O índice local de Moran e os mapas de Moran e *LISA* (*Local Indication of Spatial Association*) fazem parte do grupo de estatísticas espaciais locais, facilitando a identificação de aglomerações e valores extremos com uma análise mais individual de cada zona.

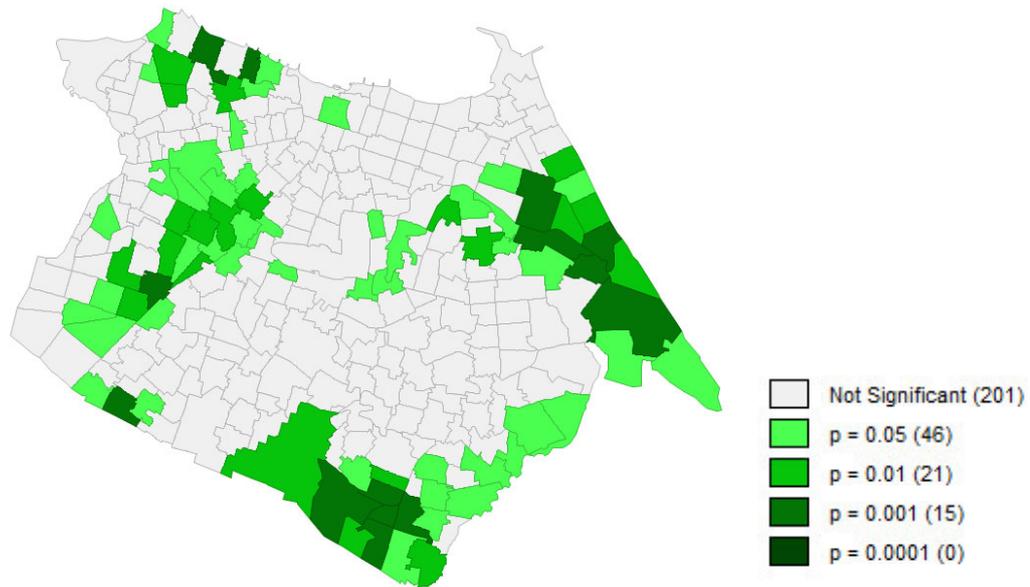
O cálculo do índice local de Moran (Equação 2.6) é similar ao índice global de Moran, a diferença é que para o local os valores são obtidos para cada zona e analisados de forma individual.

$$I_i = \frac{\sum_{j=1}^n w_{ij} * z_i * z_j}{\sum_{i=1}^n z_i^2} \quad (2.6)$$

Em que: I_i corresponde ao índice local de Moran, n é o número total de zonas, w_{ij} é o elemento correspondente da matriz de adjacência W , z_i é a diferença entre o valor do atributo na zona i e a média de todos os atributos, z_j é a diferença entre o valor dos atributos vizinhos da zona i e a média de todos os atributos.

Um dos principais aspectos dessa ferramenta é a sua significância estatística, que ocorre utilizando os mesmos métodos da validação do índice global de Moran. Como exemplo, o mapa de *LISA* (Figura 5) apresenta o índice local de Moran de acordo com seu nível de significância (sem significância, significância de 95%, significância de 99% e significância de 99,9%). A verificação de uma alta significância pode indicar a ocorrência de aglomeração, para altos ou baixos valores. Já locais sem significância estatística podem representar áreas de transição (Queiróz, 2003).

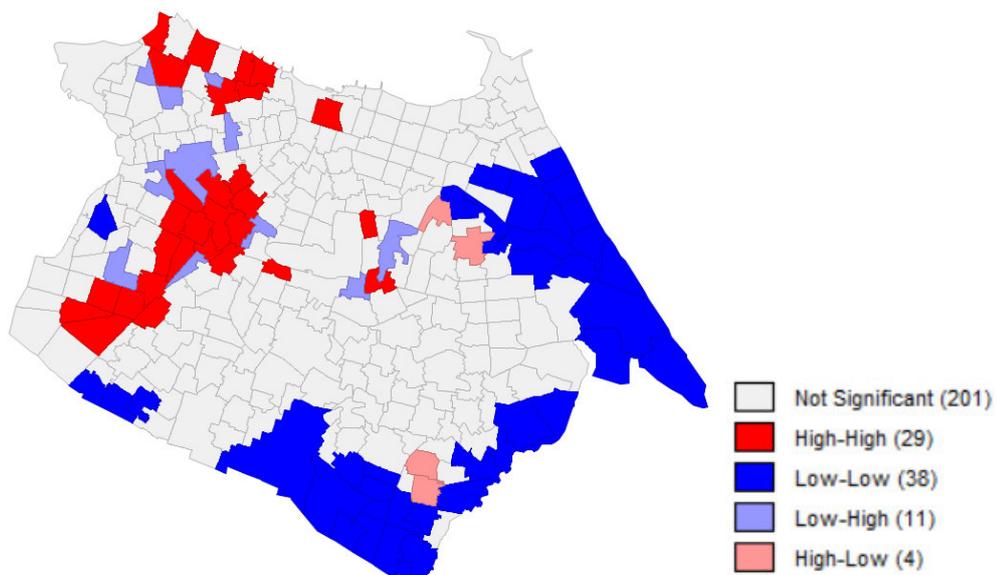
Figura 5 - Mapa de *LISA* da variável acidentes com ciclistas em 2009 a 2011 em Fortaleza



Fonte: Próprio autor (2016)

Já o mapa de Moran (Figura 6) é a representação em forma de mapa do diagrama de dispersão de Moran. Ele utiliza a classificação dos quadrantes do diagrama de dispersão (Q1 = *High-High*, Q2 = *Low-Low*, Q3 = *Low-High* e Q4 = *High-Low*) apresentando os locais que tiveram significância estatística. Os locais sem significância estatística são referenciados como não significantes.

Figura 6 - Mapa de Moran da variável acidentes com ciclistas em 2009 a 2011 em Fortaleza



Fonte: Próprio autor (2016)

2.3 Modelos de previsão de acidentes

Como citado na Seção 2.1, os MPA são formulações matemáticas, amplamente utilizadas em estudos de segurança viária, desenvolvidas a partir da associação de parâmetros e variáveis explicativas para a predição da frequência dos acidentes.

Segundo Hauer e Hakkert (1989), os primeiros MPA desenvolvidos a partir de modelos de regressão linear consideravam distribuição normal e homoscedasticidade (variância constante) para o fenômeno dos acidentes. Entretanto essas hipóteses foram consideradas não representativas para a natureza dos acidentes de trânsito. Pesquisas seguintes mostraram que a incorporação de distribuições de Poisson e binomial negativa aos modelos melhorava a aderência dos erros à característica dos acidentes que apresentam, em geral, comportamento disperso (Bonneson e McCoy, 1993; Persaud e Mucsi, 1995). Cunto (2008) cita que a utilização de modelos lineares generalizados (MLG) é uma possível alternativa para lidar com a incompatibilidade da suposição de distribuição normal de modelos tradicionais de regressão, devido ao padrão de dispersão dos dados.

2.3.1 Modelos de previsão de acidentes com distribuição binomial negativa

A utilização de MLG com a distribuição binomial negativa para representação da ocorrência de acidentes em MPA já é consagrada entre pesquisadores da área (Quddus, 2008; Cunto *et al*, 2011; Persaud e Lyon, 2012; Pirdavani *et al*, 2012; Sobreira e Cunto, 2014). A Equação 2.7 apresenta a formulação de MLG.

$$Y = \alpha * \left[\prod_i A_i^{\beta_i} \right] * e^{\sum_j (\gamma_j B_j)} \quad (2.7)$$

Em que: Y representa o número esperado de acidentes em um determinado período, A e B são vetores de variáveis independentes, α , β e γ são vetores de coeficientes do modelo.

Nos estudos citados, normalmente as variáveis de exposição foram incluídas no vetor A e as outras variáveis, no vetor B .

Para verificar que os acidentes se adequam à distribuição binomial negativa, Bonneson e McCoy (1993) desenvolveram o parâmetro de dispersão σ_d , que relaciona a variância do modelo com o grau de liberdade do modelo. Valores de σ_d próximos a um indicam aderência à distribuição. Cunto *et al* (2011) e Sobreira e Cunto (2014) utilizaram o σ_d para verificar que

os acidentes seguiam a distribuição binomial negativa e obtiveram os resultados de 1,13 e 1,42, respectivamente. A distribuição foi considerada como adequada em ambos os casos.

2.3.2 Modelos de previsão de acidentes espaciais

Como citado anteriormente, análises espaciais buscam incorporar possíveis tendências de espacialidade presente nos dados. Segundo Anselin (1992), dois tipos de modelos são aplicados na econometria espacial tradicional: modelo espacial autoregressivo (*Spatial AutoRegressive Model – SAR*) e modelo do erro espacial (*Spatial Error Model – SEM*). Esses modelos apresentam formatação clássica, ou seja, são modelos de regressão linear, por isso, é preciso que a distribuição dos erros seja normal, com média igual a zero e variância (σ^2) para que as inferências realizadas sejam representativas (Bailey e Gatrell, 1995).

A principal característica do modelo autoregressivo é a incorporação da média dos acidentes nas zonas vizinhas à estudada como um dos fatores do modelo. A Equação 2.8 exhibe a sua formulação.

$$Y_i = \rho WY_i + \beta X_i + \varepsilon, \text{ onde } \varepsilon \sim N(0, \sigma^2) \quad (2.8)$$

Em que: Y_i representa o número esperado de acidentes em um determinado período, ρ é o coeficiente espacial autoregressivo, WY_i é a média dos acidentes das zonas vizinhas, β expressa um vetor com os coeficientes das variáveis, X_i é a matriz de variáveis independentes, ε representa o erro determinado utilizando distribuição normal com média igual a zero e variância σ^2 .

O coeficiente espacial autoregressivo (ρ) verifica a autocorrelação espacial da variável dependente. Caso esse coeficiente seja significativamente estatístico, essa correlação existe (Anselin, 1992).

Por outro lado, o modelo espacial do erro incorpora o erro associado a cada zona vizinha na predição dos acidentes da zona estudada. As Equações 2.9 e 2.10 apresentam a formulação desse modelo.

$$Y_i = \beta X_i + u_i \quad (2.9)$$

$$u_i = \lambda W u_i + \varepsilon_i, \text{ onde } \varepsilon_i \sim N(0, \sigma^2) \quad (2.10)$$

Em que: Y_i representa o número esperado de acidentes em um determinado período, β expressa um vetor com os coeficientes das variáveis, X_i é a matriz de variáveis independentes, u_i é o termo espacial do erro, λ é o coeficiente autoregressivo do erro, Wu_i é a média dos erros nas zonas vizinhas, ε representa o erro determinado utilizando distribuição normal com média igual a zero variância σ^2 .

Assim como no *SAR*, o coeficiente autoregressivo do erro (λ) indica a autocorrelação espacial dos erros do modelo. Caso seja estatisticamente significativo, a correlação existe (Anselin, 1992; Quddus, 2008).

Em uma análise sobre a frequência de acidentes de trânsito em Londres-ENG, Quddus (2008) comparou resultados utilizando os modelos espaciais *SAR* e *SEM* com modelos de distribuição binomial negativa. Ele concluiu que as duas análises foram similares. Já Delmelle e Thill (2008) utilizaram esses dois modelos espaciais para verificar a ocorrência de acidentes com ciclistas na cidade de Buffalo-EUA. Os resultados foram satisfatórios. Por fim, Simões *et al* (2015) utilizaram técnicas de *SAR* e *SEM* para verificar a espacialidade de acidentes de trânsito em Lisboa-POR. Os dados não apresentaram correlação espacial, então a utilização dos modelos foi descartada.

A fim de melhorar o desempenho dos modelos espaciais, alguns autores aplicaram técnicas estatísticas mais avançadas que utilizam a distribuição de Poisson no desenvolvimento dos MPA espaciais. Os resultados mostraram que a implementação desses modelos, assim como o *SAR* e o *SEM*, gera melhores resultados do que técnicas que não incorporam a espacialidade (Hadayeghi, 2009; Hadayeghi *et al*, 2010; Pirdavani *et al*, 2012; Siddiqui *et al*, 2012).

2.3.3 Avaliação do ajuste do modelo e comparação entre modelos

Segundo Hauer (2004), a avaliação dos parâmetros dos MPA deve ser realizada com base nos resíduos produzidos por esse modelo ao comparado com a frequência observada de acidentes. Usualmente, dois ou mais modelos são desenvolvidos, para que se possa ter uma ideia de como a adição de variáveis está influenciando nos resultados dos modelos.

Os parâmetros dos modelos supracitados são calculados a partir do método da máxima verossimilhança, uma técnica que maximiza a probabilidade dos parâmetros calculados se ajustarem aos acidentes observados. Assim, uma das medidas de validação e comparação global entre modelos é a verificação do valor do logaritmo da verossimilhança, em que quanto maior esse valor, melhor o ajuste do modelo (Torres, 2016).

Akaike (1974) desenvolveu uma medida amplamente utilizada por pesquisadores na comparação da qualidade do ajuste de um MPA. O critério de informação de Akaike (*Akaike Information Criterion – AIC*) compara modelos baseando-se no equilíbrio entre o viés e a variância que a inclusão de variáveis traz a esses. Por exemplo, modelos com menos variáveis tendem a possuir menor viés e maior variância, já a inclusão de mais variáveis está relacionada ao aumento do viés do modelo (incorporação de mais características específicas do local estudado, diminuindo a efetividade da aplicação do MPA em outras regiões) e à redução da variância. O modelo que apresenta o menor valor de AIC é o mais equilibrado (Cunto *et al*, 2011). A Equação 2.11 traz a formulação do AIC.

$$AIC = -2 * \log(L) + 2 * p \quad (2.11)$$

Em que: AIC representa o critério de informação de Akaike, L é o máximo valor da função de verossimilhança do modelo estimado, p expressa o número de parâmetros do modelo.

A utilização do logaritmo da verossimilhança e do AIC só pode ser utilizado na comparação do ajuste de modelos do mesmo tipo, ou seja, não é possível comparar modelos espaciais e binomiais negativos por esses parâmetros, pois eles são calculados de forma diferente (Hadayeghi, 2009).

Segundo Costa (2013), a tarefa de inclusão das variáveis em MPA pode ser algo bastante complexo, devido ao elevado número de opções diferentes de modelos que as variáveis possibilitam. O autor cita três técnicas para facilitar a análise da inclusão das variáveis: exclusão sequencial (*backward*), inclusão sequencial (*forward*) e inclusão/exclusão alternada (*stepwise*). Geralmente, a variável base para a aplicação das técnicas representa a característica de exposição (fluxo veicular, comprimento de trechos, entre outros) para ocorrência dos acidentes (Hauer, 2004). A melhora ou piora do modelo com a inclusão ou exclusão de variáveis é medida verificando o AIC e logaritmo da verossimilhança.

Cunto *et al* (2011) utilizaram a técnica de inclusão sequencial para determinação das variáveis relacionadas à ocorrência de acidentes em interseções semaforizadas de Fortaleza. O fluxo veicular foi escolhido como variável base, e o logaritmo da verossimilhança foi utilizado como ferramenta de avaliação de melhora dos modelos.

De acordo com Hauer (2015), a utilização exclusiva de técnicas globais de validação, que avaliam o erro para o conjunto amostral, não é suficiente para examinar o ajuste do modelo. Segundo ele, a aplicação de técnicas que permitam uma análise local dos erros é

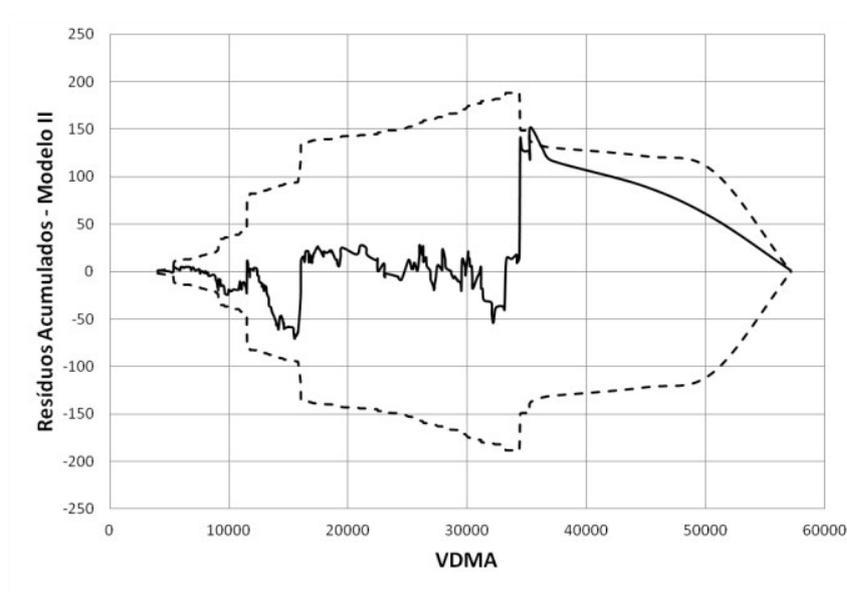
necessária. Uma dessas ferramentas é o gráfico de resíduos acumulados, denominado de *CURE plot* (*CUmulative REsiduals*).

Proposto por Hauer e Bamfo (1997), o *CURE plot* consiste na plotagem dos resíduos acumulados pelo modelo para cada observação e de duas envoltórias composta pelo valor de $\pm 2\sigma^*$ (Equação 2.12). É esperado para um modelo bem ajustado que os resíduos acumulados variem em torno do eixo das abscissas e estejam sempre dentro das duas envoltórias. A Figura 7 apresenta um exemplo de *CURE plot*.

$$\sigma_i^* = \hat{\sigma}_i(n) \sqrt{1 - \frac{\hat{\sigma}_i^2(n)}{\hat{\sigma}^2(N)}} \quad (2.12)$$

Em que: σ_i^* representa o valor da envoltória para a observação i , $\hat{\sigma}_i(n)$ é o resíduo acumulado para i , $\hat{\sigma}_i^2(n)$ é a variância acumulada para i , $\hat{\sigma}^2(N)$ é o somatório da variância da amostra.

Figura 7 - Exemplo de *CURE plot*: resíduos acumulados X fluxo veicular



Fonte: Sobreira e Cunto (2014).

Hadayeghi (2009) cita que além dos *CURE plots*, duas ferramentas são importantes para a comparação entre modelos espaciais e binomiais negativos: o erro médio absoluto (EMA – Equação 2.13) e a raiz do erro quadrático médio (REQM – Equação 2.14). Essas duas medidas são usualmente utilizadas para a verificação do ajuste global do modelo.

$$EMA = \frac{\sum_i^n |\hat{y}_i - y_i|}{n} \quad (2.13)$$

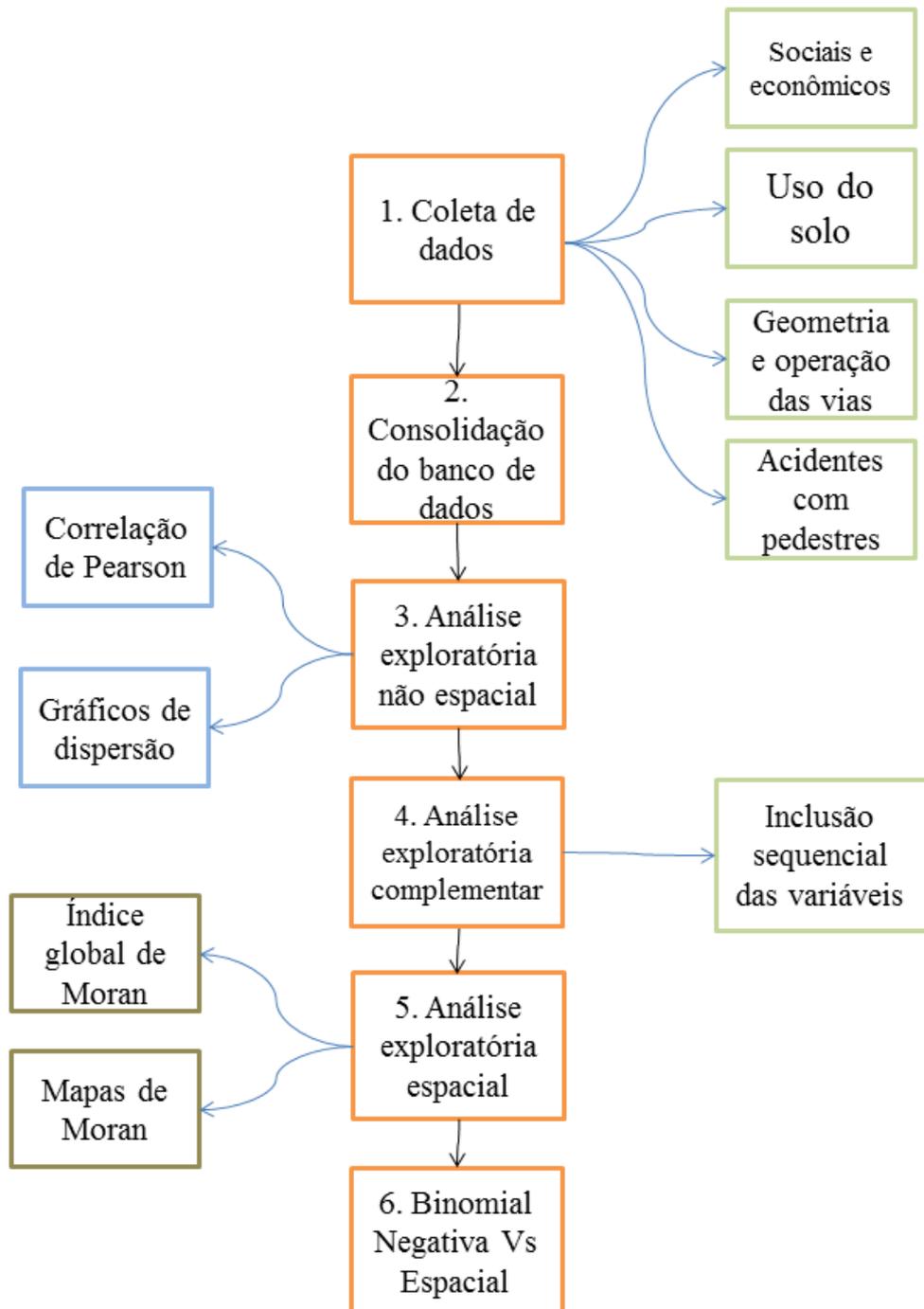
$$REQM = \sqrt{\frac{\sum_i^n (\hat{y}_i - y_i)^2}{n}} \quad (2.14)$$

Em que: \hat{y}_i representa o estimado pelo modelo, y_i é o valor observado, n é o número de elementos considerados.

3 METODOLOGIA

O desenvolvimento da análise exploratória das variáveis que estão relacionadas à frequência de atropelamentos em zonas de tráfego de Fortaleza foi baseado nas seguintes etapas descritas no fluxograma abaixo (Figura 8).

Figura 8 - Resumo da metodologia do trabalho



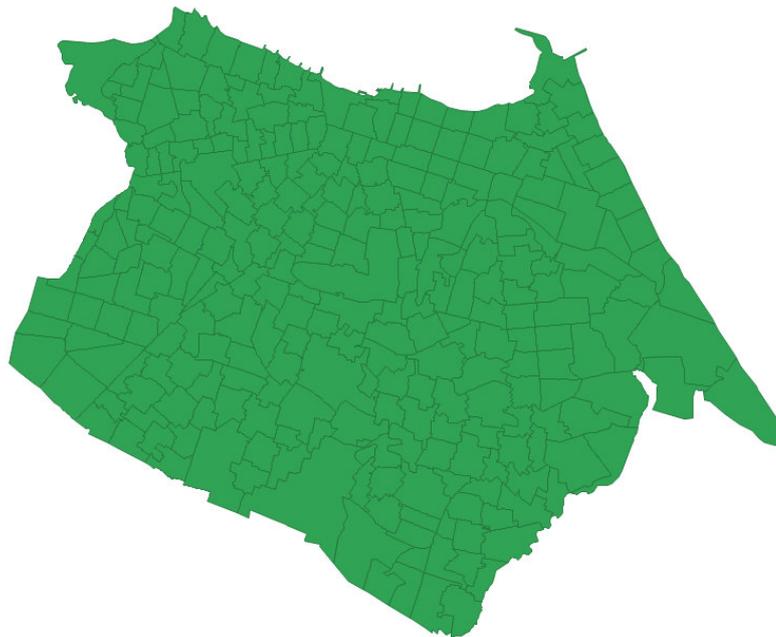
Fonte: Próprio autor (2016)

3.1 Coleta de dados

O processo de coleta de dados é uma etapa de extrema importância para o desenvolvimento da análise exploratória, visto que a qualidade e a representatividade dos resultados obtidos dependem de uma confiável coleta. Primeiramente, foi realizada uma análise das variáveis utilizadas em estudos similares de atropelamentos em zonas de tráfego para facilitar a escolha das variáveis.

No presente trabalho, o conjunto de zonas de tráfego utilizado correspondeu aquele definido no âmbito do projeto Fortaleza 2040 da Prefeitura Municipal de Fortaleza. O projeto estabeleceu 283 zonas de tráfego para a cidade baseado no conceito de corredores urbanos orientados pelo transporte as de tráfego conforme apresenta a Figura 9.

Figura 9 - 283 zonas de tráfego do Fortaleza 2040



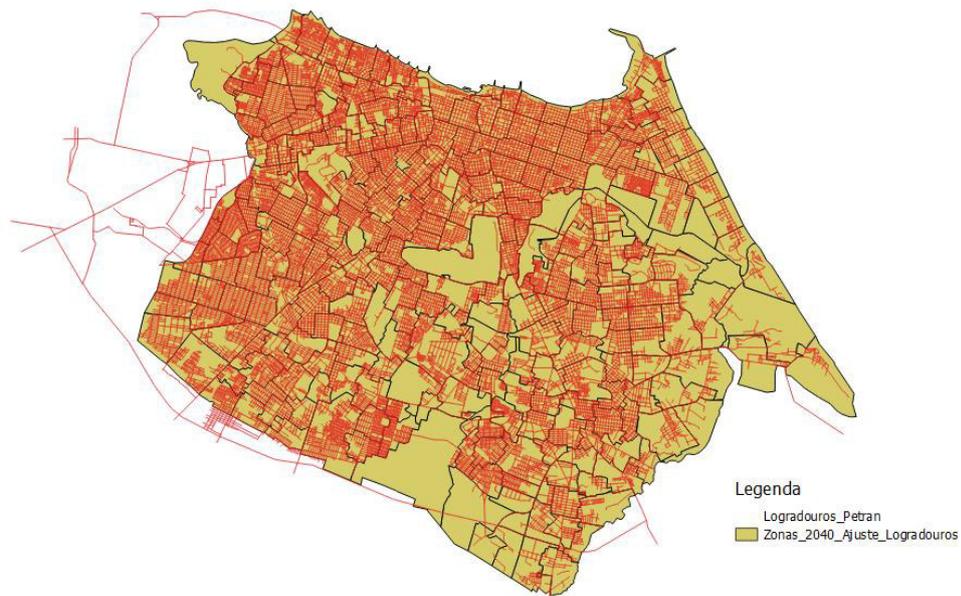
Fonte: Próprio autor (2016)

Foram coletados dados que representem: características de operação, controle, tráfego, infraestrutura e uso do solo, características socioeconômicas das regiões e dados de atropelamentos.

A obtenção das características de infraestrutura, controle de tráfego e operação foi realizada utilizando uma base de logradouros georreferenciados (Figura 10) fornecida pelo órgão gestor de trânsito de Fortaleza a Autarquia Municipal de Trânsito de Cidadania (AMC).

Essa base contém informações de: comprimento dos segmentos, número de faixas das vias, número de sentidos das vias, número de semáforos para veículos e pedestres, número de fiscalização eletrônica, comprimento de via asfaltada, número de paradas de ônibus, entre outros.

Figura 10 - Base de logradouros georreferenciados da AMC sobre as zonas de tráfego do Fortaleza 2040



Fonte: Próprio autor (2016)

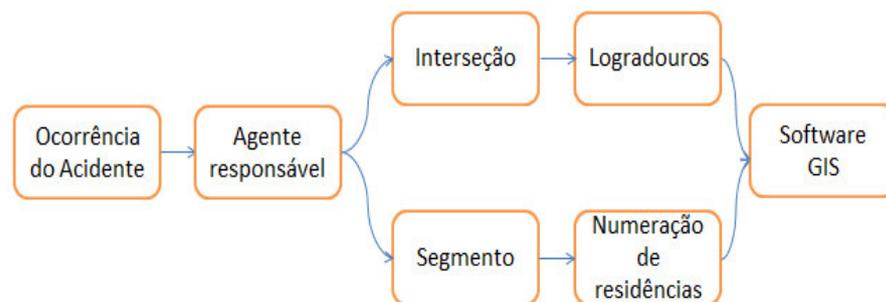
Já para os dados socioeconômicos e dados de uso do solo, as informações foram obtidas do Projeto de Mestrado do aluno Franco Matos no Departamento de Engenharia de Transportes da UFC, que atrelou informações fornecidas pelo Censo 2010 realizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE) com coleta de tipo de uso do solo utilizando ferramentas como o Google Maps e o Google Earth.

Os dados de acidentes em Fortaleza foram coletados pelo Sistema de Informações de Acidentes de Trânsito de Fortaleza (SIAT-FOR) que utiliza informações registradas por 12 órgãos responsáveis pelo atendimento da ocorrência. O SIAT-FOR consiste em um sistema georreferenciado dos acidentes incluindo informações como: severidade (com ou sem vítima), local e hora do acidente, tipos de usuários de transporte envolvidos no acidente, características do usuário (gênero, idade), entre outros.

O georreferenciamento dos acidentes se dá utilizando informações como nomes das ruas do cruzamento, para interseções, e número de alguma residência próxima à ocorrência, para segmentos. Essas informações são lançadas no software de tecnologia SIG (Sistema de

Informação Georreferenciada), que contém a informação dos logradouros, dos cruzamentos e da numeração inicial e final de cada segmento da cidade. A localização exata do acidente é feita com interpolação dessas numerações com a informada pelo agente de trânsito. De acordo com Gomes *et al* (2015), para o ano de 2011, por exemplo, o sucesso da taxa de georreferenciamento para interseções foi de 95,5%, já para segmentos com numeração, 89,8%; e para segmentos sem numeração, 0%. Em 2011, 4470 acidentes foram registrados em logradouros sem numeração, correspondendo 18,8% do total, indicando que o método de coleta desse tipo de acidentes pode ser melhorado para obtenção de dados mais confiáveis. A Figura 11 representa a metodologia de georreferenciamento dos acidentes.

Figura 11 - Metodologia de georreferenciamento dos acidentes



Fonte: Próprio autor (2016)

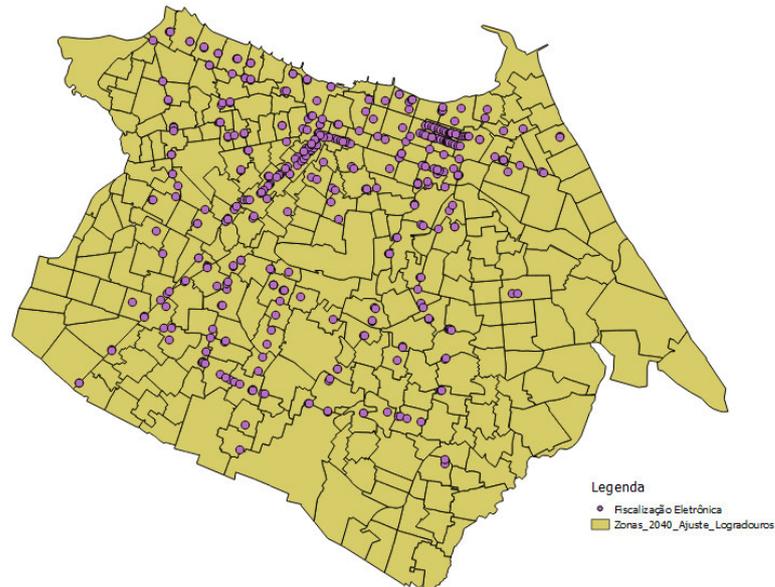
Foi utilizada uma série histórica de três anos de acidentes (2009 a 2011) para a redução de ocorrência de “zeros” nas zonas e para tentar diminuir o caráter aleatório presente nos acidentes de trânsito.

3.2 Consolidação do banco de dados

O software QGIS é uma potente ferramenta para a representação e coleta de dados georreferenciados, nele é possível visualizar espacialmente os dados de forma categórica e obter dados de acordo com regiões delimitadas. O QGIS trabalha com arquivos *shape* (.shp) que associam mapas e informações georreferenciadas de tabelas de dados. Obtidas as variáveis e suas coordenadas para toda a cidade de Fortaleza, o próximo passo metodológico é a coleta dessas para cada uma das 283 zonas de tráfego estudadas, para isso as funções do QGIS de *ponto no polígono*, que soma todos os pontos envolvidos por um perímetro de uma zona, e *somar comprimentos de linha*, que soma o comprimento de vias contidas na

delimitação de uma zona, permitem que as informações contidas nos *shapes* das variáveis sejam extraídas para o *shape* que contém as zonas de tráfego. A Figura 12 representa um mapa com as zonas e pontos de fiscalização eletrônica.

Figura 12 - Mapa com *shapes* de zonas de tráfego e fiscalização eletrônica



Fonte: Próprio autor (2016)

Os dados coletados foram alocados em uma única planilha de MS Excel, para fim de facilitar a análise exploratória dos dados, e em um único arquivo shape para a aplicação em outros softwares, visualização espacial e geração de mapas temáticos.

3.3 Análise exploratória não espacial dos dados

A análise exploratória não espacial dos dados tem a função de selecionar as variáveis que mais tem relação com a ocorrência de atropelamentos. Essa análise foi realizada em três etapas: análise do coeficiente de correlação de Pearson entre as variáveis independentes (variáveis que explicam a ocorrência do atropelamento) e dependente (atropelamentos), análise dos gráficos de dispersão entre as variáveis independentes e dependente e verificação do coeficiente de correlação de Pearson entre as variáveis independentes.

A primeira etapa foi a determinação do coeficiente de correlação de Pearson para toda a amostra de dados. O coeficiente foi obtido com auxílio do software MS Excel utilizando a função correlação do pacote de análise de dados. As variáveis que apresentaram correlação igual ou superior a 0,40 com a frequência dos atropelamentos foram selecionadas para análise

mais profunda. O valor de 0,40 do coeficiente de Pearson é o valor mínimo para uma correlação do tipo moderada, de acordo com Dancey e Reidy (2005). A partir da correlação entre as variáveis, hipóteses sobre o comportamento das mesmas com a ocorrência de acidentes foram levantadas.

Determinadas as principais variáveis, foi feita uma análise de como essas se comportam com relação aos atropelamentos utilizando gráficos de dispersão. Distribuições que apresentaram grandes variâncias, elevadas amplitudes em um pequeno intervalo de dados ou comportamentos sem tendência foram descartadas.

Por fim, foi analisada a correlação de Pearson entre as variáveis independentes escolhidas nas duas etapas anteriores. Caso exista correlação forte ($R > 0,7$) entre alguma delas, foi escolhida apenas a variável que apresentar maior correlação com os atropelamentos. Esse procedimento foi realizado para evitar o uso de variáveis que estejam caracterizando a mesma porção dos acidentes, o que aumentaria o viés da análise.

3.4 Análise exploratória complementar dos dados

A análise exploratória complementar dos dados foi realizada com a utilização da técnica de inclusão sequencial de variáveis em MPA, objetivando um maior entendimento das variáveis selecionadas no item anterior com os atropelamentos. Os modelos foram desenvolvidos utilizando a distribuição binomial negativa, a partir do software estatístico R, com uso da função *glm.nb*.

A escolha da variável base que foi utilizada no primeiro modelo foi feita considerando a característica da variável de representar a exposição ao atropelamento. Para a inclusão das demais variáveis ao modelo primitivo, foi verificada a melhor forma de realizar esse processo. Como exposto na Equação 2.7, as variáveis podem ser incluídas no vetor *A* ou *B*. Para determinar se cada variável se comporta melhor de forma direta (vetor *A*) ou a partir do número neperiano (vetor *B*), modelos considerando as duas formas serão desenvolvidos e os que apresentaram melhores resultados representaram o método de inclusão para cada variável.

A comparação do ajuste entre os modelos foi realizada com a utilização das seguintes ferramentas: *AIC*, logaritmo da verossimilhança ($-2l$), *CURE plot*, EMA e REQM. Para o *AIC*, quanto menor o valor desse coeficiente, melhor o ajuste do modelo. Em relação à verossimilhança ($-2l$), quanto maior, melhor o modelo. Além disso, foi feita uma verificação estatística do $-2l$ para checar se a diferença dos resultados é significativamente estatística. Por

fim, *CURE plots* (dispersão local) e EMA e REQM (dispersão global) foram utilizados para analisar os modelos que apresentarem melhores resultados de *AIC* e $-2l$.

Em seguida, o modelo de melhor ajuste foi utilizado para a quantificação da influência de cada variável. Uma predição do número de atropelamentos para cada zona foi realizada utilizando o valor da média das variáveis para comparação de predições com a variação de ± 1 desvio padrão para cada variável. Com isso, foi possível analisar a influência no aumento dos atropelamentos com a alteração de cada uma das variáveis.

3.5 Análise exploratória espacial dos dados

Para a verificação da ocorrência de dependência espacial das variáveis selecionadas na análise exploratória complementar dos dados foram utilizados o índice global de Moran e o mapa de Moran. O software GeoDa, que combina funções de georreferenciamento e de estatística, foi utilizado para desenvolvimento dos índices e mapas.

A primeira etapa de uma análise espacial é a determinação do modo que a zona estudada i se relaciona com as zonas vizinhas j para desenvolvimento da matriz de adjacência W . Para a determinação do tipo de contiguidade a ser considerada (*Rook* ou *Queen*) entre as zonas, foram utilizados os resíduos do MPA obtido no item anterior. O tipo de contiguidade que apresentou a maior autocorrelação espacial determinada pelo índice global de Moran foi utilizado.

O índice global de Moran foi obtido utilizando a ferramenta *Univariate Moran's I* do GeoDa. Valores do índice acima de 0,30 foram considerados indicação de elevada autocorrelação espacial (O'Sullivan e Unwin, 2010). Para a verificação da significância estatística do índice, o método da permutação foi aplicado com a utilização de 999 permutações. Foi considerado um grau de confiança de 95% para rejeição da hipótese nula de completa aleatoriedade na amostra.

O mapa de Moran foi utilizado para a realização de uma análise microscópica da amostra, podendo verificar-se as regiões de aglomeração e a quantidade de zonas que apresentam espacialidade com significância estatística. Esse é um modo de confirmar visualmente o valor obtido pelo índice global de Moran. Os mapas de Moran foram gerados utilizando a função *Univariate Local Moran's I* do GeoDa.

3.6 MPA binomial negativo Vs MPA espacial

Os modelos espaciais *SAR* e *SEM* foram desenvolvidos utilizando a função regressão do software GeoDa. O programa gera um relatório com informações dos coeficientes, erros, parâmetros e valores preditos. Foram analisados o coeficiente espacial autoregressivo (ρ) e o coeficiente espacial do erro (λ), para verificar a significância estatística da espacialidade atribuída aos acidentes e aos erros, respectivamente.

Para comparar o ajuste entre os modelos espaciais e o binomial negativo, foram utilizadas as seguintes ferramentas: *CURE plot*, *EMA* e *REQM*. Além disso, os resíduos gerados por cada modelo para cada zona foram plotados em mapas para a verificação do tipo de distribuição destes. Também foi determinado o índice de Moran global para eles, objetivando verificar se a dependência espacial está presente nos resíduos.

4. RESULTADOS E ANÁLISES DOS RESULTADOS

4.1 Apresentação dos dados

Como destacado no capítulo anterior, a coleta dos dados foi realizada utilizando informações disponibilizadas pelos órgãos públicos de Fortaleza. Os dados serão apresentados em tabelas divididas por categorias de informação, em que serão exibidos a descrição, a média, o desvio padrão e os valores máximos e mínimos de cada variável. A Tabela 3 apresenta a estatística descritiva dos atropelamentos. A Tabela 4 expõe a estatística descritiva das variáveis relacionadas à operação de tráfego. Já a Tabela 5 apresenta a estatística descritiva das variáveis de características geométricas das vias. A Tabela 6 mostra a estatística descritiva das variáveis de uso do solo. Por fim, a Tabela 7 demonstra a estatística descritiva das características socioeconômicas das zonas.

Devido à falta de informações sobre o uso do solo de algumas áreas da cidade de Fortaleza, 12 zonas de tráfego foram retiradas da análise, assim o estudo foi feito utilizando 271 zonas de tráfego.

Tabela 3 - Estatística descritiva dos atropelamentos

Variável	Descrição	Média	Desvio padrão	Mínimo	Máximo
#Acid_Ped_2009	Atropelamentos em 2009	5,43	5,72	0	40
#Acid_Ped_2010	Atropelamentos em 2010	4,53	4,66	0	31
#Acid_Ped_2011	Atropelamentos em 2011	3,84	4,39	0	32
Soma_Acid_Ped	Soma dos atropelamentos em 2009 a 2011	13,80	13,78	0	103

Fonte: Próprio autor (2016)

Analisando as variáveis de atropelamentos, pode-se verificar que para essa amostra houve uma redução, em média, na ocorrência de acidentes com o passar dos anos. Além disso, a redução do desvio padrão com o decorrer dos anos mostra um aumento na homogeneidade dos atropelamentos, em que os acidentes nas zonas ocorridos em 2011 estão mais próximos à média do que em 2009.

Tabela 4 - Estatística descritiva das variáveis de operação do tráfego

Variável	Descrição	Média	Desvio padrão	Mínimo	Máximo
#Fisc_Elet	Número de equipamentos de fiscalização eletrônica	1,52	2,95	0	30
#Dens_Fisc_Elet_Vias_Asfalt	Densidade de fiscalização eletrônica por km de vias asfaltadas	0,12	0,20	0	1,49
#Int_Semaf	Número de interseções semaforizadas	3,00	5,87	0	54
#Dens_Int_Semaf_Vias_Asfalt	Densidade de interseções semaforizadas por km de vias asfaltadas	0,23	0,38	0	3,36
#Semaf_Estag_Ped	Número de semáforos com estágio para pedestres	0,24	0,62	0	4
#Semaf_Excl_Ped	Número de semáforos exclusivos para pedestres	0,28	0,76	0	9
#Paradas_Ativas	Número de paradas ativas de ônibus	17,64	11,76	0	98
#Dens_Paradas_Vias_Asfalt	Densidade de paradas ativas de ônibus por km de vias asfaltadas	1,79	1,04	0	8,87

Fonte: Próprio autor (2016)

Tabela 5 - Estatística descritiva das variáveis de características de infraestrutura das vias

Variável	Descrição	Média	Desvio padrão	Mínimo	Máximo
Area_Zona (km ²)	Área das zonas	1,08	0,74	0,27	8,32
#Vias_<=4fx (km)	Km de vias com 4 ou menos faixas de tráfego	14,76	7,01	1,56	62,46
#Vias_>4fx (km)	Km de vias com mais de 4 faixas de tráfego	0,48	0,64	0	4,28
#Via_Total (km)	Km de vias	17,11	8,42	2,88	77,08
#Via_Asfaltada (km)	Km de vias asfaltadas	10,96	6,17	0,67	36,73
#Via_Sentido_Unico (km)	Km de vias com sentido único	11,50	6,74	0	44,73
#Via_Sentido_Duplo (km)	Km de vias com sentido duplo	3,28	5,50	0	62,08

Fonte: Próprio autor (2016)

Para a apresentação das variáveis relacionadas ao uso do solo, é importante a definição do índice de aproveitamento (I.A.). Definido pelos planos diretores das cidades, esse índice representa o valor que deve ser multiplicado pela área total do lote da construção para determinação da área total que pode ser construída.

Tabela 6 - Estatística descritiva das variáveis de uso do solo

Variável	Descrição	Média	Desvio padrão	Mínimo	Máximo
#Mat_Ens_Bas	Número de matrículas em ensino básico (ensino infantil e médio)	2.332	2.383	0	21.038
#Mat_Ens_Sup	Número de matrículas em ensino superior	496	2.204	0	28.130
#Mat_Pos_Grad	Número de matrículas em ensino de pós-graduação	28	204	0	2.571
#Soma_Matriculas	Somatório das matrículas	2.856	3.195	0	29.399
L_Indust (m2)	Área de uso do solo industrial	11.496	42.557	0	514.466
CSBD (m2)	Área de uso do solo comercial e serviços de baixa densidade. Locais I.A. < 2,0	132.052	96.484	0	603.362
CSAD (m2)	Área de uso do solo comercial e serviços de alta densidade. Locais com I.A. > 2,0	5.516	14.584	0	177.297
RBD (m2)	Área de uso do solo residencial de baixa densidade. I.A. < 0,8	229.291	175.778	0	1.276.394
RMD (m2)	Área de uso do solo residencial de média densidade. 0,8 < I.A. < 2,0	47.440	36.366	0	240.065
RAD (m2)	Área de uso do solo residencial de alta densidade. I.A. > 2,0	15.658	39.856	0	356.479
HP (m2)	Área de uso do solo de habitações precárias	3.095	7.796	0	74.986
Entropia	Índice de entropia da zona de tráfego	0,50	0,16	0	0,79

Fonte: Próprio autor (2016)

Tabela 7 - Estatística descritiva dos dados de características socioeconômicas

Variável	Descrição	Média	Desvio padrão	Mínimo	Máximo
#Dom_Baixa_Renda	Número de domicílios de baixa renda (renda per capita média entre R\$291,00 e R\$ 441,00)	1677	1438	0	8114
#Dom_Media_Renda	Número de domicílios de média renda (renda per capita média entre R\$441,00 e R\$ 641,00)	597	413	0	2511
#Dom_Alta_Renda	Número de domicílios de alta renda (renda per capita média acima de R\$641,00)	344	548	0	4479
#Domicilios	Somatório dos domicílios	2617	1795	0	10115

Fonte: Próprio autor (2016)

4.2 Análise exploratória não espacial dos dados

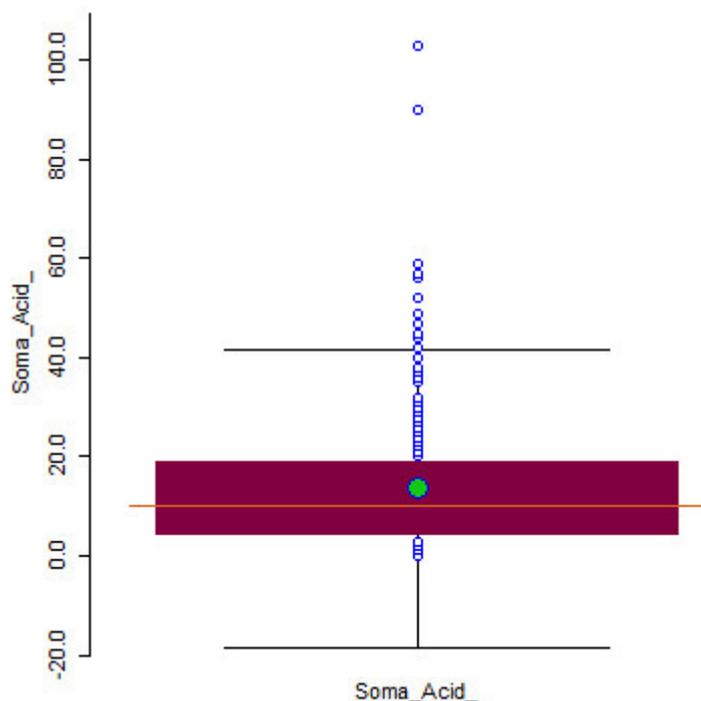
Para realizar a análise exploratória não espacial dos dados que objetiva identificar como as variáveis se comportam entre si, foram utilizadas as ferramentas da correlação de Pearson e dos gráficos de dispersão. O Anexo A apresenta a tabela com a correlação de Pearson de todas as variáveis estudadas.

Analisando a correlação de Pearson entre todas as variáveis, foram inicialmente selecionadas aquelas que apresentaram correlação igual ou superior a 0,40 com a variável da soma dos atropelamentos, com exceção da variável *índice de entropia*, que será debatida adiante. A Tabela 8 apresenta os valores da correlação de Pearson dessas variáveis.

Após a definição das variáveis que possuem maior correlação com a ocorrência de atropelamentos, foi feita a análise dos gráficos de dispersão dessas variáveis com os acidentes. O Anexo B apresenta esses gráficos de dispersão com a linha de tendência de regressão linear dos dados e o valor R^2 do ajuste da linha. A Tabela 9 apresenta um resumo da interpretação dos gráficos de dispersão com base na distribuição dos dados e na variância dos atropelamentos.

Para a análise dos gráficos de dispersão e das correlações de Pearson, a verificação de *outliers* é de extrema importância para que a característica ímpar de poucas zonas não afete a interpretação dos resultados em âmbito global. O *boxplot*, que é uma ferramenta comumente utilizada na verificação dos *outliers*, foi utilizado para a identificação de possíveis valores extremos nos atropelamentos. A Figura 13 apresenta o *boxplot* dos atropelamentos.

Figura 13 - *Boxplot* dos atropelamentos



Fonte: Próprio autor (2016)

O *boxplot* dos atropelamentos mostra que doze observações estão fora do limite superior do gráfico, entretanto apenas duas zonas de tráfego apresentam valores destoantes

dos demais com 103 e 90 atropelamentos cada. A zona em que ocorreram 103 atropelamentos gráficos representa uma área do centro da cidade de Fortaleza em que suas características são totalmente atípicas do resto da cidade, pelo fato de apresentar uma extrema concentração de comércio e serviços. Já a zona com 90 acidentes representa o bairro da Aerolândia, bairro caracterizado pela alta diversidade de uso do solo e concentração de habitantes. A elevada ocorrência de atropelamentos certamente está relacionada à grande movimentação de pessoas nessas áreas. As análises seguintes foram realizadas, sempre que possível, expondo resultados com e sem as duas zonas de valores extremos.

É importante destacar que as variáveis independentes, apesar de serem ditas independentes, normalmente possuem correlação entre si, assim, a análise de forma individual de uma delas carrega a influência de outras variáveis.

A análise dos coeficientes de Pearson em conjunto com os gráficos de dispersão permite que conclusões sejam obtidas sobre o comportamento das variáveis. Pôde-se identificar, por exemplo, que a ocorrência dos atropelamentos está relacionada à *quilometragem de vias asfaltadas* ($R = 0,56$), mas fracamente relacionada à de *vias totais* ($R = 0,25$). Isso ocorre, provavelmente, pela baixa velocidade desempenhada por veículos em vias não asfaltadas e pela maior concentração de comércios e serviços em áreas que possuem asfaltamento. Por outro lado, pôde-se verificar que as características de infraestrutura apresentaram correlação baixa com os acidentes. A *quilometragem de vias com sentido duplo* ($R = 0,03$) e a *quilometragem de vias com mais de quatro faixas de trânsito* ($R = 0,25$) que são variáveis que poderiam estar relacionadas aos atropelamentos pelo fato de dificultarem os movimentos de pedestres apresentaram correlação baixa com a variável dependente.

Tabela 8 - Correlação de Pearson entre as variáveis que apresentaram correlação igual ou superior a 0,40 e a variável de atropelamentos

	<i>Soma_Acid_Ped</i>	<i>#Int_Semaf</i>	<i>#Dens_Int_Semaf_Vias_Asfalt (int/km)</i>	<i>#Semaf_Excl_Ped</i>	<i>#Paradas_Ativas</i>	<i>#Via_Asfaltada (km)</i>	<i>#Mat_Ens_Bas</i>	<i>#Soma_Matriculas</i>	<i>CSBD (m2)</i>	<i>CSAD (m2)</i>	<i>Entropia</i>	<i>#Dom_Media_Renda</i>	<i>#Domicilios</i>
<i>Soma_Acid_Ped</i>	1,00												
<i>#Int_Semaf</i>	0,53	1,00											
<i>#Dens_Int_Semaf_Vias_Asfalt (int/km)</i>	0,45	0,93	1,00										
<i>#Semaf_Excl_Ped</i>	0,49	0,46	0,45	1,00									
<i>#Paradas_Ativas</i>	0,53	0,39	0,28	0,35	1,00								
<i>#Via_Asfaltada (km)</i>	0,56	0,37	0,21	0,19	0,71	1,00							
<i>#Mat_Ens_Bas</i>	0,55	0,36	0,26	0,47	0,60	0,50	1,00						
<i>#Soma_Matriculas</i>	0,47	0,37	0,29	0,37	0,49	0,40	0,69	1,00					
<i>CSBD (m2)</i>	0,40	0,42	0,35	0,23	0,41	0,46	0,21	0,27	1,00				
<i>CSAD (m2)</i>	0,51	0,76	0,68	0,57	0,50	0,32	0,40	0,37	0,41	1,00			
<i>Entropia</i>	0,37	0,35	0,30	0,15	0,30	0,48	0,14	0,14	0,28	0,24	1,00		
<i>#Dom_Media_Renda</i>	0,53	0,23	0,13	0,24	0,61	0,74	0,44	0,33	0,37	0,20	0,41	1,00	
<i>#Domicilios</i>	0,44	0,12	0,01	0,15	0,58	0,72	0,45	0,31	0,30	0,14	0,29	0,84	1,00

Fonte: Próprio autor (2016)

Tabela 9 - Resumo da interpretação dos gráficos de dispersão

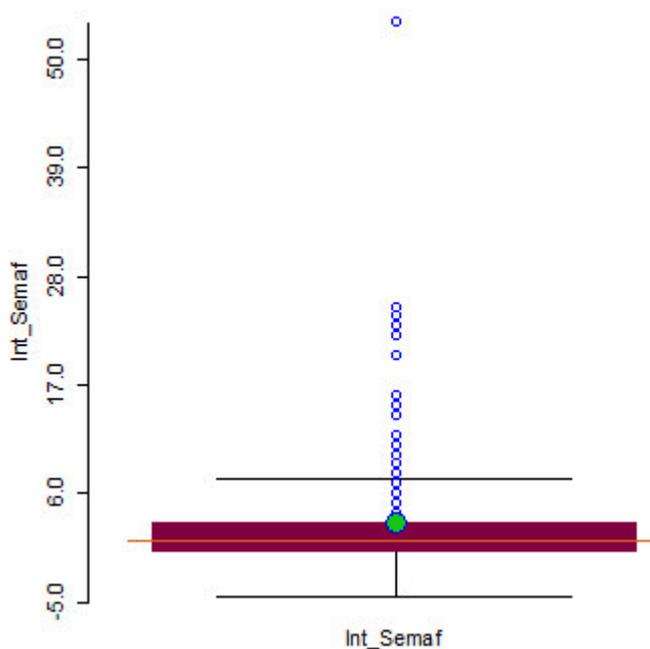
Atropelamentos Vs	R	R*	Análise da dispersão	Dispersão torna viável a consideração da variável? (Sim/Não)
#Int_Semaf	0,53	0,45	Elevada concentração das observações no intervalo de 0 a 4 interseções semaforizadas com grande variância dos atropelamentos nesse trecho	Não
#Dens_Int_Semaf_Vias_Asfalt (int/km)	0,45	0,36	Elevada concentração das observações no intervalo de 0 a 0,4 interseções semaforizadas por km com grande variância dos atropelamentos nesse trecho	Não
#Semaf_Excl_Ped	0,49	0,36	Elevada concentração das observações no intervalo de 0 e 1 semáforos exclusivos para pedestres com grande variância dos atropelamentos nesse trecho	Não
#Paradas_Ativas	0,53	0,47	Apresenta linearidade com pequeno aumento da variância com o aumento do número das paradas ativas	Sim
#Via_Asfaltada (km)	0,56	0,61	Apresenta linearidade com pequeno aumento da variância com o aumento de quilômetro de via asfaltada	Sim
#Mat_Ens_Bas	0,55	0,48	Apresenta linearidade e concentração das observações no intervalo de 0 a 2000 matrículas com pequena variância dos atropelamentos	Sim
#Soma_Matriculas	0,47	0,42	Apresenta linearidade e concentração das observações no intervalo de 0 a 4000 matrículas com pequena variância dos atropelamentos	Sim
CSBD (m2)	0,40	0,39	Apresenta pequena linearidade e elevada variância dos atropelamentos em toda a amostra	Não
CSAD (m2)	0,51	0,39	Elevada concentração das observações no intervalo de 0 a 2000 m2 de CSAD com grande variância dos atropelamentos nesse trecho	Não
Entropia	0,37	0,43	Apresenta linearidade com relativo aumento da variância dos atropelamentos com o aumento da entropia	Sim
#Dom_Media_Renda	0,53	0,60	Apresenta linearidade com pequeno aumento da variância dos atropelamentos com o aumento do número de domicílios de média renda	Sim
#Domicílios	0,44	0,51	Apresenta linearidade com pequeno aumento da variância dos atropelamentos com o aumento do número da soma de domicílios	Sim

R*: Coeficiente de Pearson desconsiderando os dois *outliers* presentes na amostra

Fonte: Próprio autor (2016)

Para a avaliação das variáveis de operação de tráfego é importante enfatizar nas variáveis *semáforos exclusivo para pedestres* ($R = 0,49$. Sem os *outliers*: $R = 0,36$) e *semáforos com estágio para pedestres* ($R = 0,36$). A presença desses tipos de semáforos pode indicar que o fluxo de pedestres é considerável nos locais que os tem, mas também indica que a travessia dos usuários é feita com segurança, por isso a baixa correlação com os atropelamentos. Já a quantidade de *paradas de ônibus ativas* ($R = 0,53$) que está altamente relacionada à movimentação de pedestres apresenta uma correlação considerável com os acidentes. O número de *interseções semaforizadas* ($R = 0,53$. Sem os *outliers*: $R = 0,45$), apesar de indicar locais com elevado fluxo de veículos e correlação moderada com os acidentes, possui grande parte de sua distribuição concentrada no intervalo de 0 a 4 semáforos, como mostra o *boxplot* da variável (Figura 14), e a variância dos atropelamentos nesse trecho é alta (ver Figura 1B, Anexo B), implicando que essa variável não é indicada para entender o comportamento dos atropelamentos para as zonas de tráfego. Além disso, é pouco provável que a influência de três interseções semaforizadas seja importante para a ocorrência de atropelamentos em zonas de 1 km^2 (média da área das zonas), essa variável é mais indicada em estudos microscópicos, em que cada interseção é analisada de forma mais profunda.

Figura 14 - *Boxplot* das interseções semaforizadas



Fonte: Próprio autor (2016)

A presença de *fiscalização eletrônica* ($R = 0,36$) não está altamente relacionada à frequência de atropelamentos, entretanto é possível que essa variável possua relação com a severidade das ocorrências, dado que a presença da fiscalização modera a velocidade dos condutores. A aplicação de fatores por unidade de área utilizados em densidade de fiscalização eletrônica, de interseções semaforizadas e de paradas de ônibus ativas não trouxe melhorias às variáveis aplicadas de forma absoluta, indicando que a concentração desses tipos de controle de operação não está relacionada à ocorrência dos atropelamentos.

O número de *matrículas de ensino básico* ($R = 0,55$. Sem os *outliers*: $R = 0,48$) apresenta elevada correlação com a variável dependente, provavelmente pelo fato dos alunos de ensino básico serem mais imaturos com relação ao comportamento no trânsito quando comparados a alunos de ensino superior e também graças à imensa maioria dos alunos que percorrem o trajeto escola-casa andando, aumentando a exposição destes a ocorrências. Em oposição ao exposto acima, alunos do *ensino superior* ($R = 0,09$) e de *pós-graduação* ($R = 0,03$) possuem maturidade e outras opções de locomoção, diminuindo a probabilidade e exposição para a ocorrência de atropelamentos. A *soma das matrículas* dos tipos supracitados ($R = 0,47$) apresenta correlação considerável com os acidentes, entretanto essa tendência é impulsionada pelas *matrículas de ensino básico*. Essas duas variáveis apresentam um coeficiente de Pearson de 0,69 entre si, por isso, apenas a variável de *ensino básico* será utilizada para representar os atropelamentos.

As variáveis relacionadas aos tipos de densidade das residências apresentaram baixa correlação com a variável resposta, possivelmente pelo fato das viagens realizadas no trajeto casa-trabalho serem realizadas utilizando veículos próprios ou transporte público e pela movimentação ao redor de locais de concentração de moradias ser geralmente menor do que em áreas de comércio e serviços.

Por outro lado, as variáveis de *comércio e serviço de baixa densidade* ($R = 0,40$. Sem os *outliers*: $R = 0,39$) e de *comércio e serviço de alta densidade* ($R = 0,51$. Sem os *outliers*: $R = 0,39$) apresentaram maior correlação com os atropelamentos, quando comparadas às variáveis de densidade de residências, provavelmente pela maior rotatividade de pessoas nesses locais, aumentando, assim, a exposição dessas. Entretanto pela alta concentração de área de *comércio e serviço de alta densidade* no intervalo de 0 a 5000 m^2 e alta variância dos atropelamentos nesse intervalo (ver Figura 9B, Anexo B), essa variável não foi considerada para futuras análises. Já *comércio e serviços de baixa densidade* apresenta menor concentração das observações do que a variável anterior, entretanto a variância presente em toda a amostra (ver Figura 8B, Anexo B) inviabiliza essa variável de caracterizar os

atropelamentos. Essas variáveis de uso do solo serão incorporadas à análise com a utilização do *índice de entropia*, que, ao combinar todas elas, pode expressar com mais qualidade a movimentação de pessoas.

Como esperado, a área de *uso do solo industrial* ($R = 0,06$) apresentou baixa correlação com os acidentes, pois, em Fortaleza, ambientes industriais estão localizados em áreas mais afastadas do centro da cidade, assim, a movimentação de pessoas no entorno dessas áreas é praticamente nula, reduzindo a probabilidade de ocorrência de atropelamentos.

Como citado acima, o *índice de entropia* ($R = 0,37$. Sem os *outliers*: $R = 0,43$) expressa a diversidade de uso do solo das zonas. Essa variável pode ser uma boa medida da movimentação de pessoas e já foi utilizada para representar o fluxo de pedestres em outros trabalhos (Ewing e Cervero, 2010). O gráfico de dispersão da variável (ver Figura 12B, Anexo B) apresenta certa linearidade e pequeno aumento da variância com o aumento do índice.

Para a análise das variáveis socioeconômicas, pôde-se verificar que *domicílios de alta renda* ($R = 0,17$) apresentaram correlação com os acidentes inferior a *domicílios de renda baixa* ($R = 0,34$) e *renda média* ($R = 0,53$. Sem os *outliers*: $R = 0,60$), possivelmente pelo menor fluxo de pedestres em zonas com domicílios de uma maior renda, motivada pelo maior uso de veículos individuais. Já zonas com domicílios de baixa e média renda apresentam a maior parte da população se deslocando por meio de transporte público, aumentando, assim, a exposição a atropelamentos. Uma hipótese para a diferença de correlação entre domicílios de média e de baixa renda é que o primeiro apresenta maior correlação com a área de comércio e serviços, indicando uma maior movimentação de pedestres.

Por fim, o *número total de domicílios* ($R = 0,44$. Sem os *outliers*: $R = 0,51$) apresentou considerável correlação com os atropelamentos, indicando que zonas com mais domicílios provavelmente são zonas mais habitadas, conseqüentemente aumentando o número de pessoas expostas a acidentes.

Realizada a análise exploratória não espacial dos dados, seis variáveis foram escolhidas como mais representativas para a ocorrência de atropelamentos: *paradas de ônibus ativas*, *quilometragem de via asfaltada*, *matrículas no ensino básico*, *índice de entropia*, *domicílios de média renda* e *número total de domicílios*. Verificando a correlação de Pearson entre essas variáveis, notou-se que o *número total de domicílios* apresenta alta correlação com *domicílios de média renda* ($R = 0,84$), portanto, como *domicílios de média renda* tem correlação com os atropelamentos maior do que o *número total de domicílios*, a variável do somatório será retirada das análises seguintes, evitando, assim, que variáveis que estejam

explicando a mesma porção dos acidentes sejam utilizadas. A Tabela 10 traz as variáveis selecionadas nessa etapa.

Tabela 10 - Variáveis selecionadas na análise exploratória não espacial dos dados

Variáveis
#Paradas_Ativas
#Via_Asfaltada (km)
#Mat_Ens_Bas
Entropia
#Dom_Media_Renda

Fonte: Próprio autor (2016).

4.3 Análise exploratória complementar dos dados

A análise exploratória complementar dos dados foi realizada com a estimação dos coeficientes de MPA das variáveis mais correlacionadas com a frequência dos atropelamentos, definidas na seção anterior. A técnica de inclusão sequencial de variáveis em MPA foi aplicada nessa etapa para uma investigação mais profunda da relação das variáveis independentes com os atropelamentos.

A variável *quilômetros de vias asfaltadas* foi escolhida para ser a formadora do modelo base, pelo fato dela ser um modo de caracterizar a exposição de pedestres aos acidentes nas zonas de tráfego. Traçando um paralelo, ela representa para modelos macroscópicos o que o fluxo de veículos, por exemplo, significa para a maioria dos MPA microscópicos.

Para a determinação do modo de inclusão das outras variáveis ao modelo base, as quatro variáveis foram testadas nas duas formas. *Paradas ativas de ônibus e domicílios de média renda* tiveram um melhor desempenho quando aplicadas de modo direto no modelo. Já *matrículas em ensino básico e entropia* apresentaram melhor ajuste quando usadas com o número neperiano. Uma das razões para isso ocorrer é a maior quantidade de valores nulos presentes nas duas últimas variáveis que, se aplicadas de forma direta, tornariam a predição nula.

A Tabela 11 traz os oito modelos desenvolvidos com seus coeficientes, desvios padrão, $-2l$ e o *AIC*. No rodapé da tabela está a indicação da significância estatística dos coeficientes obtidos. Para verificar que os atropelamentos se adequam a uma distribuição

binomial negativa, o teste do parâmetro de dispersão σ_d foi utilizado. Para o modelo base, o σ_d foi de 1,11, indicando grande aderência dos acidentes à distribuição.

Tabela 11 - Coeficientes dos modelos desenvolvidos com a aplicação da técnica inclusão sequencial de variáveis

Variáveis		Modelos							
		1	2	3	4	5	6	7	8
intercepto	Coef.	1,00	0,42*	0,75	1,12	0,72	0,50**	0,46*	0,45*
	D.P.	0,17	0,28	0,19	0,17	0,19	0,28	0,28	0,28
#Via_Asfáltada (km)	Coef.	1,09*	0,84*	0,86*	0,96*	0,94*	0,74*	0,67*	0,61*
	D.P.	0,07	0,09	0,10	0,08	0,08	0,09	0,10	0,10
#Dom_Media_Renda	Coef.	-	0,23*	-	-	-	0,21*	0,20*	0,15**
	D.P.	-	0,06	-	-	-	0,06	0,06	0,06
#Paradas_Ativas	Coef.	-	-	0,30*	-	-	-	0,13	-
	D.P.	-	-	0,09	-	-	-	0,09	-
#Mat_Ens_Bas	Coef.	-	-	-	6,77E-05*	-	6,10E-05*	5,31E-05*	7,69E-05*
	D.P.	-	-	-	1,92E-05	-	1,86E-05	1,94E-05	5,90E-05
Entropia	Coef.	-	-	-	-	1,26*	-	-	1,42*
	D.P.	-	-	-	-	0,35	-	-	0,36
-2ℓ		-1817	-1801	-1806	-1801	-1805	-1786	-1784	-1771
AIC		1823	1809	1814	1809	1813	1796	1797	1783

Significância estatística dos coeficientes: (*) 0,01; (**) 0,05; () Sem significância estatística

D.P.: Desvio padrão

Fonte: Próprio autor (2016)

Analisando os valores da Tabela 11, pode-se comparar o modelo base (modelo 1) com modelos de duas variáveis (modelos 2, 3, 4 e 5). Percebe-se que a inclusão de domicílios de média renda (modelo 2) e matrículas do ensino básico (modelo 4) acarretaram na maior melhoria do -2ℓ (16) e do AIC (14). Simonoff (2003) diz que uma diferença de AIC acima de 10 já é considerada representativa. A partir disso, foi desenvolvido o modelo com três variáveis (modelo 6). Comparando-o com os modelos 2 e 4, verifica-se uma considerável melhora com a redução de 13 pontos do AIC . Seguindo o mesmo raciocínio, o modelo 8 também foi considerado para uma avaliação mais minuciosa. Vale destacar que a diferença do -2ℓ entre os modelos foi confirmada de forma estatística com a aplicação do teste da taxa de verossimilhança.

O EMA e o REQM (Tabela 12) e *CURE plots* (Figura 15) dos modelos 2, 4, 6 e 8 foram analisados para a escolha do melhor MPA. Como esperado, o acréscimo no número de variáveis reduziu os valores dos indicadores de erros globais, visto que seus parâmetros de ajustes analisados acima foram melhorando com a inclusão das variáveis.

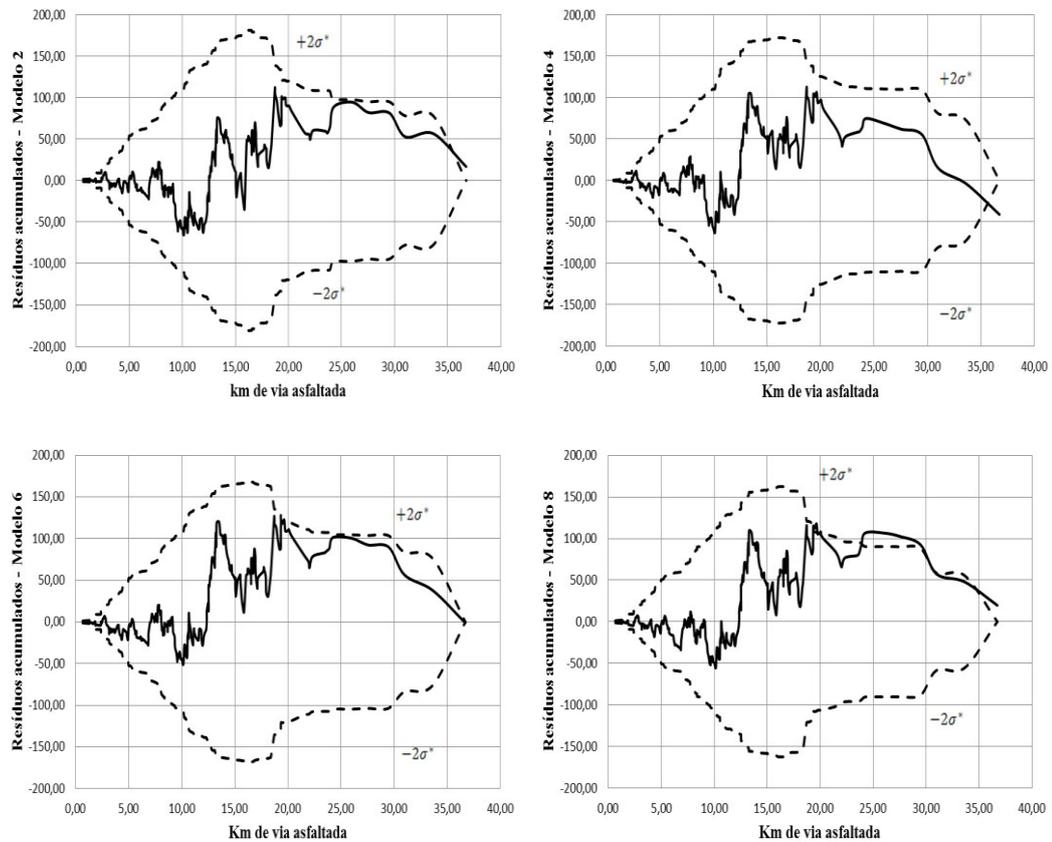
O resíduo acumulado calculado para desenvolvimento do *CURE plot* representa o número de atropelamentos observados menos o número de atropelamentos previstos pelo modelo para cada observação, ou seja, valores positivos para o resíduo acumulado indicam que o modelo está subestimando os atropelamentos. Analisando os gráficos, percebe-se que para os quatro modelos para dados entre 12 e 13 *quilômetros de via asfaltada* o previsto pelo modelo é inferior ao observado, por isso a reta cresce nesse trecho. A elevada variação da reta para dados acima de 20 quilômetros deve-se a presença de poucas observações nessa faixa. No geral, os quatro gráficos apresentam distribuição semelhante, indicando que o melhor ajuste de modelos com mais variáveis não pode ser verificado de forma gráfica para os resíduos locais. A rigor, o fato do resíduo do modelo 8 ter extrapolado a envoltória do desvio invalidaria a utilização desse, porém como o principal objetivo é entender o comportamento das variáveis, esse modelo será o utilizado para o desenvolvimento das análises em âmbito espacial.

Tabela 12 - Indicadores do ajuste global dos modelos binomiais negativos

Indicador	Modelos			
	2	4	6	8
EMA	6,85	6,93	6,65	6,50
REQM	123,13	109,30	104,55	97,38

Fonte: Próprio autor (2016)

Figura 15 - *CURE plots* dos modelos binomiais negativos



Fonte: Próprio autor (2016)

A fim de avaliar a influência de cada variável na predição dos atropelamentos, foram aplicadas, utilizando o modelo 8, diversos valores para as variáveis na equação de formação do modelo (Equação 4.1). Vale destacar que o objetivo dessa análise não é obter uma relação de causa-efeito, e sim obter mais informações sobre o comportamento das variáveis.

$$Y_i = 0,45 * Vias^{0,61} * Domicilios^{0,15} * exp^{(0,0000769Mat+1,42Entropia)} \quad (4.1)$$

Para comparação da influência de cada variável foi obtido o valor da predição com o valor da média de cada variável e, em seguida, os valores que a variação de ± 1 desvio padrão para cada variável traria às predições. A tabela 13 traz os valores para cada variável, a predição dos atropelamentos para cada variação de desvio padrão e as comparações numéricas e percentuais das predições com relação ao valor predito pela média.

Tabela 13 - Influência das variáveis na predição dos atropelamentos utilizando o modelo 8

Alteração dos valores	Variáveis				Atropelamentos estimados	Diferença numérica p/ a média	Variação p/ a média (%)
	#Via_ Asfaltada (km)	#Dom_Media_ Renda	#Mat_Ens_ Bas	Entropia			
Média	10,96	597	2332	0,50	12,52	-	-
+D.P.*	17,13	597	2332	0,50	16,49	3,96	32%
-D.P.	4,79	597	2332	0,50	7,53	-5,00	-40%
+D.P.	10,96	1010	2332	0,50	13,55	1,02	8%
-D.P.	10,96	184	2332	0,50	10,51	-2,02	-16%
+D.P.	10,96	597	4715	0,50	15,04	2,52	20%
-D.P.	10,96	597	0	0,50	10,47	-2,06	-16%
+D.P.	10,96	597	2332	0,66	15,73	3,21	26%
-D.P.	10,96	597	2332	0,34	9,97	-2,55	-20%

D.P.: Desvio padrão

Fonte: Próprio autor (2016)

Analisando a Tabela 13 é possível verificar, por exemplo, que o acréscimo de 6,17 quilômetros de vias asfaltadas aumentou em 3,96 a predição dos atropelamentos. Já a redução de aproximadamente 2332 matrículas de ensino básico em uma zona diminuiu 2,06 atropelamentos. Esses valores são úteis para quantificar a influência de cada variável.

4.4 Análise exploratória espacial dos dados

A análise da autocorrelação espacial entre os acidentes e da correlação espacial entre os acidentes e as outras variáveis se deu com a utilização do índice global de Moran e da comparação de mapas de Moran. O tipo de contiguidade utilizada para caracterizar a vizinhança das zonas foi o critério de *Queen*, pois ao aplicar o índice global de Moran para os resíduos do modelo binomial negativo 8, esse critério apresentou maior autocorrelação espacial do que o critério de *Rook*. O critério de *Queen* considera fronteiras e cantos para efetivação da vizinhança. A Tabela 14 traz os índices globais de Moran e suas significâncias estatísticas obtidas com o método da permutação.

Tabela 14 - Índice global de Moran e significância estatística do índice de Moran das variáveis

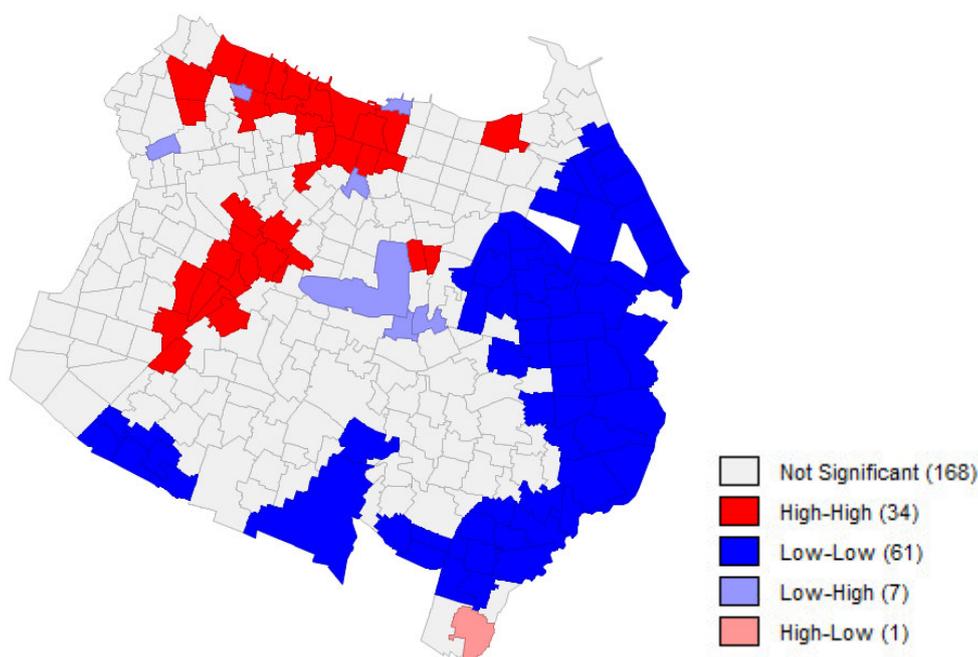
Variável	Índice global de Moran	Significância estatística (p)
Soma_Acid_Ped	0,421	0,001
#Via_Asfaltada (km)	0,287	0,001
#Mat_Ens_Bas	0,063	0,039
#Dom_Media_Renda	0,352	0,001
Entropia	0,375	0,001

Fonte: Próprio autor (2016)

O índice global de Moran dos atropelamentos indica uma forte autocorrelação dessa variável, para uma verificação mais minuciosa e localizada o mapa de Moran será analisado (Figura 16). A zona superior de aglomeração de altos valores (*high-high*) é onde estão localizados os bairros do Centro, Barra do Ceará e Pirambu, bairros de elevada concentração de domicílios de média e baixa renda e de comércios e serviços. Já a zona inferior de aglomeração de atropelamentos contém os bairros da Parangaba e Vila Pery, de mesma característica dos citados acima. Além disso, a Avenida Osório de Paiva, que possui um elevado fluxo de veículos e pessoas por conectar a região sudoeste ao centro da cidade, está nessa zona de aglomeração.

Por outro lado, a aglomeração de baixos atropelamentos (*low-low*) localizada na região leste e sudeste da cidade está relacionada a bairros que estão em desenvolvimento e ainda não apresentam elevada concentração de domicílios (Edson Queiroz) e a bairros mais afastados do centro da cidade com baixo poder de atração de viagens (Lagoa Redonda e Pedras). Vale destacar que os dados de acidentes são para os anos de 2009 a 2011, após esse período houve o desenvolvimento de algumas zonas de Fortaleza, principalmente do bairro Edson Queiroz.

Figura 16 - Mapa de Moran dos atropelamentos



Fonte: Próprio autor (2016)

Dentre as variáveis independentes, a única que não apresentou autocorrelação espacial considerável foi *matrículas no ensino básico*. Isso ocorreu, provavelmente, pela grande distribuição de colégios por toda Fortaleza, visto que geralmente as pessoas buscam estudar em locais perto de suas moradias. Realizando um comparativo dos mapas de Moran dos atropelamentos e das outras variáveis (Anexo C), verifica-se que as distribuições espaciais da *entropia e do número de domicílios de média renda* apresentam similaridades com as aglomerações de altos e baixos atropelamentos. Isso pode ser um indicativo que modelos com a presença dessas variáveis podem explicar a espacialidade dos atropelamentos sem que o fator espacial seja incorporado diretamente no modelo.

4.5 MPA binomial negativo Vs MPA espacial

Os modelos espaciais *SAR* e *SEM* foram desenvolvidos utilizando as variáveis preditoras do modelo 8 e a função regressão do GeoDa. A Tabela 15 traz os coeficientes, desvios padrão, *-2l* e o *AIC*. No rodapé da tabela está a indicação da significância estatística dos coeficientes obtidos. É importante destacar que o *-2l* e *AIC* calculados na modelagem espacial possuem formulação diferente dos parâmetros desenvolvidos para a distribuição

binomial negativa, portanto eles servirão apenas para comparação dos modelos espaciais entre si, e não para comparação de ajuste entre espaciais e binomiais negativos.

Tabela 15 - Coeficientes e parâmetros dos modelos espaciais

Variáveis		Modelos	
		SAR	SEM
Constante	Coef.	-7,50*	-2,83
	D.P.	1,83	2,45
#Via_Asfaltada (km)	Coef.	0,40*	0,45*
	D.P.	0,14	0,16
#Dom_Media_Renda	Coef.	2,90E-03	3,82E-03
	D.P.	1,90E-03	2,28E-03
#Mat_Ens_Bas	Coef.	1,75E-03*	0,00146*
	D.P.	2,70E-04	2,62E-04
Entropia	Coef.	8,34**	12,09*
	D.P.	3,99	4,32
ρ	Coef.	0,49*	-
	D.P.	0,06	-
λ	Coef.	-	0,57*
	D.P.	-	0,07
	-2ℓ	-1968	-1978
	AIC	1980	1989

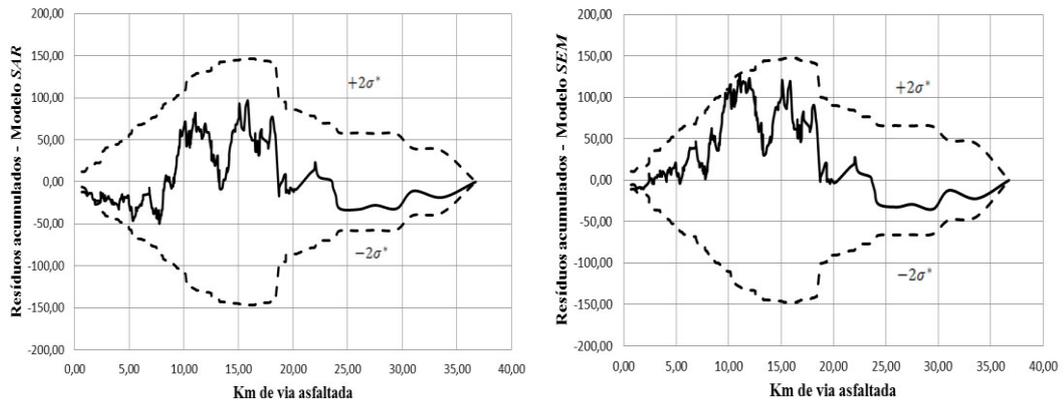
(*) 0,01; (**) 0,05; () Sem significância estatística

D.P.: Desvio padrão

Fonte: Próprio autor (2016)

Como esperado, pelo fato da dependência espacial dos atropelamentos, os parâmetros de espacialidade ρ e λ foram estatisticamente significativos. Comparando os dois modelos espaciais, verifica-se que o SAR possui um ajuste melhor dos parâmetros de -2ℓ e AIC do que o SEM. Vale ressaltar que para os dois modelos espaciais a variável de *domicílios de média renda* não foi significativa. Uma possibilidade para isso ter ocorrido seria o fato dessa variável apresentar estrutura espacial semelhante aos atropelamentos (ver Figura 16 e Figura 4C, Anexo C), assim a incorporação dos acidentes das zonas vizinhas pelo modelo espacial diminuiu a importância da mesma na explicação dos atropelamentos.

Os *CURE plots* dos modelos espaciais (Figura 17) mostram que o modelo SAR também possui melhor ajuste local do que o SEM. Verifica-se um pico nos dois gráficos entre os trechos de 10 e 20 *quilômetros de via asfaltada* em que os modelos subestimam a previsão dos atropelamentos.

Figura 17 - *CURE plots* dos modelos espaciais

Fonte: Próprio autor (2016)

Definido o *SAR* como o modelo espacial a ser utilizado, a comparação dele com o modelo 8 binomial negativo será realizada pela análise do EMA e REQM (Tabela 16), comparação dos *CURE plots* e verificação da distribuição espacial dos resíduos.

Tabela 16 - Indicadores do ajuste global do modelo 8 binomial negativo e do modelo *SAR*

Indicador	Modelos	
	8	<i>SAR</i>
EMA	6,50	5,85
REQM	97,38	79,54

Fonte: Próprio autor (2016)

Os indicadores de ajuste global dos erros mostram que o modelo espacial prevê melhor os atropelamentos do que o modelo binomial negativo. O *CURE plot* do *SAR* (Figura 17) apresenta menores amplitudes do que o gráfico do modelo 8 (Figura 15) e não extrapola em nenhum momento a envoltória do desvio $\pm 2\sigma^*$, indicando um melhor ajuste, também, em âmbito local.

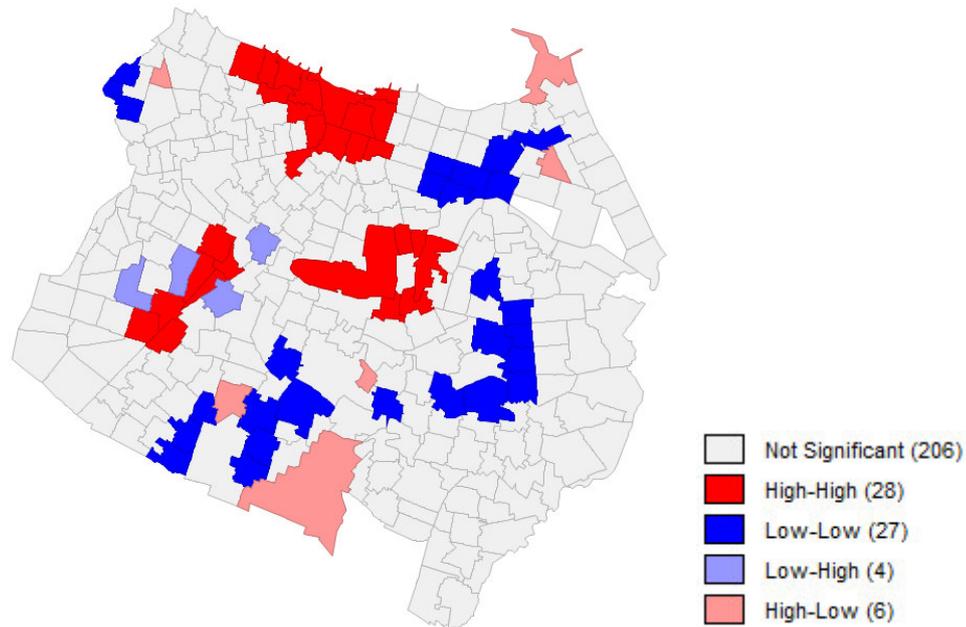
Para a verificação da distribuição espacial dos resíduos de cada modelo, serão analisados os índices globais de Moran (Tabela 17) e os mapas de Moran (Figuras 18 e 19) dos erros.

Tabela 17 - Índice global de Moran dos resíduos dos modelos 8 e *SAR*

Variável	Índice global de Moran	Significância estatística (p)
Resíduo modelo 8	0,293	0,001
Resíduo <i>SAR</i>	0,005	0,383

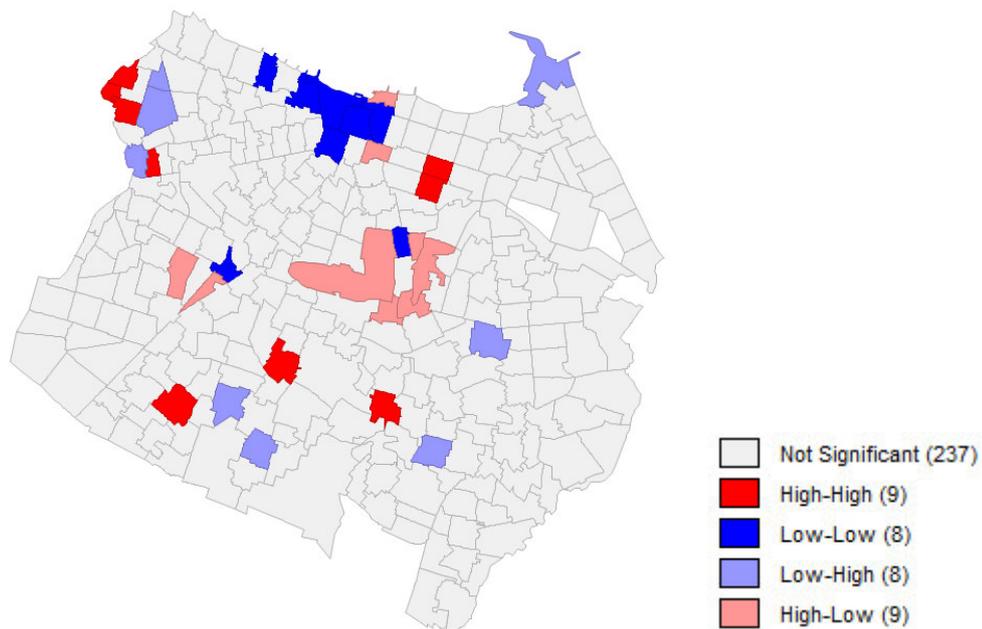
Fonte: Próprio autor (2016)

Figura 18 - Mapa de Moran dos resíduos do modelo 8



Fonte: Próprio autor (2016)

Figura 19 - Mapa de Moran dos resíduos do modelo SAR



Fonte: Próprio autor (2016)

Tanto o índice global de Moran, quanto o mapa de Moran dos resíduos do modelo 8 (Figura 18) indicam que a espacialidade dos atropelamentos não foi incorporada ao modelo, mesmo com a utilização de variáveis independentes que apresentaram aglomerações similares aos acidentes, como *entropia* e *domicílios de média renda*. A aglomeração de valores

positivos de resíduos ocorreu em locais que existe a concentração de acidentes, como mostra a Figura 16, ou seja, o modelo 8 está subestimando os acidentes para essas zonas.

Já o modelo *SAR*, ao utilizar a média dos acidentes das zonas vizinhas como parte da equação de previsão, conseguiu incorporar a espacialidade dos atropelamentos, distribuído os resíduos de forma sem padrão espacial, comprovado pela não rejeição da hipótese nula de aleatoriedade espacial dos dados analisados ($p = 0,383$) e pelo baixo número de zonas que apresentaram significância estatística no mapa de Moran.

É importante ressaltar que o modelo *SAR* considera que os atropelamentos seguem uma distribuição normal, entretanto foi realizado o teste estatístico de Anderson-Darling para verificar se essa distribuição ocorria nos acidentes. A hipótese nula de distribuição normal para os acidentes foi rejeitada. Esse fato corrobora ainda mais com a importância da inclusão de fatores espaciais em MPA desenvolvidos para áreas, visto que a incorporação da espacialidade desempenhou um papel mais significativo do que a utilização de uma distribuição que melhor caracteriza os atropelamentos dessa amostra, a binomial negativa.

Assim como realizado para o modelo 8, a influência de cada variável para o modelo *SAR* também será analisada utilizando a mesma metodologia. A formulação do modelo espacial é apresentada na Equação 4.2.

$$Y_i = 0,49 WY_i - 7,5 + 0,40 Vias + 0,0029 Domic + 0,00175 Matric + 8,34 Ent \quad (4.2)$$

O fator WY_i representa a média dos atropelamentos das zonas vizinhas à zona em estudo. Esse fator também foi incluído na análise, e a média dos atropelamentos foi utilizada como valor para o mesmo. A Tabela 18 traz a influência de cada variável na predição de atropelamentos do modelo *SAR*.

Tabela 18 - Influência de cada variável na predição de atropelamentos utilizando o modelo

SAR

Alteração dos valores	Variáveis					Atropelamentos estimados	Diferença numérica p/ a média	Variação p/ a média (%)
	WY	#Via_ Asfáltada (km)	#Dom_Media _ Renda	#Mat_Ens_ Bas	Entropia			
Média	13,80	10,96	597	2332	0,50	13,63	-	-
+D.P.*	27,58	10,96	597	2332	0,50	20,38	6,75	50%
-D.P.	0,02	10,96	597	2332	0,50	6,88	-6,75	-50%
+D.P.	13,80	17,13	597	2332	0,50	16,10	2,47	18%
-D.P.	13,80	4,79	597	2332	0,50	11,16	-2,47	-18%
+D.P.	13,80	10,96	1010	2332	0,50	14,83	1,20	9%
-D.P.	13,80	10,96	184	2332	0,50	12,43	-1,20	-9%
+D.P.	13,80	10,96	597	4715	0,50	17,80	4,17	31%
-D.P.	13,80	10,96	597	0	0,50	9,55	-4,08	-30%
+D.P.	13,80	10,96	597	2332	0,66	14,96	1,33	10%
-D.P.	13,80	10,96	597	2332	0,34	12,29	-1,33	-10%

D.P.: Desvio padrão

Fonte: Próprio autor (2016)

Analisando a Tabela 18, verifica-se que para o modelo espacial SAR a influência dos acidentes das zonas vizinhas é superior do que a de qualquer variável explicativa. Um acréscimo de aproximadamente 14 atropelamentos na média das zonas vizinhas aumenta em 6,75 a predição dos acidentes.

Comparando o comportamento das variáveis nos modelos SAR e 8, apenas a variável *matrículas de ensino básico* apresentou maior representatividade no espacial do que no 8 (variação para a média com a alteração da variável passou de 20% para 31%). Esse fato era esperado, pois a inclusão de mais uma variável ao modelo (média dos acidentes das zonas vizinhas) tende a diminuir a influência das outras.

5 CONCLUSÕES E SUGESTÕES PARA FUTUROS TRABALHOS

Este Projeto de Graduação apresentou uma análise exploratória das variáveis relacionadas a atropelamentos em 271 zonas de tráfego da cidade de Fortaleza. A análise foi dividida em três etapas: não espacial, complementar e espacial.

Com a análise exploratória não espacial, foi possível o levantamento de hipóteses sobre o comportamento das variáveis coletadas com os atropelamentos. As mais significativas foram aquelas relacionadas à geração de viagem e ao constante movimento de pedestres, aumentando, assim, a exposição desses aos acidentes. Variáveis relacionadas à geometria da via apresentaram baixa correlação com os atropelamentos. As variáveis mais relacionadas com a frequência dos atropelamentos foram: *quilômetros de via asfaltada*, *paradas de ônibus ativas*, *domicílios de média renda*, *entropia* e *matrículas do ensino básico*.

As cinco variáveis supracitadas apresentaram relação linear com os atropelamentos, ou seja, com o aumento dessas variáveis é esperado o acréscimo de acidentes nas zonas. Esse conjunto de variáveis apresentam características específicas para representarem a frequência dos atropelamentos. *Quilômetros de via asfaltada*, por exemplo, estão relacionados à exposição do pedestre ao acidente, podendo representar o fluxo de veículos e de pedestres aplicados em MPA microscópicos. Já a variável *paradas ativas de ônibus* está ligada aos locais de concentração pontual de movimentação de pedestres e, também, à dificuldade verificada em usuários no ato de atravessar a via com a presença do ônibus que está embarcando e desembarcando pessoas. *Matrículas de ensino básico* representam um grande fluxo de pedestres com pouca idade e que, em alguns casos, não possuem experiência nem conhecimento sobre as leis de trânsito, podendo aumentando o risco para a ocorrência de atropelamentos. *Domicílios de média renda* estão relacionados à presença de pessoas que se deslocam em automóveis e andando, além da concentração de comércios e serviços próximos a esses domicílios, o que aumenta a movimentação de pedestres. Por fim, o *índice de entropia* é uma variável que agrupa a diversidade dos tipos de uso do solo e pode estar relacionada à geração de viagens de pedestres.

A identificação dessas variáveis como representativas para a ocorrência de atropelamentos permite que futuras campanhas de conscientização ou medidas de melhoria da segurança viária sejam realizadas de forma mais eficaz. Pode-se, por exemplo, melhorar as condições de travessias em locais que apresentem elevadas concentrações de domicílios de média renda e paradas de ônibus ou desenvolver campanhas para a conscientização de motoristas, para que eles tenham maior atenção ao trafegarem por zonas escolares.

A análise da espacialidade dos atropelamentos indicou uma forte autocorrelação (índice global de Moran = 0,421). As zonas de aglomeração de altos valores de acidentes representaram locais com elevada concentração de domicílios de média e baixa renda e de grande variedade de uso do solo. Já as aglomerações de baixos valores de atropelamentos foram em zonas que ainda estão em desenvolvimento e em zonas que não possuem atrativos para geração da movimentação de pedestres.

MPA binomiais negativos e espaciais foram desenvolvidos para verificação de como a espacialidade influenciou na significância de parâmetros e variáveis e na distribuição espacial dos resíduos. O modelo espacial *SAR* apresentou medidas globais e locais de ajuste melhores do que o modelo binomial negativo. Além disso, pôde-se verificar dependência espacial nos resíduos do modelo binomial negativo (índice global de Moran = 0,293), algo que não ocorreu no *SAR*.

Outro indicativo da importância da aplicação do fator espacial em MPA em áreas é a utilização da distribuição normal para os atropelamentos no desenvolvimento do modelo *SAR*. Foi realizado um teste estatístico de Anderson-Darling, e a hipótese dos acidentes seguirem a distribuição normal foi rejeitada. Mesmo assim, ao considerar a espacialidade como um fator do modelo, o *SAR* desenvolveu melhores resultados do que a utilização de uma distribuição que melhor representava os acidentes, a binomial negativa.

Uma limitação desse Projeto de Graduação é que uma parcela dos atropelamentos não é georreferenciada, sendo assim, não considerada nas análises. Outra limitação é a impossibilidade de medir de forma direta o fluxo de pedestres e veículos para cada zona de tráfego. Além disso, os modelos espaciais do tipo *SAR* e *SEM* são mais adequados para dados contínuos, assim a utilização de um nível mais alto de agregação (zonas maiores) reduziria zonas com zero atropelamentos, tornando mais viável a utilização desses tipos de modelos espaciais.

Para trabalhos futuros, recomenda-se a aplicação de MPA espaciais com técnicas mais sofisticadas utilizando a distribuição de Poisson e binomial negativa. Além disso, um estudo mais profundo sobre o comportamento humano nas travessias urbanas traria mais componentes para serem agregados às análises.

6 REFERÊNCIAS BIBLIOGRÁFICAS

AKAIKE, H. *A new look at the statistical model identification*. **IEEE Transactions on Automatic Control**, vol. 19, n. 6, p. 716–723. 1974.

AMC . *Anuário estatístico de acidentes de trânsito de Fortaleza - 2011*. Autarquia Municipal de Trânsito, Serviços Públicos e Cidadania, Prefeitura Municipal de Fortaleza, Fortaleza, CE, 2011.

ANSELIN, L. *SpaceStat Tutorials*. University of California, Santa Barbara, USA. 1992.

ANSELIN, L. *Exploratory spatial data analysis in a geocomputational environment*. In: LONGLEY, P.A.; BROOKS, S.M.; MCDONNELL, R.; MACMILLAN, B. (Eds.), *Geocomputation: A Primer*. Wiley, New York, USA, p. 77-94. 1998.

BAILEY, T. C.; GATRELL, A.C. *Interactive Spatial Data Analysis*. Longman, London, England. 1995.

BONNESON, J. A.; P. T. MCCOY. *Estimation of safety at two-way stop-controlled intersections on rural highways*. **Transportation Research Record**, Washington, D.C., vol. 1401, p. 83-99. 1993.

CÂMARA, G.; M. S. CARVALHO; O. G. CRUZ e V. CORREA. *Análise de Dados de Área*. In: Fuks, S. D.; M. S. Carvalho; G. Câmara; A. M. V. Monteiro (eds.). *Análise Espacial de Dados Geográficos*. Instituto Nacional de Pesquisas Espaciais – Divisão de Processamento de Imagens, São José dos Campos, São Paulo. 2000.

CARVALHO, M. S. **Aplicação de Métodos de Análise Espacial na Caracterização de Áreas de Risco à Saúde**. Tese (Doutorado em Engenharia Biomédica) – Instituto Alberto Luiz Coimbra de Pós Graduação e Pesquisa em Engenharia, Rio de Janeiro, Brasil. 1997.

CHAMBERS, R. *Rural Development: putting the last first*. Longman Scientific and Technical Publishers, New York, 1983.

CHEN, P.; SHEN, Q. *Built environment effects on cyclist injury severity in automobile-involved bicycle crashes. Accident Analysis and Prevention*, vol. 86, p. 239–246. 2016.

COSTA, J. O. **Desenvolvimento dum modelo de previsão de acidentes**. Tese (Doutorado em Segurança Viária) - Departamento de Engenharia Civil, Faculdade de Engenharia da Universidade do Porto, Porto, Portugal. 2013.

CUNTO, F. J. C. *Assessing Safety Performance of Transportation Systems using Microscopic Simulation*. Tese (Doutorado) - Department of Civil and Env. Engineering, Universidade de Waterloo, Ontário. 2008.

CUNTO, F. J. C.; CASTRO NETO, M. M.; BARREIRA, D. S. Modelos de Previsão de Acidentes de Trânsito em Interseções Semaforizadas de Fortaleza. *Transportes*, vol.20, n.2, p. 55-62. 2011.

DANCEY, C.; REIDY, J. **Estatística Sem Matemática para Psicologia: Usando SPSS para Windows**. Porto Alegre, Artmed. 2005.

DELMELLE, E.C.; THILL, J.C. *Spatial Analysis of Adult and Youth Traffic Hazard Intensity. Transportation Research Record, Washington, D.C.*, vol. 2074, p. 31-39. 2008.

DOLATSARA, H.A. *Development of Safety Performance Functions for Non-Motorized Traffic Safety. Thesis (Mastery) – Western Michigan University, USA*. 2014.

EWING, R.; CERVERO, R. *Travel and the Built Environment. Journal of the American Planning Association*, vol. 76, n. 3, p. 265–294, 2010.

FERREIRA, S. M. P. **A Segurança Rodoviária no Processo de Planeamento de Redes de Transporte em Meio Urbano**. Tese (Doutorado em Vias de Comunicação) - Departamento de Engenharia Civil, Faculdade de Engenharia da Universidade do Porto, Porto, Portugal. 2010.

GARSON, G. David. *Statnotes: Topics in Multivariate Analysis*. 2009. Disponível em: <http://faculty.chass.ncsu.edu/garson/PA765/statnote.htm>. Acessado em 25/08/2016.

GEEDIPALLY, S. R.; LORD, D. *Investigating the effect of modeling single-vehicle and multi-vehicle crashes separately on confidence intervals of Poisson-gama models*. *Accidents Analysis and Prediction*, vol. 42, p. 1273-1282. 2010. DOI: 10.1016/j.aap.2010.02.004

GOMES, M. J. T. L., TORRES, C. A., OLIVEIRA NETO, F. M. e CUNTO, F. J. C. Análise exploratória para a modelagem da frequência de acidentes de trânsito agregados ao nível de zonas de tráfego. *Transportes*, v. 23, n. 4, p. 42-50. 2015.

GOPAUL, C.D.; GOPAUL, A.S.; SUTHERLAND, J. M.; ROSTANT, L.; EBI, K. L.; CHADEE, D. D. *The Epidemiology of Fatal Road Traffic Collisions in Trinidad and Tobago, West Indies (2000-2011)*. *Global Health Action*, vol. 9: 32518. 2016.

GUIMARÃES, P. R. B. *Métodos Quantitativos Estatísticos*. IESDE Brasil S.A., Curitiba, Brasil. 2008.

IPEA. *Acidentes de trânsito nas rodovias brasileiras: caracterização, tendências e custos para a sociedade*. Instituto de Pesquisas Econômicas Aplicadas. Brasília. 2015.

HADAYEGHI, A.; SHALABY, A.S.; PERSAUD, B. *Macro-level accident prediction models for evaluating safety of urban transportation systems*. *Transportation Research Record*, Washington, D.C., vol. 1840, n. 1, p. 87-95. 2003.

HADAYEGHI, A., SHALABY, A.S., PERSAUD, B. *Development of planning level transportation safety tools using Geographically Weighted Poisson Regression*. *Accident Analysis and Prevention*, vol. 42, p. 676-688. 2010.

HADAYEGHI, A. *Use of Advanced Techniques to Estimate Zonal Level Safety Planning Models and 39 Examine Their Temporal Transferability*. Thesis (PhD) - Department of Civil Engineering, University of Toronto. 2009.

HAUER, E. *Statistical Road Safety Modeling*. **Transportation Research Record**, Washington, D.C., vol. 1897, p. 81-87. 2004.

HAUER, E. *The Art of Regression Modeling in Road Safety*. 1ª ed.. Springer. Toronto, Canadá. 2015.

HAUER, E.; A. S. HAKKERT. *The extent and implications of incomplete accident reporting*. **Transportation Research Record**, Washington, D.C., vol. 1185, p. 1-10. 1989.

HAUER, E.; J. BAMFO. *Two Tools for Finding What Function Links the Dependent Variable to the Explanatory Variables*. In: **Conference of the International Co-operation on Theories and Concepts in Traffic Safety**, Lund, Sweden. 1997.

HASHIMOTO, T. *Spatial analysis of pedestrian accidents*. Thesis (Mastery) – University of South Florida, USA. 2005.

HERBEL, S.; MEYER, M.D.; KLEINER, B.; GAINES, D. *A primer on safety performance measures for the transportation planning process*. Report No.FHWA-HEP-09-043. 2009.

HUANG H.; ABDEL-ATY M.; DARWICHE A. *County-Level Crash Risk Analysis in Florida*. **Transportation Research Record**, vol. 2148, n. 1, 30 p. 27–37. 2010.

KIM, K.; BRUNNER, I. M.; YAMASHITA, E. Y. *Influence of Land Use, Population, Employment, and Economic Activity on Accidents*. **Transportation Research Record**, vol. 195, p. 56-64. 2006.

MEDRI, W. *Análise Exploratória de Dados*. [Apostila do curso de especialização em estatística]. Departamento de Estatística – Universidade Estadual de Londrina. 2011.

NHTSA. *Traffic Safety Facts 2015: Pedestrians*. DOT HS 812 124. 2015. National Highway Traffic Safety Administration. 2015.

NIST/SEMATECH. *e-Handbook of Statistical Methods*. Disponível em: <http://www.itl.gov/div898/handbook>. Acessado em 10/07/2016.

O'SULLIVAN, D.; UNWIN, D. J. *Geographical Information Analysis*. Wiley, Hoboken, NJ. 2003.

OUYANG, Y.; BEJLERI, I. *GIS-Based Community Level Method to Evaluate the Influence of Built Environment on Traffic Crashes*. **93rd Transportation Research Board Annual Conference**. p.18, 2014. DOI: 10.3141/2432-15

PERSAUD, B.; LYON, C. *Empirical Bayes Before–after Safety Studies: Lessons Learned from Two Decades of Experience and Future Directions*. **Accident Analysis and Prevention**, vol. 39, n. 3, p. 546-555. 2007.

PERSAUD, B.; LYON, C. *Pedestrian Collision Prediction Models for Urban Intersections*. **Transportation Research Record**, vol. 1818, p. 102-107. 2012.

PERSAUD, B.; K. MUCSI. *Microscopic accident potential models for two-lane rural roads*. **Transportation Research Record: Journal of the Transportation Research Board**, vol. 1485, p. 134-139. 1995.

PIRDAVANI, A.; BRIJS, T.; BELLEMANS, T.; WETS, G. *Spatial Analysis of Fatal and Injury Crashes in Flanders, Belgium; Application of Geographically Weighted Regression Technique*. **The 92th Annual Meeting of Transportation Research Board**, Washington, D.C., USA. 2012.

QUDDUS, M.A. *Modelling area-wide count outcomes with spatial correlation and heterogeneity: An analysis of London crash data*. **Accident Analysis and Prevention**, vol. 40, p. 1486-1497. 2008.

QUEIRÓZ, M.P. **Análise Espacial dos Acidentes de Trânsito do Município de Fortaleza**. Tese (Mestrado) - Departamento de Engenharia de Transportes, Universidade Federal do Ceará. 2003.

SIDDIQUI, C.; ABDEL-ATY, M.; CHOI, K. *Macroscopic spatial analysis of pedestrian and bicycle crashes*. **Accident Analysis and Prevention**, vol. 45, p. 382-391. 2012.

SIMONOFF J. S. *Analyzing Categorical Data. (1a. ed)*, Springer. New York, USA. 2003.

SIMÕES, P.; SHRUBSALL, S.; NATÁRIO, I. *A spatial econometrics analysis for road accidents in Lisbon. Int. J. Business Intelligence and Data Mining*, vol. 10, n. 2, p.152–173. 2015.

SOBREIRA, L. T. P.; CUNTO, F. J. C. Análise de variações temporais na estimação de acidentes em segmentos arteriais urbanos. *Transportes*, vol. 22, n. 3, p. 128-137. 2014.

TORRES, C.A.. **Avaliação do desempenho da segurança viária em redes urbanas de transportes com utilização de modelos de previsão de acidentes.** Tese (Mestrado) – Departamento de Engenharia de Transportes, Universidade Federal do Ceará. 2016.

UKKUSURI, S.; MIRANDA-MORENO, L. F.; RAMADURAI, G.; ISA, TAVAREZ, J. *The role of built environment on pedestrian crash frequency. Safety Science*, vol. 50, p. 1141–1151, 2012.

WASHINGTON, S.; SCHALKYK, I.V.; MEYER, M.; DUMBAUGH, E.; ZOLL, M. *Incorporating safety into long-range transportation planning. NCHRP Report n. 546, TRB. National Cooperative Highway Research Program, Washington D.C.* 2006.

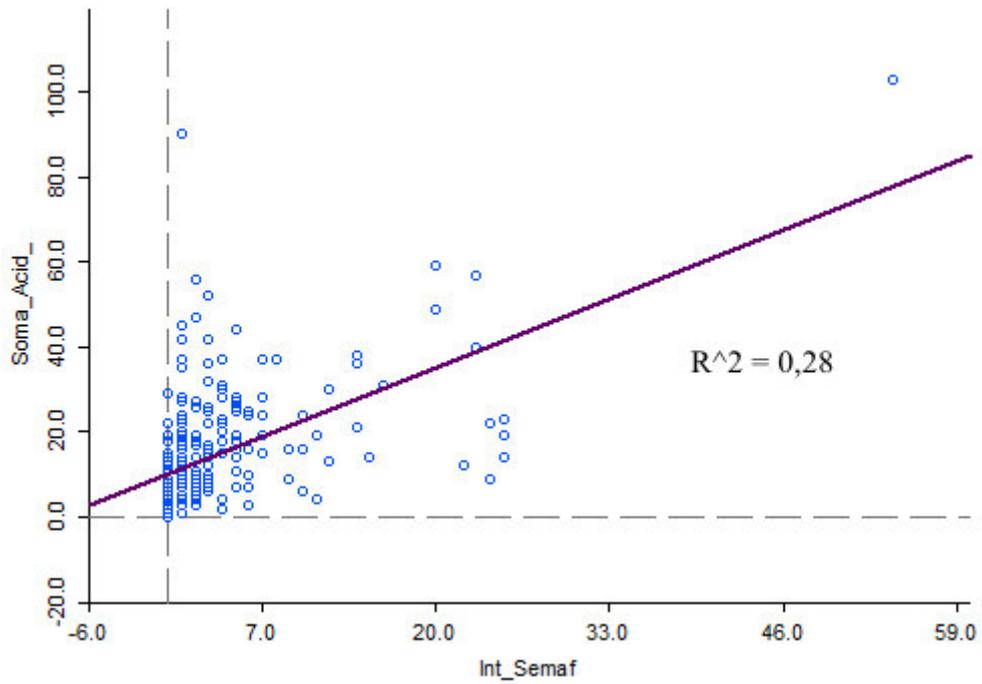
WIER, M.; WEINTRAUB, J.; HUMPHREYS, E. H.; SETO, E.; BHATIA, R. *An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning. Accident Analysis and Prevention*, vol.4, n.1, p. 137-145. 2009.

ANEXOS A – CORRELAÇÃO DE PEARSON DE TODAS AS VARIÁVEIS

	Soma_Acid_Ped	#Fisc_Elet	#Dens_Fisc_Elet_Vias_Asfalt (fisc/km)	#Int_Semaf	#Dens_Int_Semaf_Vias_Asfalt (int/km)	#Semaf_Estag_Ped	#Semaf_Excl_Ped	#Paradas_Ativas	#Dens_Paradas_Vias_Asfalt (par/km)	Area_Zona (km2)	#Vias_<=4fx (km)	#Vias_>4fx (km)	#Via_Total (km)	#Via_Asfaltada (km)	#Via_Sentido_Unico (km)	#Via_Sentido_Duplo (km)	#Mat_Ens_Bas	#Mat_Ens_Sup	#Mat_Pos_Grad	#Soma_Matriculas	L_Indust (m2)	CSBD (m2)	CSAD (m2)	RBD (m2)	RMD (m2)	RAD (m2)	HP (m2)	Entropia	#Dom_Baixa_Renda	#Dom_Media_Renda	#Dom_Alta_Renda	#Domicilios	
Soma_Acid_Ped	1,00																																
#Fisc_Elet	0,36	1,00																															
#Dens_Fisc_Elet_Vias_Asfalt (fisc/km)	0,27	0,89	1,00																														
#Int_Semaf	0,53	0,55	0,45	1,00																													
#Dens_Int_Semaf_Vias_Asfalt (int/km)	0,45	0,47	0,45	0,93	1,00																												
#Semaf_Estag_Ped	0,36	0,33	0,34	0,41	0,43	1,00																											
#Semaf_Excl_Ped	0,49	0,16	0,16	0,46	0,45	0,26	1,00																										
#Paradas_Ativas	0,53	0,24	0,11	0,39	0,28	0,16	0,35	1,00																									
#Dens_Paradas_Vias_Asfalt (par/km)	-0,06	-0,11	-0,10	-0,02	0,01	-0,05	0,08	0,28	1,00																								
Area_Zona (km2)	-0,07	-0,02	0,04	-0,10	-0,14	-0,06	0,00	0,24	0,15	1,00																							
#Vias_<=4fx (km)	0,32	0,12	-0,02	0,15	0,01	0,01	0,11	0,71	0,04	0,34	1,00																						
#Vias_>4fx (km)	0,25	0,19	0,24	0,04	0,04	0,15	0,06	0,13	-0,07	0,21	-0,10	1,00																					
#Via_Total (km)	0,25	0,10	-0,02	0,05	-0,08	-0,03	0,08	0,69	0,07	0,48	0,92	0,01	1,00																				
#Via_Asfaltada (km)	0,56	0,35	0,17	0,37	0,21	0,17	0,19	0,71	-0,32	0,11	0,73	0,13	0,65	1,00																			
#Via_Sentido_Unico (km)	0,34	0,04	-0,12	0,06	-0,06	0,01	0,11	0,45	-0,07	0,13	0,67	-0,07	0,53	0,58	1,00																		
#Via_Sentido_Duplo (km)	0,03	0,10	0,11	0,11	0,08	0,01	0,00	0,36	0,14	0,25	0,43	0,03	0,49	0,23	-0,36	1,00																	
#Mat_Ens_Bas	0,55	0,13	0,03	0,36	0,26	0,10	0,47	0,60	0,10	0,08	0,47	-0,03	0,43	0,50	0,36	0,17	1,00																
#Mat_Ens_Sup	0,09	0,17	0,16	0,14	0,14	0,10	0,02	0,05	-0,02	0,03	-0,01	0,13	-0,05	0,05	-0,01	-0,02	-0,07	1,00															
#Mat_Pos_Grad	0,11	0,23	0,27	0,06	0,07	0,25	0,07	0,10	0,10	0,07	0,02	0,15	-0,06	0,01	0,04	-0,03	-0,04	0,46	1,00														
#Soma_Matriculas	0,47	0,23	0,15	0,37	0,29	0,16	0,37	0,49	0,07	0,08	0,35	0,08	0,28	0,40	0,27	0,11	0,69	0,67	0,36	1,00													
L_Indust (m2)	0,06	-0,02	-0,03	-0,07	-0,08	-0,03	0,07	0,06	0,09	0,07	0,06	0,05	0,06	0,04	0,09	-0,02	0,03	-0,03	-0,03	0,00	1,00												
CSBD (m2)	0,40	0,30	0,21	0,42	0,35	0,21	0,23	0,41	-0,09	0,21	0,28	0,42	0,29	0,46	0,18	0,15	0,21	0,16	0,08	0,27	0,10	1,00											
CSAD (m2)	0,51	0,39	0,28	0,76	0,68	0,30	0,57	0,50	0,09	0,04	0,21	0,09	0,18	0,32	0,13	0,11	0,40	0,11	0,01	0,37	-0,03	0,41	1,00										
RBD (m2)	0,12	-0,06	-0,12	-0,15	-0,23	-0,11	-0,03	0,39	0,05	0,23	0,55	-0,09	0,53	0,39	0,53	0,04	0,29	-0,10	-0,08	0,15	0,03	0,13	-0,05	1,00									
RMD (m2)	0,33	0,12	0,02	0,11	0,03	0,09	0,12	0,54	-0,15	0,00	0,59	-0,05	0,52	0,66	0,43	0,23	0,32	-0,07	-0,04	0,19	-0,01	0,18	0,13	0,42	1,00								
RAD (m2)	0,14	0,47	0,33	0,55	0,46	0,27	-0,01	0,15	-0,16	-0,04	0,16	0,05	0,10	0,33	0,08	0,09	0,01	0,07	0,01	0,06	-0,06	0,22	0,38	-0,14	0,09	1,00							
HP (m2)	0,15	0,04	-0,03	-0,05	-0,10	-0,06	0,07	0,13	-0,03	-0,06	0,14	0,02	0,11	0,18	0,25	-0,12	0,11	-0,03	-0,02	0,06	0,14	0,02	-0,02	0,05	0,05	0,02	1,00						
Entropia	0,37	0,33	0,20	0,35	0,30	0,24	0,15	0,30	-0,17	-0,34	0,23	0,06	0,15	0,48	0,17	0,09	0,14	0,04	-0,03	0,14	0,16	0,28	0,24	-0,05	0,45	0,40	0,23	1,00					
#Dom_Baixa_Renda	0,34	-0,06	-0,12	-0,14	-0,21	-0,10	0,11	0,47	0,01	0,16	0,63	-0,06	0,62	0,52	0,47	0,22	0,42	-0,07	-0,05	0,26	0,20	0,16	-0,04	0,57	0,39	-0,16	0,23	0,08	1,00				
#Dom_Media_Renda	0,53	0,22	0,11	0,23	0,13	0,13	0,24	0,61	-0,10	0,03	0,62	-0,01	0,52	0,74	0,51	0,16	0,44	0,01	0,01	0,33	0,09	0,37	0,20	0,52	0,76	0,15	0,08	0,41	0,66	1,00			
#Dom_Alta_Renda	0,17	0,48	0,33	0,59	0,49	0,26	0,02	0,21	-0,19	-0,04	0,21	0,05	0,14	0,43	0,15	0,07	0,04	0,08	0,01	0,09	-0,05	0,30	0,40	-0,04	0,20	0,93	0,02	0,43	-0,14	0,26	1,00		
#Domicilios	0,44	0,15	0,03	0,12	0,01	0,03	0,15	0,58	-0,08	0,12	0,72	-0,04	0,66	0,72	0,54	0,23	0,45	-0,03	-0,03	0,31	0,16	0,30	0,14	0,57	0,55	0,19	0,21	0,29	0,91	0,84	0,25	1,00	

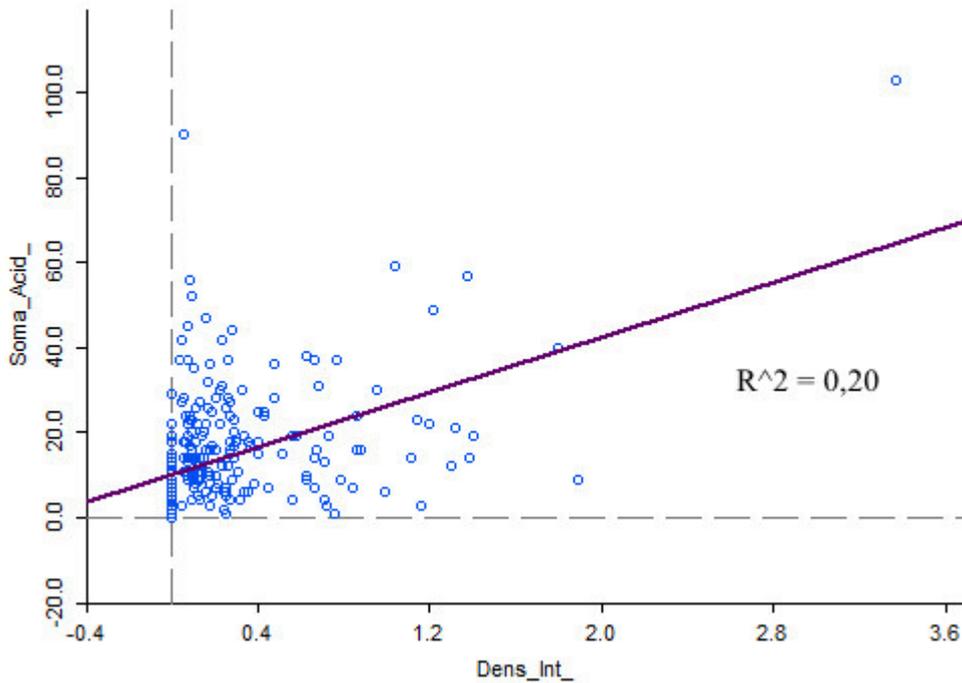
ANEXOS B – GRÁFICOS DE DISPERSÃO

Figura 1B – Gráfico de dispersão Interseções semaforizadas Vs Atropelamentos



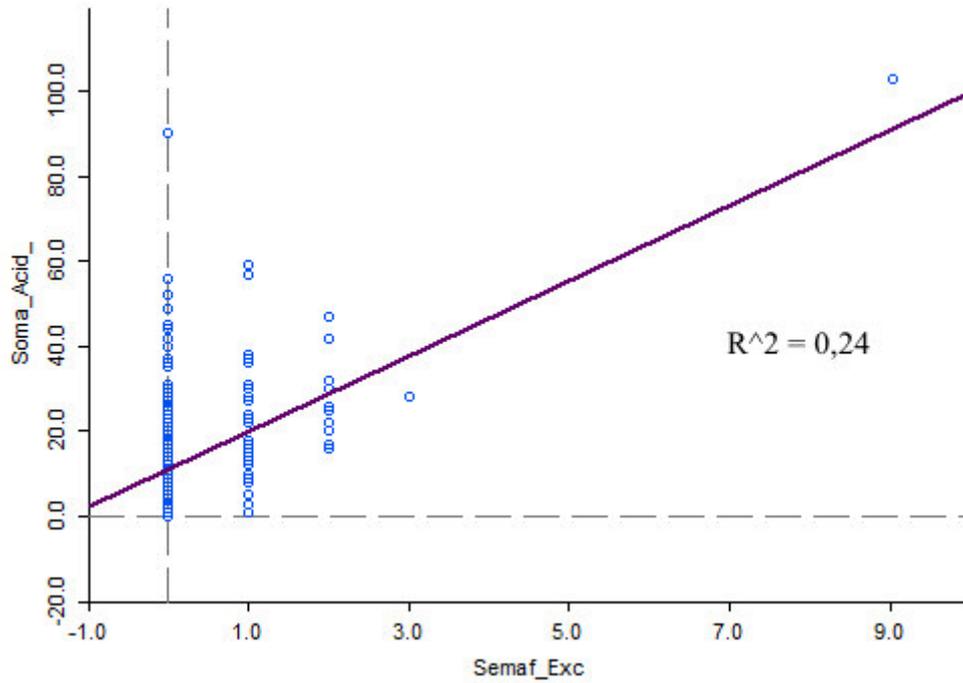
Fonte: Próprio autor (2016)

Figura 2B - Gráfico de dispersão Densidade de interseções semaforizadas Vs Atropelamentos



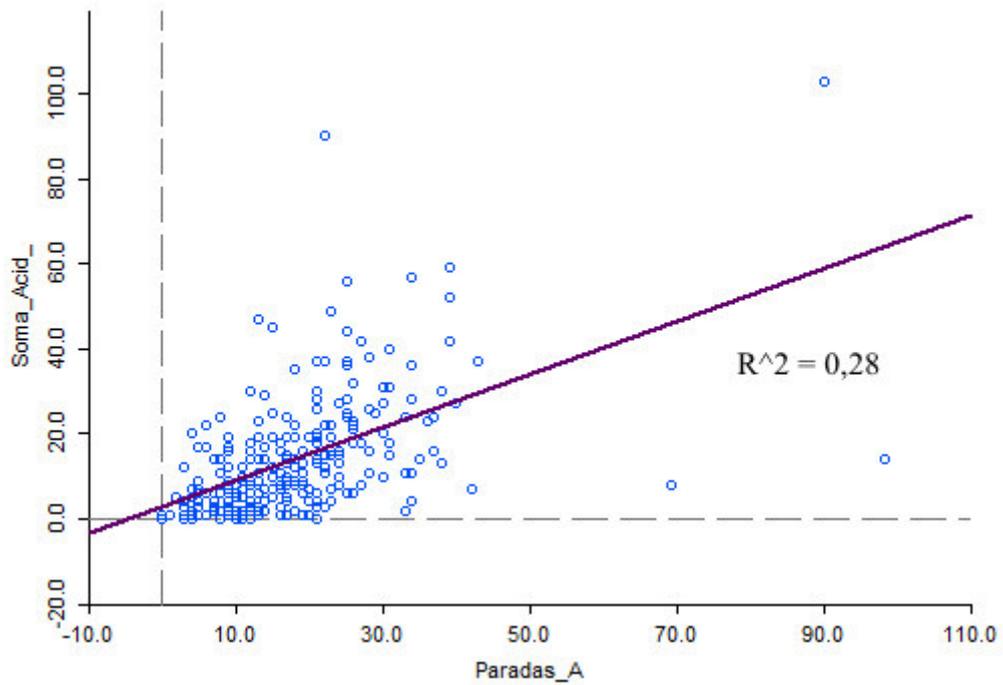
Fonte: Próprio autor (2016)

Figura 3B - Gráfico de dispersão Semáforos exclusivos para pedestres Vs Atropelamentos



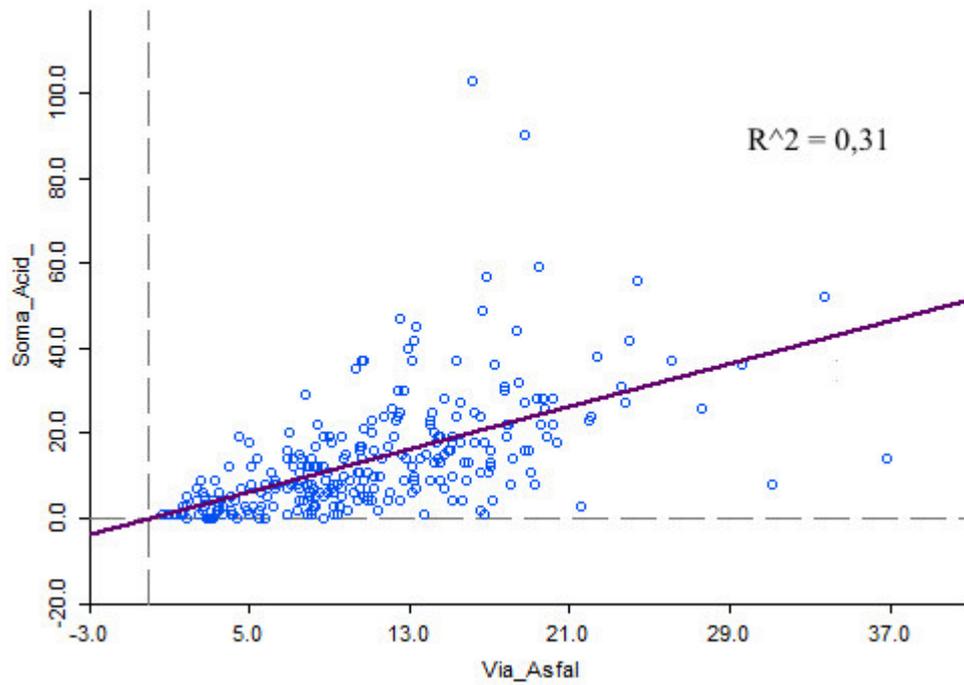
Fonte: Próprio autor (2016)

Figura 4B - Gráfico de dispersão Paradas ativas de ônibus Vs Atropelamentos



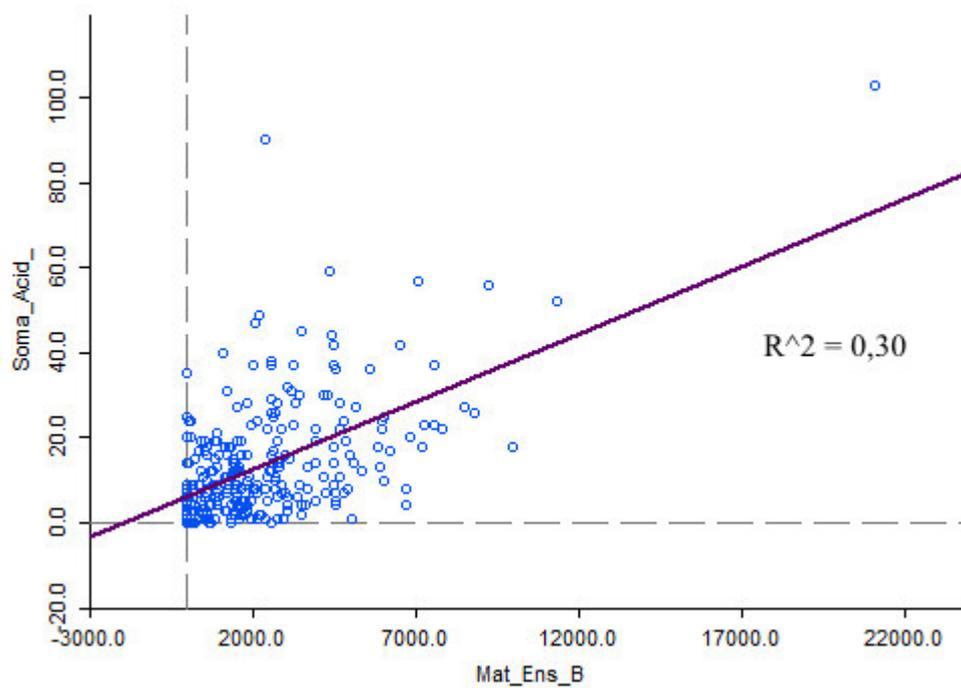
Fonte: Próprio autor (2016)

Figura 5B - Gráfico de dispersão Quilômetros de via asfaltada Vs Atropelamentos



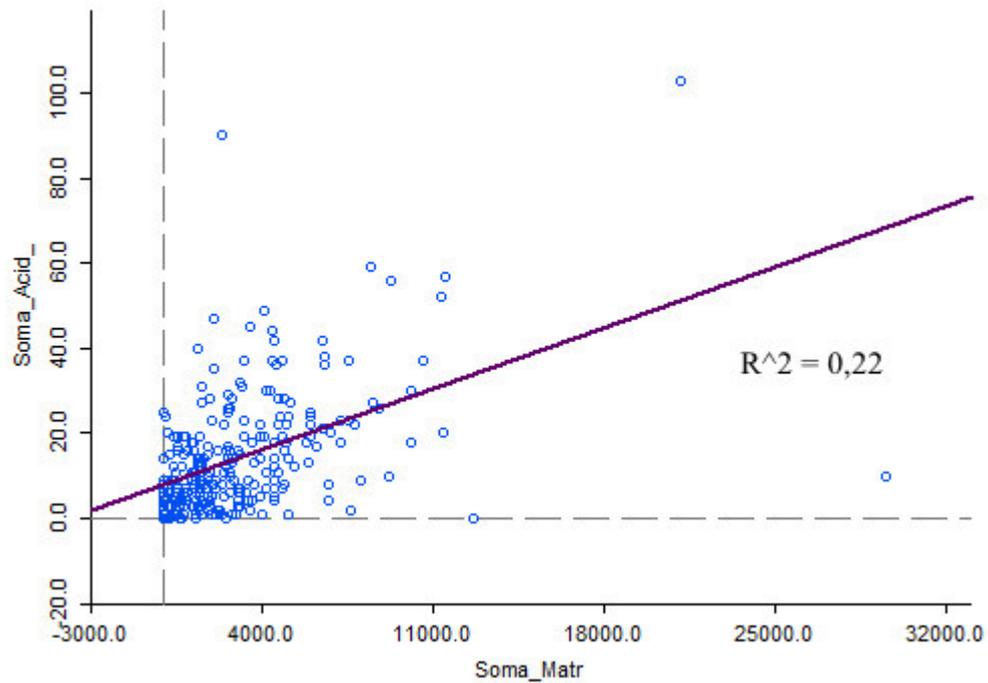
Fonte: Próprio autor (2016)

Figura 6B - Gráfico de dispersão Matrícula do ensino básico Vs Atropelamentos



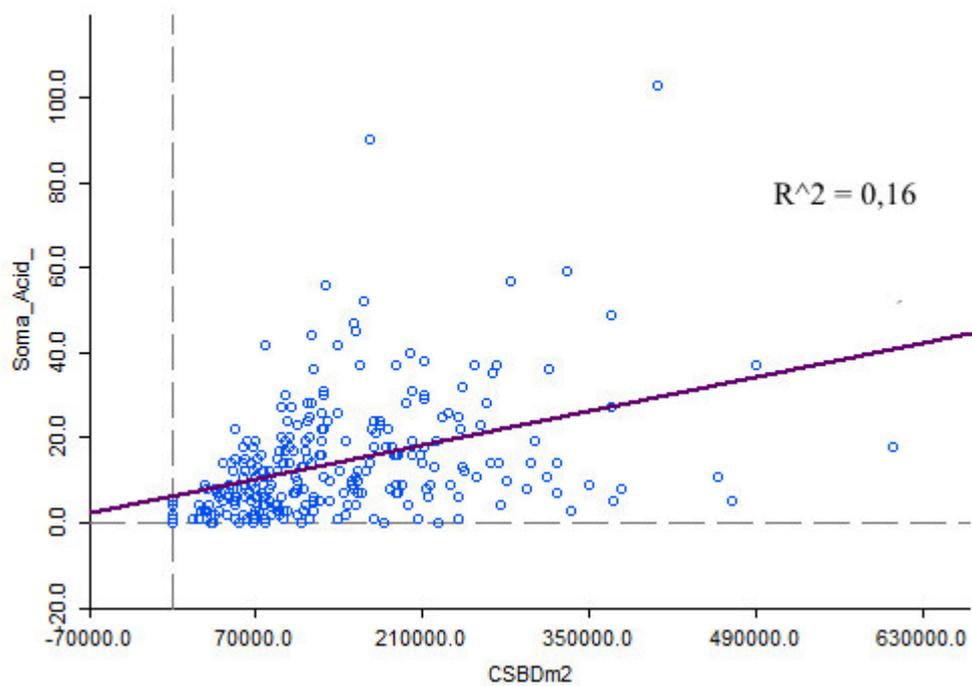
Fonte: Próprio autor (2016)

Figura 7B - Gráfico de dispersão Soma das matrículas Vs Atropelamentos



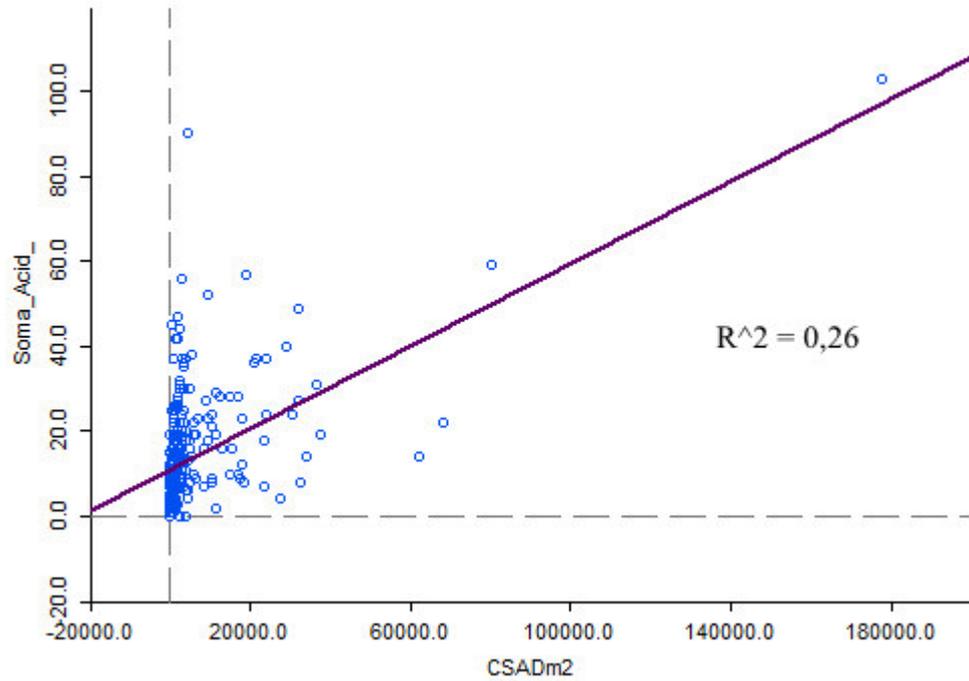
Fonte: Próprio autor (2016)

Figura 8B - Gráfico de dispersão CSBD Vs Atropelamentos



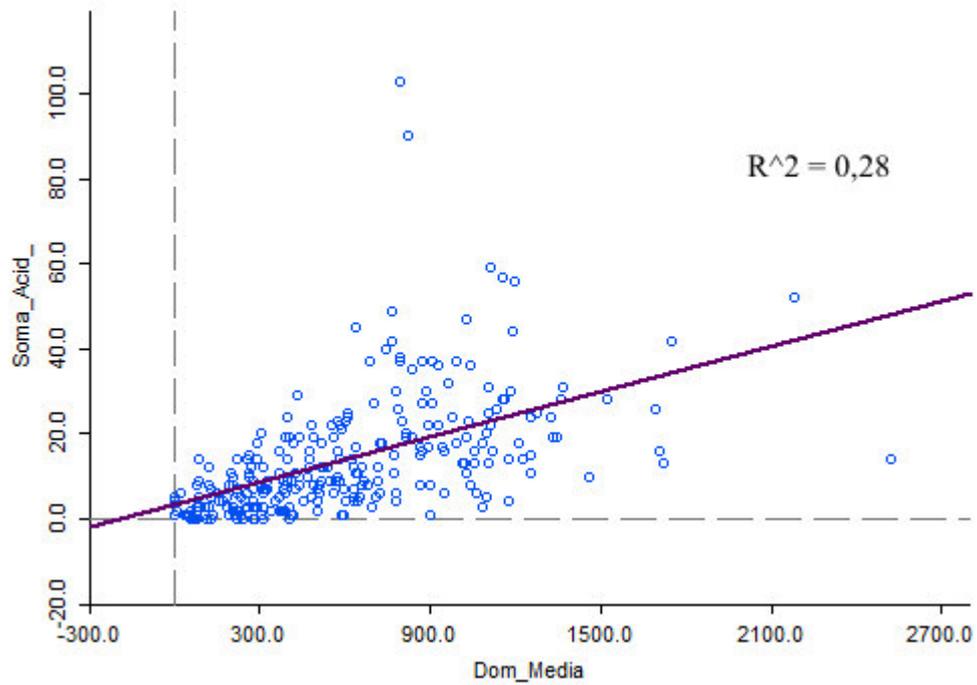
Fonte: Próprio autor (2016)

Figura 9B - Gráfico de dispersão CSAD Vs Atropelamentos



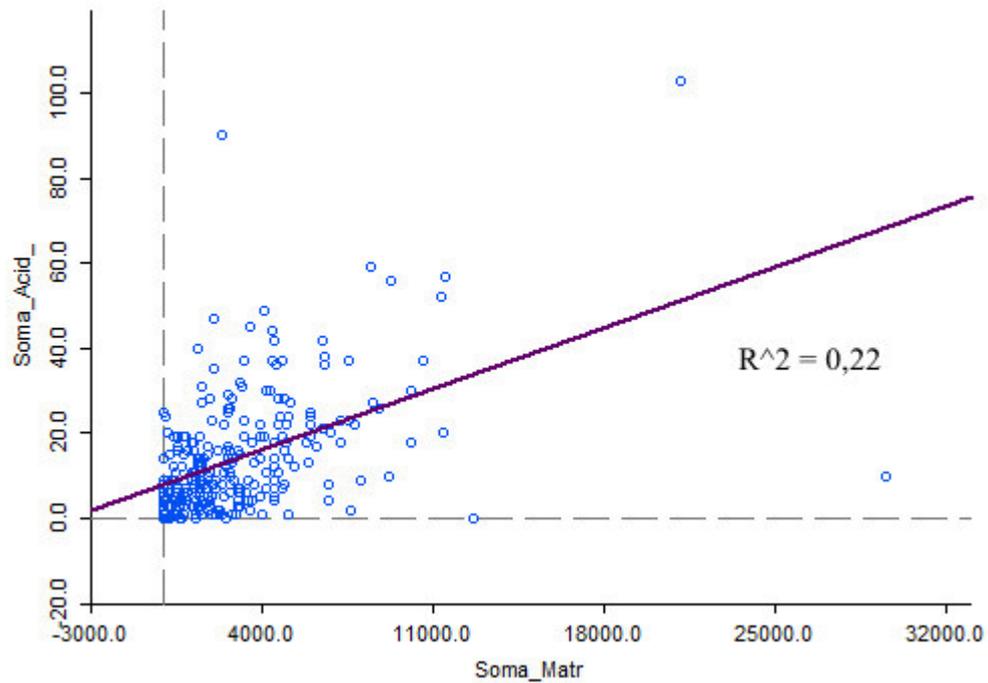
Fonte: Próprio autor (2016)

Figura 10B - Gráfico de dispersão Domicílios de média renda Vs Atropelamentos



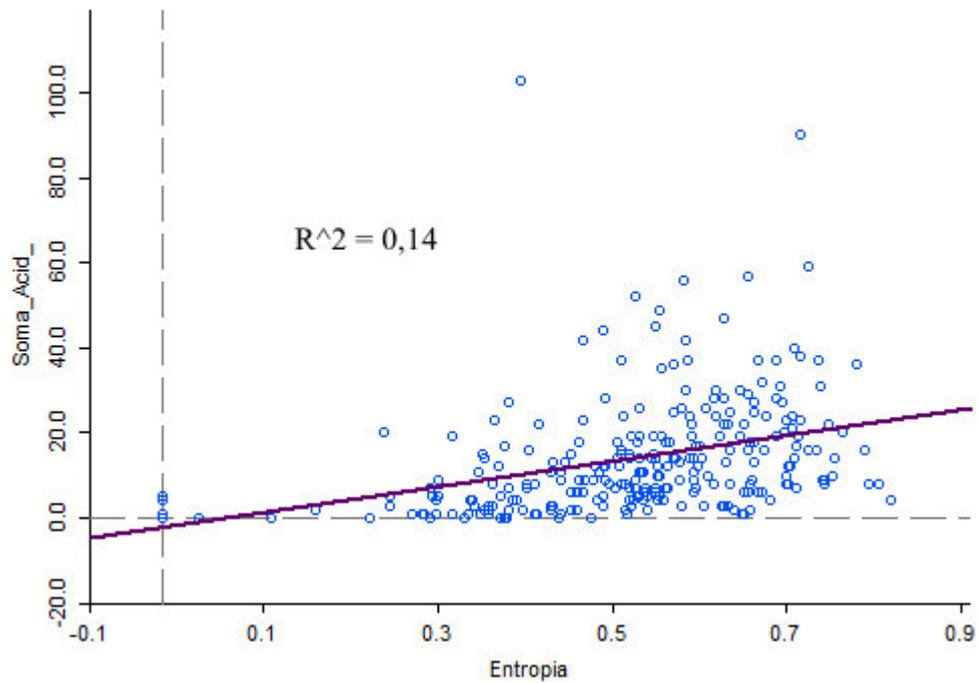
Fonte: Próprio autor (2016)

Figura 11B - Gráfico de dispersão Soma de domicílios Vs Atropelamentos



Fonte: Próprio autor (2016)

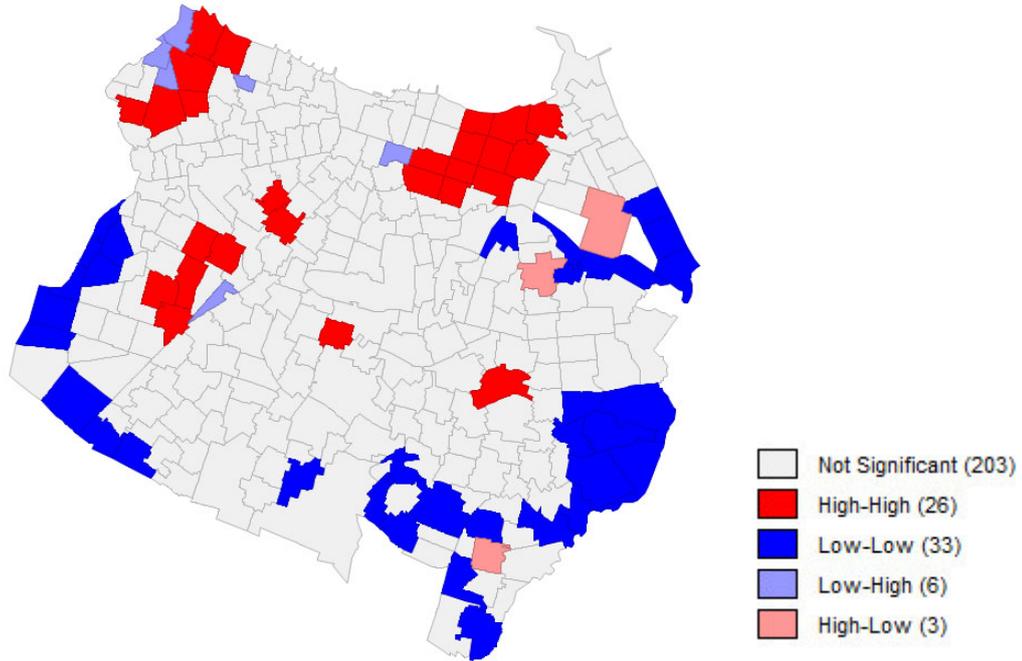
Figura 12B - Gráfico de dispersão Entropia Vs Atropelamentos



Fonte: Próprio autor (2016)

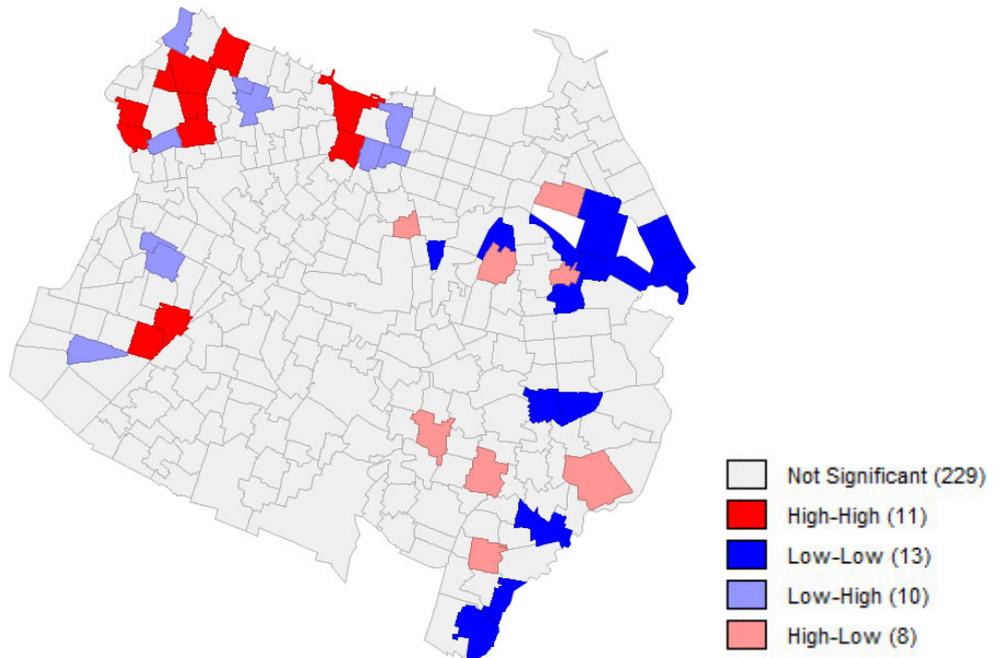
ANEXOS C – MAPAS DE MORAN

Figura 1C - Mapa de Moran da variável quilômetros de vias asfaltadas



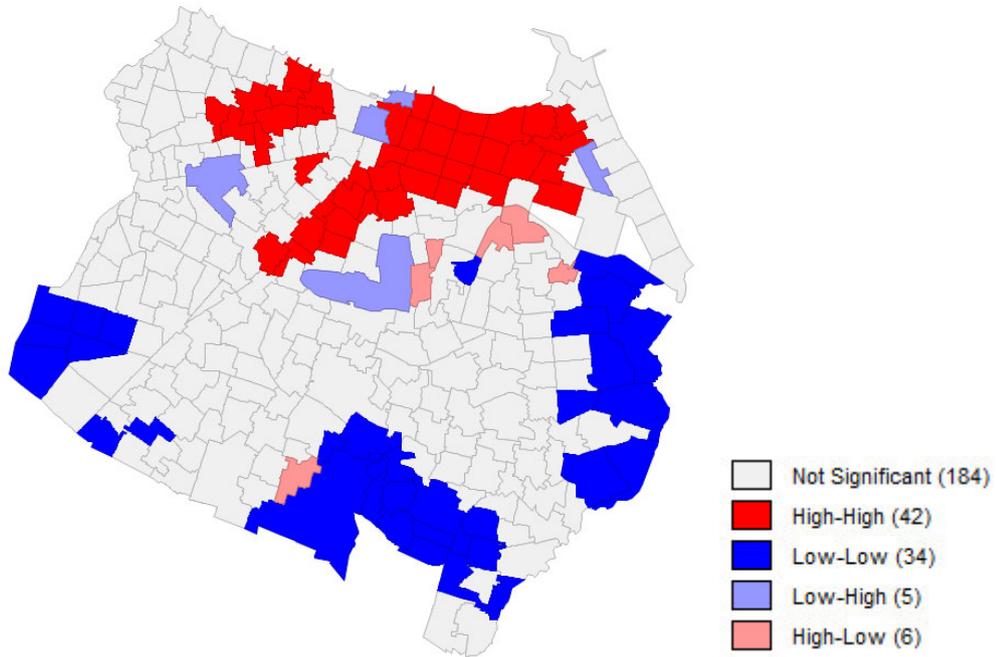
Fonte: Próprio autor (2016)

Figura 2C - Mapa de Moran da variável matrículas do ensino básico



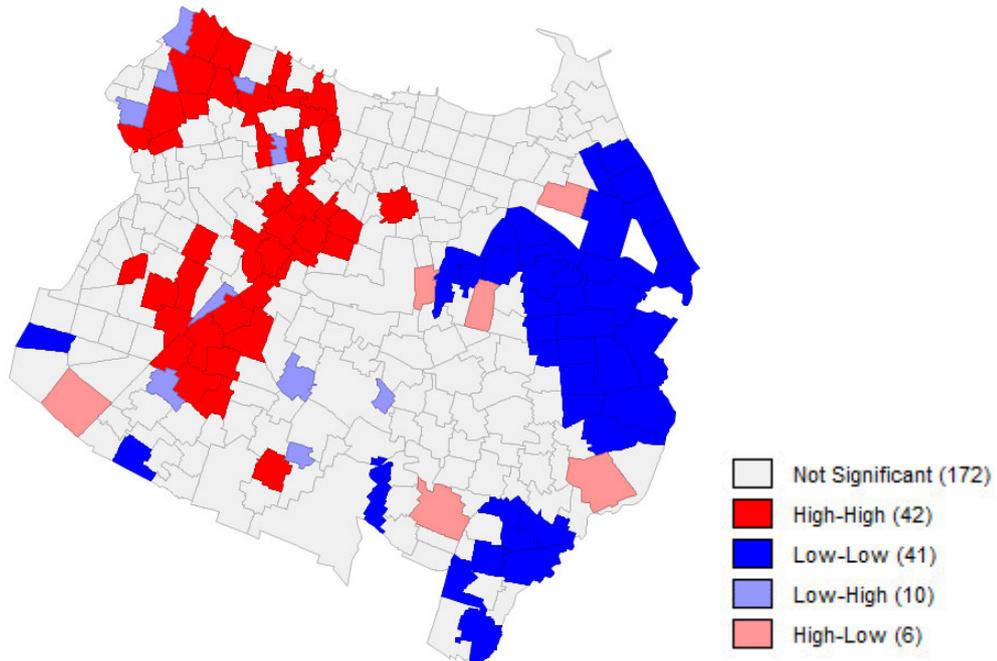
Fonte: Próprio autor (2016)

Figura 3C - Mapa de Moran da variável entropia



Fonte: Próprio autor (2016)

Figura 4C - Mapa de Moran da variável domicílios de média renda



Fonte: Próprio autor (2016)