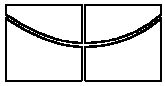

Universidade de Brasília

Instituto de Ciências Exatas

Departamento de Estatística

Dissertação de Mestrado

IDENTIFICAÇÃO DE CONGLOMERADOS ESPACIAIS EM CENÁRIOS COM VIZINHANÇAS DEFINIDAS TOPOLOGICAMENTE

por

João Ricardo Eliseu

Orientador: Prof. Dr. André Luiz Fernandes Cançado

João Ricardo Eliseu

IDENTIFICAÇÃO DE CONGLOMERADOS ESPACIAIS EM CENÁRIOS COM VIZINHANÇAS DEFINIDAS TOPOLOGICAMENTE

Dissertação apresentada ao Departamento de Estatística do Instituto de Ciências Exatas da Universidadede de Brasília como requisito parcial à obtenção do título de Mestre em Estatística.

Universidade de Brasília

Brasília, Junho de 2014

Termo de Aprovação

João Ricardo Eliseu

IDENTIFICAÇÃO DE CONGLOMERADOS ESPACIAIS EM CENÁRIOS COM VIZINHANÇAS DEFINIDAS TOPOLOGICAMENTE

Dissertação apresentada ao Departamento de Estatística do Instituto de Ciências Exatas da Universidadede de Brasília como requisito parcial à obtenção do título de Mestre em Estatística.

Data da defesa: 06 de junho de 2014

Orientador:

Prof. Dr. André Luiz Fernandes Cançado
Departamento de Estatística, UnB

Comissão Examinadora:

Prof. Dr. Luiz Henrique Duczmal,
Departamento de Estatística, UFMG

Prof. Dr. Antônio Eduardo Gomes
Departamento de Estatística, UnB

Brasília, Junho de 2014

Ficha Catalográfica

ELISEU, JOÃO RICARDO
Identificação de Conglomerados Espaciais em Cenários com Vizinhanças definidas Topologicamente , (UnB - IE, Mestre em Estatística, 2014).

Dissertação de Mestrado - Universidade de Brasília. Departamento de Estatística - Instituto de Ciências Exatas.

1. Acidente de Trânsito Fatal
2. Conglomerado Espacial
3. *Scan* Circular
4. *Scan* de Vizinhança

À minha família, especialmente meus pais,

Fátima e João, por tornarem o caminho possível

À minha namorada, Vivian Alves, pela paciência e dedicação.

Agradecimentos

- Ao professor André, pela insistência e paciência nas inúmeras reuniões. Por passar tranquilidade e conhecimento, sem os quais não obteria sucesso.
- Aos professores do Departamento de Estatística, em especial aos membros da banca examinadora.
- Aos colegas da graduação e do mestrado, pela amizade e companheirismo ao longo dos anos.
- Aos técnicos do DER/DF, pela ajuda e contribuição para a dissertação.

Sumário

Lista de Figuras

Lista de Tabelas

Resumo

IDENTIFICAÇÃO DE CONGLOMERADOS ESPACIAIS EM CENÁRIOS COM VIZINHANÇAS DEFINIDAS TOPOLOGICAMENTE

O período entre 2011 e 2020 foi definido pela ONU como a década de ações pela segurança no trânsito. A redução de acidentes de trânsito fatais é o principal foco dos órgãos da administração pública. Dessa maneira, a delimitação de regiões com alta incidência de acidentes aumentaria o foco de ações corretivas e diminuiria o gasto dos órgãos públicos, adotando ações mais eficientes. Os dados de acidentes de trânsito em rodovias distritais são consolidados pelo Sistema de Informações de Trânsito e classificados espacialmente conforme o Sistema Rodoviário do Distrito Federal. Os acidentes de trânsito são distribuídos ao longo de trechos rodoviários, que são subdivisões ao longo das rodovias. No estudo de acidentes, o critério espacial que agrupa regiões mais próximas pode apresentar resultados menos satisfatórios, dado que os acidentes de trânsito são restritos às vias de circulação de veículos e pedestres. Neste trabalho buscamos criar métodos que utilizam o conceito de vizinhança ligado à idéia de proximidade. Os métodos de vizinhança, ao contrário do *Scan* circular, agrupam regiões que fazem fronteira entre si. O método *Scan* Circular e os métodos de vizinhança foram testados para três formatos de conglomerados, para os quais dois métodos de vizinhança tiveram desempenho superior ao *Scan* Circular, sendo menos afetados pela geometria do conglomerado. Após os testes numéricos, os quatro métodos foram utilizados para identificar conglomerados de acidentes fatais no mapa rodoviário do DF, 2012.

Palavras Chave: *acidentes de trânsito, Sistema Rodoviário, Scan Circular, Scan de Vizinhança.*

Abstract

IDENTIFICATION OF SPATIAL CLUSTERS IN SCENARIOS WITH TOPOLOGICALLY DEFINED NEIGHBORHOODS

The period between 2011 and 2020 was defined by the United Nations as the decade of action for road safety. The reduction of fatal traffic accidents is the main focus of public administration. Thus, the delineation of regions with a high incidence of accidents would increase the focus of corrective actions and reduce the expenditure of public resources, adopting more efficient actions. The data of traffic accidents in district highways are consolidated by the Integrated Traffic Accident system and spatially classified by the Highway System of the Federal District. Traffic accidents are distributed along road sections, which are subdivisions along highways. In the study of traffic accidents, the spatial criterion that groups closer regions may have less satisfactory results, given that traffic accidents are restricted to traffic routes for vehicles. In this work we aim to create methods that use the concept of frontier connected to the idea of proximity. The neighborhood based methods, unlike the Circular Scan, group regions bordering each other. The Circular Scan method and the neighborhood methods were tested for three cluster shapes, for which two neighborhood methods had superior performance compared to the Circular Scan, being less affected by the geometry of the cluster. After the numerical tests, the four methods were used to identify clusters of fatalities in the road map of the Federal District, 2012.

key words: *traffic accidents, Highway System, Scan Statistic, Neighborhood Scan.*

Capítulo 1

Introdução

A Organização das Nações Unidas (ONU) promulgou o período entre 2011 e 2020 como a Década de Ações pela Segurança no Trânsito, quando uma série de medidas que visam a redução do número de vítimas decorrentes de acidentes de trânsito deverão ser implementadas. De acordo com o Sistema de Informação sobre Mortalidade, Ministério da Saúde, o número de vítimas fatais em acidentes de trânsito no Brasil foi de 43 mil, em 2010. Um fato alarmante é que os acidentes de trânsito estão entre as principais causas de internações em hospitais, sobrecarregando a saúde pública do país e onerando os cofres públicos. No Distrito Federal, segundo dados do Departamento Nacional de Trânsito (DENATRAN), o número de mortos em acidentes de trânsito em 2012 foi de 417, sendo 45,8% ocorridas em rodovias distritais. O número de feridos em acidentes de trânsito no Distrito Federal é de 10000, aproximadamente. As vias do Distrito Federal são classificadas em vias urbanas, rodovias distritais e federais, de acordo com sua circunscrição.

A rede rodoviária distrital pavimentada, sob circunscrição do DER/DF, é de aproximadamente 893,8 quilômetros (Sistema Rodoviário do Distrito Federal(2012), 2012), divididas em 81 rodovias. O tamanho da malha rodoviária distrital torna difícil a verificação de todas as ocorrências de acidentes de trânsito. Uma saída é a utilização dos trechos rodoviários como sendo regiões geográficas com probabilidade não nula do evento acidente de trânsito. Com isso, o problema passa a ser a determinação dos conglomerados críticos, onde um conglomerado é a união de regiões delimitadas ao longo da rodovia que apresentam alguma característica em comum, de acordo com os

objetivos e a abordagem do estudo. Os trechos rodoviários são espaços geométricos delimitados, definidos de acordo com critérios estabelecidos pelo SRDF e descritos na Seção 2.4. Os conglomerados são regiões formadas por um ou mais trechos rodoviários. Portanto, a menor subdivisão espacial possível de um conglomerado será o trecho rodoviário. Assim, diminuem-se os erros não amostrais de classificação espacial, uma vez que não mais se exige a localização exata do acidente, e sim que o mesmo pertença ao trecho rodoviário correto.

A união do banco de dados do Sistema Rodoviário do Distrito Federal (SRDF) com o de acidentes de trânsito permite a classificação das ocorrências por rodovia e trecho rodoviário, além de possibilitar a combinação de variáveis presentes nos dois bancos de dados. A partir da observação conjunta é possível o cálculo de medidas descritivas e a obtenção de índices que fazem uso de variáveis presentes no banco de dados de acidentes e do Sistema Rodoviário, permitindo comparar diferentes regiões. Uma possibilidade é a classificação hierárquica dos trechos rodoviários, de acordo com cada um dos índices. Porém, a classificação dos trechos com a utilização de índices, além de não apresentar validade estatística, nem sempre retrata a realidade, principalmente em trechos com pouca extensão, em que os fenômenos aleatórios podem influenciar de forma mais considerável. Além disso, a abordagem exploratória analisa a influência isolada do trecho, não sendo capaz de examinar a correlação espacial entre dois ou mais trechos.

A estatística espacial estuda métodos científicos para a análise de dados em que há dependência espacial, onde alguma referência geográfica é utilizada no modelo. O desenvolvimento computacional foi um fator que contribuiu para o desenvolvimento da estatística espacial, com aplicações em análise de experimentos agrícolas, padrões de morbidade, aplicações geofísicas, em agronomia, epidemiologia, etc. Neste trabalho, a área de aplicação será com dados de acidentes de trânsito, onde não se sabe a coordenada geográfica de cada acidente, e sim o número de acidentes pertecentes a determinado trecho, sendo um estudo com dados de área. Deseja-se descobrir se os dados de acidentes de trânsito ao longo dos trechos rodoviários estão distribuídos aleatoriamente ou se existe algum conglomerado - subconjunto conexo de trechos - que possui maior incidência de acidentes. Portanto, métodos de detecção de conglomerados são necessários a fim de se determinar regiões com número de acidentes

maior do que o valor esperado, para a mesma região.

Os testes para a detecção de conglomerados tem como finalidade identificar regiões que merecem uma investigação ou estudo detalhado, ou o uso de métodos científicos que são impraticáveis sob uma região muito grande (Besag & Newell, 1991).

Algumas análises estatísticas verificam se a taxa de incidência de dada região é diferente à taxa de outra região, entretanto, testar se uma região do mapa tem maior incidência de eventos em detrimento ao restante do mapa é incorreto, mesmo que levemos em consideração a proporção de suas populações. Outra possibilidade é listar todas as regiões passíveis de serem um conglomerado e, para cada uma delas, testar se sua taxa difere estatisticamente daquela associada com o restante do mapa sob estudo. Este segundo procedimento também é incorreto, dado que mesmo fixando a probabilidade do erro tipo I, vários testes serão significativos mesmo que a hipótese nula seja verdadeira em todos eles. O erro é que o α considerado em cada teste individual não é válido para os testes simultâneos. O fato é que a probabilidade do erro do tipo I tende a ser bem maior que o valor dos testes individuais. No capítulo de revisão bibliográfica serão apresentados vários artigos relacionados à estatística espacial em que o problema dos testes simultâneos é recorrente. Somente com o artigo de Kulldorff & Nagarwalla (1995) este problema é solucionado.

O método *Scan* Circular foi proposto por Kulldorff (1997), sendo uma técnica de detecção e inferência de conglomerados espaciais para determinado evento de interesse. O método parte do pressuposto que a população sob o risco de ocorrência é conhecida, bem como o número de casos do evento "acidente de trânsito" e a coordenada geográfica do centróide da região.

O Método de Kulldorff utiliza uma matriz quadrada das distâncias entre os centróides de cada trecho rodoviário, unindo os trechos com menor distância entre si. O tamanho do conglomerado é limitado pelo tamanho da população total sob risco, tendo a proporção da mesma pré-estabelecida. O processo iterativo varre a superfície total com janelas circulares até preencher a área correspondente ao conglomerado. O objetivo é encontrar regiões onde o número de casos na região é significativamente maior que o valor esperado para ela. Através da distribuição empírica da estatística do teste - obtida pelo método de Monte Carlo - , é feito um teste de hipóteses para definir se o conglomerado é significativo ou não.

O capítulo 2 trata dos principais conceitos sobre acidentes de trânsito, fonte de dados, e de como é feito o tratamento dos dados de acidentes. Também são descritas informações sobre o SRDF (Sistema Rodoviário do Distrito Federal(2012), 2012), documento elaborado em conformidade com o Roteiro Básico para Sistemas Rodoviários Estaduais. O SRDF apresenta a situação da malha rodoviária distrital, como a extensão da malha viária pavimentada, nomenclatura de rodovias, definição das subdivisões em cada rodovia, etc. Neste capítulo são definidos os critérios para a criação dos trechos rodoviários - subdivisões conexas -, ao longo das rodovias sob circunscrição do DER/DF, bem como as variáveis que definem e delimitam cada trecho.

O capítulo 3 contém uma pesquisa bibliográfica dos principais artigos relacionados à estatística espacial, dando ênfase aos artigos relacionados ao método *Scan* Circular de Kulldorff.

O capítulo 4 versa sobre a metodologia de pesquisa do *Scan* Circular. No capítulo 5 uma proposta alternativa de construção de conglomerado é apresentada, utilizando um critério topológico para agrupamento de regiões através da matriz de vizinhança. Dentro desse contexto, diferentes métodos de vizinhança foram apresentadas. Cada método de vizinhança procura selecionar os candidatos vizinhos de maneira específica.

No capítulo 6, três conglomerados artificiais foram criados, com diferentes geometrias. Os métodos de varredura especificados nos capítulos 4 e 5 foram testados para os 3 (três) tipos de conglomerados artificiais. A mensuração do desempenho é feita por meio do poder do teste, sensibilidade e valor preditivo positivo. Além de comparar o *Scan* Circular com os métodos de vizinhança, essas medidas buscam comprovar que além de eficientes, os métodos de vizinhança são mais apropriados para estudos com acidentes de trânsito. A seguir o trabalho observacional é encontrar regiões no mapa rodoviário do Distrito Federal com incidência significativa de acidentes de trânsito fatais, 2012. Para identificar conglomerados de acidentes, utilizamos o método *Scan* Circular de Kulldorff (1997) e os métodos de vizinhança definidos no capítulo 5.

Capítulo 2

Definições relativas ao conjunto de dados

2.1 Conceitos relativos aos acidentes de trânsito

Segundo a Associação Brasileira de Normas Técnicas (ABNT, 1989), acidente de trânsito é "todo evento não premeditado de que resulte dano em veículo ou na sua carga e/ou lesões em pessoas e/ou animais, em que pelo menos uma das partes está em movimento nas vias terrestres ou áreas abertas ao público. Pode originar-se, terminar ou envolver veículo parcialmente na via pública".

Assim sendo, os acidentes ocorridos dentro de casas, garagens, chácaras, canteiros de obras, não são considerados como acidentes de trânsito.

O conceito de acidente de trânsito como sendo evento não premeditado em áreas abertas ao público não exclui os eventos ocorridos parcialmente na via pública. Ou seja, para que não seja considerado acidente de trânsito, o acontecimento precisa ocorrer totalmente em área particular. Caso o evento ocorra parcialmente em via pública, será considerado acidente de trânsito.

2.2 Fonte de dados de acidentes de trânsito

Segundo a Confederação Nacional dos Municípios (CNM, 2009), existem três fontes distintas de dados estatísticos de acidentes de trânsito com morte: DENATRAN -

Departamento Nacional de Trânsito; DATASUS - Banco de dados do Sistema Único de Saúde/MS; e Seguros DPVAT - Danos Pessoais Causados por Veículos Automotores de Via Terrestre ou por sua Carga a Pessoas Transportadas ou Não. As três fontes são totalmente distintas e não são passíveis de comparação.

O DENATRAN - Departamento Nacional de Trânsito - elabora seus anuários estatísticos de acordo com o registro do boletim de ocorrência nas Delegacias de Polícia. Geralmente só levam em conta a morte "in loco" - com exceção de alguns estados, como o Distrito Federal - que acompanha a vítima ferida nos 30 (trinta) dias posteriores ao acidente (CNM, 2009).

Os dados estatísticos do DENATRAN geralmente subestimam o número de mortos por acidentes de trânsito porque nem sempre os acidentes com vítima são registrados nas Delegacias de Polícia. Além disso, no caso dos estados que não consideram o óbito após os 30 (trinta) dias, não são registradas as mortes que ocorrem posteriormente, no hospital (CNM, 2009).

O Ministério da Saúde, por meio do DATASUS, registra os óbitos decorrentes de acidentes de trânsito através do atendimento em unidades de saúde. É a única fonte que registra as mortes por município. A operacionalização é feita através do registro da Declaração de Óbito - D.O. -, que é exigida nas instituições de saúde dos estados e municípios. A Declaração de Óbito pode ser preenchida pelo perito do Instituto Médico Legal (IML). Tais informações passam a constar na Base Nacional do Sistema de Informações sobre Mortalidade - SIM (CNM, 2009).

O registro de óbitos originários de acidentes de trânsito, constante na base de dados do Ministério da Saúde, também contém erros que subestimam as estatísticas reais, dado que muitos mortos decorrentes de acidente de trânsito são registrados como acidentes de outra natureza.

A terceira e última fonte de dados é a Seguradora Líder dos Consórcios do Seguro DPVAT (Danos Pessoais Causados por Veículos Automotores de Via Terrestre ou por sua Carga a Pessoas Transportadas ou Não), um seguro obrigatório, instituído em 1974. Tal seguro foi instituído para amparar as vítimas de acidentes com veículo. A estatística divulgada é feita através dos seguros pagos às vítimas de acidentes de trânsito. Esta avaliação dos seguros pagos às vítimas de trânsito é a fonte mais próxima da realidade, contudo, ainda é subestimada, dado que uma boa parcela da

população desconhece o direito de receber indenização (CNM, 2009).

2.3 Tratamento dos dados

O tratamento dos dados se dá por meio da crítica, que tem por objetivo melhorar a qualidade do banco de dados, corrigindo e complementando informações. A crítica dos dados é imprescindível para que a estatística retrate de maneira mais fidedigna a realidade. O tratamento das informações que visam a consistência dos dados é a parte que demanda maior esforço, custo e tempo de trabalho.

A correção é feita através da leitura do histórico do boletim de ocorrência, que é preenchido pelo agente de trânsito ou policial militar. Dependendo do caso é necessário o contato com as delegacias, pessoas envolvidas ou até mesmo testemunhas.

De acordo com o Anuário Estatístico de Acidentes de Trânsito do Distrito Federal (2010) (2010), a crítica dos dados estatístico é dividida em:

- Crítica geral

  O sistema realiza um filtro das informações advindas dos dados da PCDF (Polícia Civil do Distrito Federal), onde é identificada alguma inconsistência interna. Alguns desses erros dizem respeito à localização do acidente (dentro ou fora do DF), acidentes com vítimas sem que sejam identificadas, colisão com menos de dois condutores ou dois veículos, condutor sem registro de CNH (Carteira Nacional de Habilitação), atropelamento sem pedestre, etc.

- Crítica do IML (Instituto Médico Legal)

  Verificação se o número de vítimas fatais constantes no IML está de acordo com os dados da PCDF. Além disso, é conferido o nome das vítimas.

- Crítica das declarações de óbito (DO)

  As Declarações de Óbito chegam mensalmente ao DETRAN/DF, então é conferido se há alguma vítima fatal que ainda não conste no banco de dados. Posteriormente a isso, são preenchidos os campos de alcoolemia, data do óbito, entre outras variáveis.

- Crítica do endereço do acidente

  No caso de rodovias distritais e vicinais, a crítica de endereço é feita de acordo com o Sistema Rodoviário do Distrito Federal, que lista todas as rodovias distritais e vicinais divididas por trecho. Para cada trecho é definido um início e final (com definição do quilômetro), com as referências.

A figura 2.1 apresenta um organograma dos dados de acidentes de trânsito:

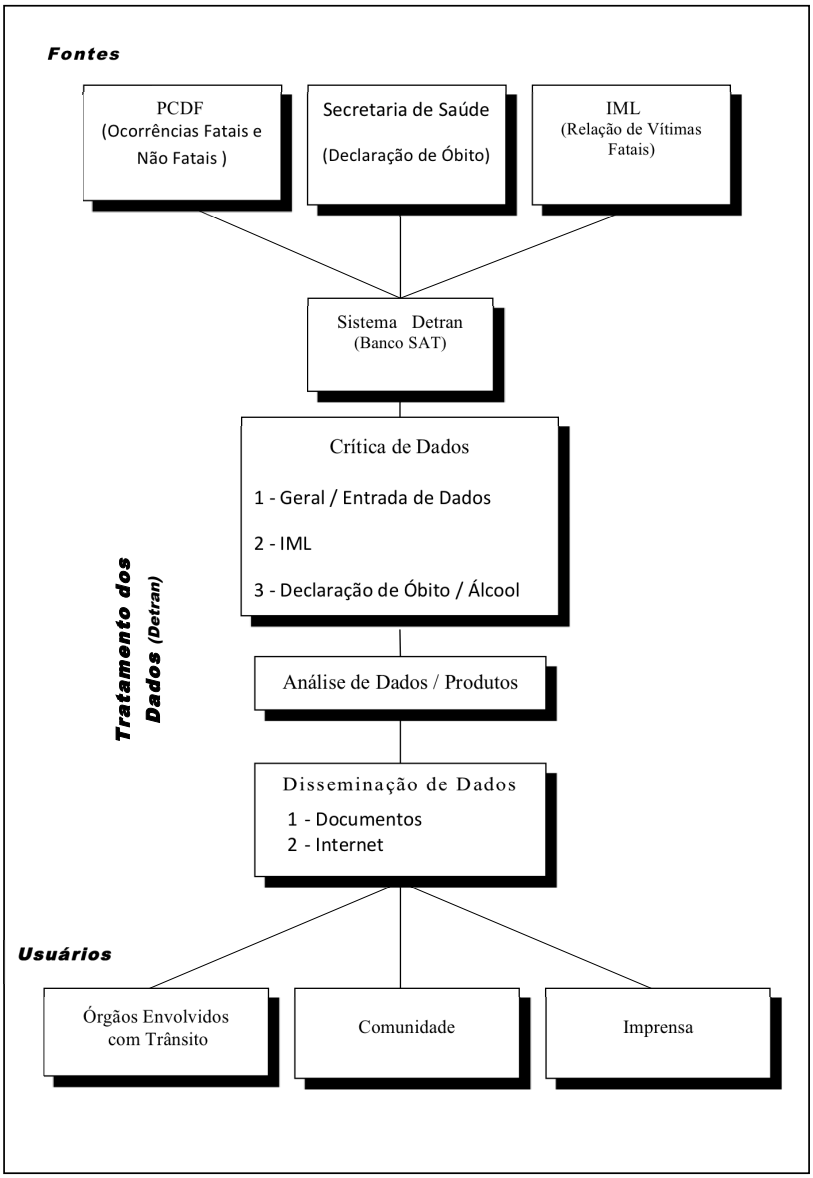

Figura 2.1: Fonte, tratamento e disseminação dos dados de acidentes de trânsito no DF.

2.4 Sistema Rodoviário do Distrito Federal

O Departamento de Estradas de Rodagem do Distrito Federal, DER/DF, apresenta anualmente o Sistema Rodoviário do Distrito Federal, que define, atualiza, exclui as rodovias e trechos rodoviários (Sistema Rodoviário do Distrito Federal(2012), 2012) e definições sobre a mesma. O Sistema Rodoviário do Distrito Federal(2012) (2012) define o Sistema Rodoviário da seguinte forma:

"O presente documento elaborado em conformidade com o Roteiro Básico para Sistemas Rodoviários Estaduais - SREs, da Diretoria de Planejamento - DP do Departamento Nacional de Infraestrutura de Transportes - DNIT, publicado em outubro de 2006, bem como com a portaria do Ministério dos Transportes - MT, de 19 de setembro de 2006.

2.4.1 Divisão das rodovias em trechos

A Rede Rodoviária do Distrito Federal é dividida em trechos rodoviários. Cada rodovia possui suas subdivisões de acordo com a característica da mesma.

Não existe descontinuidade nessas subdivisões, de acordo com o que foi explicitado pelo Sistema Rodoviário do Distrito Federal(2012) (2012) :"Este procedimento visa evitar a possibilidade de serem omitidos da listagem os trechos de rodovias distritais coincidentes com rodovias federais planejadas, e os trechos coincidentes de rodovias distritais. Os trechos de rodovias distritais existentes que coincidem com rodovias federais planejadas constam da listagem de rodovias estaduais, evitando, dessa forma, interrupção na seqüência da rodovia distrital".

2.4.2 Tráfego médio diário (TMD) por trecho

O tráfego médio diário por trecho é a estimativa do número de veículos (leves, médios e pesados) que passam em determinado trecho rodoviário. A estimativa de tráfego é feita através dos dados de equipamentos de fiscalização eletrônica, seja do tipo I (Barreira) ou do tipo II (Pardal).

A contagem de volume de tráfego desconsidera os meses atípicos (janeiro, fevereiro, junho, julho, novembro e dezembro) e os dias atípicos (segunda e sexta-feira).

Para as rodovias de mão dupla, o tráfego médio diário será dado pela soma dos tráfegos dos dois sentidos, caso exista apenas um equipamento de fiscalização eletrônica, o valor do volume médio diário será multiplicado por 2 (dois).

2.4.3 Trechos rodoviários críticos

O Sistema Rodoviário do Distrito Federal divide as rodovias distritais (sob circunscrição do Departamento de Estradas de Rodagem do Distrito Federal) em trechos rodoviários. Os seguintes critérios são utilizados para criação de trechos ao longo de rodovias:

- Pontos com grande adensamento populacional
- Entroncamentos com rodovias federais e distritais e vicinais existentes
- Locais com alteração de situação física da superfície de rolamento
- Divisa do Distrito Federal com outros estados vizinhos

Portanto, trechos rodoviários são extensões que subdividem determinada rodovia, possibilitando a criação de regiões geográficas para tratamento específico.

Assim sendo, estatísticas de acidentes por rodovia nem sempre têm resultados práticos, sendo pouco eficazes na determinação de medidas que visem a redução do número de acidentes de trânsito. Um ponto crítico pode ser fruto de projeto de engenharia viária falha, más condições do veículo, excesso de velocidade por parte do condutor, entre outros. Uma saída é a descrição de acidentes de trânsito por regiões geográficas menores, onde ao invés da determinação de rodovias críticas, fossem encontrados trechos rodoviários críticos.

2.4.4 Análise de dados de acidentes por trecho rodoviário

Com o banco de dados de acidentes de trânsito e do sistema rodoviário possuem fontes de dados distintas, formando assim bancos distintos. Com base na criação de uma variável de ligação é possível associar informações conjuntas de dois ou mais banco de dados.

O banco do sistema rodoviário tem uma variável denominada "Código do Trecho", formada por 10 (dez) dígitos, onde os três primeiros indicam o número da rodovia, o quarto define se a rodovia é distrital ou federal, o quinto e sexto dígitos identificam a unidade da federação à qual pertence a rodovia, os últimos 4 (quatro) dígitos definem o número do trecho.

A variável auxiliar "Código do Trecho" definida no parágrafo acima será criada no banco de dados de acidentes. Tal variável será formada com base nas informações presentes no próprio banco de acidentes:

- Os três primeiros dígitos indicam o número da rodovia, possuem corresponde no banco de acidentes, na variável "END_VIA" (exemplo: DF 004).
- O quarto, quinto e sexto dígitos serão sempre os mesmos: DDF, já que todas as rodovias analisadas pertencem à rodovia distrital e unidade de federação Distrito Federal.
- O sétimo dígito será sempre o zero.
- Oitavo, nono e décimo dígitos: número do trecho rodoviário. No caso em que o trecho só possuir dois algarismos, o terceiro dígito à esquerda será o número 0 (zero).

Assim sendo, a variável "Código do Trecho´´ será comum aos dois banco de dados, permitindo a junção e a utilização das informações dos mesmos. Os acidentes de trânsito por trecho rodoviário podem ser estudados de acordo com as seguintes perspectivas:

- Classificação dos trechos por ordem decrescente de periculosidade

  A última análise classifica, em ordem decrescente, os trechos com maior número de acidentes de trânsito. Todavia, os trechos rodoviários possuem tamanhos distintos, sendo assim, quanto maior o trecho, maior a chance de ocorrência de um acidente. Assim sendo, o número de acidentes num determinado trecho não define o grau de periculosidade do mesmo, já que um trecho com maior extensão possui, naturalmente, maior chance de ocorrência do evento acidente de trânsito. Outro fator ou covariável que pode aumentar o número de acidentes de trânsito

é o tráfego médio diário, já que se espera que em rodovias com maior volume de tráfego, haja maior ocorrência de acidentes. Assim sendo, desconta-se o efeito das covariáveis extensão e tráfego médio diário do trecho. Logo, fez-se o uso de índices compostos para classificação de periculosidade.

A classificação da periculosidade será feita de acordo com os seguintes índices:

1. $I = \frac{\text{Acidentes com vítimas fatais no trecho}\times 10}{\text{Extensão do trecho (em quilômetros)}}$, índice de periculosidade definido com o diferencial de que será considerado a extensão do trecho rodoviário e não a da rodovia inteira.

2. $I = \frac{\text{Acidentes com feridos no trecho}\times 10}{\text{Extensão do trecho (em quilômetros)}}$, índice que considera os acidentes com feridos.

3. $I = \frac{\text{Número de acidentes com morte}}{\text{Número de acidentes com ferido}}$, índice que considera a razão do número de acidentes com morte pelo número de acidentes com ferido, ou seja, nos casos em que o índice é maior do que 1 (um), indica que houveram mais acidentes fatais na rodovia do que acidentes com feridos. Quanto mais o índice se distanciar do valor 1 (um), no sentido de ser maior, maior a periculosidade da rodovia.

4. $I = \frac{\text{Acidentes com vítimas fatais no trecho}\times 10000}{\text{Extensão do trecho (em quilômetros)}\times\text{Veículos por dia}}$, índice que acrescenta o efeito do volume de tráfego.

O primeiro índice dá peso maior à severidade de um acidente, considerando apenas aqueles que resultaram em óbito de alguma vítima. Uma falha desse índice é que alguns trechos certamente não conterão ocorrência de acidentes de trânsito fatal, dado que são pouco extensos ou pela engenharia da via não permitir grandes velocidades.

O segundo índice considera os acidentes com feridos. Todavia, tal índice não considera o efeito da covariável tráfego.

O quarto índice é o mais completo, por considerar as covariáveis extensão e

tráfego do trecho. Dado que o objetivo do índice é retratar os trechos rodoviários mais problemáticos, desconta-se do índice os fatores extensão e tráfego, já que rodovias com um maior tráfego e extensão, tendem, naturalmente, a apresentarem mais acidentes.

Para cada índice calculado é feita uma classificação em ordem decrescente de periculosidade - quanto maior o índice, maior o grau de severidade dos trechos rodoviários listados no Sistema Rodoviário do Distrito Federal.

- Classificação dos trechos periculosos coincidentes

  Este tipo de classificação leva em conta os trechos rodoviários classificados em dois ou mais índices. Cada índice retrata ou leva em consideração algum tipo de característica. Tais trechos são considerados críticos para mais de um critério estabelecido.

Mesmo com as análises descritas na seção anterior, a detecção de regiões com elevado número de acidentes apresenta alguns problemas:

- Os índices são calculados por trecho rodoviário, não sendo possível o cálculo por regiões geográficas que englobem mais de um trecho.

- Os índices são meramente descritivos, não se podendo fazer inferência dos mesmos.

- Índices que levam em conta a extensão do trecho rodoviário são falhos quando tais trechos são muito pequenos. Qualquer acidente fatal que ocorra dentro de trechos com extensão pequena levará a valores de índices inflacionados.

- Em termos estatísticos, os índices das diversas áreas não são comparáveis já que possuem variâncias muito diferentes.

- Os índices não levam em conta o fator de aleatoriedade: suponha que num trecho de 2,5 quilômetros ocorram 2 acidentes fatais, quando apenas um era esperado. Agora suponha outro trecho com 10 quilômetros onde são esperados 4 acidentes fatais. Assuma agora que tenham ocorrido 8 acidentes fatais no segundo trecho. A análise nos moldes da seção 2.4.4, considera os dois trechos

como hierarquicamente iguais, quando na verdade o primeiro trecho descrito é menos problemático, dado que os 2 acidentes fatais podem ocorrer por mero acaso com maior probabilidade.

- Quando o valor esperado de acidentes para cada trecho rodoviário tem como critério o tráfego médio diário, os trechos com pouco tráfego são bastante afetados pela aleatoriedade.

Uma solução para encontrar regiões geográficas com incidência significativa de acidentes é a utilização de métodos de estatística espacial. Assim, é necessário um estudo bibliográfico e definição de uma metodologia adequada para o problema.

Capítulo 3

Revisão bibliográfica

Neste capítulo há uma breve revisão bibliográfica acerca dos métodos de detecção e identificação de conglomerados espaciais. Em cada um deles supomos conhecida a população em risco, os casos e suas localizações geográficas, em alguma resolução.

A população em risco são os elementos passíveis de sofrerem o evento de interesse. O sistema de coordenadas geográficas são linhas imaginárias conhecidas como paralelos e meridianos. Para localização de um ponto no globo terrestre são estipuladas coordenadas de latitude e longitude. Latitudes variam de 0 a 90 graus, para o norte ou para o sul. Longitudes variam de 0 a 180 graus, para leste ou oeste.

O objetivo principal dos métodos de scan em estatística espacial é encontrar áreas no mapa em que a incidência de casos é significativamente maior do que no restante do mapa.

3.1 Mapas Baseados em Probabilidades - Choynowski (1959)

Ao estudar a distribuição de tumores cerebrais em uma parte da Polônia, Choynowski (1959) empregou um modelo de mapa estatístico que nunca tinha sido utilizado, que segundo o mesmo, é muito útil para uma série de aplicações. O método mostra a distribuição geográfica de certo fenômeno em termos das probabilidades dos desvios observados em relação à média. Esta probabilidade é calculada sobre a suposição de que a distribuição geográfica é uniforme. Os mapas que até então eram

utilizados, baseavam-se na distribuição de algum fenômeno por uma área geográfica em termos de frequências absolutas ou percentuais.

Choynowski (1959) construiu um mapa com as taxas de tumores cerebrais em sessenta municípios ao sul da Polônia, tendo observado que algumas taxas eram muito altas ou muito baixas, comparadas com a média de toda a região. Estudando as discrepâncias nas taxas de tumores cerebrais, Choynowski (1959) notou que entre os municípios que possuíam altas taxas, todos tinham população pequena. Além disso, esses municípios se desviavam sem justificativa plausível, ou seja, não apresentavam diferença na qualidade dos cuidados médicos, na composição etária, ou em qualquer outra covariável. Assim sendo, uma pequena diferença na frequência absoluta criava uma diferença substancial nas taxas, podendo ser consequência de uma variação amostral.

Uma saída encontrada por Choynowski (1959) foi a construção de mapas que mostram a probabilidade desses incidentes sob a hipótese de que a incidência é a mesma para toda a área. Sob a hipótese de homogeneidade das taxas de tumores ao longo de toda a área, quanto maior o tamanho da população, maior deverá ser a taxa de incidência de tumores. Portanto, a taxa de tumor é proporcional ao tamanho da população de determinada região. O número de casos de tumor em cada região pode ser modelado pela distribuição de Poisson. Seja C o número total de casos na região inteira, o número esperado de casos para determinada área i será igual a Cn_i/N, onde n_i é a população da i-ésima área, N é a população total e i=1,2,....R, onde R é o número de áreas ou regiões. Seja C_i o número de casos na i-ésima região. Então:

$$C_i \sim Poisson(\mu_i), \text{onde } \mu_i = Cn_i/N$$

Como o número de casos de tumor em cada região possui distribuição de Poisson, o método é utilizado para calcular probabilidades para cada região ou área. Ou seja, o método não apresenta um teste de significância global. Uma saída no caso de um teste que busque detectar um conglomerado em toda a área é a utilização do método de Bonferroni, que garante um nível de significância global para o teste.

3.2 Agrupamento de pontos aleatórios em duas dimensões - Naus (1965a)

Este artigo apresenta uma extensão ao primeiro artigo de Naus (1965b), descrevendo a situação para o caso bidimensional, onde são traçadas N coordenadas $(x_1, y_1), ..., (x_N, y_N)$ aleatórias do quadrado unitário. Naus (1965a) define o seguinte evento: existe um sub-retângulo do quadrado unitário, com lados de comprimentos u e v, onde tais lados são paralelos aos lados do quadrado, que contém pelo menos n dos N pontos. O método persiste em varrer o quadrado unitário com um sub-retângulo, r, com lados de comprimento u e v orientados paralelamente aos lados do quadrado. Procura-se a probabilidade $P(n|N; u; v)$ deste evento. Se ao invés de definirmos u e v (eixos x e y), fixarmos u ou v, o problema é equivalente ao problema unidimensional.

O artigo de Naus (1965a) limita-se a derivar os limites inferior e superior para $P(n|N; u; v)$ de ocorrência do evento. No caso em que $n = N$, $u \geq 0$ e $v \leq 1$, a probabilidade $P(n|N; u; v)$ é dada por:

$$P(N|N; u; v) = N^2 A^{N-1} + (N-1)^2 A^N - N(N-1)A^{N-1}(u+v), \qquad (3.1)$$

onde $A = uv$.

A equação mostra que $P(N|N; u; v)$ é função da área do retângulo de scan, A, bem como função da sua forma.

A primeira aplicação apresentada no artigo trata-se de um caso de scan ou detecção retangular. Considere uma área quadrada ou retangular fixa, no qual deseja-se encontrar agrupamentos. Qual a configuração da janela de detecção que dá a maior probabilidade de encontrar um conglomerado maior? Outro questionamento é se o formato do conglomerado tem algum efeito sobre o número esperado de agrupamentos ou sobre a verossimilhança. O artigo de Naus (1965a) coloca um exemplo em que, para ambos os casos, a área de scan, A, é a mesma. Entretanto, o quadrado leva a uma maior probabilidade de agrupamento. A partir da Equação 3.2, um problema de maximização de $P(N|N; u; v)$ implica em minimizar $(u + v)$ sujeitos a $uv = A$. O retângulo com determinada área e com menor perímetro é o quadrado. Portanto, dentre as janelas de detecção retangular, a que possui a maior probabilidade de agrupamento

é a com formato quadrado.

Na segunda aplicação, cinco navios estão de forma aleatória e independente, distribuídos num quadrado de 10 graus de latitude e longitude do oceano. Qual é a probabilidade de que pelo menos quatro navios estejam dentro do quadrado com dois graus de longitude e três graus de latitude? O problema então se resume a encontrar a $P(4|5; 0.2; 0.3)$. Aplicando as inequações e resultados presentes no artigo de Naus (1965a), encontram-se os limites inferior e superior para $P(4|5; 0.2; 0.3) : 0.0086 \leq P(4|5; 0.2; 0.3) \leq 0.0278$. De acordo com o teorema de convergência, o autor afirma que a melhor estimativa é dada pelo limite inferior.

3.3 Distribuição do Tamanho do Máximo de conglomerados em uma linha - Naus (1965b)

Partindo da idéia de Choynowski (1959), o qual modelou determinado fenômeno em termos de probabilidade, não mais utilizando-se de percentuais ou frequências absolutas. Naus (1965b) trabalha em termos da distribuição de probabilidade de pontos desenhados de forma independente. Dados N pontos independentes e com distribuição uniforme entre $(0, 1)$, então $P(n|N; p)$ é a probabilidade de que o número de pontos aglomerados num intervalo de comprimento p, seja maior ou igual a n. O artigo de Naus (1965b) utiliza abordagem combinatória para encontrar $P(n|N; p)$, onde $n > N/2$ e $p \leq 1/2$. Note que $P(n|N; p)$ é um polinômio em p de ordem N.

A aplicação da técnica no contexto espacial fica limitada, dado que a probabilidade é calculada em apenas uma dimensão. Uma saída seria somar os casos por longitude ou latitude, considerando apenas um par de coordenadas geográficas e trabalhando com os dados unidimensionais. Assim sendo, assume-se que a distribuição dos dados observados é uniforme e calcula-se a probabilidade de ocorrência dos mesmos. A partir disso é possível julgar se os dados se distribuem de maneira aleatória e uniforme.

Naquele trabalho foram apresentadas duas aplicações: a primeira sobre chamadas telefônicas e a segunda sobre processo de contagem. A primeira relata sobre discagens realizadas por quinze telefones em horários distribuídos aleatoriamente ao longo de um período de um minuto. O tempo de discagem para uma chamada é de dez segundos

e deseja-se encontrar a probabilidade de que oito ou mais chamadas estarem sendo discadas ao mesmo tempo.

O segundo problema trata-se de um processo de Poisson com taxa média λ, que gera impulsos que são recebidos por um contador. O contador registra os n impulsos conjuntos em um intervalo de comprimento inferior a t. Desejamos encontrar a distribuição de tempo de espera até o contador realizar o primeiro registro. Seja a probabilidade $F_n(T;t|\lambda)$ do tempo de espera até que o primeiro registro do contador seja inferior ou igual a T. Seja N o número total de pontos em $(0,T)$, N tem distribuição de Poisson com média igual a λ.

Seja $F_n(T;t|\lambda,N)$ a probabilidade $F_n(T;t|\lambda)$ condicionada a N fixo. $F_n(T;t|\lambda,N)$ representa a probabilidade de que n ou mais dos N impulsos apareçam num subintervalo de T de comprimento t. A distribuição de Poisson condicionada ao número total N, fixo, tem distribuição uniforme. Assim sendo, o problema se resume a probabilidade condicional sobre a distribuição de N para encontrar:

$$F_n(T,t|\lambda) = \sum_{N=n}^{\infty} P(n|N;t/T)\frac{exp^{-\lambda T}(\lambda T)^N}{N!} \tag{3.2}$$

3.4 Estimador de Bayes Empírico para o Risco Relativo - Clayton e Kaldor (1987)

Um dos interesses da estatística espacial é o de mapear a incidência de algum fenômeno de interesse. No artigo de Clayton & Kaldor (1987), o evento de interesse é a mortalidade pelo câncer. O mapeamento do evento de interesse geralmente é feito por medida de risco relativo por região geográfica, denominada taxa de mortalidade geral padronizada, ou por teste estatístico que verifique a significância da diferença entre as taxas de cada região. Segundo Clayton & Kaldor (1987), nenhuma proposta é inteiramente satisfatória, propondo a construção de um estimador de Bayes empírico para o risco relativo.

Clayton & Kaldor (1987) comentam que ambas as abordagens podem deturpar a distribuição geográfica de casos de câncer, afirmando que no primeiro caso, a variação do tamanho da população não é levada em conta, obtendo estimativa imprecisa da

taxa de mortalidade geral padronizada. Por outro lado, o mapeamento com utilização de teste de significância é falho no caso em que duas regiões com taxas de mortalidade idênticas possuem populações distintas.

Seja uma região dividida em N distritos mutuamente exclusivos. Seja θ_i o risco relativo para o i-ésimo distrito a ser estimado, considere que C_i - número de casos observados dado θ_i - tenha distribuição de Poisson com média $\theta_i\mu_i$, onde μ_i é o número esperado de casos. Desta forma, a distribuição marginal de C_i permite estimar os parâmetros da distribuição conjunta dos θ_i. A esperança a posteriori de θ_i dado C_i dá a estimativa do risco relativo. O artigo apresenta três modelos mistos para a distribuição dos riscos relativos, $f(\theta)$: distribuição Gama, Log-Normal e método não paramétrico.

O modelo utilizando a distribuição Gama considera θ_i, iid, como binomial negativa e as estimativas de Bayes empíricos são calculadas de forma analítica. O modelo Log-Normal exige algumas aproximações, todavia, possibilita trabalhar com a suposição de que o logaritmo dos riscos relativos são correlacionados. O modelo não paramétrico estima a função de verossimilhança utilizando o algoritmo EM, que é uma ferramenta computacional iterativa.

Clayton & Kaldor (1987) apresentam um exemplo de câncer de lábio, na Escócia. Foram listados os casos observados e esperados de câncer de lábio entre 1975 e 1980 em cada um dos 56 condados da Escócia. A taxa de mortalidade geral padronizada (SMR, na sigla em inglês) é comparada com as estimativas de Bayes empírico dos quatro modelos apresentados no estudo. O modelo Gama e Log-Normal apresentam valores bastante próximos, com amplitude menor que o SMR. A estimativa de Bayes empírica do método não paramétrico parece diferir pouco dos dois métodos paramétricos (Gama e Log-Normal). O artigo conclui afirmando que as estimações apresentadas funcionam como modelos de alisamento do SMR e que no caso dos mapas de doenças com áreas muito pequenas é necessário o pressuposto de correlação espacial.

3.5 Um Teste para a detecção de conglomerados de Doenças - Whittemore et al. (1987)

Whittemore *et al.* (1987) definem um teste para detectar conglomerados de incidências de carcinoma de células escamosas do reto e do ânus. O teste é ilustrado em 63 casos durante o período entre 1973 e 1981. A estatística do teste é a distância média entre todos os pares de incidência da doença:

$$\delta = \frac{\sum_{i<j} \Delta(i,j)}{\begin{pmatrix} n \\ 2 \end{pmatrix}},$$

onde $\Delta(i,j)$ é a distância entre o i-ésimo e o j-ésimo evento de interesse. A normalidade assintótica, o conhecimento da média e da variância de δ sob hipótese nula são informações que tornam possíveis o cálculo da estatística do teste e do respectivo p-valor.

Whittemore *et al.* (1987) alertam sobre a dificuldade de definir conglomerados em doenças comuns, onde os conjuntos podem ser produzidos por fatores não relacionados com o processo da doença, seja na variação da distribuição da população em geral ou nos subgrupos demográficos. Whittemore *et al.* (1987) salientam que testes anteriores de scan não foram satisfatórios para o estudo de doenças crônicas como o câncer, dado que são testes projetados para determinar agrupamento espaço-temporais, simultaneamente.

Diferentemente de Choynowski (1959) que estabelece uma probabilidade para cada área, Whittemore *et al.* (1987) propõem um método para responder à questão da existência de *conglomerado* espacial, não identificando sua localização. A hipótese nula é a de que todos os membros da populacão possuem a mesma probabilidade de contrair a doença.

No exemplo apresentado no artigo, foram examinados agrupamentos por setor censitário, de 63 casos de carcinoma de células escamosas do reto e do ânus, em São Francisco, entre 1973 e 1981. Foram calculados o valor esperado e a variância de δ e a distância média entre os pares de casos. Os resultados utilizando aproximação para a distribuição *Gaussiana* (resultado assintótico) é comparado com valores em-

píricos gerados computacionalmente. Os resultados teóricos e empíricos estão em boa concordância.

3.6 Máquina de Análise Geográfica - Openshaw et al. (1988)

Diferentemente do artigo anterior de Whittemore *et al.* (1987), que apenas identificava a existência de conglomerados, o estudo de Openshaw *et al.* (1988) apresenta um procedimento denominado Máquina de Análise Geográfica (GAM, na sigla em inglês), que permite a localização espacial de um possível conglomerado.

Em 1983, houve uma preocupação de órgãos públicos com um aumento na taxa de leucemia em jovens que vivam próximo a uma usina nuclear, no norte da Inglaterra. Apesar de concluir que existia um excesso de casos da doença, não conseguiram relacioná-las com fatores biológicos devido a exposição à radiação. O artigo de Openshaw *et al.* (1988) busca uma análise mais aprofundada dos casos de leucemia, sobre uma área mais ampla. Desde então há relatos de aumento na taxa de leucemia infantil em áreas próximas a usinas nucleares. As técnicas utilizadas nesses estudos são variadas: comparação do valor observado com o valor esperado, utilização da distribuição de Poisson; alguns olharam para áreas administrativas (por exemplo, círculos eleitorais, paróquias, etc); enquanto outros desenharam círculos com raios arbitrários em torno de um ponto de origem. Alguns incluíram a idade do paciente e com diferentes períodos de duração do estudo. Portanto, os estudos apresentaram resultados distintos.

Openshaw *et al.* (1988) desenvolveram um método que independe de qualquer hipótese pré-estabelecida, sendo testado em dados relativos a crianças com leucemia, no norte da Inglaterra. Entre 1968 e 1985, 853 crianças foram diagnosticadas com Leucemia Linfoblástica Aguda (ALL, sigla em inglês), antes de completarem 15 anos. A população total de crianças da área relativa ao estudo era de 1.544.963. De todos os 812.993 círculos examinados, 1792 foram considerados significativos. Estes se distribuem basicamente em volta de 5 áreas. O código postal do endereço de cada criança e o número de crianças com menos de 15 anos eram conhecidos.

A análise foi feita utilizando uma Máquina de Análise Geográfica (GAM, sigla em inglês), que desenhou vários círculos sobrepostos de tamanhos distintos, espaçados regularmente e que cobrem a totalidade do estudo. A construção do teste exige a definição do raio do círculo r que será utilizado na busca do conglomerado. Ao longo do procedimento há variação no valor do raio pré-definido. É necessário sobrepor uma malha quadriculada sobre o mapa, onde cada um dos vértices será o centro de um círculo de raio r e para cada um dos círculos calcula-se o número de casos no interior de cada círculo. O número de casos no interior de cada círculo é comparado com o valor do 99,8° percentil da distribuição de probabilidade do número de casos sob a hipótese nula. No artigo de Openshaw *et al.* (1988), o valor do percentil sob hipótese nula é encontrado pelo método de Monte Carlo, exigindo muito de técnicas computacionais. Também é possível o cálculo do quantil de ordem 99,8 por meio da distribuição de Poisson, com valor esperado igual a Cn_i/N. Para cada valor do raio r pré-fixado, repete-se toda a análise. Assim sendo, o resultado final são vários círculos de diferentes tamanhos desenhados no mapa. Todos os círculos são individualmente significativos.

A aplicação da técnica possui um grande apelo visual, onde quanto maior a densidade de círculos sob determinada região, aumenta-se a intensidade do sombreamento. Embora cada círculo possa ser julgado individualmente, a significância para todos os círculos simultaneamente não é conhecida. A razão é que são testes simultâneos não independentes, sendo difícil realizar um imenso número de testes.

As vantagens deste método é que se trata de um método simples de entender, com grande apelo visual. As desvantagens é que se trata de um método exploratório e não inferencial devido ao problema de muitos testes simultâneos e dependentes, exige muito computacionalmente, os círculos não são comparáveis entre si, dado que as variáveis envolvidas possuem diferentes distribuições.

3.7 A detecção de conglomerados de Doenças Raras - Besag e Newell (1991)

Os testes de agrupamento de doenças raras investiga se um padrão observado de casos, em uma ou mais regiões, originou-se do acaso. O principal objetivo do artigo de Besag & Newell (1991) é identificar pequenos conglomerados de doenças, tendo como critério de parada o número de casos. O objetivo secundário é discutir algumas armadilhas na aplicação de testes de agrupamento em dados epidemiológicos. Besag & Newell (1991) comentam que existem dois tipos de testes: um geral e um focado. No geral há uma preocupação com o padrão global de uma doença ao longo de regiões grandes. Em contraste, os testes focados concentram-se em uma ou mais regiões menores selecionadas ostensivamente por algum fator associado a doença ou ao caso investigado. Nos testes gerais, a teoria estatística utilizada muitas vezes é inadequada, ao passo que em testes com foco, muitas vezes há enviesamento de seleção.

Na Seção 3 do artigo, há preocupação com o problema de detecção de conglomerado sob uma grande região, utilizando-se de testes de significâncias múltiplas. Em particular é proposto um método com os mesmos objetivos básicos de Openshaw *et al.* (1988), porém, computacionalmente menos exigente. Na Seção 4 é apresentado um exemplo com dados de crianças com Leucemia, no norte da Inglaterra, entre 1975 e 1985. Na Seção 5 são discutidas algumas limitações e modificações da técnica.

A intenção de Besag & Newell (1991) é detectar possíveis conglomerados de uma doença rara numa região geográfica extensa subdividida em pequenas zonas utilizando-se das coordenadas dos seus centróides. A hipótese nula é a de que o número observado de casos distribui-se totalmente ao acaso, ou seja, para qualquer caso particular, a probabilidade na zona i é igual a t_i/t_+, onde t_i é a população na i-ésima zona e t_+ é a população total. Primeiramente considere a região em que ocorre o caso como A_0 e as outras regiões como $A_1, A_2, ...$, determinado pelo acréscimo na distância dos seus centróides ao centróide da região A_0. Desta forma define-se:

$$D_i = \left(\sum_{j=0}^{i} y_j \right) - 1,$$

$$u_i = \left(\sum_{j=0}^{i} t_j \right) - 1,$$

Onde y_i é o número de casos na i-ésima zona e t_j é a população na j-ésima zona. De tal forma que $D_0 \leq D_1 \leq ...$ são os casos acumulados em $A_0, A_1, ...$ e $u_0 \leq u_1 \leq ...$ são as populações acumuladas correspondentes. Define-se, então, $M = min\{i : D_i \geq k\}$. Um pequeno valor observado de M indica um conglomerado em volta de A_0. Formalmente, sob H_0, se γ é o valor observado de M, o nível de significância do teste é $P(M \leq \gamma)$.

Utilizando aproximação da distribuição hipergeométrica pela distribuição de Poisson, então, a probabilidade de observar k indivíduos entre u_γ com a doença é dada por:

$$P(M \leq \gamma) = 1 - P(M < \gamma)$$

$$= 1 - \sum_{s=0}^{k-1} \frac{e^{-(u_\gamma p)} (u_\gamma p)^s}{S!}. \tag{3.3}$$

Portanto, fixando-se um valor de α, identificar todos os conglomerados que atingiram valor de significância menor ou igual a α. Há ainda uma maneira de avaliar se a quantidade de conglomerados encontrada é significativa, algo similar ao teste de Whittemore *et al.* (1987) - teste global sobre a existência de conglomerados -.

O exemplo apresenta dados de Leucemia Linfoblástica Aguda diagnosticada em crianças de 0 a 14 anos, entre 1975 e 1985. Foram incluídas variáveis de sexo, idade e o código postal do endereço para cada caso, formando assim o sistema de referências.

Conjunto de testes para detecção de conglomerados foram realizadas para k=2, 4, 6 e 8. Todavia, só foi apresentado o caso para k=4. O número de círculos significativos para α=0,05 é de 23. Tal resultado foi comparado com o número de conglomerados esperado e a verificação da significância de cada conglomerado e da região como um todo pelo método de Monte Carlo.

As vantagens do teste de Besag & Newell (1991) é que o mesmo estabiliza mais as estatísticas de testes locais, sendo visualmente agradável, e identifica os conglomerados, assim como em Openshaw *et al.* (1988). As desvantagens é que continua sendo um método exploratório e continua o problema de testes simultâneos.

3.8 Estatística *Scan* Circular

Todos os métodos de detecção de conglomerado apresentados anteriormente neste trabalho consideram a hipótese nula de que todas as pessoas possuem a mesma probabilidade de se tornarem um caso, ou seja, contrair uma doença, sofrer um acidente de trânsito, entre outros. Considerando que sob a região total não haja conglomerado, todos os indivíduos pertencentes à população possuem probabilidades iguais e não nulas. Além disso, a distribuição de probabilidade do número de casos é Poisson, o que é comumente utilizado por se tratar de processo de varredura.

A principal diferença entre eles está no processo de varredura, na definição do centro de cada círculo, que poderá ser feito através de uma malha (Openshaw *et al.* , 1988), ou utilizando-se de centróides. Uma vez definido o centro, pode-se proceder das seguintes formas:

- Como em Openshaw *et al.* (1988), que fixam o raio e varrem a área.
- Como em Besag & Newell (1991), que fixam o número de casos.

Todos os métodos fixam o seu raio de busca, onde os métodos apresentam definições distintas para o raio de janela. Openshaw *et al.* (1988) e Besag & Newell (1991) constroem seus métodos através de testes múltiplos, mas Besag & Newell (1991) definem um teste de modo a responder sob a existência global de um conglomerado.

Openshaw *et al.* (1988) e Besag & Newell (1991) utilizaram-se do método de detecção de conglomerados espaciais denominado Máquina de Análise Geográfica (GAM, na sigla em inglês), enquanto que Kulldorff utiliza-se do método de varredura circular.

3.9 Conglomerado espacial de doença: detecção e inferência - Kulldorff e Nagarwalla (1995)

Kulldorff & Nagarwalla (1995) apresentam um método de detecção e inferência para conglomerados espaciais aplicado à epidemiologia. A estatística do teste tem como base o teste da razão de verossimilhança. O teste pode detectar conglomerados de qualquer tamanho, em qualquer região estudada. Além disso, não está restrito a

conglomerados que estejam com fronteiras administrativas ou políticas pré-definidas. O teste pode ser aplicado a dados de área ou quando são conhecidas as coordenadas geográficas de cada evento individual. O artigo de Kulldorff & Nagarwalla (1995) trabalha em um conjunto de dados de ocorrência de leucemia no interior de Nova York.

Há duas abordagens principais utilizadas para análise de padrão espacial. Uma abordagem utiliza um teste estatístico baseado na medição de distâncias entre os casos de doenças, enquanto a outra se baseia em estudar a variabilidade do número de casos em certos subgrupos da região. O método de Whittemore *et al.* (1987) é um exemplo da primeira abordagem, enquanto que Choynowski (1959) é um exemplo da segunda abordagem.

O método baseado na medição de distâncias entre os eventos de interesse, como em Whittemore *et al.* (1987), é útil em aplicações onde a localização do conglomerado não é de interesse. Quando se está interessado na localização do conglomerado, os métodos descritivos de Openshaw *et al.* (1988) e Besag & Newell (1991) são mais apropriados para detecção de conglomerados, com a construção de um grande número de círculos sobrepostos e aplicação de um teste de significância para cada um destes círculos individualmente. Este método não se resume a testes de significância individuais, dado que os conglomerados são correlacionados e o procedimento de Bonferroni não resolve o problema dos testes múltiplos.

O método proposto por Kulldorff & Nagarwalla (1995) apresenta as seguintes características:

- Aborda diretamente o problema local da inferência de conglomerados detectados.
- Não se limita a buscar um conglomerado de tamanho pré-especificado.
- O teste se baseia na razão de verossimilhança.
- Há definição clara da hipótese alternativa, o que facilita estabelecer se o teste é apropriado para a especificidade do problema.
- A estatística do teste é única, o que torna desnecessária a realização de testes separados para cada possível localização do conglomerado ou para cada possível

tamanho do conglomerado.

- O teste aplica-se para dados agrupados (dados de área) ou não agrupados (dados pontuais).

Kulldorff & Nagarwalla (1995) consideram a região de estudo dividida em sub-regiões geográficas denominadas células. Para cada célula é necessária a coordenada geográfica do seu centróide, do número de indivíduos e de casos. O método não requer qualquer suposição sobre a distribuição da população dentro da célula, construindo círculos com raios distintos. Cada um dos diversos círculos assim construídos definem uma zona. Para dados não agregados, as zonas são perfeitamente circulares, isto é, os indivíduos numa zona são exatamente aqueles localizados no interior do círculo. Com dados agregados, uma zona pode ter limite irregular, havendo casos em que o círculo incluirá determinado centróide da região, mas não englobará todos os indivíduos pertencentes a esta zona. Em outros casos, o círculo não abrangerá o centróide da região, todavia haverá indivíduos da zona localizados dentro do círculo. No primeiro caso, mesmo os indivíduos fora do círculo serão incluídos, dado que o centróide da região a que pertencem está dentro do círculo. No segundo caso, mesmo havendo indivíduos dentro do círculo, o fato do centróide da região a que pertencem estar fora do círculo, faz com que os indivíduos sejam excluídos. De fato, no caso de dados agregados, como não conhecemos a localização exata de cada indivíduo, consideramos que as coordenadas de todos eles coincidem com as coordenadas do centróide da região.

A hipótese alternativa é implicitamente definida pela forma particular em que se constroem as zonas. Isto não significa que o método só funciona com hipótese alternativa exata, pelo contrário, ele dá uma indicação dos tipos de alternativas pelos quais o poder do teste é alto ou baixo. Como o raio de busca é crescente, os círculos irão incluir a região inteira, não sendo apropriado classificar um conglomerado nesta zona, mesmo que a taxa de incidência seja consideravelmente mais elevada do que fora do mesmo. Assim sendo, define-se um limite superior sobre o raio dos círculos, que varra no máximo 50% da população total. A escolha do critério de parada é feita a priori e não por tentativa e erro.

A formulação do teste de razão de verossimilhança será visto no Capítulo 4.

3.10 Estatística scan espacial - Kulldorff (1997)

No artigo de Kulldorff (1997) mantem-se a base do que foi apresentado em Kulldorff & Nagarwalla (1995), com uma estatística capaz de verificar a existência de conglomerados espaciais e sua localização aproximada. O artigo anterior de Kulldorff & Nagarwalla (1995) se limitava ao modelo Poisson. No artigo de Kulldorff (1997) o modelo binomial é incluído. A diferença no resultado entre as duas distribuições é reduzida quando o número de eventos é pequeno comparado ao tamanho da população. Para algumas das extensões, e dependendo da aplicação, a estatística *Scan* Circular pode ou não ser condicionada ao número total de pontos observados.

A estatística scan pode ser aplicada tanto a dados agregados numa determinada região geográfica quanto quando são conhecidas as coordenadas exatas de cada ocorrência do evento de interesse.

A hipótese nula é de que todos os indivíduos possuem a mesma probabilidade de vir a sofrer o evento de interesse. Neste caso a probabilidade dentro da zona (candidata a conglomerado) é igual a probabilidade fora da zona. Sendo assim, o número esperado de casos numa determinada área é proporcional ao tamanho da população na respectiva área.

Uma propriedade importante da estatística scan de Kulldorff é que rejeitando-se a hipótese nula, fixando-se os pontos dentro do conglomerado e independente da distribuição dos pontos fora dele, continua-se a rejeitar a hipótese nula.

O artigo apresenta uma aplicação com ocorrências de síndrome da morte súbita infantil na Carolina do Norte. A população de controle foi o número de nascimentos. Duas zonas são consideradas significativas segundo os dois modelos. Ao utilizar-se a covariável raça, um terceiro conglomerado significativo é revelado.

Kulldorff (1997) obtém a verossimilhança para o modelo Poisson e Bernoulli, afirmando que o modelo Bernoulli é mais natural para o conjunto de dados apresentado, onde cada nascimento corresponde a no máximo uma morte súbita infantil. Uma vez que o conjunto de dados se refere a doenças raras, o modelo Bernoulli pode ser utilizada pelo modelo Poisson. A aproximação pelo modelo Poisson é especialmente utilizada quando há covariáveis que se deseja incluir na análise.

O teste baseado na estatística scan espacial de Kulldorff (1997) possui as seguintes

vantagens:

- Leva em conta o tamanho da população em cada região, ou seja, em sua densidade populacional.
- Procura conglomerados sem especificar de antemão sua localização e tamanho.
- Além de dizer se existe conglomerado no mapa, o teste fornece a localização do conglomerado no mapa.
- Evita o problema de teste múltiplos, fornecendo um p-valor real.

3.11 Aplicação da estatística Scan de Kulldorff em criminologia: aglomerados espaciais de mortes violentas em uma região recém-urbanizada do Brasil: destaque para as disparidades sociais - Ruth Minamisava et al. (2009)

Tanto as mortes por homicídios quanto as mortes por acidentes de trânsito entre os jovens são um problema de saúde pública mundial. Neste contexto, Minamisava *et al.* (2009) procuram analisar a distribuição espacial e potenciais conglomerados de risco para mortes intencionais e não intencionais entre jovens de 15 a 24 anos, em Goiânia.

Foram coletados os dados de óbitos e endereços residenciais pelo Sistema de Informação sobre Mortalidade (SIM), do Ministério da Saúde, validados por visitas domiciliares. Dentro do universo de casos, o artigo de Minamisava *et al.* (2009) classificou cada morte da seguinte forma: acidentes de transporte, agressão e intervenção legal, excluindo os casos de óbitos por suicídio.

O Sistema de Informação Geográfica (SIG) foi utilizado para georreferenciar os endereços residenciais. O objetivo do estudo é identificar conglomerados de setores censitários com elevada taxa de mortalidade com a aplicação da estatística de varredura circular. Considerou-se que o número de casos de mortes possui distribuição de

Poisson.

Os resultados mostram que a maioria das mortes violentas entre os jovens aconteceram por lesões intencionais. Entre agosto de 2005 e agosto de 2006, 145 endereços de casos de óbitos intencionais e acidentes de transporte foram localizados e georreferenciados. Não foi encontrado conglomerado de óbitos por acidentes de transporte, significando que sua distribuição espacial ao longo do município de Goiânia é aleatória. Um conglomerado de alto risco para óbitos por trauma intencional foi detectado, com p-valor de 0,029, sendo a maioria, homicídios. A área de risco localiza-se na periferia do município, no Distrito Sanitário Noroeste, apresentando os piores indicadores de renda, de escolaridade e de condições sanitárias.

Minamisava *et al.* (2009) conclui que a associação entre mortes intencionais e desigualdades sociais mostra a necessidade de políticas sociais urgentes.

O estudo faz uso do software Satscan (V7.0.3) para análise espacial. A estatística *Scan* Circular define uma série de janelas circulares de diferentes raios sobre a área de estudo. O critério de parada para cada raio de círculo definido é de perfazer 50% da população em risco. Sob hipótese nula, a probabilidade de qualquer elemento dentro do conglomerado vir a morrer é igual a probabilidade fora do conglomerado, sendo assim, o número de casos dentro de cada setor censitário é proporcional ao tamanho da população. Para cada setor selecionado, calcula-se o valor do logarítmo da razão de verossimilhança, que compara o modelo sob H_0 com o modelo sob H_1. A estatística do teste é igual ao maior valor do logarítmo da razão de verossimilhança de todos os setores selecionados. A estatística do teste é comparada com o percentil da distribuição empírica gerada pela simulação de Monte Carlo e verifica-se se o setor é significativamente um conglomerado.

Capítulo 4

Metodologia

4.1 Estatística Scan de Kulldorff

O método *Scan* Circular proposto por Kulldorff (1997) consiste em uma técnica de detecção e inferência de conglomerados espaciais de algum evento de interesse. O método consiste em verificar se a ocorrência de determinado evento está distribuída de maneira aleatória em relação à população de interesse, ou se há um conjunto de regiões que pode ser considerada um conglomerado. Um conglomerado espacial é uma área que possui incidência de casos significativamente maior que a esperada sob hipótese de aleatoriedade das ocorrências. A procura por conglomerados espaciais pode ser utilizada tanto em situações em que o evento de interesse encontra-se agregado em regiões, quanto quando são conhecidas as coordenadas geográficas exatas do mesmo.

A hipótese nula é de que a distribuição dos casos é homogênea ao longo do mapa, ou seja, cada elemento da população possui a mesma probabilidade de ser um caso. Resumindo, o número esperado de casos em determinada área é proporcional ao tamanho da população na respectiva área.

Uma propriedade importante no método de Kulldorff (1997) é que ao rejeitarmos a hipótese nula, fixando-se a distribuição dos dados dentro do conglomerado mais provável, não importa a configuração dos pontos fora do conglomerado, continua-se a rejeitar a hipótese nula.

Para aplicação da técnica de Kulldorff (1997), faz-se necessário o conhecimento das seguintes informações, para cada área:

- População sob o risco do evento de interesse.
- Número de casos em cada região.
- Coordenada geográfica do centróide (ponto médio) de cada região.

4.2 Teste da razão de verossimilhança

O número de casos observados em determinada região é variável aleatória, dado que ao repetirmos o evento várias vezes, o número de casos não seria sempre o mesmo, ou poderia ser diferente. Dado que é variável aleatória, o número de casos possui distribuição de probabilidade.

Suponha que o número de casos de um evento de interesse possui distribuição Poisson, assim, o número de casos na i-ésima região tem a seguinte função de probabilidade:

$$f_i(c) = \begin{cases} \frac{e^{-\lambda_i}\lambda_i^c}{c!}, & \text{se } c \geq 0 \\ 0, & \text{caso contrário} \end{cases} \tag{4.1}$$

O número de casos na i-ésima região possui distribuição de Poisson com média $\lambda_i = pn_i$, onde p é a probabilidade de um indivíduo vir a ser um caso e n_i é a população na i-ésima região.

Sejam X e Y variáveis aleatórias independentes representando o número de casos nas regiões 1 e 2, respectivamente. Considere ainda que ambas as variáveis possuem distribuição de Poisson com parâmetros θ e λ, respectivamente. Assim:

$$X \sim Poisson(\theta)$$

$$Y \sim Poisson(\lambda)$$

O conjunto de regiões conexas será uma zona. Portanto, supondo que a zona z seja formada pelas regiões 1 e 2, seria razoável questionar qual é a distribuição de probabilidade do número de casos nesta zona z, bem como o seu parâmetro. Sabe-se

que o número de casos na zona z é dado pela soma do número de casos das regiões 1 e 2. Portanto, pretende-se encontrar a distribuição de $U = X + Y$.

$$f_{X,Y}(x, y) = \frac{\theta^x e^{-\theta}}{x!} \frac{\lambda^y e^{-\lambda}}{y!}, \text{ onde } x = 0, 1, 2, ..., y = 0, 1, 2, ...$$

Se X e Y são variáveis aleatórias independentes com funcões geradoras de momento $M_X(t)$ e $M_Y(t)$. Então a função geradora de momento da variável aleatória $Z = X + Y$ é dada por (Casella & Berger, 2002):

$$M_Z(t) = M_X(t)M_Y(t) = (Ee^{tX})(Ee^{tY}) = Ee^{t(X+Y)}.$$

Portanto, a distribuição marginal de $U = X + Y$ é dada pela seguinte expressão:

$f_U(u) = \frac{e^{-(\theta+\lambda)}}{u!}(\theta + \lambda)^u, u = 0, 1, 2, ...$

Este resultado nos leva ao Teorema 4.2.1.

Teorema 4.2.1. *Se $X \sim Poisson(\theta)$ e $Y \sim Poisson(\lambda)$, X e Y são independentes, então $X + Y \sim Poisson(\theta + \lambda)$.*

A partir disso, seja C_i o número de casos na i-ésima região e z uma zona, isto é, z é um conjunto de regiões. Então, o número de casos C_z na zona z é dado por:

$$C_z = \sum_{i \in z} C_i \sim Poisson\left(\sum_{i \in z} \lambda_i\right). \tag{4.2}$$

C_z também segue distribuição de Poisson com parâmetro igual a soma dos λ_i's.

Uma zona z possuirá o número de casos C_z e uma população n_z. Seja p a probabilidade de que um indivíduo pertencente à zona z venha a ser um caso. A média de casos na zona z, λ_z, será proporcional ao tamanho da população n_z e será dada pelo somatório das médias nas i regiões englobadas pela zona, isto é, $\sum_{i \in z} \lambda_i = \lambda_z = \sum_{i \in z} pn_i = p \sum_{i \in z} n_i = pn_z$.

A função de probabilidade do número de casos na zona z, com distribuição de Poisson, será dada por:

$$f_Z(c_z) = \begin{cases} \frac{e^{-\lambda_z}\lambda_z^{c_z}}{c_z!}, & \text{se } c_z \geq 0 \\ 0, & \text{caso contrário} \end{cases} \tag{4.3}$$

Sejam N e C a população e o número total de casos, respectivamente. Considere $n_{\overline{z}}$ e $c_{\overline{z}}$ como a população e o número de casos fora da zona z, respectivamente. Portanto, $n_{\overline{z}} = N - n_z$ e $c_{\overline{z}} = C - c_z$.

Para que exista um conglomerado no mapa, a probabilidade de que ocorra um caso dentro dessa zona deve ser significativamente maior do que fora da zona. Se a probabilidade de ocorrência de um caso for a mesma, independente de estar dentro ou fora da zona, então não haverá conglomerado. Uma maneira de descobrir se existe ou não um conglomerado no banco de dados é através de um teste de hipóteses. O teste sobre a existência de um conglomerado é definido com as seguintes hipóteses:

$$\begin{cases} H_0 : p = q \\ H_1 : p > q \end{cases}$$

onde, p é a probabilidade de ocorrer um caso dentro da zona e q a probabilidade de ocorrer um caso fora da zona. Caso a hipótese nula seja rejeitada, então existe uma zona z tal que a probabilidade de ocorrer um caso dentro da zona será significativamente maior do que a probabilidade de ocorrer um caso fora da zona, portanto, a zona z é significativamente um conglomerado.

A formulação do teste da razão de verossimilhança compara o modelo sob H_0 com o modelo sob H_1. A obtenção da estatística do teste é descrita a seguir:

Sob H_0, $\lambda_z = pn_z$ e $\lambda_{\overline{z}} = p(N - n_{\overline{z}})$. Logo,

$$L_0(z;p) = \frac{\lambda_z^{c_z} e^{-\lambda_z}}{c_z!} \frac{\lambda_{\overline{z}}^{c_{\overline{z}}} e^{-\lambda_{\overline{z}}}}{c_{\overline{z}}!}$$

$$= \frac{(pn_z)^{c_z} e^{-pn_z}}{c_z!} \frac{(p(N-n_z))^{(C-c_z)} e^{-p(N-n_z)}}{(C-c_z)!}$$

Utilizando $l_0(z;p) = \log L_0(z;p)$,

$$l_0(z;p) = c_z[\log p + \log n_z] - pn_z - \log c_z! + (C - c_z)[\log p + \log(N - n_z)]$$

$$-p(N - n_z) - \log[(C - c_z)!]$$

O objetivo é encontrar o ponto que maximiza a função $l_0(z;p)$. Então:

$$\frac{\partial l_0(z;p)}{\partial p} = \frac{c_z}{p} - n_z + \frac{(C - c_z)}{p} - (N - n_z) = 0 \Rightarrow$$

$$\Rightarrow \frac{C}{p} - N = 0 \Rightarrow p = \frac{C}{N}$$

Substituindo $p = \frac{C}{N}$ em $L_0(z;p)$:

$$L_0(z) = \frac{\left(\frac{C}{N}n_z\right)^{c_z} e^{-\left(\frac{C}{N}n_z\right)}}{c_z!} \frac{\left(\frac{C}{N}(N - n_z)\right)^{C-c_z} e^{-\frac{C}{N}(N-n_z)}}{(C - c_z)!}$$

Considere $\lambda_z = c\frac{n_z}{N}$. Logo tem-se que:

$$L_0(z) = \frac{\lambda_z^{c_z} e^{-\lambda_z}}{c_z!} \frac{(C - \lambda_z)^{C-c_z} e^{-(C-c_z)}}{(C - c_z)!} = \frac{\lambda_z^{c_z}(C - \lambda_z)^{C-c_z} e^{-C}}{c_z!(C - c_z)!}$$

Sob a hipótese alternativa, tem-se que:

$\lambda_z = pn_z$ e $\lambda_{\overline{z}} = q(N - n_z), p > q$

Como $p > q$, a probabilidade de ocorrer o evento de interesse dentro da zona é maior que a probabilidade de ocorrer fora dela. Sob H_1, a função de verossimilhança fica da seguinte forma:

$$L(z;p;q) = \frac{(pn_z)^{c_z} e^{-pn_z}}{c_z!} \frac{[q(N - n_z)]^{(C-c_z)} e^{-q(N-n_z)}}{(C - c_z)!}$$

Considerando $l(z;p;q) = \log L(z;p;q)$, então:

$$l(z;p;q) = c_z[\log p + \log n_z] - pn_z - \log c_z! + (C - c_z)[\log q + \log(N - n_z)]$$

$$-q(N - n_z) - \log[(C - c_z)!]$$

Assim como foi feito com a verossimilhança sob H_0, pretende-se encontrar os pontos que maximizam $l(z;p;q)$ sob H_1:

$$\frac{\partial l(z;p;q)}{\partial p} = \frac{c_z}{p} - n_z = 0 \Rightarrow p = \frac{c_z}{n_z},$$

$$\frac{\partial l}{\partial q} = \frac{(C - c_z)}{q} - (N - n_z) = 0 \Rightarrow q = \frac{(C - c_z)}{(N - n_z)}.$$

Substituindo $p = \frac{c_z}{n_z}$ e $q = \frac{(C-c_z)}{(N-n_z)}$ em $L(z;p;q)$:

$$L(z) = \frac{c_z^{c_z} e^{-c_z}}{c_z!} \frac{(C - c_z)^{(C-c_z)} e^{-(C-c_z)}}{(C - c_z)!} = \frac{c_z^{c_z} (C - c_z)^{(C-c_z)} e^{-C}}{c_z!(C - c_z)!}.$$

Após a definição da verossimilhança sob H_0 e para a hipótese alternativa H_1, a razão de verossimilhança para o modelo Poisson será dada pela seguinte equação:

$$LR(z) = \frac{L}{L_0} = \begin{cases} \left(\frac{c_z}{\lambda_z}\right)^{c_z} \left(\frac{C-c_z}{C-\lambda_z}\right)^{C-c_z}, c_z > \lambda_z \\ 1, \text{caso contrário} \end{cases} \tag{4.4}$$

A razão de verossimilhança será calculada sobre um conjunto de zonas (agrupamento de regiões). O objetivo é identificar o conglomerado mais verossímil.

A razão de verossimilhança cresce muito rapidamente. Uma maneira de diminuir a escala da curva de crescimento da função é aplicando o logaritmo em $LR(z)$. Outra vantagem é que facilita o cálculo da estatística do teste e permite utilizar a propriedade da função logarítmica, que é estritamente crescente. Portanto, o valor que maximiza $LR(z)$ também maximiza o $\log LR(z) = LLR(z)$. Aplicando o logaritmo em $LR(z)$, tem-se a seguinte equação:

$$LLR(z) = \begin{cases} c_z(\log c_z - \log \lambda_z) + (C - c_z)[\log(C - c_z) - \log(C - \lambda_z)], c_z > \lambda_z \\ 0, \text{caso contrário} \end{cases} \tag{4.5}$$

Achar o valor máximo de $LR(z)$ implica em achar o máximo de $LLR(z)$, portanto:

$$T = max_z LLR(z). \tag{4.6}$$

4.3 Construção do algoritmo para solução do problema

A partir da formulação do modelo estatístico apropriado para aplicação do método apresentado por Kulldorff (1997), definem-se os passos do algoritmo que será implementado, bem como as complementações teóricas que busquem a validação do resultado. Primeiramente será apresentada a matriz das distâncias, que é um critério específico para construção de zonas candidatas. Outro passo importante é a verificação da significância das zonas formadas e candidatas a conglomerados. A distribuição empírica da estatística do teste se dará via simulação de Monte Carlo.

4.3.1 Matriz das distâncias

Considere o mapa com n regiões, cada uma com população n_i e número de casos c_i, i=1,2,..., n. A primeira parte consiste em coletar as coordenadas geográficas do início e do final de cada região, calculando o seu centróide. O sistema de coordenadas é do tipo geográfico - latitude e longitude - e terá notação por meio de um par ordenado (x_i,y_i), onde a abcissa representa a latitude e a ordenada a longitude.

O centróide é o ponto médio do segmento de reta que tem sua localização definida pelas coordenadas geográficas do ponto inicial e final. A distância entre dois centróides é dada pela distância entre dois pontos, onde cada ponto ou centróide é definido por um par ordenado (x_i,y_i). Então a distância entre dois centróides i e j quaisquer é dada pela seguinte expressão:

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

A matriz de distâncias entre os centróides de cada região será quadrada com n linhas e n colunas, onde n é o número de centróides das n regiões. Seja $d_{i,j}$ o elemento da i-ésima linha e j-ésima coluna da matriz das distâncias. Então:

$$d_{i,j} = \begin{cases} D_{i,j}, \text{ se } i \neq j, \\ 0, \text{ se } i = j \end{cases}$$

A matriz é simétrica, isto é, para $i \neq j$, $d_{i,j} = d_{j,i}$, para i, j=1,...,n. Além disso, $d_{i,j}$ será a distância entre os centróides dos trechos i e j, logo a matriz das distâncias será dada por:

$$D = \begin{bmatrix} 0 & D_{1,2} & \dots & D_{1,j} & \dots & D_{1,n} \\ D_{2,1} & 0 & \dots & D_{2,j} & \dots & D_{2,n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ D_{i,1} & D_{i,2} & \dots & 0 & \dots & D_{i,n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ D_{n,1} & D_{n,2} & \dots & D_{n,j} & \dots & 0 \end{bmatrix} \tag{4.7}$$

A primeira coluna da matriz das distâncias será composta pelas distâncias do centróide da região 1 ao centróide das demais regiões (inclusive a distância do centróide da região 1 com ela mesma). A segunda coluna fixará a região 2 e terá a distância do centróide da mesma para as demais regiões, e assim sucessivamente.

Após o cálculo da matriz das distâncias, o próximo passo é ordenar as n colunas da matriz. O resultado 4.3.1 é a primeira coluna da matriz das distâncias:

$$\begin{bmatrix} 0 \\ D_{(2),1} \\ D_{(3),1} \\ \vdots \\ D_{(n),1} \end{bmatrix}$$

De acordo com o exemplo apresentado do vetor coluna ordenado, temos que $D_{(n),1} > ... > D_{(3),1} > D_{(2),1}$. A primeira zona candidata será formada unicamente pela região 1, $Z = \{1\}$. A segunda zona é formada pela região 1 e pela região mais próxima, correspondendo à distância $D_{(2),1}$. Esta zona será representada por $Z = \{1, (2)\}$, que é a união da região 1 com a região mais próxima. As demais zonas são, então, obtidas sucessivamente adicionando à zona atual a região mais próxima. Para cada zona formada, será calculado o valor de LLR, conforme equação 4.6. O algoritmo adiciona regiões à zona atual até que a população da zona não ultrapasse 50% da população total. Esse processo é repetido para cada vetor coluna, de modo que o primeiro vetor coluna inicia pela primeira região, o segundo vetor inicia pela segunda região, e assim sucessivamente. Os valores da $LLR(z)$, para cada zona z considerada, são armazenados em uma matriz L com dimensão $n \times n$. O LLR será igual a zero em duas ocasiões:

- População da zona corresponda a mais de 50% da população total.
- Número de casos observado é menor que o número de casos esperado.

Ao final do processo, estamos interessados na zona z que produz o maior valor de LLR(z), isto é,

$$z = \arg\max_{z} LLR(z)$$

A zona z é a estimativa de máxima verossimilhança para o conglomerado, sendo chamada de zona mais verossímil.

4.3.2 Verificação da significância do conglomerado

A forma analítica de testar a significância da estatística do teste T é através da comparação do seu valor com os de uma distribuição conhecida, fixando-se um nível de significância. Todavia, não é possível obter, analiticamente, a distribuição da estatística do teste sob a hipótese nula. Além disso, a aproximação assintótica usual por uma distribuição qui-quadrado da transformação -2$log\lambda$ não é válida, pois

as condições de regularidade não são satisfeitas. Uma saída encontrada por Kulldorff & Nagarwalla (1995) é utilizar procedimento computacional para simulação da distribuição empírica de T.

O método computacional utilizado será a simulação de Monte Carlo, que foi desenvolvido na época da segunda guerra mundial, conjuntamente com o projeto Manhattan, por Von Neumann e Fermi. O método de Monte Carlo (MMC) tem esse nome devido ao fato de Monte Carlo ser uma região com muitos cassinos e de que esse método era muito utilizado em jogos de azar (Hammersley & Handscombr, 1964).

Segundo Metropolis & Ulam (1949), o MMC consiste basicamente em gerar aleatoriamente sucessivas amostras (variáveis aleatórias) que são testadas contra um modelo estatístico, no caso, uma distribuição de probabilidade. Este método fornece uma estimativa do valor esperado e um provável erro para a estimativa, que é inversamente proporcional ao número de réplicas. Logo, quanto maior o número de réplicas, menor será o erro.

A idéia é simular sob a hipótese de que o número esperado de casos em determinada região é proporcional à sua população. O número esperado de casos é igual a pn_i, onde n_i é a população na região i e p, a probabilidade de um indivíduo ser um caso. Na simulação será aplicada a mesma probabilidade de um indivíduo da população vir a ser um caso, ou seja, $p = \frac{C}{N}$. Portanto, o número esperado de casos no i-ésimo trecho será dado por:

$$\mu_i = \frac{Cn_i}{N}, \tag{4.8}$$

onde N é o tráfego total.

Serão geradas m réplicas, onde em cada réplica, C casos são distribuídos aleatoriamente no mapa, sob a hipótese nula de proporcionalidade quanto às populações. Para cada réplica gerada obtém-se o valor da estatística T conforme o procedimento descrito na seção 4.2. Assim, ao final teremos $T_1, \ldots, T_m$, uma amostra de tamanho m para a estatística de teste sob H_0.

Após a geração da distribuição empírica da estatística do teste é possível compará-la ao valor observado de T - isto é, ao valor de T obtido para os dados observados. O problema de testar a hipótese se resume a verificar se:

$$P(X > T) \leq \alpha,$$

onde X é uma v.a. cuja distribuição é estimada pela distribuição empírica obtida, T é o valor da estatística de teste para os dados observados e α é o nível de significância.

4.4 Definição do algoritmo Scan Circular

Para aplicação do algoritmo *Scan* circular, listam-se as seguintes ações:

1. Associar cada evento de interesse a uma região no mapa.
2. Inserir a coordenada geográfica do centróide de cada região, bem como a sua população.
3. Calcular o número esperado de casos, de acordo com a Equação 4.8.
4. Calcular a matriz das distâncias conforme 4.7.
5. Ordenar cada vetor coluna em ordem crescente de distâncias e a partir daí construir as zonas.
6. Para cada zona z candidata a conglomerado, obter o valor de $LLR(z)$.
7. Após calcular o $LLR(z)$ para todas as zonas, encontra-se o maior valor entre os elementos da matriz $L_{n,n}$, como descrito na Equação 4.6.
8. O próximo passo é gerar a distribuição empírica da estatística do teste considerando que todos os elementos da população possuem a mesma probabilidade de vir a ser um caso, ou seja, o valor esperado de casos em cada região é proporcional a população nesta região. São geradas várias distribuições de casos sob H_0 e para cada uma calculamos o valor da estatística do teste. Com esses valores, temos a distribuição empírica da estatística do teste sob H_0.
9. A estatística do teste $T = max_z LLR(z)$ é comparada com a distribuição empírica encontrada através da simulação de Monte Carlo.

10. Calcula-se o $p-valor$ do teste e, fixando-se um nível de significância α, verifica-se a significância da estatística T.

11. Caso p-valor$<\alpha$, rejeita-se a hipótese nula de que a probabilidade dentro da zona é igual a probabilidade fora da zona. Confirmando que a zona mais verossímil é estatisticamente um conglomerado.

Capítulo 5

Metodologia Scan de vizinhança

Conforme descrito no capítulo 4, a tarefa de encontrar conglomerados passa pelo critério de formação destas zonas candidatas a conglomerados. Conforme descrito na seção 4.3.1, a matriz das distâncias é um critério utilizado por Kulldorff (1997) para construção de zonas. A matriz das distâncias agrupa regiões pelo critério da menor distância entre os casos observados, ou no caso de dados agregados, a menor distância entre os centróides das regiões. Portanto, regiões com centróides mais próximos serão agrupadas.

A dependência espacial está ligada à distribuição dos dados ao longo das subdivisões territoriais, fazendo com que o valor de determinado atributo se assemelhe mais aos seus vizinhos em detrimento ao restante do conjunto amostral. O conceito de vizinhança está intimamente ligado a idéia de proximidade, que pode ser exemplificado pela distância linear ou pela zona fronteiriça.

Uma pergunta válida seria qual estratégia de construção de conglomerados é mais apropriada? Em alguns cenários, a estrutura de vizinhança entres as regiões do mapa pode ser mais conveniente, ao invés da simples adoção da distância euclidiana. Por exemplo, Duczmal *et al.* (2010) propõem uma nova ferramenta para testar a hipótese de agrupamento local, tendo como critério fatores ambientais. A conectividade entre regiões é reforçada ou enfraquecida de acordo com determinadas características de interesse. A probabilidade de detecção de conglomerados é aumentada ou diminuída de acordo com alterações feitas na estrutura de vizinhança, referentes à seleção de características ambientais.

5.1 Matriz de vizinhança

Com base nas coordenadas geográficas do centróide das regiões do mapa é possível determinar quais regiões são vizinhas entre si. Uma matriz com valores binários 0 ou 1 é criada, onde 0 indica que não há vizinhança e 1, o contrário.

A matriz de vizinhança leva em conta as características topológicas das regiões para a formação de zonas, unindo regiões adjacentes, que possuem fronteira contígua entre si. Seja a matriz de vizinhança $V = \{v_{i,j}\}$, então:

$$v_{i,j} = \begin{cases} 1, \text{ se trecho } i \text{ é vizinho do trecho } j. \\ 0, \text{ caso contrário} \end{cases} \tag{5.1}$$

A matriz de vizinhança será definida da seguinte forma:

$$V = \begin{bmatrix} 0 & v_{1,2} & \dots & v_{1,j} & \dots & v_{1,n} \\ v_{2,1} & 0 & \dots & v_{2,j} & \dots & v_{2,n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ v_{i,1} & v_{i,2} & \dots & 0 & \dots & v_{i,n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ v_{n,1} & v_{n,2} & \dots & v_{n,j} & \dots & 0 \end{bmatrix}, \tag{5.2}$$

Com $v_{i,j}$ podendo ser 0 ou 1.

A matriz de vizinhança possui n linhas e n colunas, onde n representa o número de regiões implantadas. A diagonal da matriz V será igual a zero, por convenção. A matriz V é simétrica, dado que $v_{i,j} = v_{j,i}$.

A primeira coluna da matriz de vizinhança será composta pela determinação de adjacência da região 1 com as demais regiões. A segunda coluna da matriz representa a estrutura de adjacência tendo a segunda região fixa. O mesmo procedimento é feito até a n-ésima região.

Considere l um vetor de listas contendo regiões que formam fronteiras. Portanto, $l_{\{1\}}$ conterá uma lista com todas as regiões vizinhas da primeira região, $l_{\{2\}}$ conterá uma lista com todas as regiões vizinhas a segunda região, e assim sucessivamente.

Caso a região não seja vizinha de nenhuma outra região, $l = \oslash$.

Considere ainda que as zonas $Z_{\{1\}}$ e $Z_{\{2\}}$ se agrupem formando a zona $Z_{\{1,2\}}$, então $l_{\{1,2\}}$ denotará uma lista com os vizinhos dessa zona resultante. Assim sendo, l representará a lista com regiões candidatas a formar uma nova zona.

5.2 Critérios de vizinhança para construção de conglomerados

Para fins de comparação com o *Scan* Circular de Kulldorff (1997), foram implementados três métodos de vizinhança: *Scan* de Vizinhança Aleatória, *Scan* de Vizinhança Otimizada e *Scan* de Vizinhança Proporcional. Cada método possui critério distinto de seleção de regiões adjacentes.

5.2.1 *Scan* de vizinhança aleatória

Este método seleciona as zonas candidatas aleatoriamente, com igual probabilidade.

Para aplicação do algoritmo de vizinhança aleatória listam-se as seguintes ações:

1. A primeira zona candidata será formada pela região 1, $Z_{\{1\}}$. Caso $Z_{\{1\}}$ possua apenas um vizinho, então a zona resultante será igual a $Z_{\{1\}} \cup l_{\{1\}}$, onde $l_{\{1\}}$ terá apenas um elemento ou região. Caso a zona $Z_{\{1\}}$ possua mais de um vizinho, escolhe-se aleatoriamente um elemento de $l_{\{1\}}$, com igual probabilidade. Essa região 2, que necessariamente é vizinha de $Z_{\{1\}}$ é agrupada, formando a zona $Z_{\{1,2\}}$.

2. A partir da lista $l_{\{1,2\}}$, escolhe-se aleatoriamente um elemento, com igual probabilidade. Essa região 3, que necessariamente é vizinha de $Z_{\{1,2\}}$ é agrupada, formando a zona $Z_{\{1,2,3\}}$. Caso $l_{\{1,2\}}$ possua apenas uma elemento, então $Z_{\{1,2,3\}} = Z_{\{1,2\}} \cup l_{\{1,2\}}$

3. O procedimento é repetido até que Z contenha todas as regiões do mapa. O programa para se $l = \oslash$.

4. O algoritmo é reiniciado com a primeira zona candidata sendo a região 2, $Z_{\{2\}}$. Os procedimentos anteriores são repetidos até que Z contenha novamente todas as regiões do mapa ou $l = \oslash$.

5. O algoritmo se reinicia começando com a zona $Z_{\{3\}}$, e assim sucessivamente, até completar todas as n regiões do mapa.

5.2.2 *Scan* de vizinhança otimizada

Uma outra forma de selecionar vizinhos foi implementada adicionando regiões que incorporam maior valor do *log* da razão de verossimilhança à zona inicial. O algoritmo possui os seguintes passos:

1. A primeira zona candidata será formada pela região 1, $Z_{\{1\}}$. Caso $Z_{\{1\}}$ possua apenas um vizinho, então a zona resultante será igual a $Z_{\{1\}} \cup l_{\{1\}}$, onde $l_{\{1\}}$ terá apenas um elemento ou região. Caso a zona $Z_{\{1\}}$ possua mais de um vizinho, escolhe-se o elemento de $l_{\{1\}}$ que ao unir-se com $Z_{\{1\}}$ agregue maior valor de LLR a $Z_{\{1\}}$, conforme equação 5.3. Essa região 2 é agrupada formando a zona $Z_{\{1,2\}}$.

2. A partir da lista $l_{\{1,2\}}$, escolhe-se o elemento que ao agrupar-se com a zona inicial $Z_{\{1,2\}}$ apresente maior valor de LLR. Essa região 3, que necessariamente é vizinha de $Z_{\{1,2\}}$, é agrupada, formando a zona $Z_{\{1,2,3\}}$. Caso $l_{\{1,2\}}$ possua apenas uma elemento, então $Z_{\{1,2,3\}} = Z_{\{1,2\}} \cup l_{\{1,2\}}$

3. Caso todas as zonas candidatas apresentem valor de LLR iguais a zero (número de casos na zona for menor que o número esperado), então seleciona-se a região candidata aleatoriamente, conforme subseção 5.2.1. Caso existam duas ou mais zonas candidatas com valores iguais de LLR, então seleciona-se tais elementos aleatoriamente, conforme subseção 5.2.1.

4. O procedimento é repetido até que Z contenha todas as regiões do mapa ou $l = \oslash$.

5. O algoritmo é reiniciado com a primeira zona candidata igual a região 2, $Z_{\{2\}}$. Os passos anteriores são realizados até que Z contenha novamente todas as

regiões do mapa. O procedimento é encerrado se $l = \oslash$.

6. O algoritmo agora se reinicia com a região 3, e assim sucessivamente, até completar todas as regiões do mapa.

5.2.3 *Scan* de vizinhança proporcional

O *Scan* de vizinhança proporcional adiciona regiões selecionadas aleatoriamente, com probabilidade proporcional ao *log* da razão de verossimilhança das zonas resultantes.

1. A primeira zona candidata será formada pela região 1, $Z_{\{1\}}$. Caso $Z_{\{1\}}$ possua apenas um vizinho, então a zona resultante será igual a $Z_{\{1\}} \cup l_{\{1\}}$. Caso a zona $Z_{\{1\}}$ possua mais de um vizinho, escolhe-se aleatoriamente um elemento de $l_{\{1\}}$, com probabilidade proporcional ao valor do LLR das zonas resultantes. A primeira zona resultante é a união da primeira região de $l_{\{1\}}$ com a zona $Z_{\{1\}}$. A segunda zona resultante é a união da segunda região de $l_{\{1\}}$ com a zona $Z_{\{1\}}$. E assim sucessivamente. A região 2 escolhida é agrupada, formando a zona $Z_{\{1,2\}}$.

2. A partir da lista $l_{\{1,2\}}$, escolhe-se um elemento com probabilidade proporcional ao valor de LLR das zonas resultantes, conforme definido no item anterior. Essa região 3, que necessariamente é vizinha de $Z_{\{1,2\}}$ é agrupada, formando a zona $Z_{\{1,2,3\}}$. Caso $l_{\{1,2\}}$ possua apenas uma elemento, então $Z_{\{1,2,3\}} = Z_{\{1,2\}} \cup l_{\{1,2\}}$.

3. Caso todas as zonas resultantes apresentem valores de LLR idênticos ou iguais a zero (número de casos na zona for menor que o número esperado), então seleciona-se a zona candidata aleatoriamente, conforme subseção 5.2.1.

4. Caso alguma zona resultante possua LLR igual a zero, então acrescenta-se uma unidade a todos os valores de LLR das zonas resultantes e seleciona-se com probabilidade proporcional a esses valores.

5. O procedimento é repetido até que Z contenha todas as regiões do mapa ou $l = \oslash$.

6. O algoritmo é reiniciado com a primeira zona candidata igual a região 2, $Z_{\{2\}}$. Os procedimentos descritos anteriormente são repetidos até que Z contenha novamente todas as regiões do mapa ou $l = \oslash$.

7. O algoritmo agora se reinicia começando com a região 3, e assim sucessivamente, até completar todas as regiões do mapa.

5.3 Teste da razão de verossimilhança

O teste da razão de verossimilhança para o método *Scan* de vizinhança é idêntico ao teste para o método *Scan* circular de Kulldorff & Nagarwalla (1995), conforme descrito na seção 4.2.

O resultado final do logaritmo da razão de verossimilhança para o método *Scan* de vizinhança é dado pela equação 5.3:

$$LLR(z) = \begin{cases} c_z(\log c_z - \log \mu_z) + (C - c_z)[\log(C - c_z) - \log(C - \mu_z)], c_z > \mu_z \\ 0, \text{caso contrário} \end{cases} \tag{5.3}$$

A estatística do teste também é a mesma, de acordo com a equação 5.4:

$$T = max_z LLR(z). \tag{5.4}$$

5.3.1 Verificação da significância do conglomerado

A verificação da significância do conglomerado pelo método *Scan* de vizinhança é feita de maneira similar ao apresentado na subseção 4.3.2. A estatística do teste para os dados observados é comparada com o percentil da distribuição empírica, sob H_0, gerada pela simulação de Monte Carlo. São geradas várias réplicas da estatística do teste sob a suposição de que o número esperado de casos em cada região é proporcional ao tamanho de sua população.

Sejam t e m inteiros positivos e $p_1, ..., p_t$ probabilidades satisfazendo $0 \leq p_i \leq 1$, i=1,2,...t e $\sum_{i=1}^{t} p_i = 1$. Considere que $(C_1, ..., C_t)$ é a variável aleatória com

distribuição multinomial, com m eventos de interesse. A função de probabilidade conjunta de $(C_1, ..., C_t)$ é dada por:

$$f(c_1, ..., c_t) = \frac{m!}{c_1!...c_t!} p_1^{c_1} ....p_t^{c_t} = m! \prod_{i=1}^{t} \frac{p_i^{c_i}}{c_i!}, \tag{5.5}$$

Onde:

t = Total de regiões.

m=Total de casos do evento de interesse.

p_i=Probabilidade de ocorrência do evento de interesse em cada região.

C_i=Variável aleatória representando o número de casos na i-ésima região.

Serão geradas k réplicas, onde em cada réplica, m casos com distribuição multinomial são gerados e distribuídos ao longo do mapa, sob hipótese de proporcionalidade quanto às populações. Os parâmetros da distribuição são o número total de casos m e as probabilidades em cada região $p_i = \frac{n_i}{N}$.

Onde:

p_i=Probabilidade na i-ésima região.

n_i=Tráfego na i-ésima região.

N=Somatório do tráfego em todas as regiões.

Por meio da simulação de Monte Carlo serão gerados um número grande de réplicas, onde cada réplica resultará em um valor de estatística do teste, conforme a Equação 5.4. No final, compara-se o percentil da distribuição das k réplicas com o valor da estatística do teste T dos valores observados. O problema de testar a hipótese se resume a verificar se $P(X > T) \leq \alpha$,.

Onde X é o percentil da v.a. estimada pela distribuição empírica obtida, T é o valor da estatística de teste para os dados observados e α é o nível de significância.

5.4 Definição do algoritmo Scan de vizinhança

Para aplicação do algoritmo *Scan* de Vizinhança, listam-se as seguintes ações:

1. Calcular a matriz de adjacência conforme seção 5.1.

2. A partir da matriz de adjacência criar vetor de listas $l_{\{1\}}$, $l_{\{2\}}$, ... $l_{\{n\}}$, com os vizinhos de cada região.

3. O primeiro elemento do vetor coluna será a região 1, sendo a primeira zona candidata a conglomerado. Calcular o valor de LLR, conforme equação 5.3.

4. Utilizando um dos critérios descritos na subseção 5.2, selecionar elemento da lista $l_{\{1\}}$, formando a zona $Z_{\{1,2\}}$. Calcular o valor de LLR para $Z_{\{1,2\}}$.

5. Escolhendo um dos critérios definidos na subseção 5.2, selecionar elemento da lista $l_{\{1,2\}}$ e adicionar a $Z_{\{1,2\}}$, onde:

$$l_{\{1,2\}} = (l_{\{1\}} \cup l_{\{2\}}) - (l_{\{1\}} \cup l_{\{2\}}) \cap Z_{\{1,2\}} \tag{5.6}$$

6. Calcular LLR para $Z_{\{1,2,3\}}$.

7. Encontrar lista $l_{\{1,2,3\}}$ com regiões adjacentes à zona $Z_{\{1,2,3\}}$:

$$l_{\{1,2,3\}} = (l_{\{1\}} \cup l_{\{2\}} \cup l_{\{3\}}) - (l_{\{1\}} \cup l_{\{2\}} \cup l_{\{3\}}) \cap Z_{\{1,2,3\}} \tag{5.7}$$

8. Utilizando um dos critérios de vizinhança, selecionar um elemento de $l_{\{1,2,3\}}$ formando a zona $Z_{\{1,2,3,4\}}$. Calcular LLR. O mesmo procedimento é repetido até que Z contenha todas as regiões, $Z_{\{1,2,3,....,n\}}$. Calcular LLR para todas as zonas.

9. O algoritmo se encerra se $l = \oslash$ (zona não possui vizinho) ou se Z contém todas as regiões do mapa.

10. O primeiro vetor coluna com as zonas candidatas é armazenado. O mesmo procedimento é repetido até que o n-ésimo vetor coluna seja preenchido, formando assim as matrizes $M_{\{n,n\}}$ e $L_{\{n,n\}}$ com as zonas candidatas a conglomerado e valores de LLR, respectivamente.

11. Utilizando distribuição multinomial e através da simulação de Monte Carlo gerar distribuição empírica da estatística do teste, E. Ver detalhadamente na subseção 5.3.1.

12. Fixando-se um nível de significância, α, encontrar quais zonas possuem valor de LLR significativos, ou seja, $LLR > P(E < \epsilon) = 1 - \alpha$.

13. A zona que possuir maior valor de LLR será o conglomerado primário. Seguindo ordem descrescente de LLR, o conglomerado secundário será aquele que tiver LLR significativo e não possuir intersecção com o conglomerado primário. O procedimento é repetido até se encontrar todos os conglomerados significativos e sem intersecção entre si.

14. Calcular o $p - valor$ do teste para todos os conglomerados.

Capítulo 6

Resultados

A matriz das distâncias pode apresentar resultados menos satisfatórios em estudos envolvendo acidentes de trânsito, dado que os acidentes de trânsito são restritos às vias de circulação de veículos, pessoas e animais. Tais ocorrências não se distribuem ao longo de todo o mapa rodoviário, sendo restritas ao percurso das rodovias do Sistema Rodoviário do Distrito Federal(2012) (2012). A malha rodoviária é definida por segmentos de retas contíguos.

Dependendo do objetivo do estudo, a proximidade geográfica pode ter menor relevância, não sendo a melhor forma de estimar a variabilidade espacial dos dados. No estudo de acidentes de trânsito em rodovias, a matriz das distâncias pode selecionar trechos rodoviários pertencentes a diferentes rodovias, com características bastante distintas, preterindo algum trecho adjacente com característica bastante similar. Duas rodovias podem ser paralelas e bastante próximas, entretanto podem possuir geometria, velocidade operacional e característica de tráfego variadas.

Como exemplo podemos citar o final da DF-004 (EPNA) - trecho da Ponte Presidente Médice à entrada da DF-051 - e o início da DF-025 (EPDB) - trecho da entrada da DF-047 ao acesso à Ponte Presidente Médici - que segundo o critério da distância euclidiana formariam uma zona, quando na verdade são trechos separados pelo Lago Paranoá e com velocidade regulamentar distinta. Nem sempre a distância entre trechos é uma medida de dependência espacial apropriada, principalmente quando o evento de interesse é restrito a segmentos contínuos. Lembrando que a noção de autocorrelação espacial está associada a idéia de similaridade entre regiões geográficas.

Uma solução alternativa é aplicar uma estrutura de vizinhança que leva em conta a conectividade entre as regiões e a topologia da malha rodoviária, onde um trecho rodoviário é vizinho de outro se o mesmo compartilha uma fronteira. Deseja-se implementar algoritmo que considere maior correlação espacial em trechos rodoviários contíguos.

6.1 Testes numéricos

Para que possamos aplicar os métodos de vizinhança com dados de acidentes de trânsito são necessários mecanismos que possibilitem verificar a eficiência dos mesmos. Além disso, é preciso comparar o desempenho dos métodos de vizinhança com o método de varredura circular de Kulldorff (1997), definidos nos capítulos 5 e 4, respectivamente. A verificação de desempenho dos métodos de varredura é realizada para conglomerados com diferentes geometrias, com uso dos dados do Sistema Rodoviário do Distrito Federal(2012) (2012) e de acidentes de trânsito fatais, 2012. A metodologia de mensuração do desempenho foi feita seguindo os conceitos presentes nos artigos de Kulldorff *et al.* (2003) e Huang *et al.* (2007).

6.1.1 Poder do teste, valor preditivo positivo e sensibilidade

Comumente as estatísticas de varredura são utilizadas para detecção de conglomerados de doenças. Para dados de mortalidade, o modelo Poisson é mais comum, onde o evento de interesse é quantitativo discreto. O artigo de Huang *et al.* (2007) propõe um modelo exponencial para lidar com dados contínuos em análise de sobrevivência. O método também funciona para funções de sobrevivência com distribuição gama e log-normal. A investigação do desempenho do método foi feito por meio do cálculo do poder do teste, sensibilidade e valor preditivo positivo. Para testar a eficiência do método proposto, os dados de sobrevivência foram gerados aleatoriamente. Diferentes conjuntos de dados foram gerados a partir das distribuições exponencial, gama e log-normal, com diferentes médias e variâncias. Para as localizações geográficas foram utilizados dados reais da localização de residências dos homens diagnosticados com câncer de próstata. Foram criados 10000 conjuntos de dados simulados para

cada modelo de probabilidade sob hipótese alternativa. Para cada um desses conjuntos de dados simulados foram gerados 999 permutações aleatórias para obtenção de p-valores. Para cada modelo, o poder do teste é estimado da seguinte forma:

$$\text{Poder do teste} = \frac{\text{Número de simulações com } p\text{-valor} < 0.05}{10000} \tag{6.1}$$

O poder do teste não fornece informações sobre a precisão geográfica do conglomerado detectado. Para avaliar a precisão do conglomerado detectado, Huang *et al.* (2007) definiram a sensibilidade como sendo a proporção de indivíduos do conglomerado verdadeiro "capturado" pelo conglomerado detectado:

$$\frac{1}{S}\sum_{s=1}^{S}\frac{\text{População da intersecção do conglomerado verdadeiro e detectado na } i\text{-ésima simulação}}{\text{População do conglomerado verdadeiro na } i\text{-ésima simulação}}, \tag{6.2}$$

onde S é o número total de simulações.

Huang *et al.* (2007) definiram o valor preditivo positivo como sendo a proporção de indivíduos no conglomerado detectado pertencente ao conglomerado verdadeiro:

$$\frac{1}{S}\sum_{s=1}^{S}\frac{\text{População da intersecção do conglomerado verdadeiro e detectado na } i\text{-ésima simulação}}{\text{População do conglomerado detectado na } i\text{-ésima simulação}}, \tag{6.3}$$

onde S é o número total de simulações.

6.1.2 Geração de Conglomerados artificiais

O conglomerado artificial é construído atribuindo maior probabilidade de ocorrência do evento de interesse dentro do conglomerado. A probabilidade dentro do conglomerado artificial será sempre maior e constante paras as regiões internas a ele. Primeiramente definem-se os trechos rodoviários (regiões) que formarão o conglomerado artificial. Posteriormente, as probabilidades dentro do conglomerado são calculadas através de um teste binomial, utilizando-se de aproximação para a distribuição Normal. As probabilidades foram estimadas de forma que a hipótese nula seja rejeitada com probabilidade igual a 0,999.

$$P(LLR(z) > T_{\text{crítico}}) = 0,999 \tag{6.4}$$

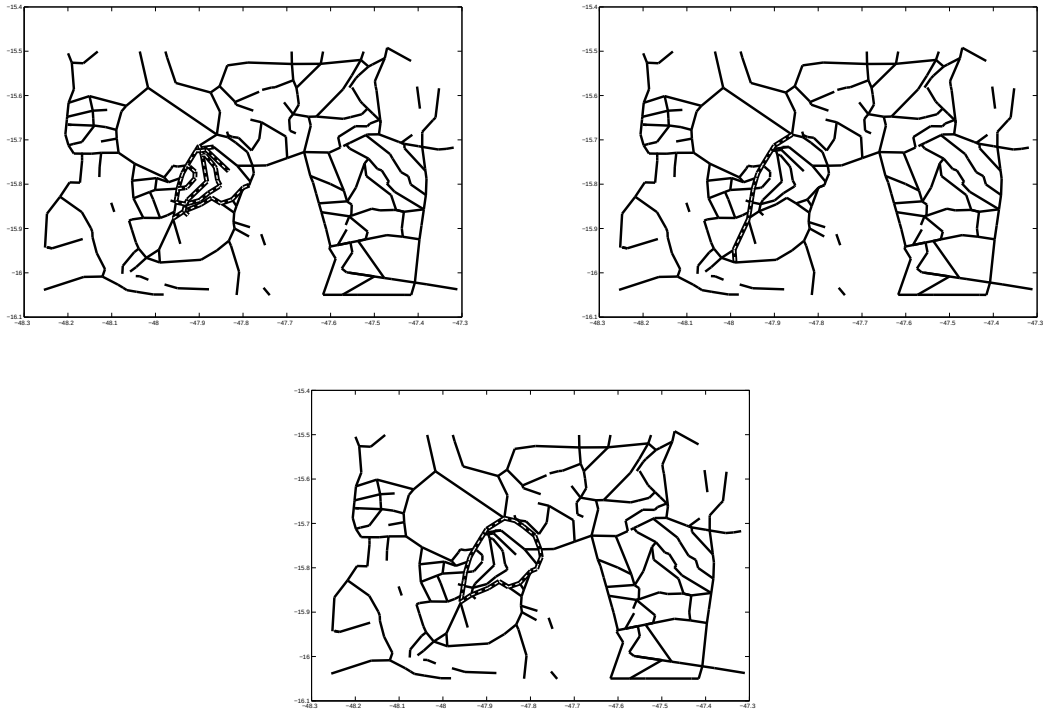

Figura 6.1: Conglomerados gerados: Superior à esquerda: Conglomerado artificial I. Superior à direita: Conglomerado artificial II. Inferior: Conglomerado artificial III.

Observe que a equação 6.4 diz respeito ao poder do teste do método em que sabemos exatamente a localização do conglomerado que desejamos que seja detectado (Kulldorff *et al.* , 2003).

De acordo com a figura 6.1.2 definiu-se geometricamente 3 (três) conglomerados artificiais. Conglomerado artificial I: Plano Piloto, geometria circular com regiões internas. Conglomerado artificial II: DF-003, formado por várias retas conexas. Conglomerado artificial III: esfera oca, geometria aproximadamente circular, sem trechos internos.

6.1.3 Comparação dos métodos de varredura via conglomerado artificial

As medidas descritas na subseção 6.1.1 serão calculadas para testar o comportamento de cada método de varredura para dados gerados artificialmente, com geometria distintas. Dado que conhecemos à priori o conglomerado verdadeiro, qual será o

poder de detecção dos métodos de varredura? Qual a sensibilidade de cada método quando alteramos a geometria do conglomerado? Cada método de varredura será testado para 3 (três) diferentes conglomerados artificiais.

Considere que queiramos testar o desempenho do método *Scan* Circular para o conglomerado artificial I. Os seguintes passos são realizados:

- Para o conglomerado artificial I são geradas m réplicas, cada réplica consiste em distribuir os C casos aleatoriamente ao longo dos k trechos rodoviários, aplicando maior probabilidade de ocorrência do evento de interesse dentro do conglomerado artificial. A probabilidade dentro do conglomerado é estabelecida conforme explicitado na subseção 6.1.2.

- De posse dos valores observados na i-ésima réplica, obtemos o valor da estatística do teste T, conforme a equação 5.4. A estatística do teste T da i-ésima réplica é comparada com o 95º percentil da estatística do teste gerada sob H_0. Se a estatística T for maior que o 95º percentil da estatística sob H_0 (valor crítico), significa dizer que o método *Scan* Circular foi capaz de detectar a existência de conglomerado.

- Se o método foi capaz de detectar a existência de conglomerado, então o poder do teste é igual a 1 (um) e a sensibilidade e o valor preditivo positivo são calculados.

- Caso contrário, o poder do teste para a i-ésima réplica é igual a 0, não sendo possível o cálculo da sensibilidade e VPP (valor preditivo positivo).

- O procedimento é repetido para as m réplicas do conglomerado artificial I, para o qual o poder do teste será dado pela proporção de vezes em que o valor da estatística do teste supera o valor crítico sob H_0, conforme equação 6.1.

- A sensibilidade e o valor preditivo positivo são calculados conforme as equações 6.2 e 6.3, considerando que o número total de simulações S será dado pelo número de vezes em que o poder do teste for igual a 1 (um).

- O mesmo cálculo é repetido para os conglomerados artificiais II e III.

- O procedimento é repetido para o *Scan* de Vizinhança Aleatória, *Scan* de Vizinhança Otimizada e *Scan* de Vizinhança Proporcional.

A tabela 6.1.3 mostra o resultado de desempenho dos métodos de varredura:

Tabela 6.1: Medidas de desempenho dos métodos de varredura

MÉTODOS	ARTIFICIAL I			ARTIFICIAL II			ARTIFICIAL III		
	Poder	Sens.	VPP	Poder	Sens.	VPP	Poder	Sens.	VPP
S.C	0,86	0,80	0,83	0,30	0,30	0,57	0,24	0,30	0,61
S.V.A	0,76	0,68	0,77	0,56	0,58	0,56	0,53	0,56	0,58
S.V.O	0,33	0,73	0,56	0,41	0,86	0,39	0,24	0,76	0,47
S.V.P	0,76	0,79	0,69	0,64	0,73	0,50	0,48	0,65	0,56

Observando os resultados numéricos, o *Scan* Circular (S.C) de Kulldorff (1997) apresenta melhor resultado para o conglomerado artificial I, o que já era esperado, pois o mesmo foi construído para se comportar bem para este tipo de conglomerado. O *Scan* de Vizinhança Aleatória (S.V.A) e Proporcional apresentaram resultados satisfatórios para o conglomerado artificial I, com uma pequena vantagem para o *Scan* de Vizinhança Proporcional (S.V.P). O *Scan* de Vizinhança Otimizada (S.V.O) foi o que apresentou o pior resultado.

No conglomerado artificial II, o *Scan* Circular teve um resultado muito ruim, com poder do teste e sensibilidade iguais a 0,3. O *Scan* de Vizinhança Proporcional foi a técnica com melhor desempenho, com poder do teste bem acima dos outros, com sensibilidade igual a 0,73. O *Scan* de Vizinhança Aleatória se comportou de maneira satisfatória, com valores sempre acima de 0,5. O *Scan* de Vizinhança Otimizada, apesar de apresentar alto valor de sensibilidade, foi insatisfatório no poder do teste e VPP.

O conglomerado artificial III trouxe problemas para o *Scan* Circular, onde o mesmo apresentou valor de poder do teste igual a 0,24 e sensibilidade igual a 0,3. O *Scan* de Vizinhança Aleatória e Proporcional foram os métodos que melhor se comportaram para esse conglomerado. Os métodos possuem VPP bastante próximos. O *Scan* de Vizinhança Proporcional possui maior sensibilidade e menor poder do teste do que o *Scan* de Vizinhança aleatória.

Os resultados numéricos demonstraram que o *Scan* Circular não se comporta bem para os conglomerados do tipo II e III, principalmente no que tange à sensibilidade e ao poder do teste. Para esses conglomerados, o *Scan* Circular possui pouco poder de identificação, além de não conseguir detectar grande proporção do conglomerado verdadeiro. Apesar do *Scan* Circular não possuir muitos falsos positivos (VPP não é tão baixo), o mesmo identifica pouca proporção do conglomerado verdadeiro.

Apesar dos métodos S.V.A. e S.V.P. não serem os melhores para o conglomerado artificial I, possuem resultados satisfatórios. Para os conglomerados artificiais II e III, o S.V.A e S.V.P. são bem mais apropriados do que o S.C. Sendo assim, o S.V.A. e S.V.P. são métodos mais robustos para detecção de conglomerados, podendo ser usados nos 3 (três) tipos de conglomerados.

6.1.4 Aplicação com dados de acidentes de trânsito, 2012

O Sistema Rodoviário do Distrito Federal(2012) (2012) conta com 403 trechos rodoviários, sendo 386 trechos implantados (onde há fluxo de veículos), com malha rodoviária com extensão aproximada de 1800 quilômetros (pavimentadas e não pavimentadas). O número de acidentes fatais gira em torno de 200 ao ano. Portanto, a tarefa de encontrar pontos críticos torna-se bastante trabalhosa e com alto custo. Com o passar dos anos, análises descritivas e utilização de índices se mostraram ineficientes. Além disso, vale ressaltar que seus resultados são não inferenciais e não possuem validade estatística.

No estudo de acidentes, o evento de interesse são os acidentes de trânsito fatais. Tais eventos estão distribuídos ao longo de regiões denominadas trechos rodoviários, classificados como dados de área. A população é o tráfego médio de veículos diário. Além disso foram utilizadas coordenadas geográficas do centróide de cada trecho rodoviário.

O estudo visa a identificação de conglomerados de acidentes de trânsito na malha rodoviária distrital, com utilização de dados de 2012. Com base numa revisão bibliográfica em estatística espacial, decidiu-se pela utilização da estatística *Scan* Circular de Kulldorff (1997). Posteriormente, implementamos métodos alternativos ao *Scan* Circular, utilizando critérios distintos de construção de conglomerados, conforme ex-

plicitado no capítulo 5. Todos os algoritmos foram feitos no software estatístico *R*.

A Figura 6.2 representa a malha rodoviária em conformidade com o Sistema Rodoviário do Distrito Federal(2012) (2012):

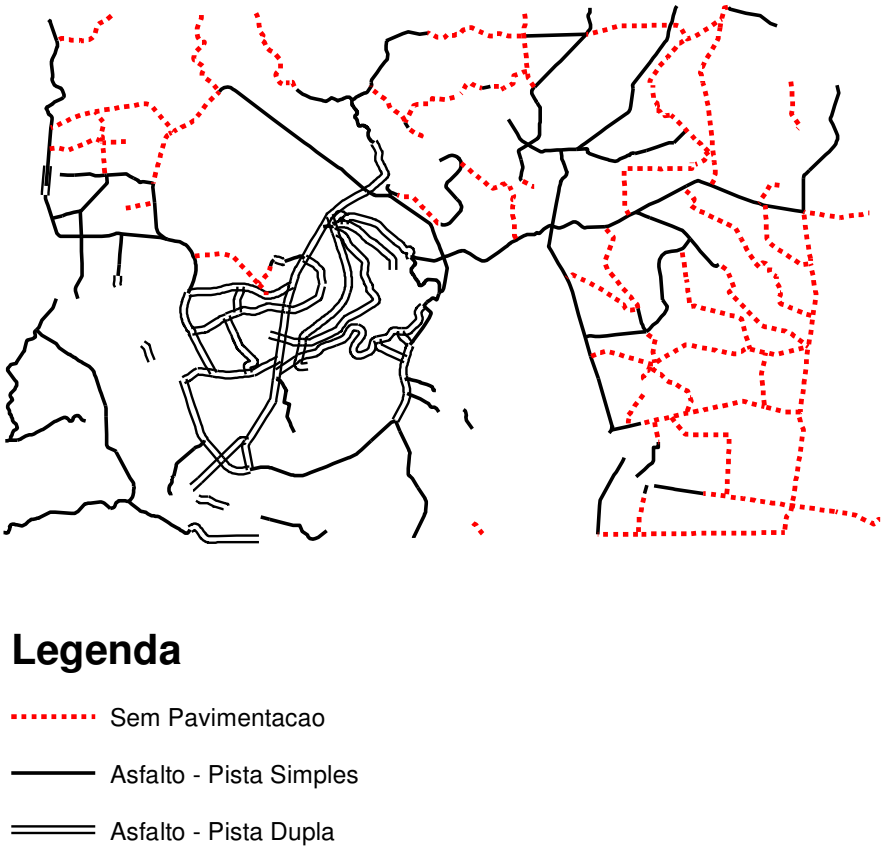

Figura 6.2: Malha rodoviária sob circunscrição do DER/DF.

6.2 Preparação da base de dados

A estatística *Scan* de Kulldorff utiliza-se do banco de dados de acidentes de trânsito, consolidados em conjunto pelo DER e DETRAN do DF; das coordenadas geográficas do início e final dos 386 trechos rodoviários com fluxo veicular; além do volume médio diário de tráfego de cada trecho rodoviário implantado.

6.2.1 Obtenção da base de dados georreferenciada

Os dados de coordenadas geográficas dos 386 trechos rodoviários foram coletados a partir da base $SICAD$ 1:10.000. As coordenadas de alguns trechos foram coletadas posteriormente, por meio de coleta de campo feita por técnicos do DER/DF. A base $SICAD$ são as cartas (mapas) com o desenho das rodovias, curvas de nível, limites de parques e edificações; perfazendo toda a área do DF. A base $SICAD$ encontra-se em arquivo gráfico. A partir do software Microstation é possível desenhar o traçado das rodovias. As coordenadas encontram-se na base grau-decimal.

6.2.2 Obtenção dos dados de acidentes de trânsito fatais

O banco de dados de acidentes de trânsito é resultado do Sistema de Informações de Trânsito, que envolve vários órgãos do Distrito Federal (DETRAN, PCDF, PMDF, DER e SES). Os principais alicerces desse sistema são as regras da Associação Brasileira de Normas Técnicas, $ABNT$.

A coleta dos dados é feita através de quatro fontes oficiais (Polícia Civil, Instituto Médico Legal, Secretaria de Saúde e Instituto de Criminalística), com cruzamento das informações de todas as instituições.

As informações preenchidas nas Delegacias Policiais, através do boletim de ocorrência, são as únicas fontes para obtenção dos dados de acidentes de trânsito com vítima não fatal. O formulário (boletim de ocorrência) é único, tanto para o registro dos acidentes, como para as demais ocorrências.

Do Instituto de Médico Legal são obtidas informações adicionais sobre as vítimas fatais no trânsito, tais quais a dosagem de alcoolemia e a verificação do óbito após a data do acidente.

O número de acidentes fatais nem sempre é o mais adequado para representar a incidência de acidentes ao longo do mapa rodoviário, uma vez que os trechos rodoviários variam de tamanho e de volume de tráfego.

Uma forma de analisar descritivamente a distribuição de casos de acidentes é através da utilização de índices. Os índices são compostos pelas covariáveis VMD (volume médio diário) de veículos e extensão do trecho rodoviário.

As figuras 6.3 e 6.4 ilustram a distribuição dos índices 1 e 4, conforme visto na

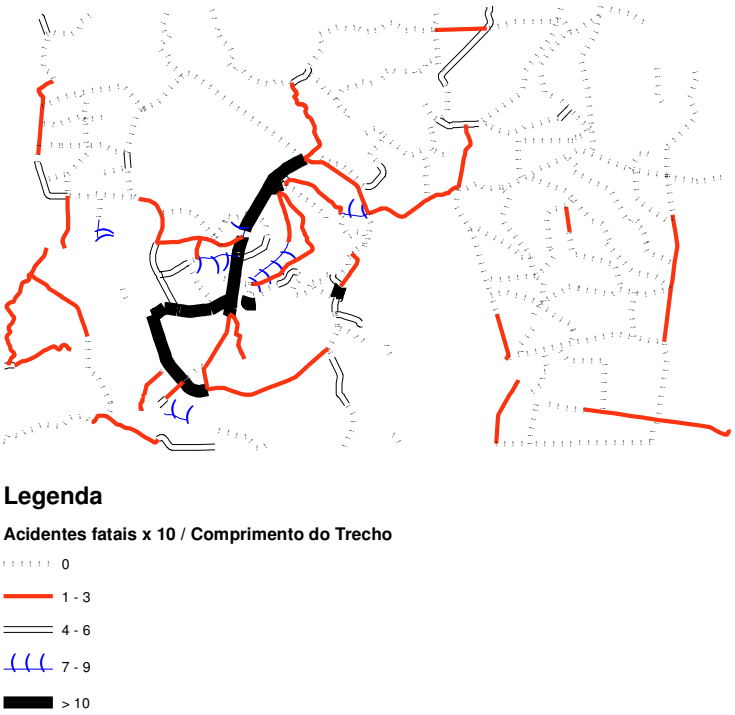

Figura 6.3: Incidência de acidentes fatais por comprimento do trecho.

subseção 2.4.4.

6.2.3 Obtenção dos VMD's por trecho rodoviário

O volume médio diário (VMD) é a estimativa do total de tráfego diário dentro de um determinado trecho rodoviário. Esse valor pode ser estimado pelo registro veicular dos equipamentos de fiscalização eletrônica, bem como pela contagem manual realizada pelos técnicos do DER/DF.

Toda obra na área de trânsito é feita através de estudos de engenharia que viabilizem e justifiquem tal gasto público, portanto, sempre que necessário, técnicos da área de estatística coletam dados sobre o fluxo de pedestre e de veículos. Além disso, estudos de impacto de trânsito, denominados Pólos Geradores de Trânsito também demandam tais informações.

Além das demandas ligadas à área de planejamento urbano, o Sistema Rodoviário do Distrito Federal(2012) (2012) também requer os VMD's por trecho rodoviário, o

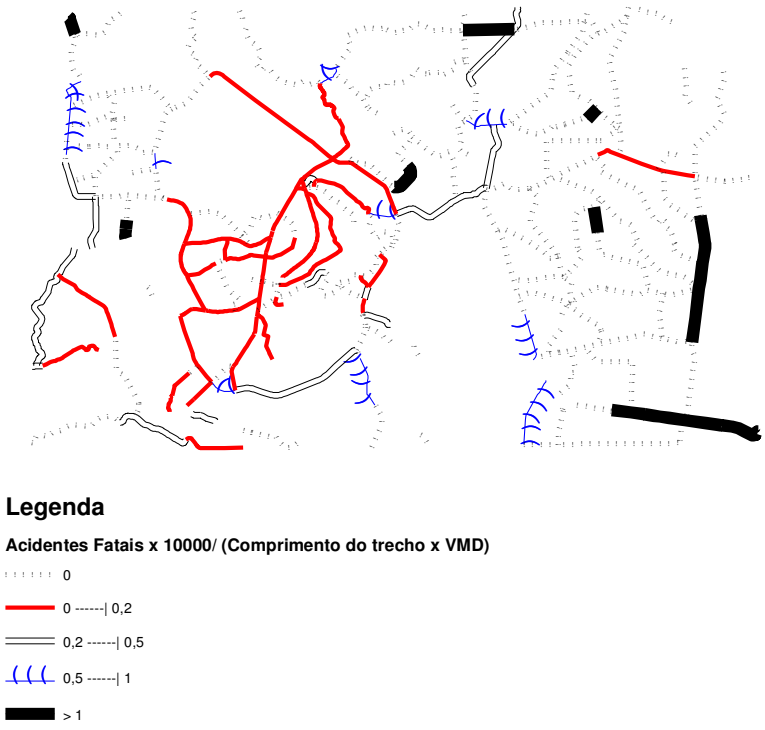

Figura 6.4: Incidência de acidentes fatais por comprimento do trecho e tráfego.

que nem sempre é possível devido ao número elevado de trechos rodoviários implantados. Os dados estatísticos levantados pelos técnicos em campo não são contabilizados nas 24 horas, sendo restritos aos horários de pico da manhã e da tarde.

A estimativa do total de tráfego por trecho depende basicamente da natureza e tempo de coleta:

- Coleta de tráfego por meio dos equipamentos de fiscalização

  Necessidade de verificação se houve falha no equipamento em algum intervalo de tempo. A seguir, o total de tráfego diário para o ano é calculado através do volume médio de veículos desconsiderando os meses e dias atípicos. Se o trecho é duplicado, o VMD será dado pela soma dos equipamentos no sentido crescente e decrescente. Caso só haja um equipamento, o VMD será multiplicado por 2.

- Coleta de tráfego manual

  Esse tipo de coleta necessita de expansão dos dados para 24 horas. Primeiramente define-se um ponto de contagem 24 horas, geralmente um equipamento de fiscalização eletrônica, e aplica-se a mesma proporção de tráfego para os dados da coleta manual. A coleta manual sempre é feita considerando os dois sentidos da rodovia e horários de pico da manhã e da tarde.

Tanto para a coleta eletrônica quanto para a manual é necessário atualização dos dados. Há vários métodos de previsão de tráfego. Neste trabalho a atualização se deu através da aplicação do crescimento anual da frota de veículos licenciados. Portanto, um volume de tráfego de 2010 foi atualizado aplicando primeiramente o crescimento percentual da frota de 2010. Posteriormente aplicou-se o crescimento da frota de 2011, resultando numa estimativa para os dados de 2012.

O trabalho observacional consiste na utilização dos 4 (quatro) métodos de varredura para identificação de conglomerados no mapa rodoviário do Distrito Federal, 2012. Além da técnica de varredura circular de Kulldorff (1997), foram implementadas técnicas de varredura que fazem uso da estrutura de vizinhança para construção das zonas candidatas a conglomerados.

A figura 6.6 ilustra os conglomerados detectados por cada um dos métodos de varredura.

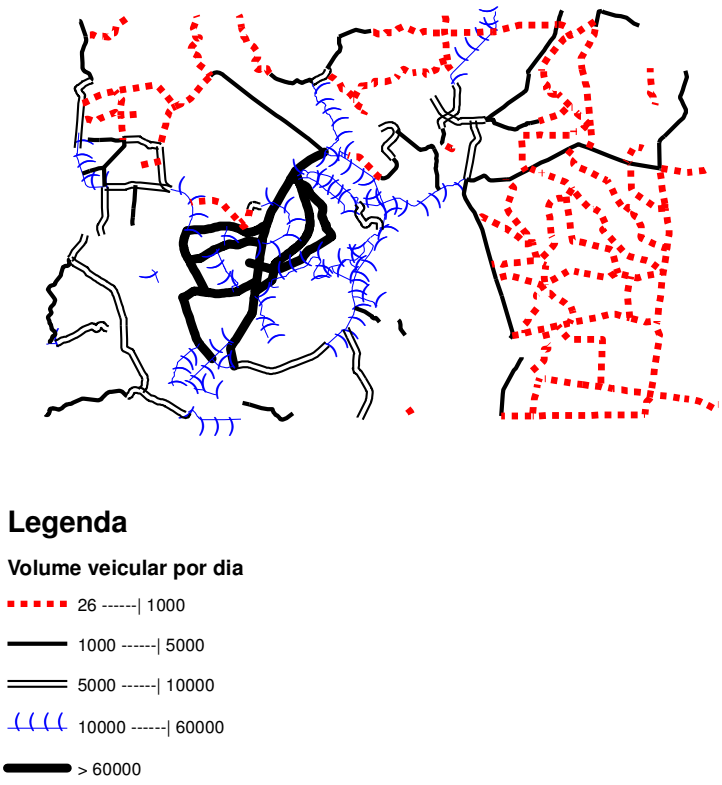

Figura 6.5: Mapa rodoviário com a estimativa do volume médio diário de veículos, 2012

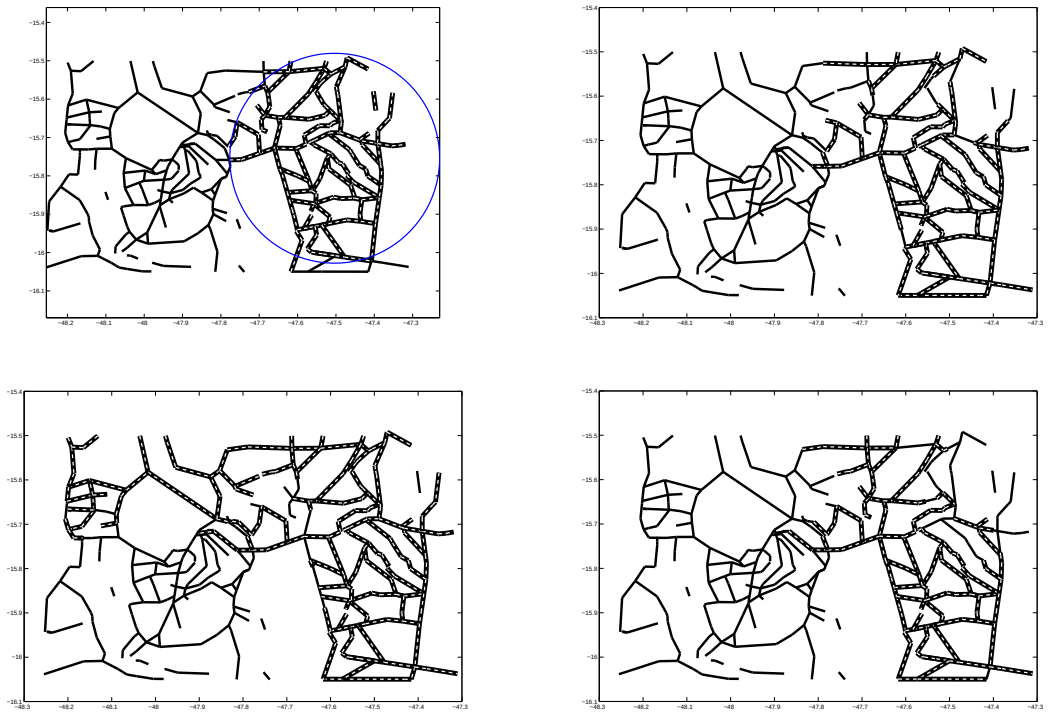

Figura 6.6: Conglomerados de acidentes fatais na malha rodoviária do DF, 2012. Superior à esquerda: *Scan* Circular. Superior à direita: *Scan* de Vizinhança Aleatória. Inferior à esquerda: *Scan* de Vizinhança Otimizada. Inferior à direita: *Scan* de Vizinhança Proporcional.

Todos os métodos detectaram rodovias pertencentes a região administrativa de São Sebastião, localizada na parte sudeste do mapa, inclusive a DF-100, que fica na parte leste extrema, bem próxima à divisa DF/GO. Essas rodovias, em quase sua totalidade, têm característica de serem não pavimentadas, bastante distantes do centro de Brasília e próximas ao 4º distrito rodoviário do DER/DF. São rodovias não duplicadas. Além disso, não há equipamentos de fiscalização eletrônica de velocidade, bem como câmeras de monitoramento de tráfego. Por serem rodovias distantes do centro do DF, a fiscalização por agentes de trânsito se tornam dispensiosas e com grande dificuldade de locomoção, devido à falta de pavimento asfáltico. Os métodos de vizinhança detectaram a DF-295 como conglomerado, já o *Scan* Circular (S.C) não detectou nenhum trecho dessa rodovia como conglomerado. Diferentemente dos outros métodos, o *Scan* de Vizinhança Proporcional (S.V.P) não detectou o trecho da DF-310 com início na entrada DF-250 até a entrada da VC-177(A). O *Scan* de Vizinhança Otimizada (S.V.O) detectou toda a DF-100 como conglomerado, já o S.C. só deixou de detectar o último trecho da DF-100, com início na entrada da DF-285 e final na entrada da DF-295. Já o S.V.A. não detectou os dois primeiros trechos da DF-100, que vai da entrada da BR-020 até entrada da VC-141. O S.V.P só identificou o trecho da DF-100 que vai da entrada da VC-169(A) até a entrada da DF-285. O método S.C. e S.V.P. detectaram toda a DF-130 como conglomerado. O S.V.O. e S.V.A não identificaram todos os trechos da DF-130. O S.V.O só identificou 7 (sete) trechos da DF-130.

Os métodos também detectaram conglomerados localizados na parte nordeste do mapa, correspondendo à região administrativa de Planaltina, região sob circunscrição do 1º distrito rodoviário do DER/DF. A maioria das rodovias dessa região são não pavimentadas, raro adensamento populacional e com pouco fluxo veicular. Todas rodovias são de mão simples. A DF-205 alterna trechos pavimentados e não pavimentados, sua geometria é de longas retas, sendo cortada por pelo menos quatro outras rodovias. A DF-128, apesar de ser asfaltada, não é duplicada e apresenta grande fluxo de veículos advindos de Planaltina e do Arapoanga. Algumas rodovias pavimentadas podem ser destacadas, como a DF-100, DF-130, DF-230, DF-250 e DF-345. Essas DF's possuem maior fluxo veicular e passam por regiões com maior adensamento populacional, como as regiões do Arapoanga e Vale do Amanhecer. O *Scan* Circular

(S.C) detectou menos trechos da DF-205, rodovia que corta de oeste a leste a parte norte do mapa. O *Scan* de Vizinhança Aleatória (S.V.A) identificou o trecho não pavimentado da DF-205 que começa no final da pavimentação e vai até a entrada da DF-131. O *Scan* de Vizinhança Proporcional (S.V.P) detectou, além desse trecho, o trecho pavimentado que vai da VC-201(B) até o final da pavimentação. O *Scan* de Vizinhança Otimizada (S.V.O) detectou como conglomerado toda a DF-205. A DF-110 também apresentou divergência quanto à detecção dos 4 (quatro) métodos. O S.C. identificou toda a DF-110 como conglomerado, já o S.V.A não identificou o trecho que vai da entrada da DF-205 e vai até a entrada da VC-113. O S.V.P. não identificou nenhum trecho da DF-110. O S.C. detectou toda a DF-405 e DF-410 como conglomerado, enquanto que o S.V.A. só identificou um trecho da DF-405 e DF-410, um que vai da entrada da VC-113 até a entrada da DF-205 e outro que vai da entrada da VC-127 até entrada da DF-230, respectivamente. O S.V.P. não identificou nenhum conglomerado na DF-405 e DF-410. O S.C. identificou toda a DF-250 como conglomerado. O S.V.A não identificou um trecho da DF-250, o que começa na entrada da DF-110 até entrada da VC-151. O S.V.P não identificou o trecho que vai da entrada da DF-320 e até entrada da DF-110, além do trecho que vai da entrada da VC-155 até a divisa DF/GO.

A parte noroeste do mapa correspondente às regiões do Lago Oeste e Brazlândia só foi detectado pelo *Scan* de Vizinhança Otimizada, que detectou a parte noroeste da DF-205, a parte final da DF-001 que passa pelo Lago Oeste. Além disso, o S.V.O. identificou trechos próximos a Brazlândia, como a DF-206, DF-220, DF-415, DF-430, DF-435 e DF-445. Essas DF's estão sob circunscrição do 5º distrito rodoviário. A DF-430 liga Brazlândia a DF-001, com tráfego de veículos em torno de 5000 veículos por dia e sendo toda pavimentada. A DF-445 é parcialmente pavimentada, com baixo tráfego veicular, ligando a DF-180 à DF-220.

Da região central do mapa, o S.V.O. foi o que identificou o maior número de trechos como conglomerados. O S.V.O. identificou 5 (cinco) trechos contíguos da DF-001, que vai da entrada da DF-430 até a entrada da DF-442. O S.V.O. também identificou 2 (dois) trechos da DF-005, 1 (um) trecho da DF-006 e 1 (um) da DF-015. O S.V.A. e S.V.O. e S.V.P identificaram o mesmo trecho da DF-001, que vai da entrada da VC-263 até a entrada da DF-250, trecho próximo ao Itapoã, Lago Norte e Paranoá.

O S.C. não identificou nenhum conglomerado na DF-001. O S.V.P. identificou 3 (três) trechos como conglomerado na parte central do mapa, 1 (um) na DF-001, 1 (um) na DF-005 e 1 (um) na DF-015. Apesar dos métodos terem identificado a DF-330 e parte da DF-440, detectaram poucos conglomerados próximos ao Plano Piloto (região central do mapa), com destaque para o S.V.P. e S.V.O. que identificaram conglomerados na DF-015, DF-005 e DF-001.

Os métodos S.C., S.V.A. e S.V.P apresentaram conglomerados na mesma região do mapa, com a diferença que o S.C. não detectou alguns trechos isolados da região sudeste e nordeste do mapa, além de não ter identificado trechos na parte central do mapa. O S.V.O foi o que apresentou o conglomerado com maior número de trechos rodoviários, principalmente por ter identificado trechos na parte noroeste do mapa. O S.V.P é o que possui o conglomerado com menor número de trechos rodoviários, tendo identificado trechos nas regiões sudeste, nordeste e central do mapa rodoviário.

A tabela 6.2.3 apresenta resultados descritivos dos métodos de varredura, onde "O" é o número de observado de casos e "E" é o número esperado de casos.

Tabela 6.2: Resultado métodos de varredura

Métodos	Total de trechos	Tráfego total	O.	E.	$\frac{O}{E}$	LLR	p-valor
S.C	166	263584	30	7,54	3,98	20,48	$< 0,001$
S.V.A	163	268493	33	7,68	4,29	24,76	$< 0,001$
S.V.O	217	446133	54	12,77	4,22	42,27	$< 0,001$
S.V.P	145	276038	34	7,90	4,30	25,63	$< 0,001$

O S.V.P foi o que apresentou maior valor da razão $\frac{O}{E}$. Também foi o método que identificou o conglomerado com o menor número de trechos rodoviários. O S.C foi o que apresentou o menor número de valores observados, possuindo também o conglomerado com menor tráfego. O S.V.O foi o método que identificou o conglomerado com maior número de trechos rodoviários, maior tráfego, maior valor observado e maior valor de LLR.

A tabela 6.2.3 apresenta os resultados comparativos entre os conglomerados detectados pelos métodos de varredura e a dimensão da região do estudo (mapa rodoviário, 2012).

Os conglomerados detectados pelos métodos de varredura englobam em média 44,75% do total do número de trechos rodoviários. Apesar dos conglomerados detectados possuírem pequena proporção do total de casos, apresentam pouco tráfego, o

Tabela 6.3: Dimensão dos conglomerados detectados

Métodos	$\frac{\text{Total de trechos do conglomerado}}{\text{Total de trechos}}$	$\frac{\text{Tráfego do conglomerado}}{\text{Tráfego total}}$	$\frac{\text{Número de casos no conglomerado}}{\text{Número total de casos}}$
S.C	0,43	0,042	0,1685
S.V.A	0,42	0,043	0,1853
S.V.O	0,56	0,071	0,3033
S.V.P	0,37	0,044	0,1910

que os reforçam como possíveis conglomerados de acidentes.

Os conglomerados detectados ocupam grande proporção do mapa rodoviário, 2012. Mesmo englobando muitas regiões do mapa, o tráfego total das regiões formadoras dos conglomerados apresentam percentual pequeno em relação ao tráfego total, ficando a sua maioria abaixo de 5%.

Capítulo 7

Conclusão

Os métodos de vizinhança tomam partido da estrutura intrínseca da malha rodoviária. Esses métodos consideram, na construção de zonas candidatas, apenas os trechos rodoviários que têm ligação com a zona atual. O *Scan* Circular agrupa regiões pelo critério de menor distância entre os centróides das regiões, utilizando-se da matriz das distâncias. Comparando o desempenho dos métodos de varredura, o *Scan* Circular se mostrou mais sensível a alterações na geometria do conglomerado. Os métodos de vizinhança são mais robustos a essas alterações e apresentaram resultados satisfatórios para os três tipos de conglomerados gerados artificialmente. A princípio três métodos de vizinhança foram implementados, cada um com critério específico de seleção de regiões adjacentes. A forma de escolha das regiões vizinhas afeta pouco a eficiência do método. O mais relevante para o desempenho do método é o critério de construção de zonas candidatas a conglomerados. Ao invés da seleção de regiões mais próximas, com a utilização da matriz das distâncias, escolhemos as regiões que fazem fronteira, por meio da matriz de vizinhança. O *Scan* Circular (S.C.) se comporta bem na detecção de conglomerados com formato circular, todavia não obteve bom desempenho nos conglomerados artificiais II e III. Já o *Scan* de Vizinhança Aleatória (S.V.A.) e *Scan* de Vizinhança Proporcional (S.V.P.) tiveram desempenho bom para os 3 (três) formatos de conglomerados. O *Scan* de Vizinhança Otimizada (S.V.O.) não apresentou resultado satisfatório para nenhum dos formatos de conglomerado, sendo um algoritmo que provavelmente falha por sempre escolher um máximo local.

O S.C. vai anexando os trechos rodoviários mais próximos, enquanto que os méto-

dos de vizinhança escolhem sempre um trecho rodoviário adjacente à zona inicial. O S.V.A. escolhe um vizinho de forma aleatória, com probabilidade igual para todas as regiões candidatas. O S.V.A. é o algoritmo mais simples dos métodos de vizinhança. O S.V.O. sempre escolhe o candidato que incorpora maior valor do *log* da razão de verossimilhança, não importando se a escolha será boa para a iteração seguinte, tomando decisões com base na iteração corrente. O S.V.O. é um um algoritmo guloso ou ganancioso. O S.V.P. é um algoritmo que seleciona os candidatos de forma aleatória, todavia ele aplica probabilidades de seleção proporcionais ao *log* da razão de verossimilhança da zona resultante. Sendo assim, nem sempre toma a melhor decisão, podendo optar por uma escolha não-ótima à priori, para na iteração seguinte conseguir resultados mais satisfatórios. É o meio termo entre o algoritmo que escolhe os vizinhos de forma totalmente aleatória (S.V.A.) e aquele que seleciona sempre o melhor vizinho (S.V.O.). O S.V.P. é o algoritmo mais complexo entre todos os algoritmos de varredura. Em cada iteração, o método S.V.P. calcula o *log* da razão de verossimilhança para as zonas resultantes, selecionando aleatoriamente com probabilidades proporcionais a esses valores de verossimilhança. O S.V.A. e o S.V.P. sempre serão algoritmos não determinísticos. O S.C. são algoritmos determinísticos, já que com o mesmo conjunto de dados, seu resultado sempre será o mesmo. O S.V.O. também é determinístico, salvo em casos de empates entres as zonas candidatas. Nesses casos, a escolha é aleatória.

Os *Scan* Circular e os métodos *Scan* de Vizinhança identificaram conglomerados de acidentes fatais bastante similares, com pequenas particularidades descritas no capítulo 6. Os métodos identificaram regiões predominantemente no sudeste e nordeste do mapa rodoviário. O método *Scan* circular não detectou nenhum trecho pertencente à parte central do mapa, diferentemente dos métodos de vizinhança, que detectaram algumas regiões próximas ao Plano Piloto. O *Scan* de vizinhança otimizada foi o único a identificar trechos rodoviários pertencentes à parte norte e noroeste do mapa rodoviário. As rodovias pertencentes às regiões sudeste e nordeste têm em comum o fato de serem rodovias, em sua maioria, não pavimentadas, com pouco tráfego veicular e distantes do centro do Plano Piloto. A parte sudeste do mapa certamente é a região mais isolada, com características rurais. As regiões sudeste e nordeste do mapa possuem poucos equipamentos de fiscalização e todas as rodovias são não duplicadas.

O teste numérico demonstrou que os métodos de vizinhança são mais robustos quanto ao formato do conglomerado, além de serem mais apropriados para estudos associados a acidentes de trânsito e a topologia da malha rodoviária. A detecção e identificação espacial do conglomerado permitiu reduzir aproximadamente dois terços do número de trechos rodoviários passíveis de serem analisados. Com a determinação espacial de trechos críticos de acidentes, políticas públicas podem ser melhor direcionadas visando a redução no número de acidentes de trânsito com alta severidade.

Um próximo passo do trabalho poderá ser a alteração da lógica de programação dos métodos de vizinhança, até mesmo do tipo de linguagem de programação, a fim de tornar os mesmos mais rápidos, devido a desvantagens dos mesmos diante do *Scan* Circular quanto à velocidade de simulação. Outro passo poderá ser testá-los para outros formatos de conglomerados e para outros conjuntos de dados reais, a fim de verificar o seu grau de robustez.

Referências Bibliográficas

ABNT. 1989. *Pesquisa de Acidentes de Trânsito*.

Anuário Estatístico de Acidentes de Trânsito do Distrito Federal (2010). 2010. *Anuário Estatístico de Acidentes de Trânsito do Distrito Federal*.

Besag, Julian, & Newell, James. 1991. The Detection of Clusters in Rare Diseases. *Journal of the Royal Statistical Society*, **154**(1), 143–155.

Casella, George, & Berger, Roger L. 2002. *Statistical Inference*. China: Thomson.

Choynowski, M. 1959. Maps based on probabilities. *Journal of the American Statistical Association*, **54**(286), 385–388.

Clayton, David, & Kaldor, John. 1987. Empirical Bayes Estimates of Age-standardized Relative Risks for Use in Disease Mapping. *Biometrics*, **43**(3), 671 – 681.

CNM. 2009. *Mapeamento das Mortes por Acidentes de Trânsito*.

Duczmal, Luiz, Tavares, Ricardo, Patil, Ganapati, & Cançado, André. 2010. Testing spatial cluster occurrence in maps equipped with environmentally defined structures. *Environ Ecol Stat*, **17**, 183–202.

Hammersley, J.M., & Handscombr, D. C. 1964. *Monte Carlo Methods*. england: Methuen.

Huang, Lan, Kulldorff, Martin, & Gregorio, David. 2007. A Spatial Scan Statistic for Survival Data. *Biometrics*, **63**, 109 – 118.

Kulldorff, M. 1997. Spatial scan statistic. *Communications in Statistics - Theory and Methods*, **26**(6), 1481–1496.

Kulldorff, M., & Nagarwalla, Neville. 1995. Spatial disease clusters:Detection and inference. *Statistics in medicine*, **14**, 799–810.

Kulldorff, Martin, Tango, Toshiro, & Park, Peter J. 2003. Power comparisons for disease clustering tests. *Computational Statistics & Data Analysis*, **42**, 665 – 684.

Metropolis, Nicholas, & Ulam, S. 1949. The Monte Carlo Method. *American Statistical Association*, **44**, 335 – 341.

Minamisava, Ruth, Nouer, Simonne S., de Morais Neto, Otaliba L., Melo, Lúcia Kamila, & Andrade, Ana Lúcia SS. 2009. Spatial clusters of violent deaths in a newly urbanized region of Brazil: highlighting the social disparities. *International journal of health geographics*, **14**, 8–66.

Naus, Joseph I. 1965a. Clustering of random points in two dimensions. *Biometrics*, **52**(1/2), 263 – 267.

Naus, Joseph I. 1965b. The distribution of the size of the maximum cluster of points on a line. *Journal of the American Statistical Association*, **60**(310), 532–538.

Openshaw, S., Charlton, M., Craft, A. W., & Birch, J. M. 1988. Investigation of leukaemia clusters by use of a geographical analysis machine. *The Lancet*, **74**(1), 272–273.

Sistema Rodoviário do Distrito Federal(2012). 2012. *Sistema Rodoviário do Distrito Federal*.

Whittemore, Alice S., Friend, Nina, Brown, Byron W., JR, & Holly, Elizabeth A. 1987. A test to detect clusters of disease. *Biometrika Trust*, **74**(3), 631 – 635.

http://www.curitiba.pr.gov.br/noticias/acidentes-de-transito-causam-43-mil-mortes-por-ano-no-brasil/28652

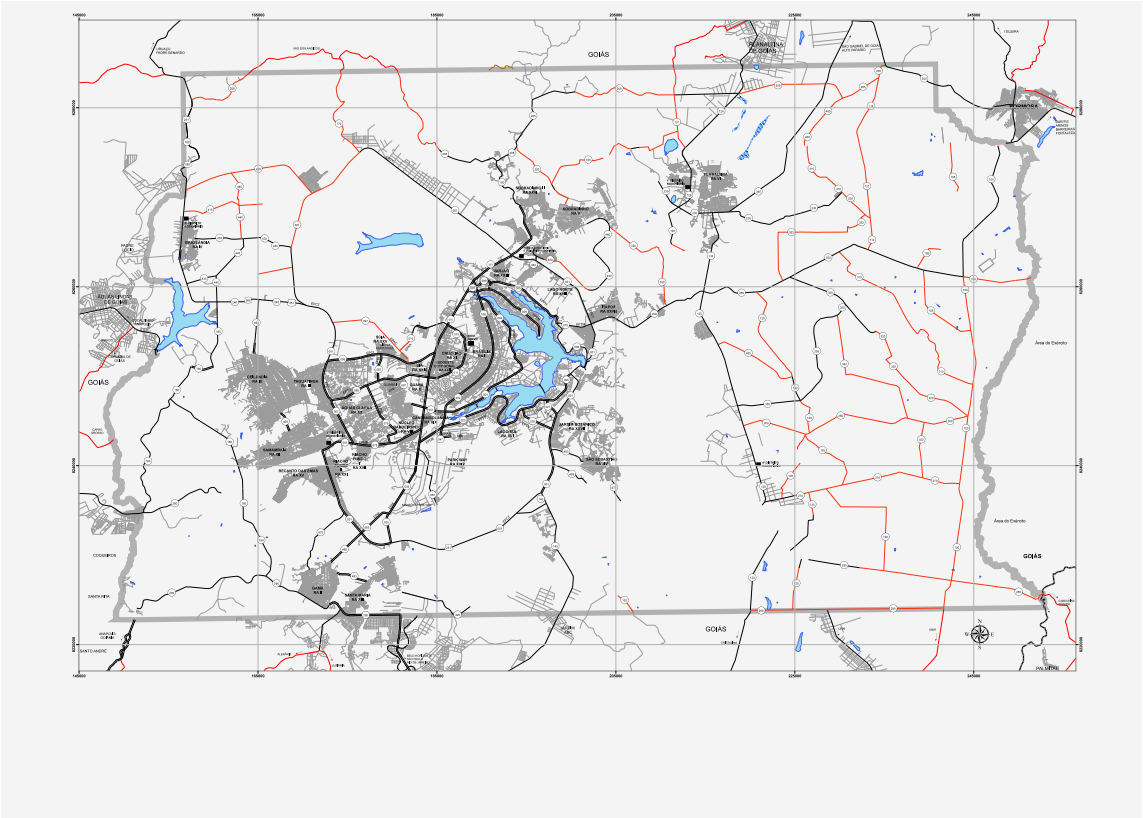

Figura 7.1: Mapa rodoviário do Distrito Federal, 2012

Tabela 7.1: Resultado conglomerados de acidentes fatais, 2012

Scan Circular		Scan Vizinhança Aleatória		Scan Vizinhança Otimizada		Scan Vizinhança Proporcional	
111	DF-100, entrada BR-020 até entrada da DF-230	4	DF-001, entrada VC-263 até entrada DF-250	1	DF-001, entrada BR-010 até entrada DF-442	4	DF-001, entrada VC-263 até entrada DF-250
112	DF-100, entrada DF-230 até entrada da VC-141	80	DF-015, entrada DF-005 até entrada DF-250	4	DF-001, entrada VC-263 até entrada DF-250	61	DF-005, entrada DF-442 até entrada DF-015
113	DF-100, entrada VC-141 até entrada da VC-145	113	DF-100, entrada VC-141 até entrada da VC-145	29	DF-001, entrada DF-435 até entrada DF-430	80	DF-015, entrada DF-005 até entrada DF-250
114	DF-100, entrada VC-145 até entrada da VC-143	114	DF-100, entrada VC-145 até entrada da VC-143	30	DF-001, entrada DF-430 até entrada DF-415	117	DF-100, entrada VC-169(A) até entrada da DF-105
115	DF-100, entrada VC-143 até entrada da DF-250	115	DF-100, entrada VC-143 até entrada da DF-250	31	DF-001, entrada DF-415 até entrada DF-220	118	DF-100, entrada DF-105 até entrada da VC-169(B)
116	DF-100, entrada DF-250 até entrada da VC-169(A)	116	DF-100, entrada DF-250 até entrada da VC-169(A)	32	DF-001, entrada DF-220 até entrada DF-170	119	DF-100, entrada VC-169(B) até entrada da DF-310
117	DF-100, entrada VC-169(A) até entrada da DF-105	117	DF-100, entrada VC-169(A) até entrada da DF-105	33	DF-001, entrada DF-170 até entrada BR-010	120	DF-100, entrada DF-310 até entrada da DF-320
118	DF-100, entrada DF-105 até entrada da VC-169(B)	118	DF-100, entrada DF-105 até entrada da VC-169(B)	60	DF-005, entrada DF-006 até entrada DF-442	121	DF-100, entrada DF-320 até entrada da VC-421
119	DF-100, entrada VC-169(B) até entrada da DF-310	119	DF-100, entrada VC-169(B) até entrada da DF-310	61	DF-005, entrada DF-442 até entrada DF-015	122	DF-100, entrada VC-421 até entrada da DF-270
120	DF-100, entrada DF-310 até entrada da DF-320	120	DF-100, entrada DF-310 até entrada da DF-320	63	DF-006, entrada DF-007 até entrada DF-005	123	DF-100, entrada DF-270 até entrada da DF-285
121	DF-100, entrada DF-320 até entrada da VC-421	121	DF-100, entrada DF-320 até entrada da VC-421	80	DF-015, entrada DF-005 até entrada DF-250	126	DF-105, entrada VC-143 até entrada DF-250(A)
122	DF-100, entrada VC-421 até entrada da DF-270	122	DF-100, entrada VC-421 até entrada da DF-270	111	DF-100, entrada BR-020 até entrada da DF-230	128	DF-105, entrada DF-250(B) até entrada VC-155
123	DF-100, entrada DF-270 até entrada da DF-285	123	DF-100, entrada DF-270 até entrada da DF-285	112	DF-100, entrada DF-230 até entrada da VC-141	129	DF-105, entrada VC-155 até entrada DF-100
125	DF-105, entrada BR-020 até Fim do trecho implantado	124	DF-100, entrada DF-285 até entrada DF-295	113	DF-100, entrada VC-141 até entrada da VC-145	136	DF-120, entrada DF-250 até entrada DF-455
126	DF-105, entrada VC-143 até entrada DF-250(A)	126	DF-105, entrada VC-143 até entrada DF-250(A)	114	DF-100, entrada VC-145 até entrada da VC-143	137	DF-120, entrada DF-455 até entrada DF-355
127	DF-105, entrada DF-250(A) até entrada DF-250(B)	127	DF-105, entrada DF-250(A) até entrada DF-250(B)	115	DF-100, entrada VC-143 até entrada da DF-250	138	DF-120, entrada DF-355 até entrada VC-419
128	DF-105, entrada DF-250(B) até entrada VC-155	128	DF-105, entrada DF-250(B) até entrada VC-155	116	DF-100, entrada DF-250 até entrada da VC-169(A)	139	DF-120, entrada VC-419 até entrada DF-260(A)
129	DF-105, entrada VC-155 até entrada DF-100	129	DF-105, entrada VC-155 até entrada DF-100	117	DF-100, entrada VC-169(A) até entrada da DF-105	140	DF-120, entrada DF-260(A) até entrada DF-260(B)
130	DF-110, entrada DF-205 até VC-113	131	DF-110, entrada VC-113 até BR-020	118	DF-100, entrada DF-105 até entrada da VC-169(B)	141	DF-120, entrada DF-125 até entrada VC-427
131	DF-110, entrada VC-113 até BR-020	132	DF-110, entrada BR-020 até VC-121	119	DF-100, entrada VC-169(B) até entrada da DF-310	142	DF-120, entrada VC-427 até entrada DF-270(A)
132	DF-110, entrada BR-020 até VC-121	133	DF-110, entrada VC-121 até DF-230(A)	120	DF-100, entrada DF-310 até entrada da DF-320	143	DF-120, entrada DF-270(A) até entrada DF-270(B)
133	DF-110, entrada VC-121 até DF-230(A)	134	DF-110, entrada DF-230(A) até DF-230(B)	121	DF-100, entrada DF-320 até entrada da VC-421	144	DF-120, entrada DF-270(B) até entrada DF-285
134	DF-110, entrada DF-230(A) até DF-230(B)	135	DF-110, entrada DF-230(B) até DF-250	122	DF-100, entrada VC-421 até entrada da DF-270	145	DF-125, entrada DF-120 até início de trecho planejado
135	DF-110, entrada DF-230(B) até DF-250	136	DF-120, entrada DF-250 até entrada DF-455	123	DF-100, entrada DF-270 até entrada da DF-285	146	DF-125, fim de trecho planejado até entrada DF-270 (A)
136	DF-120, entrada DF-250 até entrada DF-455	137	DF-120, entrada DF-455 até entrada DF-355	124	DF-100, entrada DF-285 até entrada DF-295	147	DF-125, entrada DF-270(A) até fim do trecho planejado
137	DF-120, entrada DF-455 até entrada DF-355	138	DF-120, entrada DF-355 até entrada VC-419	126	DF-105, entrada VC-143 até entrada DF-250(A)	148	DF-125, fim do trecho planejado até entrada DF-270(B)
138	DF-120, entrada DF-355 até entrada VC-419	139	DF-120, entrada VC-419 até entrada DF-260(A)	127	DF-105, entrada DF-250(A) até entrada DF-250(B)	149	DF-125, entrada DF-270(B) até córrego Lamarão
139	DF-120, entrada VC-419 até entrada DF-260(A)	140	DF-120, entrada DF-260(A) até entrada DF-260(B)	128	DF-105, entrada DF-250(B) até entrada VC-155	150	DF-125, córrego Lamarão até entrada BR-251 (A)
140	DF-120, entrada DF-260(A) até entrada DF-260(B)	141	DF-120, entrada DF-125 até entrada VC-427	129	DF-105, entrada VC-155 até entrada DF-100	151	DF-125, entrada BR-251(A) até entrada DF-251(B)
141	DF-120, entrada DF-125 até entrada VC-427	142	DF-120, entrada VC-427 até entrada DF-270(A)	130	DF-110, entrada DF-205 até VC-113	152	DF-125, entrada DF-251(B) até fim do trecho planejado
142	DF-120, entrada VC-427 até entrada DF-270(A)	143	DF-120, entrada DF-270(A) até entrada DF-270(B)	131	DF-110, entrada VC-113 até BR-020	153	DF-125, fim do trecho pavimentado até entrada DF-295
143	DF-120, entrada DF-270(A) até entrada DF-270(B)	144	DF-120, entrada DF-270(B) até entrada DF-285	132	DF-110, entrada BR-020 até VC-121	154	DF-128, divisa GO/DF até entrada DF-205
144	DF-120, entrada DF-270(B) até entrada DF-285	145	DF-125, entrada DF-120 até início de trecho planejado	133	DF-110, entrada VC-121 até DF-230(A)	155	DF-128, entrada DF-205 até entrada DF-131
145	DF-125, entrada DF-120 até início de trecho planejado	146	DF-125, fim de trecho planejado até entrada DF-270 (A)	134	DF-110, entrada DF-230(A) até DF-230(B)	156	DF-128, entrada DF-131 até entrada BR-010
146	DF-125, fim de trecho planejado até entrada DF-270 (A)	147	DF-125, entrada DF-270(A) até fim do trecho planejado	135	DF-110, entrada DF-230(B) até DF-250	157	DF-128, entrada BR-010 até acesso a Planaltina
147	DF-125, entrada DF-270(A) até fim do trecho planejado	148	DF-125, fim do trecho planejado até entrada DF-270(B)	136	DF-120, entrada DF-250 até entrada DF-455	158	DF-128, acesso a Planaltina até entrada DF-230
148	DF-125, fim do trecho planejado até entrada DF-270(B)	149	DF-125, entrada DF-270(B) até córrego Lamarão	137	DF-120, entrada DF-455 até entrada DF-355	159	DF-128, entrada DF-230 até acesso ao instituto federal de Brasília
149	DF-125, entrada DF-270(B) até córrego Lamarão	150	DF-125, córrego Lamarão até entrada BR-251 (A)	138	DF-120, entrada DF-355 até entrada VC-419	160	DF-128, acesso ao instituto federal de Brasília até entrada DF-444

Scan Circular		Scan Vizinhança Aleatória		Scan Vizinhança Otimizada		Scan Vizinhança Proporcional	
150	DF-125, córrego Lamarão até entrada BR-251 (A)	151	DF-125, entrada BR-251(A) até entrada DF-251(B)	139	DF-120, entrada VC-419 até entrada DF-260(A)	161	DF-128, entrada DF-444 até Pedra Fundamental
151	DF-125, entrada BR-251(A) até entrada DF-251(B)	152	DF-125, entrada DF-251(B) até fim do trecho planejado	140	DF-120, entrada DF-260(A) até entrada DF-260(B)	162	DF-130, entrada DF-230 até acesso ao Vale do Amanhecer
152	DF-125, entrada DF-251(B) até fim do trecho planejado	153	DF-125, fim do trecho pavimentado até entrada DF-295	141	DF-120, entrada DF-125 até entrada VC-427	163	DF-130, acesso ao Vale do Amanhecer até entrada DF-250
154	DF-128, divisa GO/DF até entrada DF-205	154	DF-128, divisa GO/DF até entrada DF-205	142	DF-120, entrada VC-427 até entrada DF-270(A)	164	DF-130, entrada DF-250 até entrada DF-455
155	DF-128, entrada DF-205 até entrada DF-131	155	DF-128, entrada DF-205 até entrada DF-131	143	DF-120, entrada DF-270(A) até entrada DF-270(B)	165	DF-130, entrada DF-455 até entrada VC-411
156	DF-128, entrada DF-131 até entrada BR-010	156	DF-128, entrada DF-131 até entrada BR-010	144	DF-120, entrada DF-270(B) até entrada DF-285	166	DF-130, entrada VC-411 até entrada VC-413
157	DF-128, entrada BR-010 até acesso a Planaltina	159	DF-128, entrada DF-230 até acesso ao instituto federal de Brasília	145	DF-125, entrada DF-120 até início de trecho planejado	167	DF-130, entrada VC-413 até entrada DF-355
158	DF-128, acesso a Planaltina até entrada DF-230	162	DF-130, entrada DF-230 até acesso ao Vale do Amanhecer	146	DF-125, fim de trecho planejado até entrada DF-270 (A)	168	DF-130, entrada DF-355 até entrada DF-260
159	DF-128, entrada DF-230 até acesso ao instituto federal de Brasília	163	DF-130, acesso ao Vale do Amanhecer até entrada DF-250	147	DF-125, entrada DF-270(A) até fim do trecho planejado	169	DF-130, entrada DF-260 até entrada VC-401
160	DF-128, acesso ao instituto federal de Brasília até entrada DF-444	164	DF-130, entrada DF-250 até entrada DF-455	148	DF-125, fim do trecho planejado até entrada DF-270(B)	170	DF-130, entrada VC-401 até acesso ao 4º Distrito Rodoviário
161	DF-128, entrada DF-444 até Pedra Fundamental	165	DF-130, entrada DF-455 até entrada VC-411	149	DF-125, entrada DF-270(B) até córrego Lamarão	171	DF-130, acesso ao 4º Distrito Rodoviário até entrada DF-270
162	DF-130, entrada DF-230 até acesso ao Vale do Amanhecer	166	DF-130, entrada VC-411 até entrada VC-413	150	DF-125, córrego Lamarão até entrada BR-251 (A)	172	DF-130, entrada DF-270 até entrada BR-251(A)
163	DF-130, acesso ao Vale do Amanhecer até entrada DF-250	167	DF-130, entrada VC-413 até entrada DF-355	151	DF-125, entrada BR-251(A) até entrada DF-251(B)	173	DF-130, entrada BR-251(A) até entrada BR-251(B)
164	DF-130, entrada DF-250 até entrada DF-455	168	DF-130, entrada DF-355 até entrada DF-260	152	DF-125, entrada DF-251(B) até fim do trecho planejado	174	DF-130, entrada BR-251(B) até entrada DF-295
165	DF-130, entrada DF-455 até entrada VC-411	169	DF-130, entrada DF-260 até entrada VC-401	153	DF-125, fim do trecho pavimentado até entrada DF-295	175	DF-131, divisa GO/DF até entrada DF-205
166	DF-130, entrada VC-411 até entrada VC-413	170	DF-130, entrada VC-401 até acesso ao 4º Distrito Rodoviário	154	DF-128, divisa GO/DF até entrada DF-205	176	DF-131, entrada DF-205 até entrada DF-335
167	DF-130, entrada VC-413 até entrada DF-355	173	DF-130, entrada BR-251(A) até entrada BR-251(B)	155	DF-128, entrada DF-205 até entrada DF-131	177	DF-131, entrada DF-355 até entrada DF-128
168	DF-130, entrada DF-355 até entrada DF-260	174	DF-130, entrada BR-251(B) até entrada DF-295	156	DF-128, entrada DF-131 até entrada BR-010	214	DF-205, entrada VC-201(B) até fim da pavimentação
169	DF-130, entrada DF-260 até entrada VC-401	177	DF-131, entrada DF-355 até entrada DF-128	162	DF-130, entrada DF-230 até acesso ao Vale do Amanhecer	215	DF-205, fim da pavimentação até entrada DF-131
170	DF-130, entrada VC-401 até acesso ao 4º Distrito Rodoviário	215	DF-205, fim da pavimentação até entrada DF-131	165	DF-130, entrada DF-455 até entrada VC-411	216	DF-205, entrada DF-131 até entrada DF-128
171	DF-130, acesso ao 4º Distrito Rodoviário até entrada DF-270	216	DF-205, entrada DF-131 até entrada DF-128	170	DF-130, entrada VC-401 até acesso ao 4º Distrito Rodoviário	218	DF-205, entrada DF-345 até entrada DF-405
172	DF-130, entrada DF-270 até entrada BR-251(A)	217	DF-205, entrada DF-128 até entrada DF-345	171	DF-130, acesso ao 4º Distrito Rodoviário até entrada DF-270	219	DF-205, entrada DF-405 até entrada VC-103
173	DF-130, entrada BR-251(A) até entrada BR-251(B)	218	DF-205, entrada DF-345 até entrada DF-405	172	DF-130, entrada DF-270 até entrada BR-251(A)	227	DF-230, entrada BR-010 até entrada DF-128
174	DF-130, entrada BR-251(B) até entrada DF-295	219	DF-205, entrada DF-405 até entrada VC-103	173	DF-130, entrada BR-251(A) até entrada BR-251(B)	228	DF-230, entrada DF-128 até entrada DF-130
177	DF-131, entrada DF-355 até entrada DF-128	220	DF-205, entrada VC-103 até entrada DF-110	174	DF-130, entrada BR-251(B) até entrada DF-295	229	DF-230, entrada DF-130 até entrada DF-345
216	DF-205, entrada DF-131 até entrada DF-128	221	DF-205, entrada DF-110 até entrada GO-430	175	DF-131, divisa GO/DF até entrada DF-205	230	DF-230, entrada DF-345 até entrada VC-137(A)
217	DF-205, entrada DF-128 até entrada DF-345	222	DF-205, entrada GO-430 até divisa DF/GO	176	DF-131, entrada DF-205 até entrada DF-335	231	DF-230, entrada VC-137(A) até fim do trecho implementado
218	DF-205, entrada DF-345 até entrada DF-405	227	DF-230, entrada BR-010 até entrada DF-128	177	DF-131, entrada DF-355 até entrada DF-128	233	DF-230, entrada VC-127 até entrada DF-410
219	DF-205, entrada DF-405 até entrada VC-103	228	DF-230, entrada DF-128 até entrada DF-130	182	DF-150, entrada BR-010 até acesso a Sobradinho II	234	DF-230, entrada DF-410 até entrada VC-139
220	DF-205, entrada VC-103 até entrada DF-110	229	DF-230, entrada DF-130 até entrada DF-345	183	DF-150, acesso a Sobradinho II até entrada DF-205	235	DF-230, entrada VC-139 até entrada DF-353
221	DF-205, entrada DF-110 até entrada GO-430	230	DF-230, entrada DF-345 até entrada VC-137(A)	184	DF-170, divisa GO/DF até entrada DF-001	241	DF-250, entrada DF-001 até entrada DF-456
222	DF-205, entrada GO-430 até divisa DF/GO	231	DF-230, entrada VC-137(A) até fim do trecho implementado	185	DF-180, entrada BR-080(A) até entrada DF-206	242	DF-250, entrada DF-456 até entrada DF-330
227	DF-230, entrada BR-010 até entrada DF-128	232	DF-230, fim do trecho implementado até entrada VC-127	186	DF-180, entrada DF-206 até entrada VC-511	243	DF-250, entrada DF-330 até entrada DF-130
228	DF-230, entrada DF-128 até entrada DF-130	233	DF-230, entrada VC-127 até entrada DF-410	187	DF-180, entrada VC-511 até entrada DF-220	244	DF-250, entrada DF-130 até entrada DF-120
229	DF-230, entrada DF-130 até entrada DF-345	234	DF-230, entrada DF-410 até entrada VC-139	188	DF-180, entrada DF-220 até entrada DF-415	245	DF-250, entrada DF-120 até entrada VC-129(A)
230	DF-230, entrada DF-345 até entrada VC-137(A)	235	DF-230, entrada VC-139 até entrada DF-353	191	DF-180, entrada VC-541 até entrada DF-435	246	DF-250, entrada VC-129(A) até entrada DF-353
231	DF-230, entrada VC-137(A) até fim do trecho implementado	236	DF-230, entrada DF-353 até entrada DF-110(A)	192	DF-180, entrada DF-435 até entrada VC-547	247	DF-250, entrada DF-353 até entrada DF-320
232	DF-230, fim do trecho implementado até entrada VC-127	237	DF-230, entrada DF-110(A) até entrada DF-110(B)	193	DF-180, entrada VC-547 até entrada BR-080(B)	252	DF-250, entrada DF-110 até entrada VC-151
233	DF-230, entrada VC-127 até entrada DF-410	241	DF-250, entrada DF-001 até entrada DF-456	209	DF-205, divisa GO/DF até entrada VC-201 (A)	253	DF-250, entrada VC-151 até entrada DF-105(A)
234	DF-230, entrada DF-410 até entrada VC-139	242	DF-250, entrada DF-456 até entrada DF-330	210	DF-205, entrada VC-201(A) até início do trecho pavimentado	254	DF-250, entrada DF-105(A) até entrada DF-105(B)
235	DF-230, entrada VC-139 até entrada DF-353	243	DF-250, entrada DF-330 até entrada DF-130	211	DF-205, início do trecho pavimentado até entrada DF-150	255	DF-250, entrada DF-105(B) até entrada VC-155

Scan Circular		Scan Vizinhança Aleatória		Scan Vizinhança Otimizada		Scan Vizinhança Proporcional	
236	DF-230, entrada DF-353 até entrada DF-110(A)	244	DF-250, entrada DF-130 até entrada DF-120	212	DF-205, entrada DF-150 até entrada DF-326	260	DF-260, entrada DF-130 até entrada DF-120(A)
237	DF-230, entrada DF-110(A) até entrada DF-110(B)	245	DF-250, entrada DF-120 até entrada VC-129(A)	213	DF-205, entrada DF-326 até entrada VC-201(B)	261	DF-260, entrada DF-120(A) até entrada DF-120(B)
241	DF-250, entrada DF-001 até entrada DF-456	246	DF-250, entrada VC-129(A) até entrada DF-353	214	DF-205, entrada VC-201(B) até fim da pavimentação	262	DF-260, entrada DF-120(B) até entrada VC-407
242	DF-250, entrada DF-456 até entrada DF-330	247	DF-250, entrada DF-353 até entrada DF-320	215	DF-205, fim da pavimentação até entrada DF-131	263	DF-260, entrada VC-407 até entrada VC-419
243	DF-250, entrada DF-330 até entrada DF-130	248	DF-250, entrada DF-320 até entrada VC-129(B)	216	DF-205, entrada DF-131 até entrada DF-128	264	DF-260, entrada VC-419 até entrada VC-423
244	DF-250, entrada DF-130 até entrada DF-120	249	DF-250, entrada VC-129(B) até entrada DF-310	217	DF-205, entrada DF-128 até entrada DF-345	265	DF-260, entrada VC-423 até entrada DF-322(A)
245	DF-250, entrada DF-120 até entrada VC-129(A)	250	DF-250, entrada DF-310 até entrada VC-133	218	DF-205, entrada DF-345 até entrada DF-405	266	DF-260, entrada DF-322(A) até entrada DF-322(B)
246	DF-250, entrada VC-129(A) até entrada DF-353	251	DF-250, entrada VC-133 até entrada DF-110	219	DF-205, entrada DF-405 até entrada VC-103	267	DF-260, entrada DF-322(B) até entrada DF-100
247	DF-250, entrada DF-353 até entrada DF-320	253	DF-250, entrada VC-151 até entrada DF-105(A)	220	DF-205, entrada VC-103 até entrada DF-110	268	DF-270, entrada DF-130 até entrada DF-125(A)
248	DF-250, entrada DF-320 até entrada VC-129(B)	254	DF-250, entrada DF-105(A) até entrada DF-105(B)	221	DF-205, entrada DF-110 até entrada GO-430	269	DF-270, entrada DF-125(A) até fim do trecho pavimentado
249	DF-250, entrada VC-129(B) até entrada DF-310	255	DF-250, entrada DF-105(B) até entrada VC-155	222	DF-205, entrada GO-430 até divisa DF/GO	270	DF-270, fim do trecho pavimentado até entrada DF-125(B)
250	DF-250, entrada DF-310 até entrada VC-133	256	DF-250, entrada VC-155 até entrada DF-100	223	DF-206, entrada BR-080 até entrada VC-505	271	DF-270, entrada DF-125(B) até entrada DF-120(A)
251	DF-250, entrada VC-133 até entrada DF-110	257	DF-250, entrada DF-100 até entrada VC-159	224	DF-206, entrada VC-505 até divisa DF/GO	272	DF-270, entrada DF-120(A) até entrada DF-120(B)
252	DF-250, entrada DF-110 até entrada VC-151	258	DF-250, entrada VC-159 até entrada VC-145	225	DF-220, entrada BR-080 até entrada DF-445	273	DF-270, entrada DF-120(B) até entrada VC-407
253	DF-250, entrada VC-151 até entrada DF-105(A)	259	DF-250, entrada VC-145 até entrada BR-479	226	DF-220, entrada DF-445 até entrada DF-001	274	DF-270, entrada VC-407 até entrada DF-322
254	DF-250, entrada DF-105(A) até entrada DF-105(B)	260	DF-260, entrada DF-130 até entrada DF-120(A)	228	DF-230, entrada DF-128 até entrada DF-130	275	DF-270, entrada DF-322 até entrada DF-100
255	DF-250, entrada DF-105(B) até entrada VC-155	261	DF-260, entrada DF-120(A) até entrada DF-120(B)	229	DF-230, entrada DF-130 até entrada DF-345	279	DF-285, entrada BR-251 até entrada VC-441
256	DF-250, entrada VC-155 até entrada DF-100	262	DF-260, entrada DF-120(B) até entrada VC-407	230	DF-230, entrada DF-345 até entrada VC-137(A)	280	DF-285, entrada VC-441 até fim do trecho pavimentado
257	DF-250, entrada DF-100 até entrada VC-159	263	DF-260, entrada VC-407 até entrada VC-419	231	DF-230, entrada VC-137(A) até fim do trecho implementado	281	DF-285, fim do trecho pavimentado até DF-120
258	DF-250, entrada VC-159 até entrada VC-145	264	DF-260, entrada VC-419 até entrada VC-423	232	DF-230, fim do trecho implementado até entrada VC-127	282	DF-285, entrada DF-120 até VC-447
259	DF-250, entrada VC-145 até entrada BR-479	265	DF-260, entrada VC-423 até entrada DF-322(A)	233	DF-230, entrada VC-127 até entrada DF-410	283	DF-285, entrada VC-447 até DF-100
260	DF-260, entrada DF-130 até entrada DF-120(A)	266	DF-260, entrada DF-322(A) até entrada DF-322(B)	234	DF-230, entrada DF-410 até entrada VC-139	284	DF-285, entrada DF-100 até entrada VC-461
261	DF-260, entrada DF-120(A) até entrada DF-120(B)	267	DF-260, entrada DF-322(B) até entrada DF-100	235	DF-230, entrada VC-139 até entrada DF-353	285	DF-285, entrada VC-461 até divisa DF/MG
262	DF-260, entrada DF-120(B) até entrada VC-407	268	DF-270, entrada DF-130 até entrada DF-125(A)	236	DF-230, entrada DF-353 até entrada DF-110(A)	294	DF-295, entrada DF-130 até entrada DF-125
263	DF-260, entrada VC-407 até entrada VC-419	269	DF-270, entrada DF-125(A) até fim do trecho pavimentado	237	DF-230, entrada DF-110(A) até entrada DF-110(B)	295	DF-295, entrada DF-125 até entrada BR-251
264	DF-260, entrada VC-419 até entrada VC-423	270	DF-270, fim do trecho pavimentado até entrada DF-125(B)	241	DF-250, entrada DF-001 até entrada DF-456	296	DF-295, entrada BR-251 até entrada VC-471
265	DF-260, entrada VC-423 até entrada DF-322(A)	271	DF-270, entrada DF-125(B) até entrada DF-120(A)	242	DF-250, entrada DF-456 até entrada DF-330	297	DF-295, entrada VC-471 até entrada DF-100
266	DF-260, entrada DF-322(A) até entrada DF-322(B)	272	DF-270, entrada DF-120(A) até entrada DF-120(B)	243	DF-250, entrada DF-330 até entrada DF-130	302	DF-310, entrada VC-177(A) até VC-177(B)
267	DF-260, entrada DF-322(B) até entrada DF-100	273	DF-270, entrada DF-120(B) até entrada VC-407	244	DF-250, entrada DF-130 até entrada DF-120	303	DF-310, entrada VC-177(B) até DF-100
268	DF-270, entrada DF-130 até entrada DF-125(A)	274	DF-270, entrada VC-407 até entrada DF-322	245	DF-250, entrada DF-120 até entrada VC-129(A)	304	DF-320, entrada DF-250 até VC-403
269	DF-270, entrada DF-125(A) até fim do trecho pavimentado	275	DF-270, entrada DF-322 até entrada DF-100	246	DF-250, entrada VC-129(A) até entrada DF-353	305	DF-320, entrada VC-403 até DF-355
270	DF-270, fim do trecho pavimentado até entrada DF-125(B)	279	DF-285, entrada BR-251 até entrada VC-441	247	DF-250, entrada DF-353 até entrada DF-320	306	DF-320, entrada DF-355 até VC-165
271	DF-270, entrada DF-125(B) até entrada DF-120(A)	280	DF-285, entrada VC-441 até fim do trecho pavimentado	248	DF-250, entrada DF-320 até entrada VC-129(B)	307	DF-320, entrada VC-165 até fim do trecho pavimentado
272	DF-270, entrada DF-120(A) até entrada DF-120(B)	281	DF-285, fim do trecho pavimentado até DF-120	249	DF-250, entrada VC-129(B) até entrada DF-310	308	DF-320, fim do trecho pavimentado até entrada VC-173
273	DF-270, entrada DF-120(B) até entrada VC-407	282	DF-285, entrada DF-120 até VC-447	250	DF-250, entrada DF-310 até entrada VC-133	309	DF-320, entrada VC-173 até entrada VC-409
274	DF-270, entrada VC-407 até entrada DF-322	283	DF-285, entrada VC-447 até DF-100	251	DF-250, entrada VC-133 até entrada DF-110	310	DF-320, entrada VC-409 até entrada VC-417
275	DF-270, entrada DF-322 até entrada DF-100	284	DF-285, entrada DF-100 até entrada VC-461	252	DF-250, entrada DF-110 até entrada VC-151	311	DF-320, entrada VC-417 até entrada DF-100
279	DF-285, entrada BR-251 até entrada VC-441	285	DF-285, entrada VC-461 até divisa DF/MG	253	DF-250, entrada VC-151 até entrada DF-105(A)	312	DF-322, entrada DF-355 até entrada VC-409

Scan Circular		Scan Vizinhança Aleatória		Scan Vizinhança Otimizada		Scan Vizinhança Proporcional	
280	DF-285, entrada VC-441 até fim do trecho pavimentado	294	DF-295, entrada DF-130 até entrada DF-125	254	DF-250, entrada DF-105(A) até entrada DF-105(B)	313	DF-322, entrada VC-409 até entrada VC-417
281	DF-285, fim do trecho pavimentado até DF-120	295	DF-295, entrada DF-125 até entrada BR-251	255	DF-250, entrada DF-105(B) até entrada VC-155	314	DF-322, entrada VC-417 até entrada DF-260(A)
282	DF-285, entrada DF-120 até VC-447	296	DF-295, entrada BR-251 até entrada VC-471	256	DF-250, entrada VC-155 até entrada DF-100	315	DF-322, entrada DF-260(A) até entrada DF-260(B)
283	DF-285, entrada VC-447 até DF-100	297	DF-295, entrada VC-471 até entrada DF-100	257	DF-250, entrada DF-100 até entrada VC-159	316	DF-322, entrada DF-260(B) até entrada VC-421
298	DF-310, entrada DF-250 até VC-151	298	DF-310, entrada DF-250 até VC-151	258	DF-250, entrada VC-159 até entrada VC-145	317	DF-322, entrada VC-421 até entrada DF-270
299	DF-310, entrada VC-151 até VC-165	299	DF-310, entrada VC-151 até VC-165	259	DF-250, entrada VC-145 até entrada BR-479	321	DF-330, entrada DF-440 até entrada DF-444
300	DF-310, entrada VC-165 até VC-173	300	DF-310, entrada VC-165 até VC-173	260	DF-260, entrada DF-130 até entrada DF-120(A)	322	DF-330, entrada DF-444 até entrada DF-250
301	DF-310, entrada VC-173 até VC-177(A)	301	DF-310, entrada VC-173 até VC-177(A)	261	DF-260, entrada DF-120(A) até entrada DF-120(B)	324	DF-335, fim do trecho planejado até acesso a universidade
302	DF-310, entrada VC-177(A) até VC-177(B)	302	DF-310, entrada VC-177(A) até VC-177(B)	262	DF-260, entrada DF-120(B) até entrada VC-407	325	DF-335, acesso a universidade até fim de trecho pavimentado
303	DF-310, entrada VC-177(B) até DF-100	303	DF-310, entrada VC-177(B) até DF-100	263	DF-260, entrada VC-407 até entrada VC-419	326	DF-335, fim do trecho pavimentado até entrada DF-131
304	DF-320, entrada DF-250 até VC-403	304	DF-320, entrada DF-250 até VC-403	264	DF-260, entrada VC-419 até entrada VC-423	327	DF-345, entrada BR-010(A) até entrada DF-205
305	DF-320, entrada VC-403 até DF-355	305	DF-320, entrada VC-403 até DF-355	265	DF-260, entrada VC-423 até entrada DF-322(A)	328	DF-345, entrada DF-205 até entrada VC-111
306	DF-320, entrada DF-355 até VC-165	306	DF-320, entrada DF-355 até VC-165	266	DF-260, entrada DF-322(A) até entrada DF-322(B)	329	DF-345, entrada VC-111 até entrada BR-010(B)
307	DF-320, entrada VC-165 até fim do trecho pavimentado	307	DF-320, entrada VC-165 até fim do trecho pavimentado	267	DF-260, entrada DF-322(B) até entrada DF-100	330	DF-345, entrada BR-010(B) até entrada DF-230
308	DF-320, fim do trecho pavimentado até entrada VC-173	308	DF-320, fim do trecho pavimentado até entrada VC-173	268	DF-270, entrada DF-130 até entrada DF-125(A)	331	DF-353, entrada DF-250 até entrada VC-129
309	DF-320, entrada VC-173 até entrada VC-409	309	DF-320, entrada VC-173 até entrada VC-409	269	DF-270, entrada DF-125(A) até fim do trecho pavimentado	332	DF-353, entrada VC-129 até entrada VC-123
310	DF-320, entrada VC-409 até entrada VC-417	310	DF-320, entrada VC-409 até entrada VC-417	270	DF-270, fim do trecho pavimentado até entrada DF-125(B)	333	DF-353, entrada VC-123 até entrada VC-133
311	DF-320, entrada VC-417 até entrada DF-100	311	DF-320, entrada VC-417 até entrada DF-100	271	DF-270, entrada DF-125(B) até entrada DF-120(A)	334	DF-353, entrada VC-133 até entrada VC-127
312	DF-322, entrada DF-355 até entrada VC-409	312	DF-322, entrada DF-355 até entrada VC-409	272	DF-270, entrada DF-120(A) até entrada DF-120(B)	335	DF-353, entrada VC-127 até entrada DF-230
313	DF-322, entrada VC-409 até entrada VC-417	313	DF-322, entrada VC-409 até entrada VC-417	273	DF-270, entrada DF-120(B) até entrada VC-407	336	DF-355, entrada DF-130 até entrada DF-120
314	DF-322, entrada VC-417 até entrada DF-260(A)	314	DF-322, entrada VC-417 até entrada DF-260(A)	274	DF-270, entrada VC-407 até entrada DF-322	337	DF-355, entrada DF-120 até entrada VC-403
315	DF-322, entrada DF-260(A) até entrada DF-260(B)	315	DF-322, entrada DF-260(A) até entrada DF-260(B)	275	DF-270, entrada DF-322 até entrada DF-100	338	DF-355, entrada VC-403 até entrada DF-322
316	DF-322, entrada DF-260(B) até entrada VC-421	316	DF-322, entrada DF-260(B) até entrada VC-421	279	DF-285, entrada BR-251 até entrada VC-441	339	DF-355, entrada DF-322 até entrada DF-320
317	DF-322, entrada VC-421 até entrada DF-270	317	DF-322, entrada VC-421 até entrada DF-270	280	DF-285, entrada VC-441 até fim do trecho pavimentado	359	DF-440, VC-249 até entrada VC-263
321	DF-330, entrada DF-440 até entrada DF-444	321	DF-330, entrada DF-440 até entrada DF-444	281	DF-285, fim do trecho pavimentado até DF-120	360	DF-440, entrada VC-263 até entrada VC-257
322	DF-330, entrada DF-444 até entrada DF-250	322	DF-330, entrada DF-444 até entrada DF-250	282	DF-285, entrada DF-120 até VC-447	361	DF-440, entrada VC-257 até entrada DF-330
326	DF-335, fim do trecho pavimentado até entrada DF-131	327	DF-345, entrada BR-010(A) até entrada DF-205	283	DF-285, entrada VC-447 até DF-100	372	DF-455, entrada DF-130 até entrada Vc-413
327	DF-345, entrada BR-010(A) até entrada DF-205	328	DF-345, entrada DF-205 até entrada VC-111	284	DF-285, entrada DF-100 até entrada VC-461	373	DF-455, entrada VC-413 até entrada DF-120
328	DF-345, entrada DF-205 até entrada VC-111	329	DF-345, entrada VC-111 até entrada BR-010(B)	285	DF-285, entrada VC-461 até divisa DF/MG		
329	DF-345, entrada VC-111 até entrada BR-010(B)	330	DF-345, entrada BR-010(B) até entrada DF-230	294	DF-295, entrada DF-130 até entrada DF-125		
330	DF-345, entrada BR-010(B) até entrada DF-230	331	DF-353, entrada DF-250 até entrada VC-129	295	DF-295, entrada DF-125 até entrada BR-251		
331	DF-353, entrada DF-250 até entrada VC-129	332	DF-353, entrada VC-129 até entrada VC-123	296	DF-295, entrada BR-251 até entrada VC-471		
332	DF-353, entrada VC-129 até entrada VC-123	333	DF-353, entrada VC-123 até entrada VC-133	297	DF-295, entrada VC-471 até entrada DF-100		
333	DF-353, entrada VC-123 até entrada VC-133	334	DF-353, entrada VC-133 até entrada VC-127	298	DF-310, entrada DF-250 até VC-151		
334	DF-353, entrada VC-133 até entrada VC-127	335	DF-353, entrada VC-127 até entrada DF-230	299	DF-310, entrada VC-151 até VC-165		
335	DF-353, entrada VC-127 até entrada DF-230	336	DF-355, entrada DF-130 até entrada DF-120	300	DF-310, entrada VC-165 até VC-173		
336	DF-355, entrada DF-130 até entrada DF-120	337	DF-355, entrada DF-120 até entrada VC-403	301	DF-310, entrada VC-173 até VC-177(A)		

Scan Circular		Scan Vizinhança Aleatória		Scan Vizinhança Otimizada	
337	DF-355, entrada DF-120 até entrada VC-403	338	DF-355, entrada VC-403 até entrada DF-322	302	DF-310, entrada VC-177(A) até VC-177(B)
338	DF-355, entrada VC-403 até entrada DF-322	339	DF-355, entrada DF-322 até entrada DF-320	303	DF-310, entrada VC-177(B) até DF-100
339	DF-355, entrada DF-322 até entrada DF-320	342	DF-405, entrada VC-113 até entrada DF-205	304	DF-320, entrada DF-250 até VC-403
340	DF-405, entrada BR-020 até entrada VC-111	345	DF-410, entrada VC-127 até entrada DF-230	305	DF-320, entrada VC-403 até DF-355
341	DF-405, entrada VC-111 até entrada VC-113	360	DF-440, entrada VC-263 até entrada VC-257	306	DF-320, entrada DF-355 até VC-165
342	DF-405, entrada VC-113 até entrada DF-205	361	DF-440, entrada VC-257 até entrada DF-330	307	DF-320, entrada VC-165 até fim do trecho pavimentado
343	DF-410, entrada BR-020 até entrada VC-121	362	DF-440, entrada DF-330 até acesso BR-010	308	DF-320, fim do trecho pavimentado até entrada VC-173
344	DF-410, entrada VC-121 até entrada VC-127	372	DF-455, entrada DF-130 até entrada Vc-413	309	DF-320, entrada VC-173 até entrada VC-409
345	DF-410, entrada VC-127 até entrada DF-230	373	DF-455, entrada VC-413 até entrada DF-120	310	DF-320, entrada VC-409 até entrada VC-417
360	DF-440, entrada VC-263 até entrada VC-257			311	DF-320, entrada VC-417 até entrada DF-100
361	DF-440, entrada VC-257 até entrada DF-330			312	DF-322, entrada DF-355 até entrada VC-409
372	DF-455, entrada DF-130 até entrada Vc-413			313	DF-322, entrada VC-409 até entrada VC-417
373	DF-455, entrada VC-413 até entrada DF-120			314	DF-322, entrada VC-417 até entrada DF-260(A)
				315	DF-322, entrada DF-260(A) até entrada DF-260(B)
				316	DF-322, entrada DF-260(B) até entrada VC-421
				317	DF-322, entrada VC-421 até entrada DF-270
				318	DF-326, entrada DF-205 até entrada DF-335
				319	DF-326, entrada DF-335 até entrada VC-215
				320	DF-326, entrada VC-215 até acesso a Sobradinho
				321	DF-330, entrada DF-440 até entrada DF-444
				322	DF-330, entrada DF-444 até entrada DF-250
				323	DF-335, entrada DF-326 até fim do trecho implantado
				324	DF-335, fim do trecho planejado até acesso a universidade
				325	DF-335, acesso a universidade até fim de trecho pavimentado
				326	DF-335, fim do trecho pavimentado até entrada DF-131
				327	DF-345, entrada BR-010(A) até entrada DF-205
				328	DF-345, entrada DF-205 até entrada VC-111
				329	DF-345, entrada VC-111 até entrada BR-010(B)
				330	DF-345, entrada BR-010(B) até entrada DF-230
				331	DF-353, entrada DF-250 até entrada VC-129
				332	DF-353, entrada VC-129 até entrada VC-123
				333	DF-353, entrada VC-123 até entrada VC-133

Scan Circular		Scan Vizinhança Aleatória		Scan Vizinhança Otimizada	
				334	DF-353, entrada VC-133 até entrada VC-127
				335	DF-353, entrada VC-127 até entrada DF-230
				336	DF-355, entrada DF-130 até entrada DF-120
				337	DF-355, entrada DF-120 até entrada VC-403
				338	DF-355, entrada VC-403 até entrada DF-322
				339	DF-355, entrada DF-322 até entrada DF-320
				340	DF-405, entrada BR-020 até entrada VC-111
				341	DF-405, entrada VC-111 até entrada VC-113
				342	DF-405, entrada VC-113 até entrada DF-205
				343	DF-410, entrada BR-020 até entrada VC-121
				344	DF-410, entrada VC-121 até entrada VC-127
				345	DF-410, entrada VC-127 até entrada DF-230
				346	DF-415, entrada BR-080 até entrada DF-445
				347	DF-415, entrada DF-445 até início de trecho planejado
				348	DF-430, entrada Brazlândia até fim de trecho pavimentado
				349	DF-430, fim de trecho pavimentado entrada DF-445(A)
				353	DF-435, entrada BR-080 até entrada VC-547
				354	DF-435, entrada VC-547 até entrada DF-445
				355	DF-435, fim do trecho planejado até entrada DF-001
				358	DF-440, fim de trecho pavimentado até entrada VC-249
				359	DF-440, VC-249 até entrada VC-263
				360	DF-440, entrada VC-263 até entrada VC-257
				361	DF-440, entrada VC-257 até entrada DF-330
				364	DF-445, entrada BR-080 até entrada DF-435
				365	DF-445, entrada DF-435 até entrada DF-430(A)
				366	DF-445, entrada DF-430(A) até entrada DF-430(B)
				367	DF-445, entrada DF-430(B) até entrada VC-527
				368	DF-445, entrada VC-527 até entrada DF-415
				369	DF-445, entrada DF-415 até entrada DF-220
				372	DF-455, entrada DF-130 até entrada Vc-413
				373	DF-455, entrada VC-413 até entrada DF-120