



Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Estatística

**ANÁLISE DA DISTRIBUIÇÃO DA
CRIMINALIDADE NO DISTRITO FEDERAL
UTILIZANDO ESTATÍSTICA ESPACIAL**

Bruno Magalhães dos Santos

09/0107969

Orientador: Prof.^o Dr. Alan Ricardo da Silva

Brasília
2013

Bruno Magalhães dos Santos

09/0107969

**ANÁLISE DA DISTRIBUIÇÃO DA
CRIMINALIDADE NO DISTRITO FEDERAL
UTILIZANDO ESTATÍSTICA ESPACIAL**

Relatório apresentado à disciplina Estágio Supervisionado II do curso de graduação em Estatística, Departamento de Estatística, Instituto de Exatas, Universidade de Brasília, como parte dos requisitos necessários para o grau de Bacharel em Estatística.

Orientador: Prof. Dr. Alan Ricardo da Silva

Brasília

2013

Agradecimentos

Ao SAS *Institute* Brasil por possibilitar a utilização desse software por meio de parceria acadêmica com o Departamento de Estatística da UnB. A Secretaria de Segurança Pública pelo banco de dados dos boletins de ocorrência do Distrito Federal.

Agradeço aos meus familiares pelo apoio, atenção e paciência destinados a mim em todos os momentos cruciais deste período. Aos meus colegas de curso cujo a companhia e amizade me ajudaram em muitos momentos de decisão. Aos professores por todo conhecimento transmitido e em especial ao meu professor orientador, Alan Ricardo da Silva, pela ajuda e tempo dedicados a este trabalho.

Bruno Magalhães dos Santos

Resumo

A criminalidade em Brasília, para muitos estudiosos e sociólogos, é herança do crescimento desordenado praticado quando da sua construção, e da política de alguns governos que doavam lotes aos imigrantes que desembargavam no Distrito Federal. A Secretaria de Estado de Segurança Pública do DF, busca combater a criminalidade através de diversas medidas e ações, buscando apoio na análise espacial para identificação, nos mapas, das regiões mais críticas.

Este trabalho busca verificar quais são as áreas de maior foco para os crimes selecionados no banco de dados da SSP/DF nos anos de 2010 e 2011. Mais do que isso, busca através da linguagem SAS/IML utilizar os três tipos de técnicas de análise espacial (pontual, área e krigagem), para atestar a existência ou não de padrões pontuais e a movimentação da distribuição da criminalidade com o passar do tempo.

A distribuição das ocorrências para os crimes estudados foi identificada pela análise pontual. Além disso, verificou-se a existência de áreas de foco e suas localizações no mapa. Devido a alta heterogeneidade dos dados não foi possível obter resultados confiáveis para as técnicas de área e krigagem.

Sumário

RESUMO	3
1 Introdução	6
1.1 Objetivo	7
2 A Criminalidade do Distrito Federal	9
2.1 Introdução	9
2.2 Os Problemas na Obtenção de Dados	10
3 Estatística Espacial	11
3.1 Introdução	11
3.2 Tipos de dados	11
3.3 Conceitos Básicos em Análise Espacial	12
3.3.1 Dependência Espacial	12
3.3.2 Autocorrelação Espacial	12
3.3.3 Estacionaridade e Isotropia	13
3.4 Análise Espacial de Eventos	13
3.4.1 Análise Pontual	13
3.4.2 Análise de Áreas	22

3.4.3	Análise de Superfície: Krigagem	30
4	Material e Métodos	37
4.1	Introdução	37
4.1.1	Sistema de Coordenadas Geográficas	37
4.1.2	Latitude e Longitude	38
4.1.3	Algoritmo IML	38
4.1.4	Método para Transformação dos Endereços	39
4.1.5	Métodos para Análise Pontual	39
4.1.6	Métodos para Análise de Áreas	40
4.1.7	Métodos para Análise da Krigagem	40
5	Análise dos Resultados	42
5.1	Introdução	42
5.2	Análise Descritiva	42
5.3	Análise Pontual	50
5.4	Análise de Áreas	56
5.5	Análise da Krigagem	63
6	Conclusões	66
	Referências	68

Capítulo 1

Introdução

A Secretaria de Estado de Segurança Pública (SSP/DF) é o órgão do Governo do Distrito Federal, criado em 11 de outubro de 1979, que coordena o Sistema de Segurança Pública, composto pela; I - Polícia Militar; II - Polícia Civil; III - Corpo de Bombeiros Militares; IV - Departamento de Trânsito. Em 03 de julho de 2002, por meio da Lei 2.997, houve sua reestruturação devido a necessidade de adequação das atividades de Segurança Pública. O órgão adotou a atual denominação de Secretaria de Estado de Segurança Pública e Defesa Social, um ato distintamente de valorização da proteção do cidadão e da comunidade (GDF, 2012).

Tem como principais objetivos: I - Centralizar e gerenciar as informações decorrentes do atendimento emergencial, para subsidiar o planejamento, a tomada de decisões e a efetiva ação dos Órgãos de Segurança; II - Contribuir para uma maior agilidade no atendimento ao cidadão tendo como consequência a melhoria da ordem pública e da defesa da coletividade; III - Registrar, autorizar, controlar e acompanhar o atendimento das ocorrências emergenciais na área de atuação da Polícia Militar, Polícia Judiciária Civil, Corpo de Bombeiros Militar e Departamento de Trânsito (SSP/DF, 2012).

Normalmente, os boletins de ocorrências, armazenados no banco de dados da SSP/DF (2012), apresentam os dados referentes aos crimes e os endereços dos locais de ocorrência (rua, quadra, bloco, etc). Para que possa ser feita uma análise espacial estatística será necessário que estes endereços sejam representados através de coordenadas geográficas.

Dessa forma, este trabalho buscará atribuir as coordenadas geográficas aos endereços, com o objetivo de analisar o comportamento da distribuição de criminalidade no Distrito Federal. Para tal, serão utilizadas técnicas de Estatística Espacial como a análise pontual, análise de área e Krigagem. De tal forma, cada uma das ocorrências será representada por um ponto ou fará parte de uma mancha em um mapa, onde ocorrerá a análise dos focos dos diversos tipos de crimes registrados. Será feita a verificação da movimentação dos focos de criminalidade no território do Distrito Federal.

1.1 Objetivo

O objetivo geral do trabalho é analisar a distribuição de criminalidade quanto ao seu crescimento no território do Distrito Federal.

Os objetivos específicos são:

- Utilizar estatística espacial, mais especificamente análise pontual de dados, análise de áreas e Krigagem;
- Realizar as análises utilizando o *software* SAS 9.2 com o procedimento IML;
- Verificar quais áreas do Distrito Federal são focos de criminalidades para os

tipos de crime selecionados.

Capítulo 2

A Criminalidade do Distrito Federal

2.1 Introdução

Apesar das diversas medidas e ações tomadas pela Secretaria de Segurança Pública do DF para o combate da criminalidade, os registros dos mesmos cresceram cerca de 14,6%, de acordo com dados divulgados pela mesma SSP/DF (2012) no primeiro semestre do ano de 2012 em comparação com o mesmo período do ano passado. Crimes conhecidos como sequestro relâmpago, furto de veículos e crimes contra a dignidade sexual (categoria que inclui estupro e abuso sexual) tiveram altas consideráveis, cerca de 45,6%, 73,6% e 42,7% respectivamente (SSP/DF (2012)).

Em uma tentativa de intensificar o policiamento em regiões com altas taxas de criminalidade, a SSP/DF buscou apoio na análise espacial, ou seja, o banco de dados da SSP/DF foi totalmente georreferenciado, possibilitando identificar no mapa as regiões mais críticas.

2.2 Os Problemas na Obtenção de Dados

O banco de dados cedido para este trabalho, pela própria SSP/DF, apresenta o registro das ocorrências dos diversos crimes ocorridos para os anos de 2010 e 2011 no Distrito Federal, sem o georreferenciamento dos locais onde ocorreram tais delitos. Este banco é originado pela alimentação dos boletins de ocorrência - B.O., realizados pelas vítimas.

Para realizar o georreferenciamento dos crimes foi necessário trabalhar com o campo CEP, que contém os endereços, registrados através dos boletins de ocorrência. Estes podem ser feitos presencialmente nas diversas delegacias ou postos de atendimento, sendo o endereçamento realizado de modo generalizado, não apresentando um padrão em seu preenchimento. Erros grosseiros, também, foram detectados, como por exemplo, endereços registrados errados. Uma solução para este problema seria a criação de um registro padronizado dos logradouros, podendo ser utilizado, por exemplo, o adotado pela Empresa Brasileira de Correios - ECT, em um sistema informatizado.

O próximo Capítulo mostrará como a Estatística Espacial pode auxiliar na tomada de decisão, quando da localização dos eventos de criminalidade.

Capítulo 3

Estatística Espacial

3.1 Introdução

A estatística espacial é o ramo da estatística que nos permite utilizar, além dos conceitos clássicos, o espaço geográfico associado ao conceito de variável aleatória. Utilizando a estatística espacial é possível identificar, localizar e visualizar fenômenos no espaço e assim modelá-los compreendendo a distribuição espacial dos dados.

Segundo Monteiro et al. (2004), “a ênfase da Análise Espacial é mensurar propriedades e relacionamentos, levando em conta a localização espacial do fenômeno em estudo de forma explícita.”

3.2 Tipos de dados

Os dados presentes em análise espacial podem ser divididos em três categorias:

- *Padrões pontuais*, onde ocorrências do fenômeno em estudo são expressas através de pontos localizados no espaço, denominados processos pontuais. As localização de crimes são um bom exemplo deste tipo de dado.
- *Superfícies Contínuas*, são estimadas a partir de um conjunto de amostras,

podendo ser regularmente ou irregularmente distribuídas. Mapas resultantes de levantamentos de recursos naturais são ótimos exemplos deste tipo de dado.

- *Áreas com Contagens*, referem-se aos dados associados a levantamentos populacionais e que originalmente se referem a indivíduos localizados em pontos específicos do espaço. Normalmente, estes dados se encontram agregados em unidades de análise, como setores censitários, devido a confidencialidade dos dados.

3.3 Conceitos Básicos em Análise Espacial

3.3.1 Dependência Espacial

Segundo Tobler (1970), a primeira lei da geografia diz: “Todas as coisas são parecidas, mas coisas mais próximas se parecem mais que coisas distantes”. Seguindo esse pensamento, define-se dependência espacial como sendo uma relação entre ocorrências que depende da distância, ou seja, ocorrências próximas normalmente deveriam ter características semelhantes e à medida que aumentamos a distância entre elas, suas características deveriam variar.

3.3.2 Autocorrelação Espacial

Segundo Monteiro et al. (2004), este termo foi derivado do conceito estatístico de correlação, sendo utilizado para mensurar o relacionamento entre duas variáveis aleatórias. O prefixo “auto” indica que a medida de correlação é realizada com a mesma variável aleatória, medida em locais distintos do espaço. Para medir a autocorrelação espacial pode-se utilizar diferentes indicadores, todos baseados na

mesma idéia: verificar como varia a dependência espacial a partir da comparação entre os valores de uma amostra e de seus vizinhos.

3.3.3 Estacionaridade e Isotropia

Para que seja possível conceituar as idéias de *estacionaridade* e *isotropia* é preciso primeiro definir os conceitos de efeitos de 1ª e 2ª ordem:

- Efeito de 1ª ordem é o valor esperado, ou seja, a média do processo no espaço;
- Efeito de 2ª ordem é a covariância entre as áreas s_i e s_j .

Tendo esses dois conceitos em mente, define-se que um processo é considerado *estacionário* quando os efeitos de 1ª e 2ª ordem são constantes em toda a região de estudo, isto é, não existe tendência. Este mesmo processo pode ser classificado como *estacionário de segunda ordem* se o valor esperado do processo estocástico for constante para toda uma região de estudo. Um processo é considerado *isotrópico*, se além de *estacionário*, o efeito de 2ª ordem depende somente da distância entre os pontos e não da direção entre eles. Caso a covariância dependa da direção, o fenômeno é classificado como *anisotrópico*.

3.4 Análise Espacial de Eventos

3.4.1 Análise Pontual

A análise pontual é definida como o estudo dos processos pontuais e tem como principal objetivo a identificação da existência ou não de algum tipo de padrão espacial. Após os pontos serem localizados no mapa, espera-se que estes sigam um dos três tipos de padrões:

- *Regular*: a distância média entre os pontos tende a ser constante, mostrando regularidade no espaçamento entre os pontos.
- *Aglomerado*: nota-se maior concentrações de pontos em determinada área em relação ao restante do espaço.
- *Aleatório*: os pontos observados no território estão distribuídos de tal forma que não apresentam nenhum modelo de distribuição.

De forma simples, os processos pontuais são definidos como o conjunto de pontos distribuídos irregularmente em uma região, tendo suas localizações geradas através de um método estocástico. Devido a este processo ser estocástico ele pode ser descrito em termos dos efeitos de 1ª e 2ª ordem (intensidade e dependência espacial do processo, respectivamente).

Segundo Carvalho and Câmara (2004), podemos exemplificar esta idéia considerando um conjunto de pontos (u_1, u_2, \dots) numa determinada região A onde ocorreram os eventos. O processo pontual é modelado considerando sub-regiões S em A através de sua esperança $E[N(S)]$ e a covariância $C[N(S_i), N(S_j)]$, onde $N(S)$ denota o número de eventos em S . Devido ao objetivo da análise ser a estimação das prováveis localizações de ocorrência de determinados eventos, essas estatísticas devem ser inferidas considerando o valor limite da quantidade de eventos por área. Este valor limite corresponde a esperança de $N(S)$ para uma pequena região du em torno do ponto u , quando essa tende a zero. Esta esperança é a propriedade de 1ª

ordem, ou seja, denomina a intensidade do processo, sendo esta definida como:

$$\lambda(u) = \lim_{|du| \rightarrow 0} \left\{ \frac{E[N(du)]}{|du|} \right\} \quad (3.1)$$

Para medirmos as propriedades de 2ª ordem, consideramos a intensidade conjunta $\lambda(u_i, u_j)$ entre duas regiões infinitesimais $|du_i|, |du_j|$ que contém os pontos u_i e u_j . Definimos esta sendo:

$$\lambda(d(u_i), d(u_j)) = \lim_{du_i, du_j \rightarrow 0} \left\{ \frac{C[N(du_i), N(du_j)]}{du_i, du_j} \right\} \quad (3.2)$$

Para que possa ser feita a análise dos padrões pontuais, é necessário tomar como referência um modelo teórico, que servirá de base para criação de métodos que servirão para a verificação da significância dos resultados exploratórios. O modelo teórico mais simples é o CSR (*complete spatial randomness*). Este modelo divide a região em estudo A em sub-áreas S_i e modela a distribuição de eventos pontuais como um processo aleatório $\{Z_i(u_i), u_i \in S_i : i = 1, \dots, n\}$, sendo $Z_i(u_i)$ o número de casos que ocorreram na subárea S_i . Neste modelo, considera-se que a posição de cada caso ocorrido é independente e que os casos tem igual probabilidade de ocorrência em toda a região de estudo. Desta forma, podemos estabelecer uma comparação entre uma distribuição totalmente aleatória e os dados coletados.

Outra forma de se avaliar padrões pontuais é medindo a distância média entre os pontos e detectando a presença ou não de aglomerados. Isto pode ser feito, através do cálculo da *distância euclidiana* utilizando as coordenadas de cada ponto no mapa:

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (3.3)$$

Os aglomerados podem ser analisados através das técnicas de aglomerados

hierárquicos, sendo o método mais usual o do vizinho mais próximo. Este método agrupa pontos dentro da área de estudo, que estão a uma distância igual ou menor que h até que se tenha uma quantidade mínima de pontos pré-estabelecida, denominando estes de aglomerados vizinhos de primeira ordem. Estes são tidos como pontos individuais e a técnica é repetida gerando aglomerados de ordem superior até que se alcance um aglomerado singular ou não seja mais possível atender as condições de distância mínima e/ou de quantidade de pontos mínimas por aglomerado.

Segundo Carvalho and Câmara (2004), esta técnica estima uma função de distribuição cumulativa $\hat{G}(h)$, sendo h a distância entre os eventos na região de análise, e pode ser estimada empiricamente por:

$$\hat{G}(h) = \frac{\#(d(u_i, u_j) \leq h)}{n} \quad (3.4)$$

sendo $\#$ a função contagem e $\hat{G}(h)$ o valor normalizado acumulado para uma distância h correspondente à soma dos vizinhos mais próximos de cada evento, cuja distância é menor ou igual a h , dividido pelo número de eventos na região (Carvalho and Câmara, 2004).

Carvalho and Câmara (2004) ainda comentam que o gráfico dos resultados desta função $\hat{G}(h)$ ainda pode ser usada como um método exploratório para se verificar a existência de interação entre os eventos. Se o crescimento do gráfico para valores pequenos de distância for muito rápido, esta situação indica que os eventos possuem interação entre si nestas faixas de distância. Caso o gráfico apresente valores pequenos no começo e venha a crescer aceleradamente para valores maiores de distância, estes eventos possuem uma distribuição mais regular.

Segundo Carvalho and Câmara (2004), a análise de vizinhança pode, através da criação de envelopes de simulação da distribuição CSR, comparar estatisticamente a distribuição dos eventos observados com o esperado na hipótese do próprio CSR obtendo a significância dos desvios.

A função de distribuição $G(w)$, na hipótese de CSR, seria dada por um processo de *Poisson* (Equação 3.5).

$$G(h) = 1 - e^{-\lambda\pi h^2} h \geq 0 \quad (3.5)$$

Assumindo a CSR, a estimação simulada para a distribuição $G(w)$ é dada por:

$$\bar{G}(h) = \frac{\sum_i^k \hat{G}_i(h)}{k} \quad (3.6)$$

onde $G_i(h)$, $i = 1, 2, \dots, k$, são funções de distribuição empíricas, estimadas a partir de k simulações independentes e uniformemente distribuídas dos n eventos, na hipótese de CSR (Carvalho and Câmara, 2004). Para se testar a condição de aleatoriedade são calculados os envelopes de simulação superior e inferior. Estes são representados pelas Equações (3.7) e (3.8), respectivamente.

$$U(h) = \max\{\hat{G}_i(h)\}, i = 1, \dots, k \quad (3.7)$$

$$L(h) = \min\{\hat{G}_i(h)\}, i = 1, \dots, k \quad (3.8)$$

Com o gráfico da distribuição $\hat{G}_i(h)$ *versus* a distribuição $\bar{G}(h)$ e mais os envelopes inferior e superior, é possível medir a significância dos desvios relativos a aleatoriedade. Se a condição de CSR for válida para os dados observados, o gráfico $\hat{G}_i(h)$ *versus* $\bar{G}(h)$ deverá ser praticamente linear com ângulo de 45° . Caso os dados apresentem agrupamentos, os traçados no gráfico estarão acima da linha de 45° .

Para os padrões regulares os traçados ficarão abaixo da linha de 45° (Carvalho and Câmara, 2004).

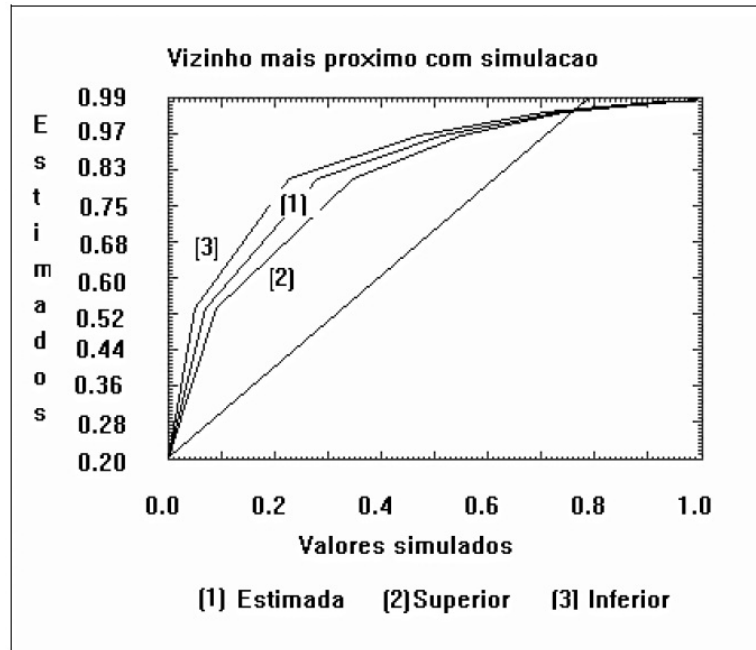


Figura 3.1: Vizinho mais próximo com simulação.
 Fonte: Câmara et al. (2004)

Apesar deste método apresentar uma indicação inicial da distribuição espacial, como mostra a Figura 3.1, ele apenas considera escalas pequenas. Para escalas maiores é recomendado o uso da *função K de Ripley*. Para o processo univariado a *função K* é dada pela equação:

$$\lambda K(h) = E(\# \text{ eventos contidos a uma distância } h \text{ de um evento arbitrário}) \quad (3.9)$$

sendo $E()$ o operador da estimativa, e λ o número médio de eventos por unidade de área, assumida constante na região. Uma estimativa de $K(h)$ é calculada como:

$$\hat{K}(h) = \frac{A}{n^2} \sum_i^n \sum_{j, i \neq j}^n \frac{I_h(d_{ij})}{w_{ij}} \quad (3.10)$$

sendo A a área da região, $I_h(d_{ij})$ uma função indicadora tendo valor 1 quando $(d_{ij}) \leq h$ e 0 caso contrário, n o número de eventos observados e w_{ij} é a proporção da circunferência do círculo centrado no evento i que está dentro da região, sendo esta correção devida ao efeito de borda.

Segundo Carvalho and Câmara (2004), a função K é utilizada como ferramenta exploratória na comparação entre $\hat{K}(h)$, função empírica, e $\bar{K}(h)$, função resultante de um processo espacial aleatório de padrão de pontos. Se o processo for aleatório, $K(h)$ é dado por πh^2 . Dessa forma, pode-se comparar a estimativa K de um conjunto de dados observados com πh^2 , por meio de um gráfico em função de $\hat{L}(h)$, sendo esta definida por:

$$\hat{L}(h) = \sqrt{\frac{\hat{K}(h)}{\pi}} - h \quad (3.11)$$

De acordo com Carvalho and Câmara (2004), o gráfico da função $\hat{L}(h)$ em função da distância h , indica atração espacial entre eventos ou agrupamentos para valores positivos e indica repulsão espacial ou regularidade para valores negativos. Como no método do vizinho mais próximo, também é possível calcular os envelopes superior e inferior e realizar a análise destes e do gráfico da função. Na análise do gráfico dos envelopes e da distribuição, picos positivos que estão acima do envelope superior indicam agrupamentos na escala e picos negativos abaixo do envelope inferior indicam regularidade na escala (Figura 3.2).

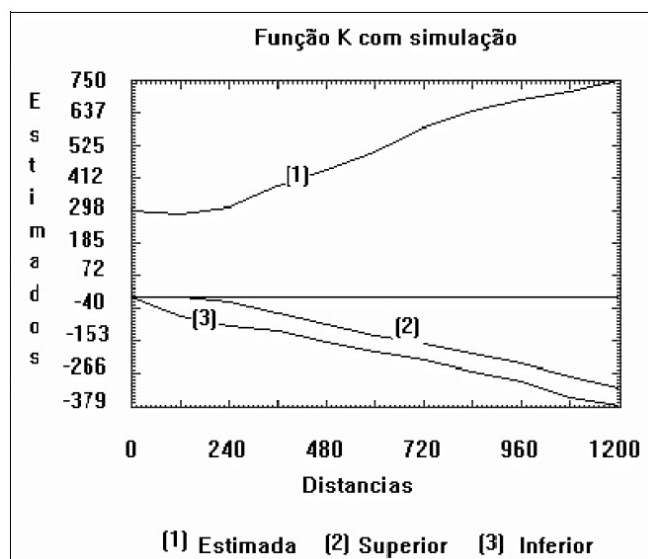


Figura 3.2: *Função K* com simulação.
 Fonte: Câmara et al. (2004)

Estimador de Intensidade

Existem diversos procedimentos para a estimação da intensidade pontual do processo, efeito de primeira ordem, no intuito de estudar o comportamento de padrões pontuais, sendo o mais prático a estimativa de *Kernel*. Esta técnica estatística não-paramétrica estima uma superfície de densidade através de uma função bi-dimensional, que é ajustada sobre os eventos em estudo, conseguindo identificar visualmente um valor proporcional à intensidade destes por unidade de área. Utilizando-se das distâncias de cada ponto em relação a uma localização de interesse, é possível indentificar concentrações/aglomerados, explorar e demonstrar padrões pontuais no espaço, criando uma superfície contínua de risco para a ocorrência destes eventos. Segundo Bailey and Gatrell (1995) e Carvalho and Câmara (2004), esta técnica possibilita filtrar a variabilidade de um conjunto de dados sem alterar suas características locais, sendo este estimador útil no fornecimento de uma visão geral da

distribuição de primeira ordem dos eventos. A Figura 3.3 ilustra a idéia da estimativa de Kernel.

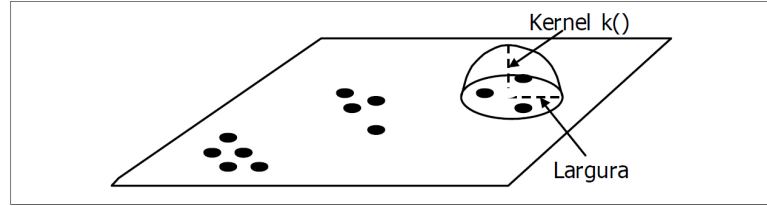


Figura 3.3: Estimador de intensidade de distribuição de pontos.
 Fonte: Câmara et al. (2004)

Segundo Carvalho and Câmara (2004), para que possa ser feita a estimativa de *Kernel* é necessário definir dois parâmetros. O primeiro é o τ , parâmetro que representa o raio de influência que define a vizinhança do ponto que será interpolado e controla o alisamento da superfície que será criada. Este raio de influência define a área centrada no ponto de estimação u , que indentifica quantos eventos u_i contribuem para a estimativa da função de intensidade. O segundo, é a função de estimação k (*Kernel*) que tem a propriedade de suavizar o fenômeno. Esta função, normalmente utiliza funções de terceira ou quarta ordem. Utilizando estes parâmetros, se uma região u contida em A e (u_1, u_2, \dots, u_n) são localizações de n eventos observados, então a intensidade estimada λ em u é dada por:

$$\lambda_\tau(u) = \frac{1}{\tau^2} \sum_{i=1}^n k \left[\frac{d(u_i, u)}{\tau} \right] \quad (3.12)$$

Esta função representa cada ponto u , presente na grade localizada em A , como um função trimensional. A partir desta idéia, são calculadas as distâncias para cada ponto u_i presente na área de influência que contribuem para a estimativa de intensidade de u de acordo com seu valor em relação ao próprio ponto u (Figura

3.4).

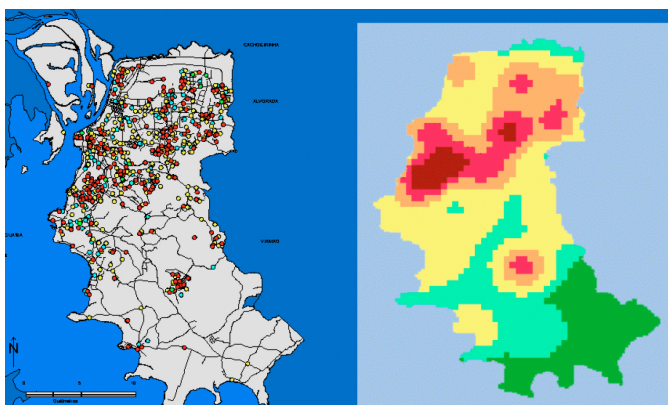


Figura 3.4: À direita distribuição de casos de mortalidade por causas externas em Porto Alegre em 1996 e à esquerda o estimador de intensidade de *Kernel*.

Fonte: Câmara et al. (2004)

3.4.2 Análise de Áreas

A análise espacial de áreas compreende os métodos cuja localização está associada a áreas delimitadas por polígonos. Uma forma usual de apresentação de dados agregados por áreas é pelo uso de mapas coloridos com o padrão espacial do fenômeno (Cruz et al., 2004).

Um passo inicial para este tipo de análise é a análise exploratória dos dados. Esta tem papel fundamental no desenvolvimento das etapas da modelagem estatística espacial, sendo esta sensível ao tipo de distribuição, à presença de valores extremos e à ausência de estacionariedade. A forma mais simples de visualização de valores extremos é através dos mapas, lembrando que o resultado desta varia de acordo com o corte feito na variável em estudo. Segundo Cruz et al. (2004), os *SIGs*, normalmente, utilizam um desses três métodos de corte:

- *intervalos iguais*, em que os valores de máximo e mínimo são divididos pelo

número de classes.

- *percentis*, usado para definição de classes obrigando a alocação dos polígonos em quantidades iguais de cores.
- *desvios padrões*, no qual a distribuição da variável é disposta em gradação de cores diferentes para valores acima e abaixo da média, faz a suposição da normalidade da distribuição da variável, o que em diversos casos, não representa a realidade.

Segundo Cruz et al. (2004), para que se possa estimar a variabilidade entre os dados de área, utiliza-se a matriz de proximidade espacial ou de vizinhança. Esta matriz $W^1(n \times n)$ é construída através de um conjunto de n áreas A_1, \dots, A_n , onde cada um dos w_{ij} representa uma medida de proximidade entre A_i e A_j . Esta medida de proximidade pode ser calculada utilizando os seguintes critérios:

- $w_{ij} = 1$, se o centróide, centro geométrico, de A_i está a uma distância pré-definida de A_j , caso contrário $w_{ij} = 0$;
- $w_{ij} = 1$, se A_i compartilha um lado comum com A_j , caso contrário $w_{ij} = 0$;
- $w_{ij} = l_{ij}/l_i$, onde l_{ij} é o comprimento da fronteira entre A_i e A_j e l_i é o perímetro de A_i .

Devido a matriz de proximidade ser utilizada no cálculo de indicadores na análise exploratória, é útil normalizar suas linhas, para que a soma destas seja igual a um. Pode-se também, generalizar a matriz de proximidade para vizinhos de maior ordem e assim construir as matrizes W^1, W^2, \dots, W^n .

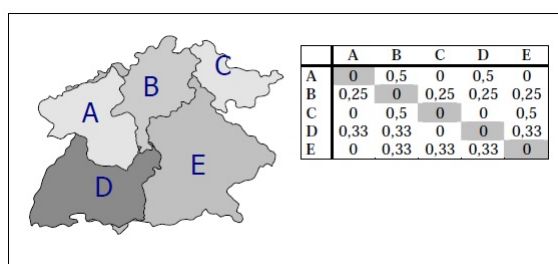


Figura 3.5: Matriz de proximidade espacial de primeira ordem, normalizada pelas linhas.

Fonte: Câmara et al. (2004)

Uma técnica para se explorar a variação da tendência espacial dos dados é a média móvel espacial. Através do cálculo da média dos valores dos vizinhos, esta técnica reduz a variabilidade espacial devido ao operador produzir uma superfície com menor flutuação que os dados originais. De acordo com Cruz et al. (2004), a média móvel $\hat{\mu}_i$ associada ao atributo Z_i , relativo à i -ésima área, pode ser calculada a partir dos elementos w_{ij} presentes na matriz de proximidade $W^{(1)}$, tomando-se a média dos vizinhos como mostra a Equação 3.13 e a Figura 3.6.

$$\hat{\mu}_i = \sum_{j=1}^n w_{ij} z_i \quad (3.13)$$

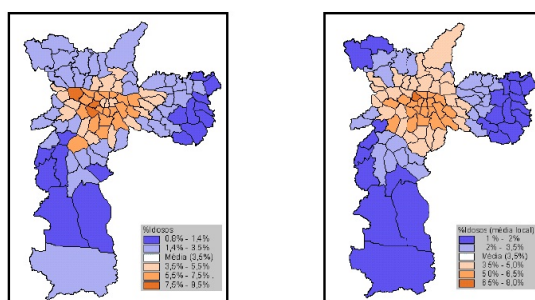


Figura 3.6: Distribuição dos idosos na cidade de São Paulo (censo de 1991). À esquerda, apresentação dos valores por distribuição estatística. À direita, média móvel local.

Fonte: Câmara et al. (2004)

Outra parte da análise exploratória é a análise da autocorrelação espacial entre áreas para indentificar a estrutura de correlação espacial que melhor descreva os dados. Neste caso, as ferramentas utilizadas são o índice global de *Moran* e o índice *C* de *Geary*. Segundo Câmara et al. (2004), o índice global *I* de *Moran* é a autocorrelação, considerando apenas o primeiro vizinho e é dado por:

$$I = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (z_i - \bar{z})(z_j - \bar{z})}{\sum_{i=1}^n (z_i - \bar{z})^2} \quad (3.14)$$

sendo n o número de áreas, z_i o valor do atributo presente na área i , \bar{z} a média do atributo na região de estudo e w_{ij} os elementos da matriz normalizada de proximidade espacial. Para se estimar a função de autocorrelação para cada ordem de vizinhança o cálculo de matrizes de proximidade de ordem maior que um, é feito de acordo com:

$$I^k = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij}^k (z_i - \bar{z})(z_j - \bar{z})}{\sum_{i=1}^n (z_i - \bar{z})^2} \quad (3.15)$$

Este índice testa a hipótese nula de independência espacial, $I = 0$. Para valores de I diferentes de zero é necessário estabelecer sua validade estatística. Para isso, pode-se tomar duas abordagens: associa-se a este valor uma distribuição estatística ou utiliza-se um teste de pseudo-significância. Este teste, gera diferentes permutações dos valores de atributos associados às regiões, sendo cada permutação um novo arranjo espacial, onde os valores estão redistribuídos entre as áreas. Como apenas um dos arranjos corresponde à situação observada, pode-se construir uma distribuição empírica de I , e se o valor do índice I , medido originalmente, corresponder a um valor extremo da distribuição simulada, então há significância estatística para este valor (Cruz et al., 2004). Este índice possui uma hipótese implícita de es-

tacionariedade de primeira e segunda ordem, caso não seja atingido o pré-requisito de estacionariedade, este índice perde sua validade.

Uma maneira adicional de visualização da dependência espacial é o diagrama de espalhamento de *Moran*. Este é construído a partir dos valores normalizados dos atributos, permitindo assim, analisar o comportamento da variabilidade espacial dos mesmos. O objetivo é construir um gráfico bidimensional onde os valores normalizados do atributo, z , são comparados com as médias dos seus vizinhos, wz .

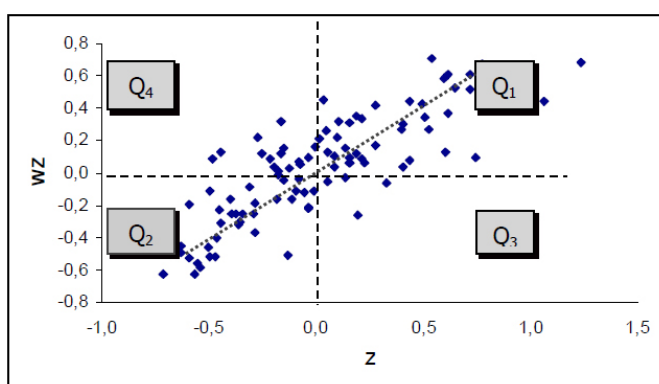


Figura 3.7: Diagrama de Espalhamento de *Moran* para o índice de exclusão/inclusão social de São Paulo, censo de 1991.

Fonte: Câmara et al. (2004)

O índice I de *Moran* é equivalente ao coeficiente de regressão linear, que inclina a reta de regressão de wz e Z . De acordo com a Figura 3.7, $Q1$ representa valores positivos e médias positivas, $Q2$ valores negativos e médias negativas, $Q3$ valores positivos e médias negativas e $Q4$ valores negativos e médias positivas. $Q1$ e $Q2$ representam pontos de associação espacial positiva, a localização possui vizinhos com valores semelhantes, enquanto que $Q3$ e $Q4$ indicam pontos de associação negativa, a localização possui vizinhos com valores distintos.

O diagrama de espalhamento de *Moran* também pode ser representado na

forma de um mapa bidimensional (Figura 3.8), onde cada polígono é apresentado indicando-se seu quadrante no diagrama de espalhamento. Os quadrantes $Q1$, $Q2$, $Q3$ e $Q4$ são representados pela legenda “Alto-Alto”, “Baixo-Baixo”, “Alto-Baixo” e “Baixo-Alto”, respectivamente.

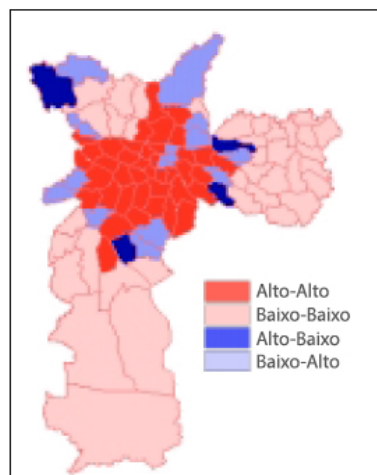


Figura 3.8: Diagrama de Espalhamento de *Moran* para o índice de exclusão/inclusão social de São Paulo, censo de 1991.

Fonte: Câmara et al. (2004)

Quando a análise possui um grande número de áreas, podem ocorrer diferentes tipos de associação espacial e máximos locais de autocorrelação espacial, onde a dependência espacial é ainda mais pronunciada. Dessa forma, os indicadores globais não são recomendados devido a apresentarem somente um valor como medida de associação espacial para todo os dados. Para que se possa estudar estes múltiplos padrões com maior detalhe são usados os indicadores locais de associação espacial. Estes indicadores produzem valores específicos para cada área, permitindo assim, a identificação de agrupamentos (Cruz et al., 2004). O índice local de *Moran* para

cada área i , a partir dos valores padronizados z_i , é calculado através:

$$I_i = \frac{z_i \sum_{j=1}^n w_{ij} z_j}{\sum_{j=1}^n z_j^2} \quad (3.16)$$

A significância estatística do índice de local *Moran* é feita de forma similar ao do índice global, uma vez que determinada a significância estatística destes índices locais é útil gerar um mapa (Figura 3.9), indicando as regiões onde a correlação local é significativamente diferente do resto dos dados.

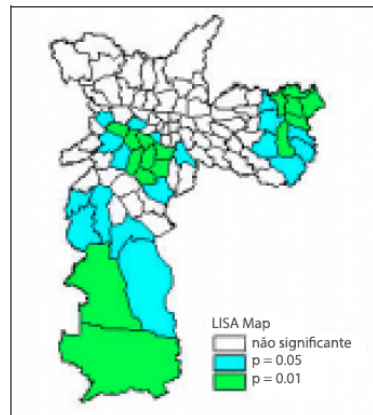


Figura 3.9: Indicador de autocorrelação espacial para o índice de exclusão/inclusão social de São Paulo (censo de 1991). Apenas os valores com significância maior que 95% estão mostrados.

Fonte: Câmara et al. (2004)

O teste C de *Geary* difere do teste I de *Moran* por utilizar a diferença entre os pares, enquanto que *Moran* utiliza a diferença entre cada ponto e a média global. A fórmula da estatística C de *Geary* é descrita como:

$$C = \frac{(n-1) \sum_{i=1}^n \sum_{j=1}^n w_{ij} (z_i - z_j)^2}{\sum_{i=1}^n \sum_{j=1}^n w_{ij} \sum_{i=1}^n z_i^2} \quad (3.17)$$

Modelos de Regressão com Efeitos Espaciais Globais

Segundo Cruz et al. (2004), o uso de modelos de regressão é uma das formas mais comuns de estudo para dados de área. Para que se possa incluir os efeitos

espaciais em modelos de regressão, supõe-se que é possível capturar a estrutura de correlação espacial num único parâmetro, que é adicionado ao modelo de regressão tradicional. A autocorrelação espacial ignorada é atribuída à variável dependente Y . Este método é denominado como modelo espacial autoregressivo misto (ou do inglês *SAR*, *Spatial Autoregressive*, ou ainda como *spatial lag model*), dado que se considera a dependência espacial através da adição ao modelo de regressão de um novo termo na forma de uma relação espacial para a variável dependente. Isto pode ser expresso por:

$$Y = \rho WY + X\beta + \epsilon \quad (3.18)$$

onde W é a matriz de proximidade espacial, WY expressa a dependência espacial em Y , ρ é o coeficiente espacial autoregressivo e X são covariáveis. Este modelo possui como hipótese nula a não existência de autocorrelação, $\rho = 0$, (Cruz et al., 2004). Em função das componentes individuais, este modelo pode ser dado por:

$$y_i = \rho \left(\sum_j w_{ij} y_j \right) + \sum_{i=1} x_i \beta_i + \epsilon_i \quad (3.19)$$

Estimação de Modelos Contínuos a partir de Dados de Área

Utiliza-se um estimador de intensidade (*Kernel*) para nos fornecer uma primeira aproximação da distribuição espacial do fenômeno ou variável para toda a posição espacial (X,Y) , cujo valor queremos estimar. Considerando o modelo de variação espacial contínua, que supõe um processo estocástico $Z(x), x \in A, A \subset R^2$, cujos valores podem ser conhecidos em todos os pontos da área de estudo. O estimador de intensidade será calculado a partir dos valores z_1, \dots, z_n contidos num raio de

tamanho τ . Esse estimador é definido como:

$$\hat{z}_i = \frac{\sum_{j=1}^n k\left(\frac{d_{ij}}{\tau}\right) z_j}{\sum_{j=1}^n k\left(\frac{d_{ij}}{\tau}\right)}, d_{ij} \leq \tau \quad (3.20)$$

sendo a função $k()$ um interpolador não-paramétrico. Caso as observações nas áreas representem dados de contagem, o estimador acima não é apropriado. Recomenda-se o uso da Equação 3.21.

$$\hat{z}_i = \frac{1}{\pi\tau^2} \sum_{j=1}^n k\left(\frac{d_{ij}}{\tau}\right) z_j, d_{ij} \leq \tau \quad (3.21)$$

3.4.3 Análise de Superfície: Krigagem

Os modelos estatísticos de efeitos locais e globais (modelos de *Krigagem*) são modelos onde cada ponto da superfície é estimado a partir da interpolação das amostras mais próximas, utilizando um estimador estatístico.

Este método se diferencia dos outros métodos de interpolação devido a estimação da matriz de covariância espacial, que determina os pesos atribuídos às diferentes amostras, o tratamento da redundância dos dados, a vizinhança a ser considerada no procedimento inferencial e o erro associado ao valor estimado. Além disso, a *krigagem* também fornece estimativas com propriedades de não tendenciosidade e eficiência (Camargo et al., 2004).

Variograma

O *variograma* pode ser usado como indicador da dependência espacial. Associa-se um valor único do atributo de cada área a um ponto, normalmente o centro geométrico ou populacional do polígono, e a partir deste ponto calcula-se a função

variograma, sendo duas vezes a função do *semivariograma*.

$$2\hat{\gamma}(h) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} [z(x_i) - z(x_i + h)]^2 \quad (3.22)$$

sendo $\hat{\gamma}(h)$ o *semivariograma* estimado e $N(h)$ o número de pares de valores medidos, $z(x)$ e $z(x + h)$, separados pelo vetor \vec{h} .

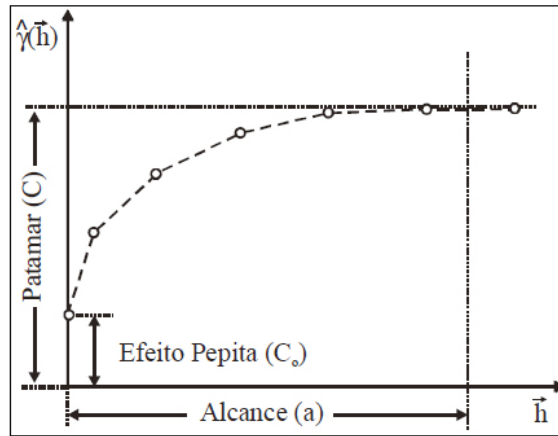


Figura 3.10: Parâmetros do *variograma*.
Fonte: Câmara et al. (2004)

Como mostra a Figura 3.10, o *Alcance (a)* representa a distância dentro da qual as amostras apresentam-se correlacionadas espacialmente. O *Patamar (C)* é o valor do *semivariograma* correspondente a seu *Alcance (a)*. Quando o dado é não-estacionário, o *variograma* não se estabiliza e continua crescendo junto com o *Alcance(a)*. O *Efeito Pepita (C_0)* é o valor da semivariância para a distância zero e representa a componente da variabilidade espacial que não pode ser relacionado com uma causa específica (Camargo et al., 2004).

A Krigagem

A idéia da *Krigagem* está baseada no conceito de variável regionalizada. Variáveis regionalizadas são variáveis distribuídas no espaço, ou tempo, e tem seus valores

considerados como realizações de uma função aleatória, ou processo aleatório ou estocástico. A variação espacial de uma variável regionalizada pode ser descrita por:

$$Z(x) = \mu(x) + \varepsilon'(x) + \varepsilon'' \quad (3.23)$$

onde x é o vetor de posições no espaço, $\mu(x)$ é uma função determinística que descreve a componente estrutural de Z em x , $\varepsilon'(x)$ é um termo estocástico, correlacionado, que varia localmente e ε'' é um ruído aleatório não correlacionado, com distribuição normal com média zero e variância σ^2 .

Uma hipótese a ser feita sobre o comportamento da variável regionalizada é que $\mu(x)$, média do fenômeno, é constante em toda região de estudo. Esta gera os modelos de *Krigagem ordinária*, onde $\mu(x)$ é constante e denotada por m . Desta forma assume-se que o valor esperado da função aleatória Z nas posições x e $x + h$ são iguais a m e que o valor esperado de sua diferença é nulo. Admite-se também, estacionariedade de segunda ordem para o fenômeno em estudo, ou seja, a covariância e a variância entre dois pontos $Z(x)$ e $Z(x + h)$ existe e depende somente da distância h .

$$C(h) = COV[Z(x), Z(x + h)] = E[Z(x).Z(x + h)] - m^2 \quad (3.24)$$

$$Var(Z(x)) = E[Z^2(x)] - 2E[Z(x)].m + m^2 = E[Z^2(x)] - m^2 = C(0) \quad (3.25)$$

As Equações 3.24 e 3.25 implicam que a determinação da função $C(h)$ é suficiente para caracterizar a variável regionalizada, ou seja, a função $C(h)$ permite caracterizar o termo $\varepsilon'(x)$. Para se determinar a função $C(h)$ utiliza-se a função *variograma* definida como:

$$2\gamma(h) = E[Z(x) - Z(x + h)]^2 = E[Z^2(x)] - 2E[Z(x).Z(x + h)] + E[Z^2(x + h)] \quad (3.26)$$

De acordo com as Equações 3.24 e 3.25 têm-se, respectivamente, que

$$E[Z(x).Z(x+h)] = C(h) + m^2 \quad (3.27)$$

$$E[Z^2(x)] = E[Z^2(x+h)] = C(0) + m^2 \quad (3.28)$$

Utilizando 3.27 e 3.28 em 3.26, obtém-se:

$$\gamma(h) = C(0) - C(h) \quad (3.29)$$

A Equação 3.29 demonstra que sob a hipótese de estacionariedade de segunda ordem o $\gamma(h)$, *semivariograma*, e a covariância são formas alternativas de caracterizar a autocorrelação dos pares $Z(x)$ e $Z(x+h)$.

A Figura 3.10, representa o gráfico do semivariograma experimental, $\gamma(\hat{h})$, que é formado por uma série de valores sobre os quais se objetiva ajustar uma função. Para que as estimativas da *Krigagem* sejam mais exatas é necessário que o modelo ajustado represente a tendência de $\gamma(\hat{h})$ em relação a h .

O procedimento de ajuste é feito de forma iterativa, fazendo-se um primeiro ajuste e, em seguida, verificando-se a adequação do modelo. Dependendo do ajuste obtido, pode-se ou não redefinir o modelo, até se obter um que seja considerado satisfatório (Camargo et al., 2004).

Os modelos básicos ou modelos isotrópicos estão divididos em dois tipos: modelos com patamar ou transitivos e modelos sem patamar. Alguns dos modelos do primeiro tipo atingem o *Patamar* (C) de forma assintótica e tem o *Alcance* (a) definido como a distância correspondente a 95% do patamar. Os modelos transitivos mais utilizados são o modelo esférico, exponencial e gaussiano. Os modelos do

segundo tipo não atingem o patamar e continuam a crescer a medida que a distancia aumenta. Este tipo de modelo é utilizado usualmente para dados que possuem capacidade infinita de dispersão.

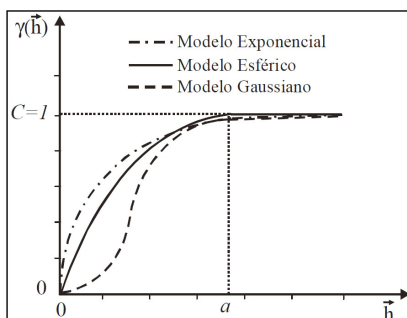


Figura 3.11: Representação gráfica de *semivariogramas* experimentais e modelos teóricos.

Fonte: Câmara et al. (2004)

Quando a variação espacial dos fenômenos não pode ser simplesmente explicada pelos modelos básicos do *semivariograma*, utilizam-se os modelos aninhados. Estes modelos são as combinações de modelos simples e em muitos casos conseguem explicar a variação dos fenômenos decorrentes da combinação de fatores independentes de formação.

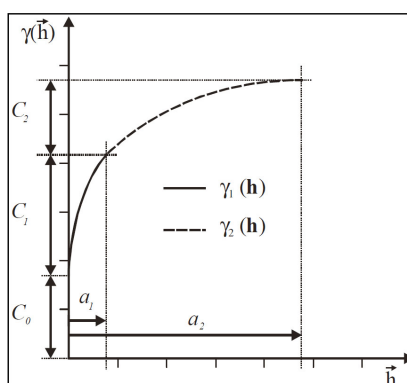


Figura 3.12: Representação gráfica de um modelo duplo esférico.

Fonte: Câmara et al. (2004)

Krigagem Ordinária

Camargo et al. (2004) explicam *Krigagem ordinária* através do seguinte exemplo: considere uma região onde se observe um fenômeno Z , em n pontos distintos, com coordenadas representadas pelo vetor x . Tem-se então, um conjunto de valores $z(x_i), i = 1, \dots, n$, onde x_i identifica uma posição em duas dimensões. O objetivo é estimar o valor do fenômeno Z em um ponto c . O valor desconhecido de $Z(x_0)$ pode ser estimado a partir de uma combinação linear dos n valores observados, adicionado a um parâmetro λ_0 como mostra a Equação 3.30.

$$Z^*(x_0) = \lambda_0 + \sum_{i=1}^n \lambda_i Z(x_i) \quad (3.30)$$

Deseja-se um estimador não tendencioso, ou seja,

$$E[Z(x_0) - Z^*(x_0)] = 0 \quad (3.31)$$

Através da Equação 3.31 supõe-se que as médias sejam iguais e obtém-se:

$$E[Z(x_0)] = E[\lambda_0 + \sum_{i=1}^n \lambda_i Z(x_i)] \Rightarrow m = \lambda_0 + \sum_{i=1}^n \lambda_i m \quad (3.32)$$

Logo, para que a igualdade da Equação 3.32 seja satisfeita temos que $\lambda_0 = 0$, $\sum_{i=1}^n \lambda_i = 1$. Essas duas condições aplicadas na Equação 3.30 geram 3.33.

$$Z^*(x_0) = \sum_{i=1}^n \lambda_i Z(x_i), \text{ com } \sum_{i=1}^n \lambda_i = 1 \quad (3.33)$$

Minimizando a variância do erro ($Var[Z(x_0) - Z^*(x_0)]$) de acordo com a condição de $\sum_{i=1}^n \lambda_i = 1$, os pesos λ_i são obtidos a partir do seguinte sistema de equações, denominado sistema de *Krigeagem ordinária*:

$$\begin{cases} \sum_{j=1}^n \lambda_j C(x_i, x_j) - \alpha = C(x_i, x_0), & \text{para } i = 1, \dots, n \\ \sum_{j=1}^n \lambda_j = 1 \end{cases} \quad (3.34)$$

sendo $C(x_i, x_j)$ e $C(x_i, x_0)$ a semivariância entre os pontos x_i e x_j e entre os pontos x_i e x_0 , respectivamente, e α o multiplicador de *Lagrange* necessário para a minimização da variância do erro.

A variância da *Krigagem ordinária* (σ_{ko}^2), corresponde a variância minimizada do erro e é definida por:

$$\sigma_{ko}^2 = Var[Z(x) - Z^*(x_0)] = C(0) - \sum_{i=1}^n \lambda_i C(x_i, x_0) - \alpha \quad (3.35)$$

De acordo com Camargo et al. (2004), quando as equações acima forem usadas, os valores interpolados irão coincidir com os valores dos pontos amostrais, isso demonstra que a *Krigagem ordinária* é um interpolador exato.

Capítulo 4

Material e Métodos

4.1 Introdução

Neste Capítulo são apresentados os materiais e métodos a serem usados no trabalho. O material a ser utilizado é composto pelos boletins de ocorrência dos crimes dos anos de 2010 e 2011 presentes no banco da SSP/DF, cedido pela própria. Dentro deste banco serão analisados os crimes de **Furto** (de um modo geral) e **Lesão corporal** através dos métodos de análise da estatística espacial. Para isso, será empregado o sistema de coordenadas lat/long para identificar cada ocorrência no mapa. Tanto a análise quanto a obtenção destas coordenadas serão feitas via um algoritmo implementado em linguagem IML do *software* SAS 9.2.

4.1.1 Sistema de Coordenadas Geográficas

O sistema de mapeamento da Terra através de coordenadas geográficas expressa qualquer posição horizontal no planeta através de duas das três coordenadas existentes num sistema esférico de coordenadas, alinhadas com o eixo de rotação da Terra.

4.1.2 Latitude e Longitude

O sistema de coordenadas geográficas é um sistema que se baseia nos conceitos de Latitude e Longitude. A latitude é um valor numérico que varia entre 0° (Equador) a 90° (pólos) e descreve a posição de um ponto situado a norte ou a sul do Equador (linha perpendicular ao eixo terrestre que divide o planeta em dois hemisférios: norte e sul). A longitude é outro valor numérico que descreve a posições de pontos situados a leste ou a oeste do meridiano de Greenwich variando de 0° (meridiano de Greenwich) a 180° (a leste ou a oeste). O meridiano de Greenwich está situado na Inglaterra e este é definido como o meridiano central.

Um sistema de coordenadas que utiliza a latitude e a longitude possui seu ponto de origem situado no ponto de 0° de latitude e 0° de longitude e este sistema pode ser dividido em quatro quadrantes: o primeiro sendo NE (latitude e longitude positivas), o segundo NW (latitude positiva e longitude negativa), o terceiro SE (latitude negativa e longitude positiva) e o quarto sendo SW (latitude e longitude negativas). As coordenadas geográficas são o conjunto de meridianos e paralelos que formam uma grade de linhas imaginárias ao redor do planeta. Cada grau de latitude e longitude pode ser transformado em 60 minutos e cada minuto em 60 segundos. Através desta transformação é possível aumentar a precisão de uma localização em qualquer lugar da superfície terrestre.

4.1.3 Algoritmo IML

Os algoritmos a seguir foram construídos utilizando o procedimento IML (*Interactive Matrix Language*), do *software* SAS 9.2, devido aos códigos implementados

no IML serem facilmente transformados para outras linguagens de programação.

4.1.4 Método para Transformação dos Endereços

O método utilizado para realizar a transformação dos endereços será a **Macro GoogleMaps**. Esta macro aplica os endereços físicos presentes em uma variável macro na *url* da busca do site do *Google maps* e armazena as coordenadas Lat/Long encontradas para um *data*.

4.1.5 Métodos para Análise Pontual

Nesta parte da análise, primeiramente, serão traçadas as ocorrências no mapa e será analisada a existência ou não de padrões pontuais aparentes ao passar da série temporal.

Em seguida, será utilizado a função de intensidade de *Kernell* para gerar a distribuição das ocorrências no mapa, através da contagem de todos os eventos dentro da região de influência, ponderando-os pela distância de cada uma das localizações de interesse.

Seguindo com a análise, será utilizado o método do vizinho mais próximo para verificar de forma exploratória, como as interações entre as ocorrências se comportam de acordo com o aumento da distância entre estas.

Para complementar a análise, será feito o método de Envelopes de Simulação. O procedimento de Envelope de Simulação pode ser dividido em duas partes: a primeira, sendo composta pelo cálculo dos valores da distribuição simulada $\overline{G}(h)$ e a segunda, sendo o gráfico da distribuição estimada $\hat{G}(h)$ versus a distribuição simulada $\overline{G}(h)$ para se testar aleatoriedade na distribuição dos eventos.

4.1.6 Métodos para Análise de Áreas

Para a análise de áreas será necessária a agregação dos eventos presentes em cada setor censitário, em seu centróide, gerando a frequência destes por área.

Em seguida, será calculado o índice global de *Moran*. Este índice analisa a autocorrelação espacial entre as áreas no intuito de identificar a estrutura de correlação espacial que melhor descreve os dados. Como análise adicional para a visualização da dependência espacial será feito o diagrama de espalhamento de *Moran*, tanto pelo gráfico quanto pelo mapa.

Para finalizar a análise de áreas, será feito o modelo autoregressivo misto (SAR, *Spatial AutoRegressive*, ou *spatial lag model*), no intuito de identificar um modelo com bom ajuste entre os valores preditos e os valores observados da variável dependente. Este tipo de modelo leva em consideração que a autocorrelação espacial é ignorada e adicionada na variável dependente Y .

4.1.7 Métodos para Análise da Krigagem

Primeiramente, nesta análise, será desenvolvida a superfície de tendência para os eventos em estudo, fornecendo assim uma idéia de como os dados estão distribuídos no espaço.

Finalizada a parte da análise descritiva, parte-se para a busca do modelo teórico que melhor explique a distribuição dos dados através do variograma e semivariograma. O intuito das técnicas do variograma e semivariograma é a busca dos melhores parâmetros: *lag*, *lagdistance*, *nugget*(pepita), *range*(alcance) e *sille*(patamar).

Como passo final, será feita a krigagem ordinária, utilizando o modelo teórico

encontrado para a criação da superfície de densidade dos dados. Esta técnica parte da hipótese que a média é constante em toda a região em estudo, e nos apresenta estimativas não tendenciosas, eficientes. Como análise adicional, será feita a superfície de densidade do erro padrão para demonstrar a eficiência das estimativas.

Capítulo 5

Análise dos Resultados

5.1 Introdução

Neste capítulo serão apresentados os resultados obtidos das análises. Inicialmente, foi realizada uma análise descritiva dos dados, com o intuito de verificar a distribuição das ocorrências no espaço. Foram descritos os resultados obtidos dos três tipos de análise espacial (pontual, área e krigagem). Todos os resultados foram gerados utilizando o *software* SAS 9.2.

5.2 Análise Descritiva

Na análise descritiva, desenhamos os pontos que representam as ocorrências do tipo de crime furto no mapa do Distrito Federal, indentificamos em quais áreas existem aglomerados destes eventos e calculamos a tabela de frequência das ocorrências no decorer do tempo.

A Figura 5.1 demonstra as localizações das ocorrências do crime de furto para os anos de 2010 e 2011, levando em conta os meses de janeiro, junho e dezembro, no intuito de indentificar se existe movimentação quanto aos locais que possuem maior foco de ocorrências.

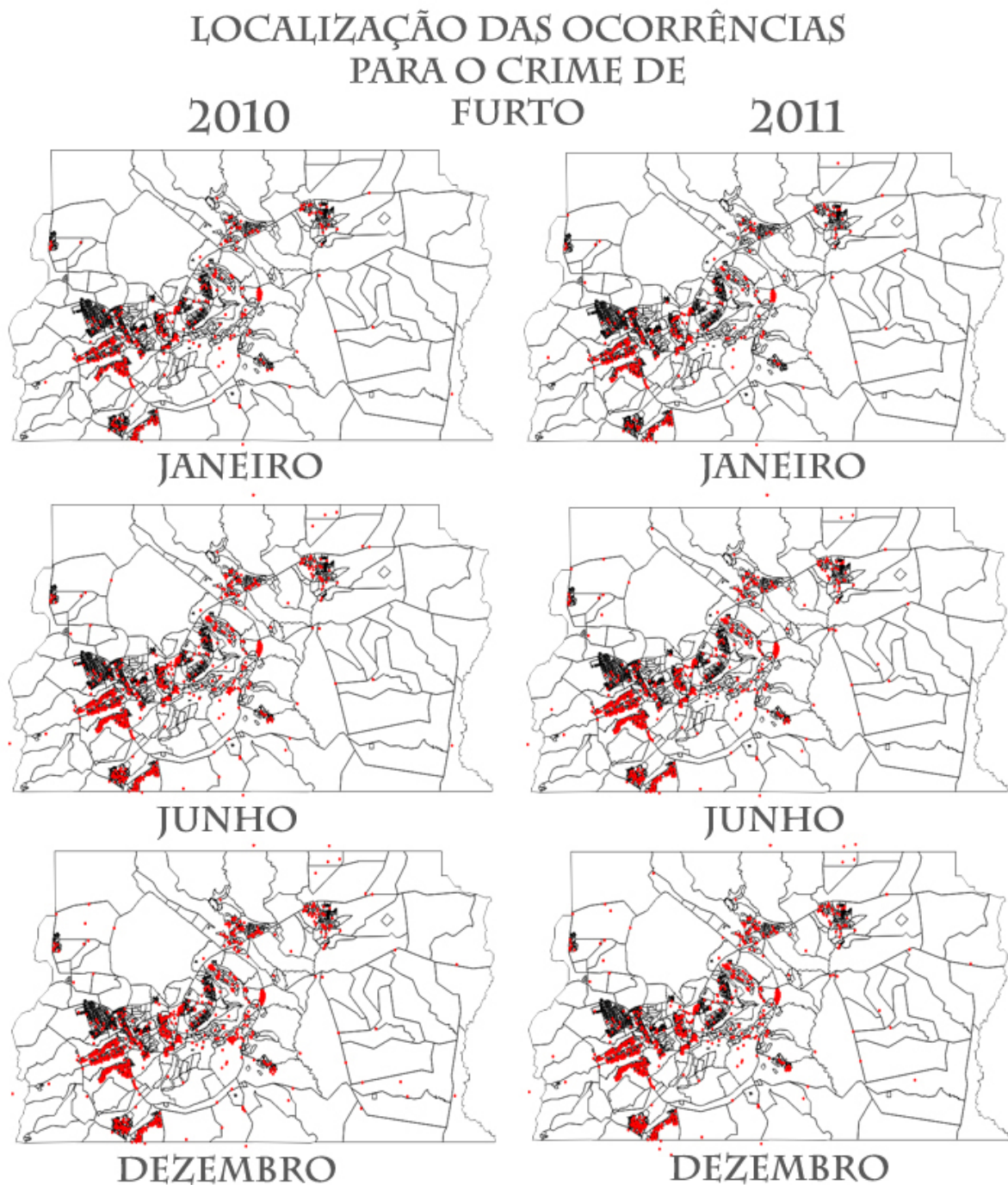


Figura 5.1: Localização das ocorrências para o crime de furto para os anos de 2010 e 2011 no Distrito Federal.

No mapeamento das localizações do crime de furto foram observados aglomerados de pontos nas regiões administrativas de Candangolândia, Núcleo Bandeirante,

Paranoá, Taguatinga, Itapoã, Gama e Samambaia, nos anos de 2010 e 2011. Com o decorrer dos meses, nota-se o surgimento de novos locais de ocorrência, entretanto, os focos de ocorrência acentuada permaneceram os mesmos.

As Tabelas 5.1 e 5.2 demonstram a frequência acumulada das ocorrências dos crimes de furto em 2010 e 2011, respectivamente, para os meses de estudo.

Tabela 5.1: **Tabela de Frequência Acumulada para o Crime de Furto em 2010**

LOCAL	Janeiro	Junho	Dezembro
ÁGUAS CLARAS	176	828	1803
BRASÍLIA	1535	7128	13615
BRAZLÂNDIA	38	263	554
CANDANGOLÂNDIA	24	88	269
CEILÂNDIA	456	2319	4697
CRUZEIRO	52	289	570
ESTRUTURAL	41	188	417
FERCAL	8	25	58
GAMA	211	1094	2228
GUARÁ	235	1177	2641
ITAPOÃ	30	189	435
JARDIM BOTÂNICO	12	62	135
LAGO NORTE	61	260	576
LAGO SUL	123	528	1076
NÚCLEO BANDEIRANTE	86	458	980
PARANOÁ	63	336	707
PARK WAY	6	36	71
PLANALTINA	204	1020	2252
RECANTO DAS EMAS	89	482	1015
RIACHO FUNDO	75	307	599
RIACHO FUNDO 2	23	92	183
SAMAMBAIA	208	1047	2310
SANTA MARIA	120	557	1203
SAO SEBASTIÃO	55	301	653
SIA	89	412	867
SOBRADINHO	128	702	1423
SOBRADINHO 2	68	347	696
SUDOESTE	85	305	627
TAGUATINGA	459	2269	4727
VARJÃO	.	.	1

Tabela 5.2: Tabela de Frequência Acumulada para o Crime de Furto em 2011

LOCAL	Janeiro	Junho	Dezembro
ÁGUAS CLARAS	207	1080	2185
BRASÍLIA	916	5951	11418
BRAZLÂNDIA	45	318	533
CANDANGOLÂNDIA	27	144	314
CEILÂNDIA	322	2005	3901
CRUZEIRO	59	257	506
ESTRUTURAL	36	215	546
FERCAL	5	37	73
GAMA	229	1136	2346
GUARÁ	240	1197	2195
ITAPOÃ	49	180	333
JARDIM BOTÂNICO	13	77	151
LAGO NORTE	55	263	523
LAGO SUL	87	486	937
NÚCLEO BANDEIRANTE	77	486	959
PARANOÁ	64	447	748
PARK WAY	9	48	85
PLANALTINA	222	1197	2139
RECANTO DAS EMAS	103	591	1214
RIACHO FUNDO	54	316	683
RIACHO FUNDO 2	20	119	247
SAMAMBAIA	238	1297	2491
SANTA MARIA	120	627	1147
SAO SEBASTIÃO	50	313	563
SIA	103	515	951
SOBRADINHO	132	688	1450
SOBRADINHO 2	59	327	585
SUDOESTE	51	279	519
TAGUATINGA	407	2193	4302
VARJÃO	.	.	1

Através das tabelas de frequência acumulada observamos a maior concentração de ocorrências na região administrativa de Brasília, aproximadamente 30%, em todos os meses de estudo. Outras focos de ocorrência que se destacaram se encontram nas regiões administrativas de Ceilândia, Taguatinga, Guará, Gama, Samambaia e Planaltina.

Realizando o mesmo procedimento para o crimes de lesão corporal encontramos resultados semelhantes aos do crime de furto, como mostra a Figura 5.2 e as Tabelas 5.3 e 5.4.

Tabela 5.3: Tabela de Frequência Acumulada para o Crime de Lesão Corporal em 2010

LOCAL	Janeiro	Junho	Dezembro
ÁGUAS CLARAS	46	304	603
BRASÍLIA	163	1198	2512
BRAZLÂNDIA	25	166	339
CANDANGOLÂNDIA	17	89	178
CEILÂNDIA	136	730	1569
CRUZEIRO	17	87	169
ESTRUTURAL	25	184	381
FERCAL	1	19	41
GAMA	72	467	1036
GUARÁ	68	385	835
ITAPOÃ	16	116	245
JARDIM BOTANICO	3	32	74
LAGO NORTE	5	70	161
LAGO SUL	25	135	246
NÚCLEO BANDEIRANTE	24	158	291
PARANOA	26	177	355
PARK WAY	1	18	40
PLANALTINA	92	540	1208
RECANTO DAS EMAS	60	375	816
RIACHO FUNDO	26	176	346
RIACHO FUNDO 2	17	79	147
SAMAMBAIA	86	492	1118
SANTA MARIA	60	380	878
SAO SEBASTIÃO	46	306	636
SIA	19	107	245
SOBRADINHO	55	283	599
SOBRADINHO 2	23	170	394
SUDOESTE	12	72	143
TAGUATINGA	130	832	1823
VARJÃO	.	.	1

Tabela 5.4: Tabela de Frequência Acumulada para o Crime de Lesão Corporal em 2011

LOCAL	Janeiro	Junho	Dezembro
ÁGUAS CLARAS	47	268	563
BRASÍLIA	175	1222	2303
BRAZLÂNDIA	26	162	326
CANDANGOLÂNDIA	16	75	141
CEILÂNDIA	128	748	1538
CRUZEIRO	25	105	193
ESTRUTURAL	31	166	327
FERCAL	3	22	52
GAMA	94	522	991
GUARÁ	72	353	684
ITAPOÃ	20	103	205
JARDIM BOTANICO	6	38	69
LAGO NORTE	9	58	123
LAGO SUL	19	124	246
NÚCLEO BANDEIRANTE	19	128	268
PARANOIA	33	198	377
PARK WAY	.	28	44
PLANALTINA	114	569	1107
RECANTO DAS EMAS	83	406	769
RIACHO FUNDO	31	169	345
RIACHO FUNDO 2	9	100	177
SAMAMBAIA	82	562	1078
SANTA MARIA	68	409	780
SAO SEBASTIÃO	72	358	665
SIA	20	115	220
SOBRADINHO	36	290	553
SOBRADINHO 2	28	212	412
SUDOESTE	13	79	149
TAGUATINGA	122	829	1601
VARJÃO	.	.	1

LOCALIZAÇÃO DAS OCORRÊNCIAS
PARA O CRIME DE
LESÃO CORPORAL

2010

2011



JANEIRO

JANEIRO



JUNHO

JUNHO



DEZEMBRO

DEZEMBRO

Figura 5.2: Localização das ocorrências para o crime de lesão corporal para os anos de 2010 e 2011 no Distrito Federal.

5.3 Análise Pontual

Para termos uma visualização mais precisa da forma como a intensidade das ocorrências está distribuída pelo território, utilizamos a função de intensidade de *Kernel*, para traçar a distribuição das ocorrências para os crimes selecionados. Começamos pelo crime de furto, como mostram as Figuras 5.3 e 5.4.

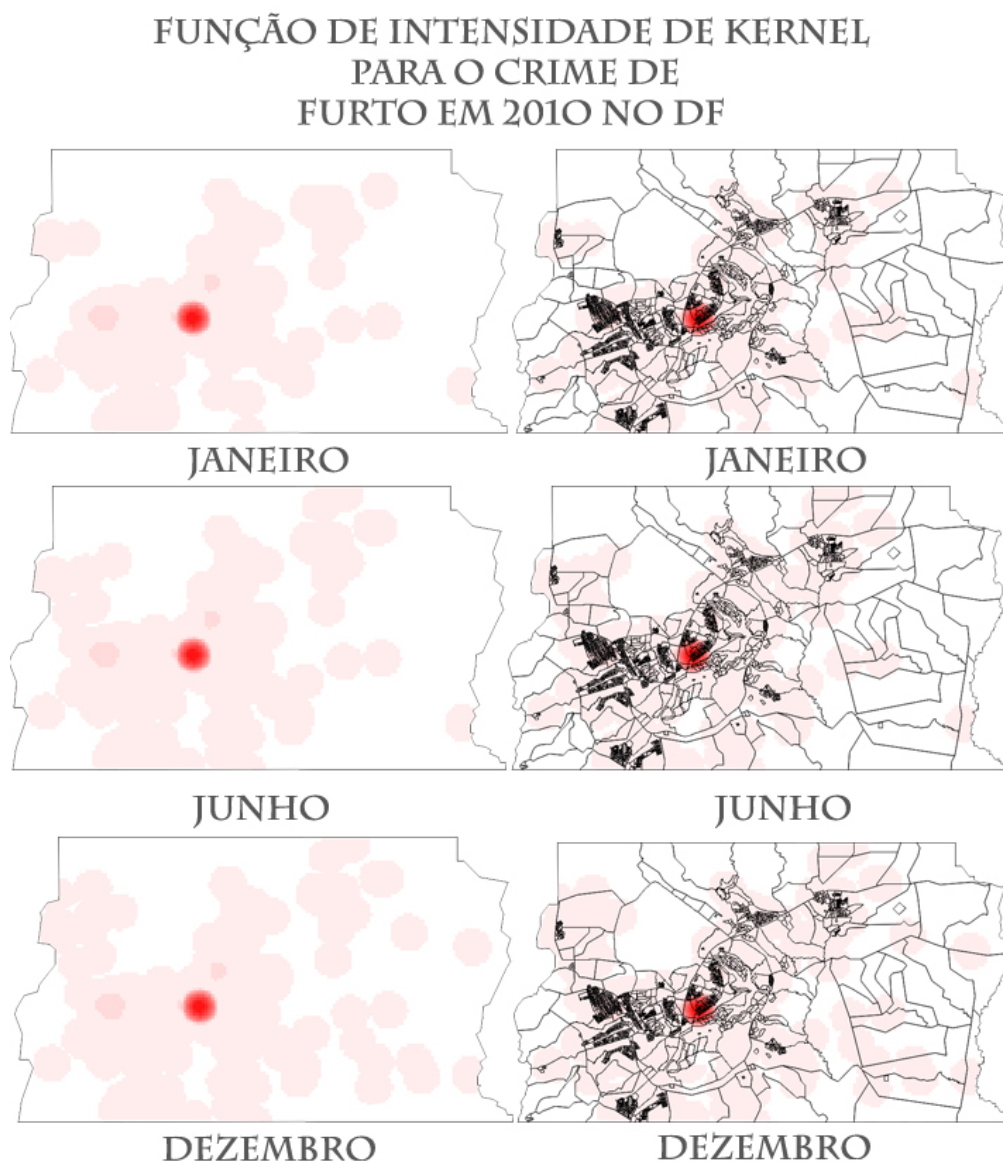


Figura 5.3: Distribuição das ocorrências para o crime de furto para os anos de 2010 no Distrito Federal utilizando a função de intensidade de *Kernel*.

FUNÇÃO DE INTENSIDADE DE KERNEL PARA O CRIME DE FURTO EM 2011 NO DF

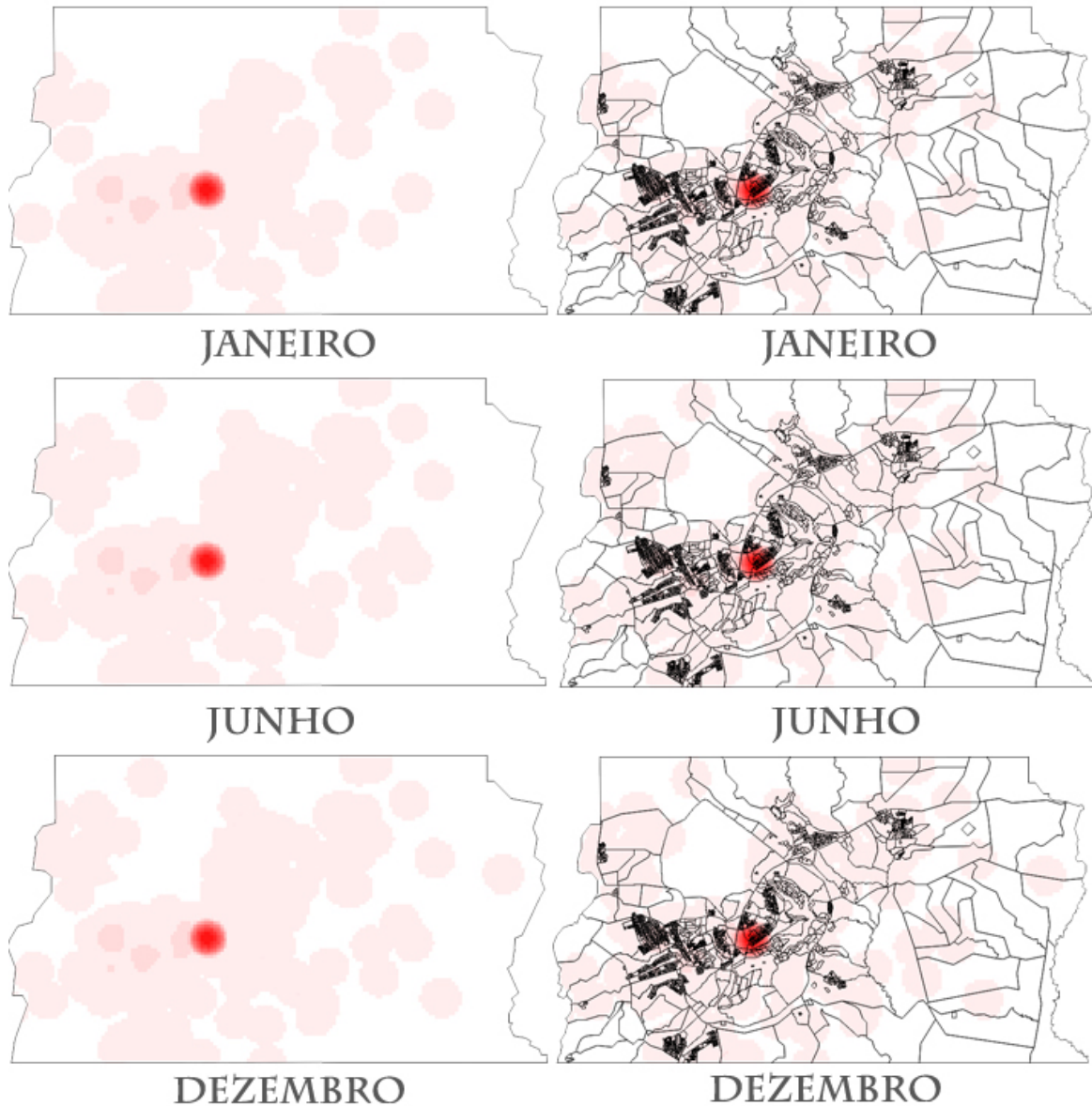


Figura 5.4: Distribuição das ocorrências para o crime de furto para os anos de 2011 no Distrito Federal utilizando a função de intensidade de *Kernel*.

Observamos que a distribuição das ocorrências tem como principal foco a região administrativa de Brasília, em 2010 e 2011, em todos os meses de estudo. Outros focos de ocorrência, que aparecem no mapa, localizam-se nas regiões administrati-

vas de Ceilândia e Taguatinga. As regiões administrativas do Guar e Samambaia apresentam maior intensidade no ano de 2011. Comparando ambos os anos, vemos o aparecimento de novos locais de ocorrncia no DF, com intensidade menor.

Como prximo passo, utilizamos o mtodo do vizinho mais prximo. Ao fazer o grfico da distribuio da funo cumulativa emprica $\hat{G}(h)$ para o crime de furto, temos uma idia da distncia mdia em qual ocorre interao entre as ocorrncias. Devido aos resultados deste mtodo variarem pouco para o crime de furto, nos meses de estudo nos anos de 2010 e 2011, apresentamos apenas um dos resultados (Figura 5.5).



Figura 5.5: Mtodo do Vizinho Mais Prximo para o crime de furto no Distrito Federal.

Notamos um crescimento acelerado para valores pequenos de distncia, isto nos mostra que h interao entre os eventos, caracterizando agrupamentos nestas esca-

las.

O último passo realizado na análise pontual foi a utilização do método dos envelopes de simulação. A Figura 5.6 mostra o gráfico gerado da distribuição estimada $\hat{G}(h)$ versus a distribuição simulada $\bar{G}(h)$, testando se existe ou não aleatoriedade na distribuição dos eventos. Novamente, devido aos valores dos resultados variarem pouco nos três meses de estudo nos anos de 2010 e 2011, apresentamos apenas um dos resultados (Figura 5.6).

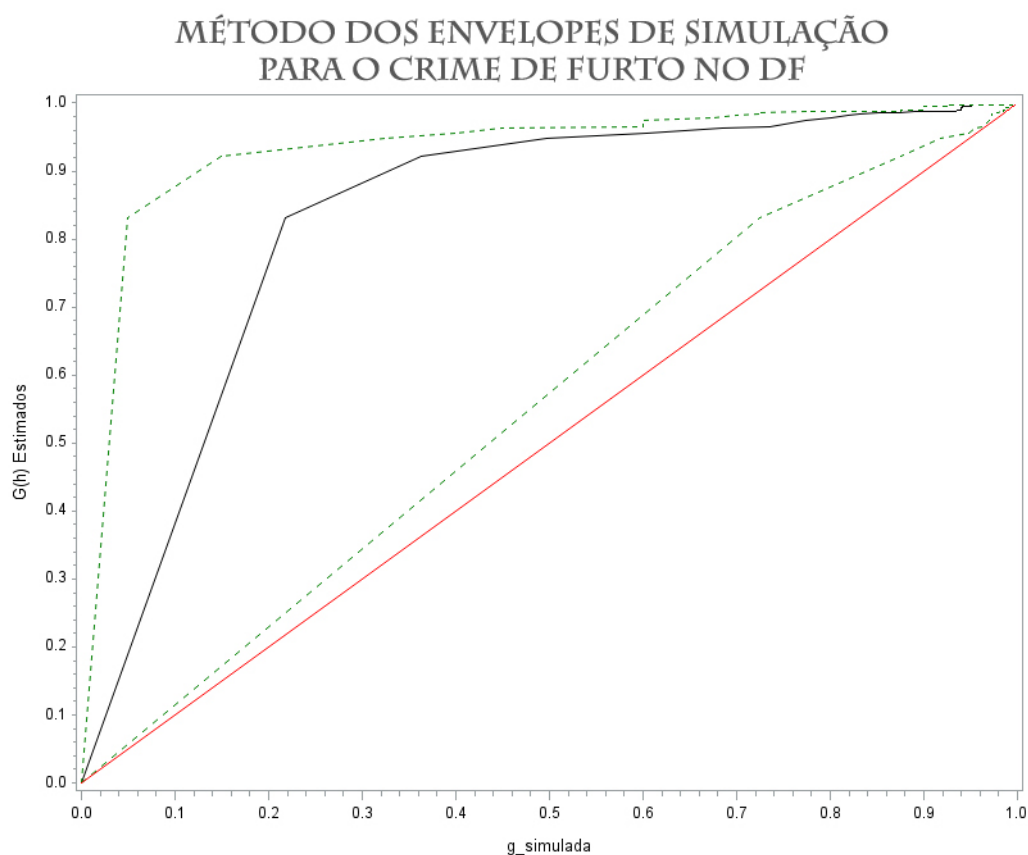


Figura 5.6: Método dos Envelopes de Simulação para o crime de furto no Distrito Federal.

Devido aos traçados no gráfico estarem acima da linha de 45° (Linha vermelha), temos evidências que os dados apresentam tendências para agrupamentos.

Em seguida, analisamos o crime de lesão corporal, conforme a Figura 5.7, que mostra os resultados da função de intensidade de *Kernell* para o crime em estudo nos meses de janeiro, junho e dezembro de 2010 e 2011.

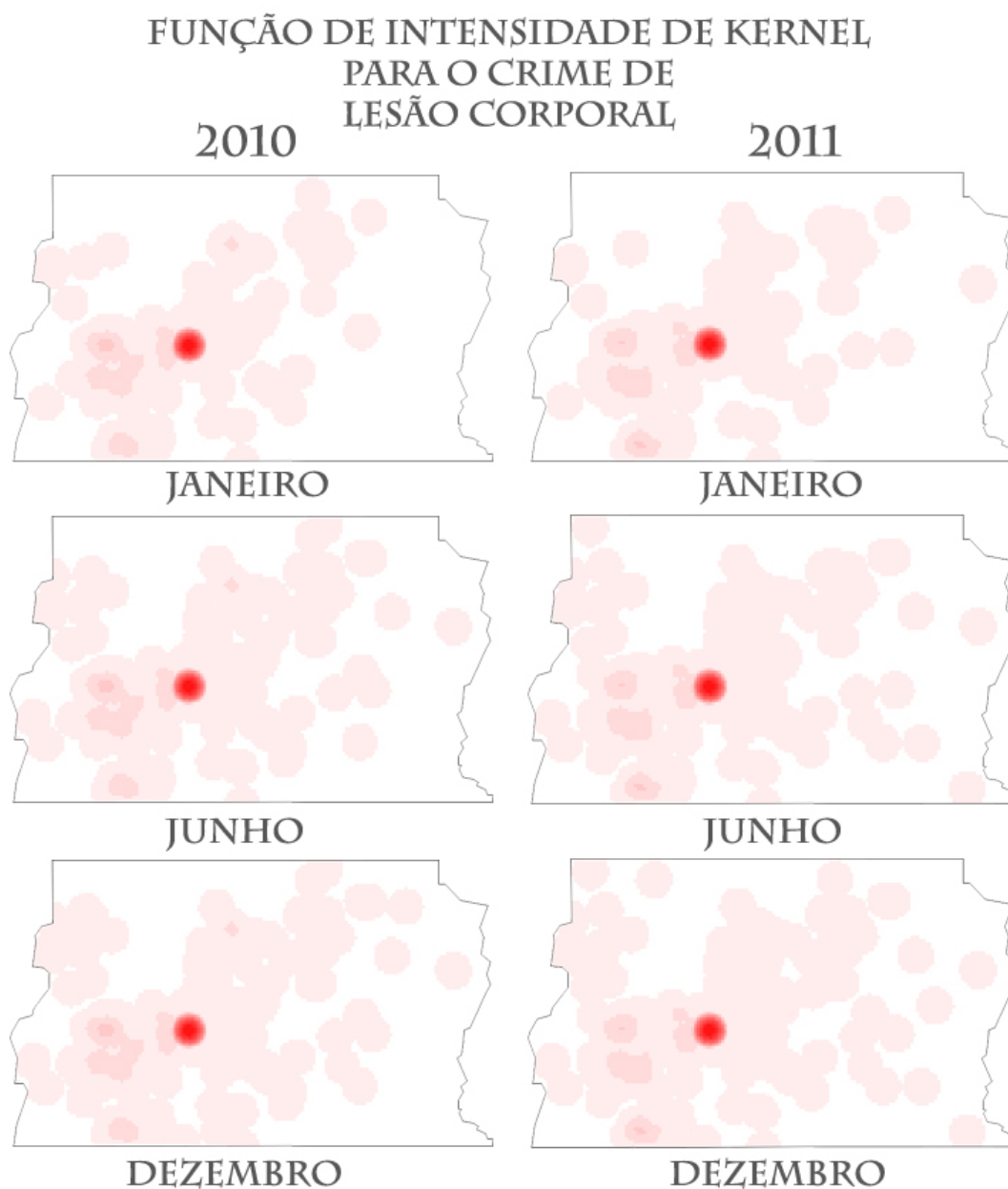


Figura 5.7: Distribuição das ocorrências para o crime de lesão corporal para os anos de 2010 e 2011 no Distrito Federal utilizando a função de intensidade de *Kernell*.

A Figura 5.7 mostra que a região administrativa de Brasília, em 2010 e 2011,

apresenta a maior intensidade de ocorrências entre todas as regiões do Distrito Federal. As áreas que pertencem as regiões administrativas de Ceilândia, Taguatinga, Samambaia, Sobradinho e Gama apresentam comportamento diferenciado para os anos em estudo. Em 2010, os valores da função de *Kernel* para as regiões de Sobradinho e Taguatinga apresentam maior intensidade. Em 2011, o Gama apresenta um crescimento do valor da função de intensidade.

Realizando os métodos do vizinho mais próximo e dos envelopes de simulação, encontramos resultados semelhantes entre os meses em estudo nos anos de 2010 e 2011, logo, apresentamos apenas um resultado de cada método.



Figura 5.8: Método do Vizinho Mais Próximo para o crime de lesão corporal no Distrito Federal.

Na Figura 5.8, notamos um acelerado crescimento para valores de pequena distância, isto nos mostra que há interação entre os eventos, caracterizando agrupamentos nestas escalas.

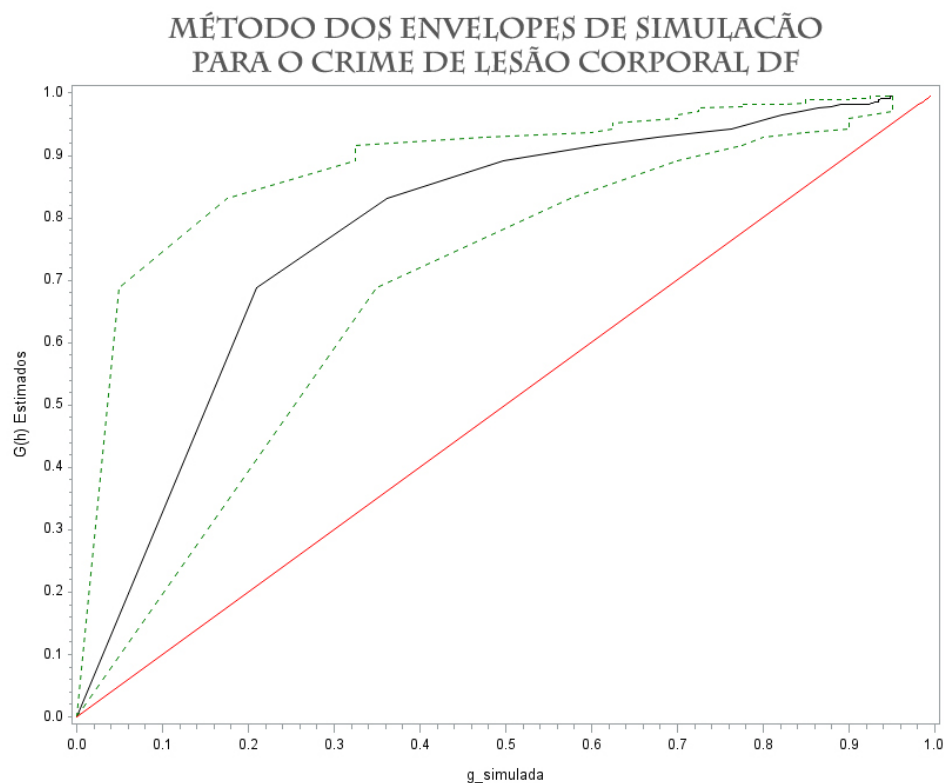


Figura 5.9: Método dos Envelopes de Simulação para o crime de lesão corporal no Distrito Federal.

A Figura 5.9 mostra que devido aos traçados no gráfico estarem acima da linha de 45° (Linha vermelha), temos evidências que os dados apresentam tendências para agrupamentos.

5.4 Análise de Áreas

Primeiramente, para a realização da análise de áreas, foi necessária à agregação dos eventos presentes em cada setor censitário, em seu centróide, gerando a

frequência destes por área.

Em seguida, calculamos o índice global de *Moran* para o crime de furto em 2010 e 2011, no Distrito Federal, para os meses de janeiro, junho e dezembro.

Tabela 5.5: **I de Moran para as ocorrências do crime de Furto em 2010**

Mês	I	p-valor
Janeiro	0.0041114658	0.4559610012
Junho	0.0034186053	0.4520537418
Dezembro	0.0032503999	0.4506855556

Tabela 5.6: **I de Moran para as ocorrências do crime de Furto em 2011**

Mês	I	p-valor
Janeiro	0.0062493007	0.4404857786
Junho	0.0038478683	0.448197191
Dezembro	0.0024630598	0.459094459

De acordo com as Tabelas 5.5 e 5.6, verificamos que os valores do índice *I* de *Moran* encontrados para o referido crime, tanto em 2010 quanto em 2011 nos meses em estudo, são positivos e baixos. Em seguida, realizamos o teste de significância para este índice associando este a distribuição normal, verificando a existência ou não de correlação espacial. De acordo com os altos valores do p-valor encontrados, não possuímos evidências para a rejeição da hipótese nula, ou seja, não rejeitamos a hipótese de independência espacial para nenhum dos três meses, nos dois anos em estudo.

Como análise adicional para a visualização da falta de dependência espacial entre os dados, foi feito o diagrama de espalhamento de *Moran*, tanto pelo gráfico quanto pelo mapa.

DIAGRAMA DE ESPALHAMENTO DE MORAN PARA O CRIME DE FURTO NO DISTRITO FEDERAL

2010

2011

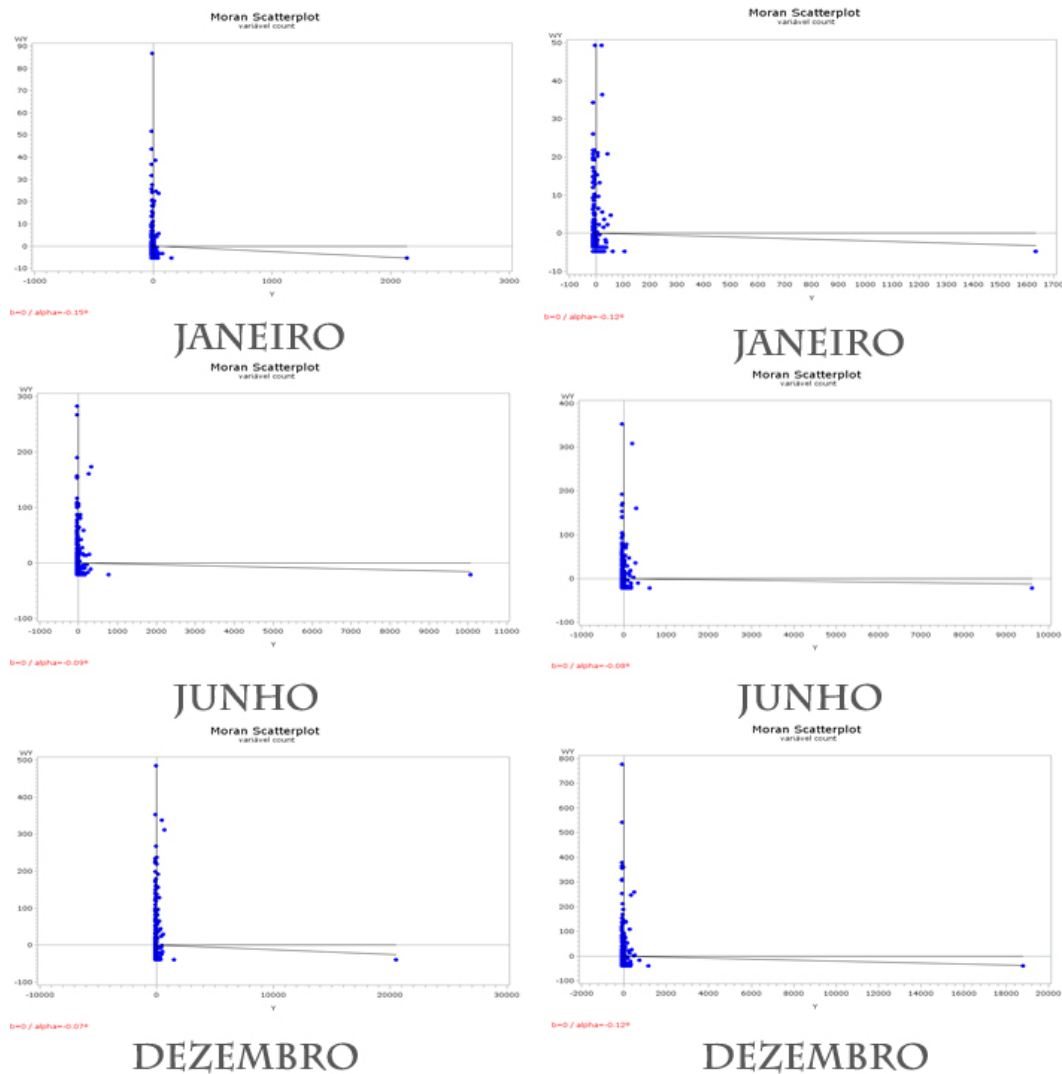


Figura 5.10: Diagrama de Espalhamento de *Moran* para o crime de Furto no Distrito Federal em 2010 e 2011.

A Figura 5.10 demonstra forte agrupamento dos dados perto do zero, o que corrobora a hipótese de independência espacial entre estes, como demonstrado pelo índice I de *Moran*.

DIAGRAMA DE ESPALHAMENTO BIDIMENSIONAL DE MORAN PARA O CRIME DE FURTO NO DISTRITO FEDERAL

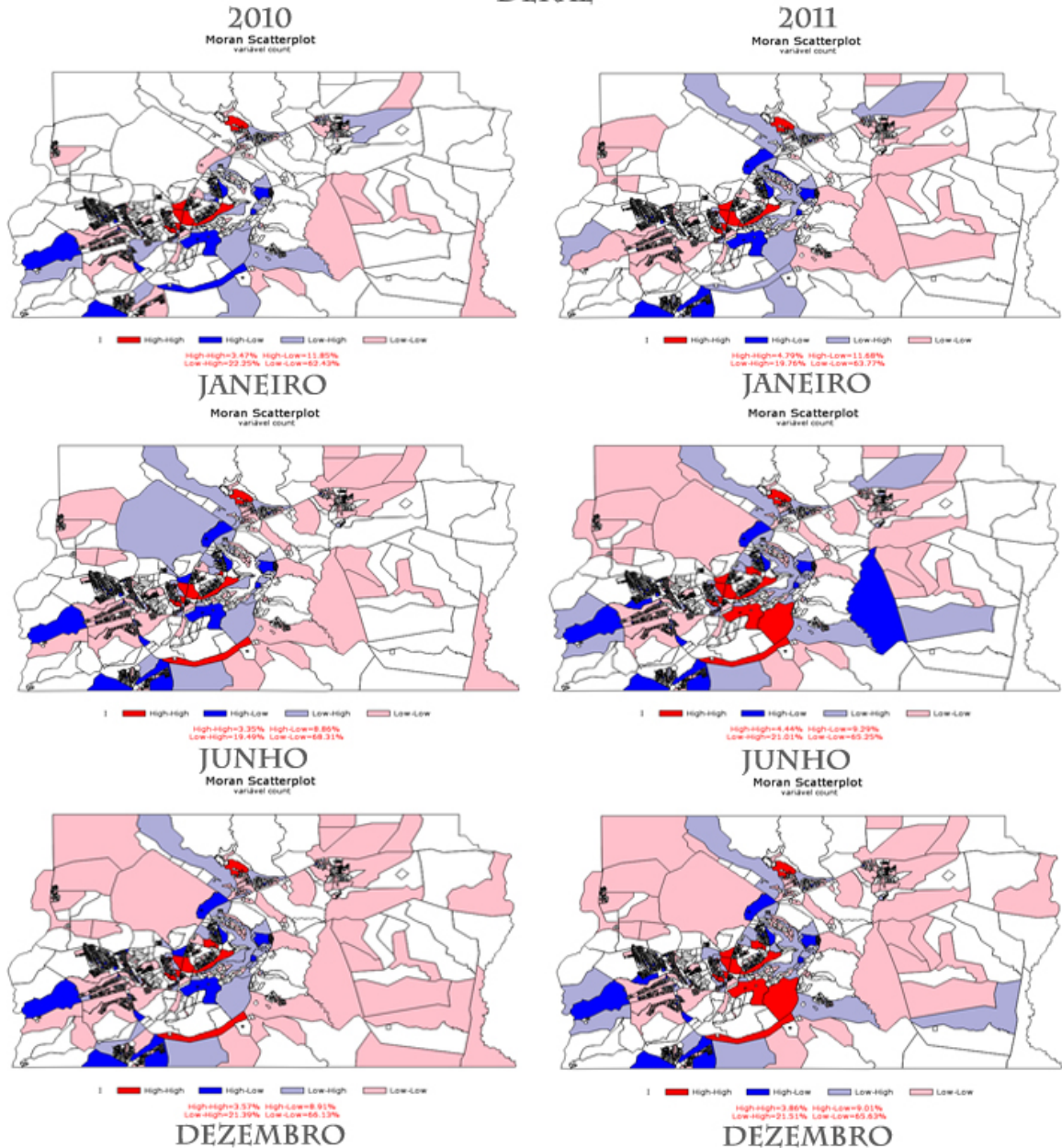


Figura 5.11: Diagrama de Espalhamento de *Moran* na forma de mapa temático bidimensional para o crime de Furto no Distrito Federal em 2010 e 2011.

A Figura 5.11 mostra no mapa que não existe uma formação específica de pólos, logo este resultado corrobora novamente a hipótese de independência espacial entre os dados, como demonstrado pelo índice I de *Moran*.

Para finalizar a análise de áreas deste tipo de crime, seria feito o modelo autorregressivo misto (SAR, *Spatial AutoRegressive*, ou *spatial lag model*), no intuito de identificar um modelo com bom ajuste entre valores preditos e os valores observados da variável dependente, mas devido a constatação de independência espacial nos passos anteriores, esta etapa não foi realizada.

Na sequência, realizamos a análise para o crime de lesão corporal. As Tabelas 5.7 e 5.8 apresentam os resultados dos valores do índice I de *Moran* e seus respectivos p-valores para o teste de independência espacial.

Tabela 5.7: I de Moran para as ocorrências do crime de Lesão Corporal em 2010

Mês	I	p-valor
Janeiro	0.0075519895	0.4427351445
Junho	0.0022737565	0.4625868673
Dezembro	0.0011702397	0.4716932915

Tabela 5.8: I de Moran para as ocorrências do crime de Lesão Corporal em 2010

Mês	I	p-valor
Janeiro	0.0065789474	0.4466371061
Junho	0.0024451791	0.4608805075
Dezembro	0.0013413261	0.4691819974

Observamos que os resultados destes p-valores são muito parecidos com os encontrados para o crime de furto, logo, chegamos a mesma conclusão de independência espacial para os dados deste crime.

Fazendo os gráficos de espalhamento de *Moran* para o crime de lesão corporal, novamente, encontramos resultados muito semelhantes ao do crime de furto. As Figuras 5.12 e 5.12 corroboram a hipótese de independência espacial para os dados do crime de lesão corporal.

DIAGRAMA DE ESPALHAMENTO DE MORAN PARA O CRIME DE FURTO NO DISTRITO FEDERAL,

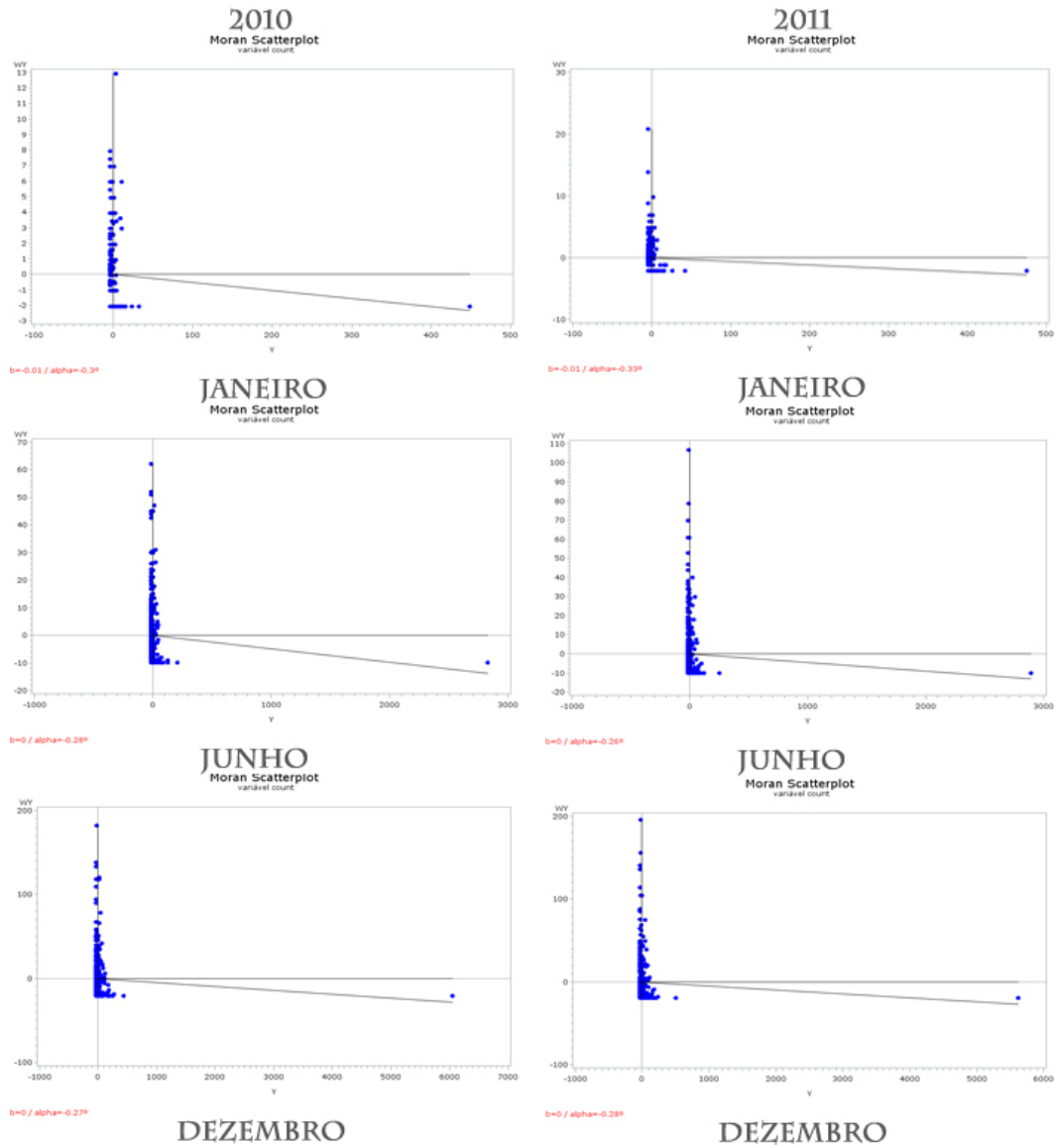


Figura 5.12: Diagrama de Espalhamento de *Moran* para o crime de Lesão Corporal no Distrito Federal em 2010 e 2011.

DIAGRAMA DE ESPALHAMENTO BIDIMENSIONAL DE MORAN PARA O CRIME DE LESÃO CORPORAL NO DISTRITO FEDERAL

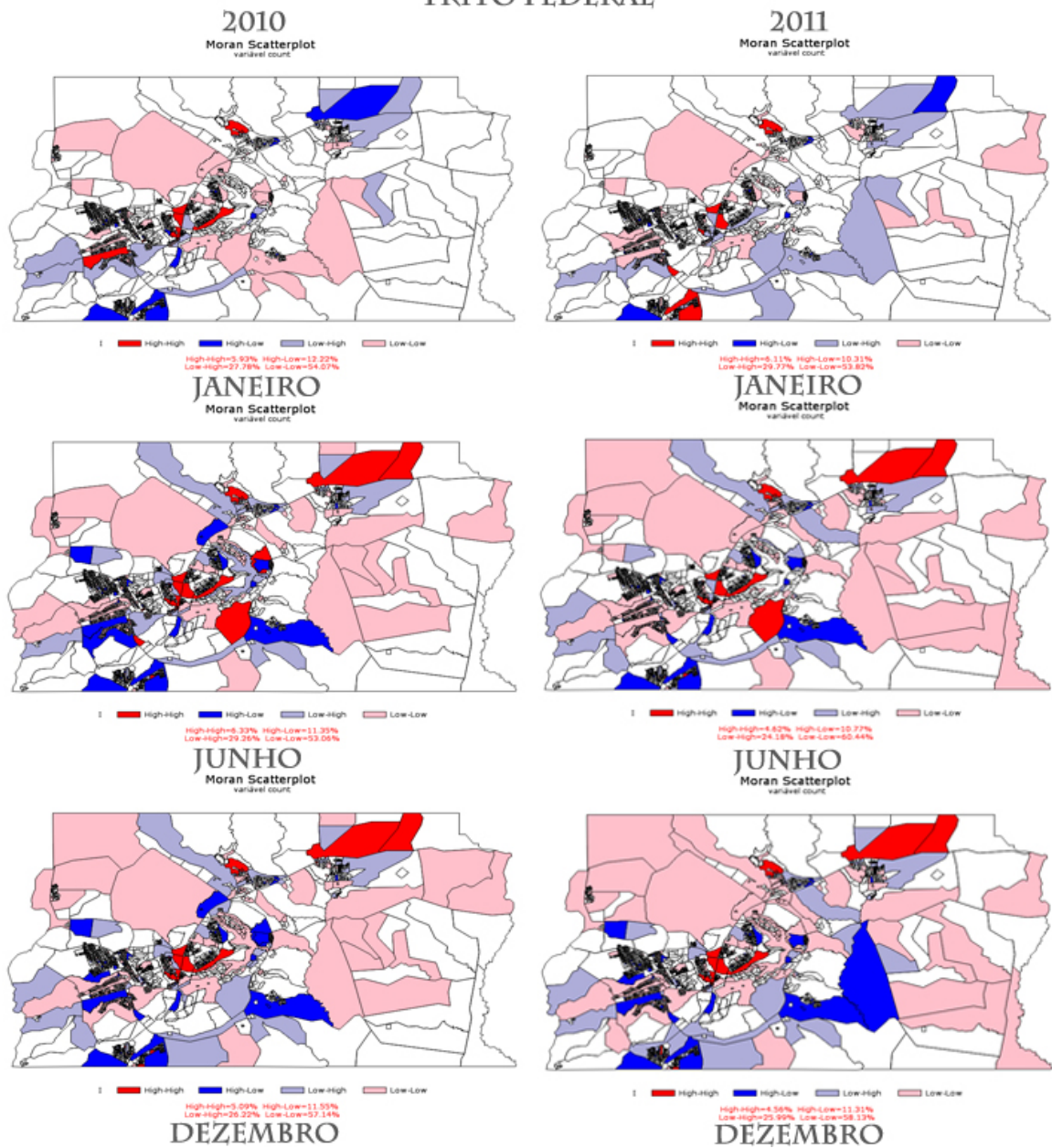


Figura 5.13: Diagrama de Espalhamento de *Moran* na forma de mapa temático bidimensional para o crime de Lesão corporal no Distrito Federal em 2010 e 2011.

5.5 Análise da Krigagem

Na análise da krigagem, não obtivemos boas estimativas para as superfícies interpoladas geradas, como nos mostra a Figura 5.14, referente as ocorrências do crime de furto em janeiro de 2010.

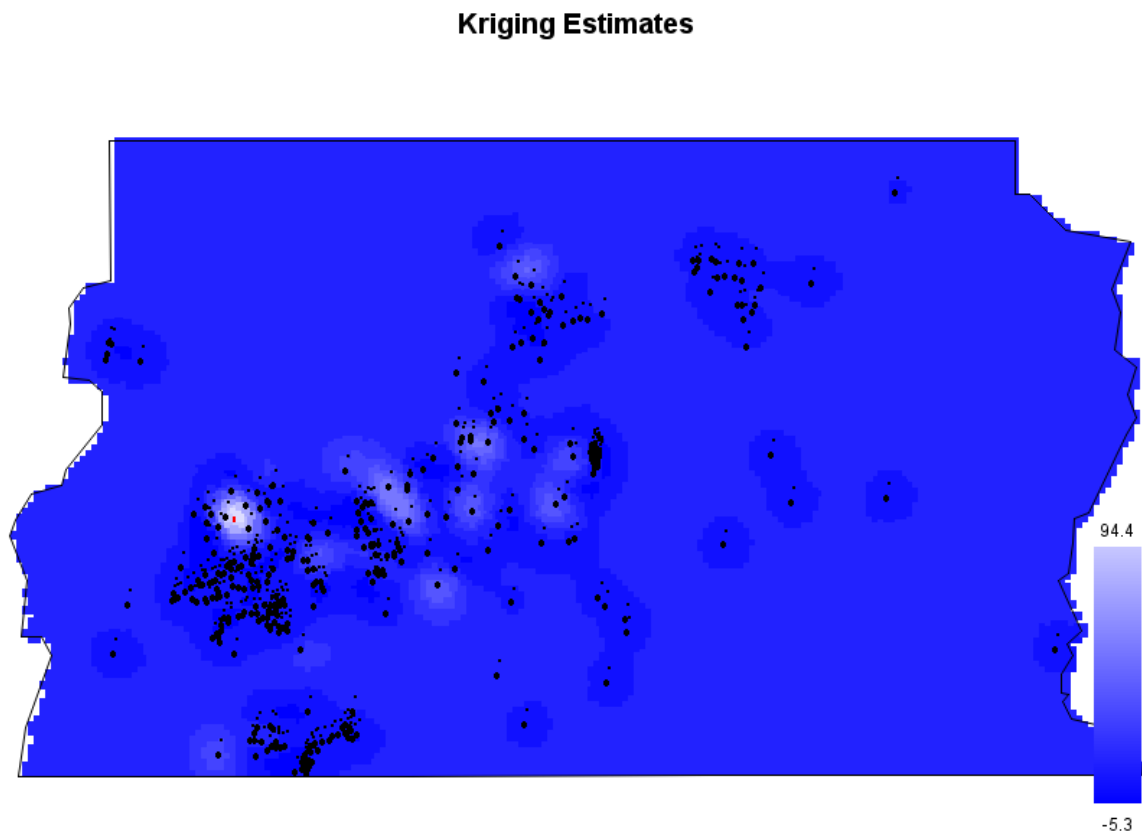


Figura 5.14: Superfície interpolada gerada pela técnica da krigagem para o crime de Furto em Janeiro de 2010 no Distrito Federal.

A técnica da krigagem foi desenvolvida para análise de efeitos naturais, ou seja, efeito com variação suave. Os valores negativos para as estimativas se deram devido a alta heterogeneidade dos dados, que por sua vez, violou a hipótese de média constante da krigagem ordinária.

Tomemos como outro exemplo as ocorrências do crime de furto em janeiro de

2010, apenas para região administrativa de Taguatinga. A Figura 5.15 demonstra a frequência da quantidade de ocorrência em cada setor censitário e mostra que existem valores muito baixos de ocorrências, como 14, em alguns setores enquanto que em outros, existem valores muito altos, como 166, criando assim uma alta discrepância entre áreas adjacentes.

Frequência da quantidade de ocorrências por setor censitário para o crime de Furto em Janeiro de 2010 na RA de Taguatinga

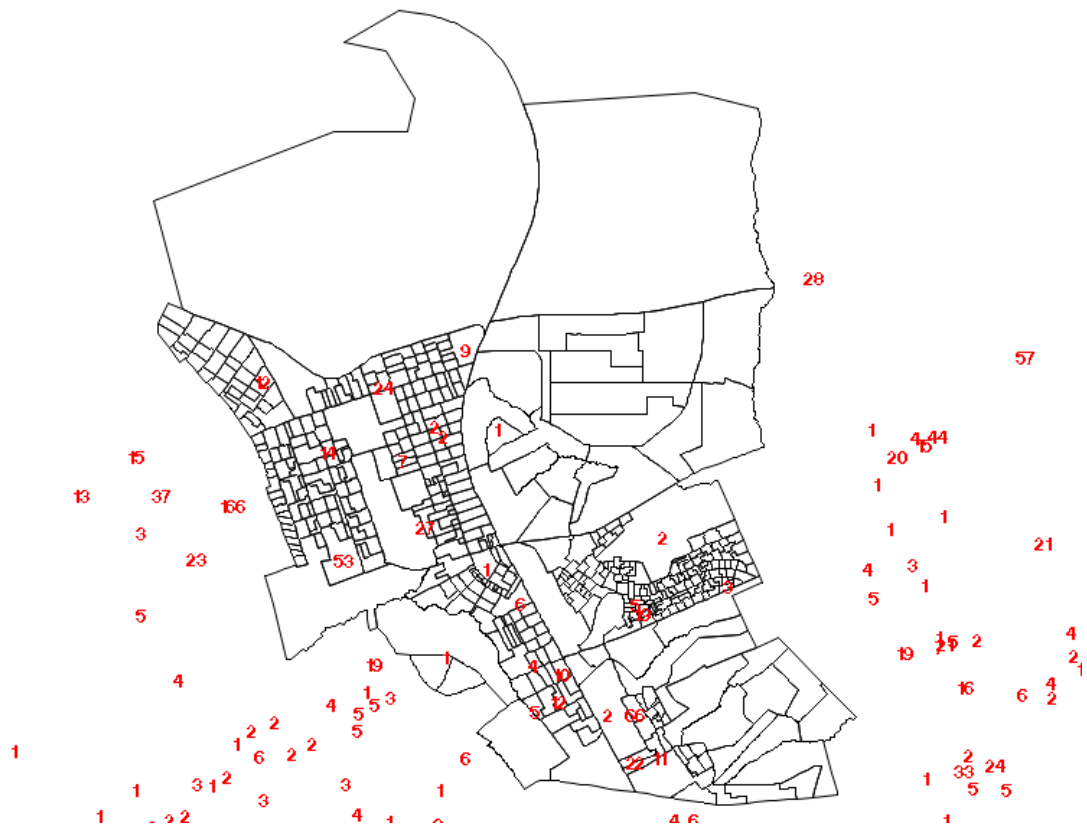


Figura 5.15

Outra forma de verificar este problema é através da superfície de tendência (Figura 5.16). A presença de picos e vales, criando uma superfície acidentada, nos comprova a heterogeneidade dos dados e a impossibilidade de se adequar um modelo explicativo para estes, não podendo gerar estimativas precisas.

Superfície de Tendência para o crime de Furto em Janeiro de 2010 em Taguatinga

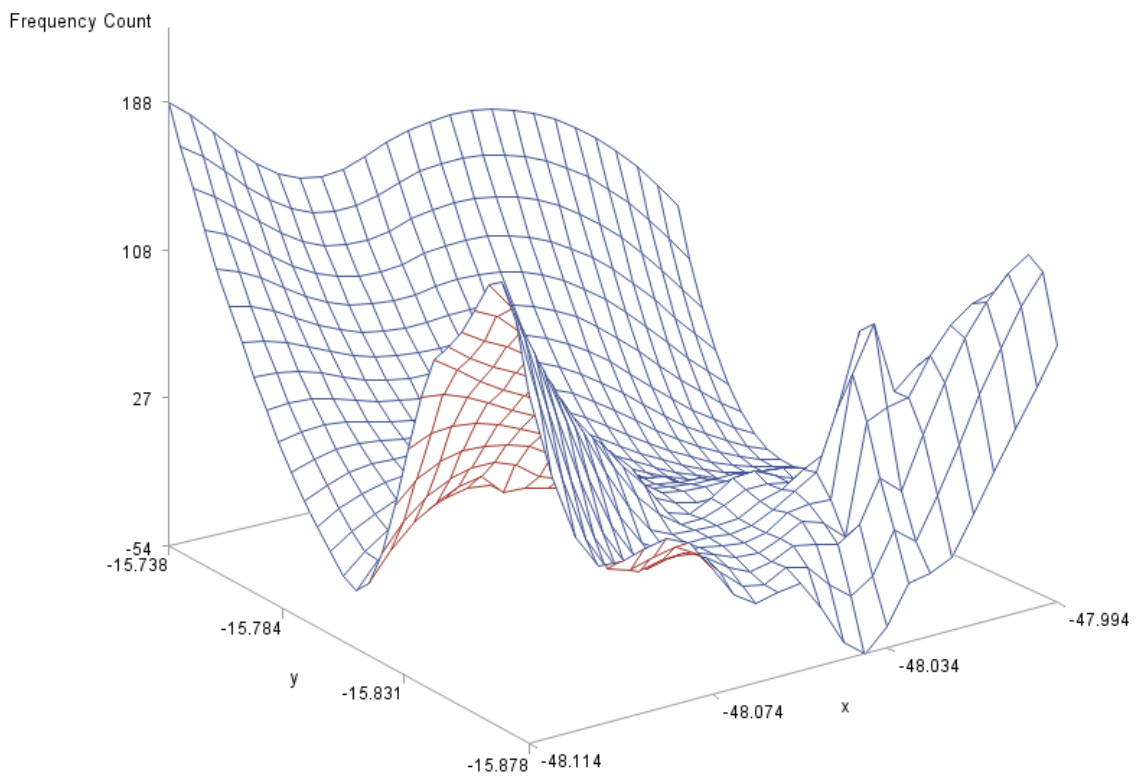


Figura 5.16

Capítulo 6

Conclusões

Através da análise espacial das ocorrências de dois crimes, comprovamos a existência de focos de criminalidade destes, no território do Distrito Federal no decorrer dos anos de 2010 e 2011.

A *Macro GoogleMaps*, utilizada para o georreferenciamento do banco de dados da SSP/DF não se mostrou adequada na realização do posicionamento dos endereços no mapa, havendo perda de precisão na marcação de endereços, e conseqüentemente acúmulo de ocorrências em um mesmo par de coordenadas.

As Figuras 5.3 e 5.4, presentes na parte de análise pontual, mostraram que para o crime de furto, a distribuição das ocorrências tem como principal foco a região administrativa de Brasília, em 2010 e 2011, em todos os meses de estudo. Outros focos de ocorrência, que aparecem no mapa, localizam-se nas regiões administrativas de Ceilândia e Taguatinga. As regiões administrativas do Guar e Samambaia apresentam maior intensidade no ano de 2011. Comparando ambos os anos, vemos o aparecimento de novos locais de ocorrência no DF, com intensidade menor.

Para o crime de leso corporal, a Figura 5.7 mostra que a regio administrativa de Braslia, em 2010 e 2011, apresenta a maior intensidade de ocorrncias entre todas

a regiões do Distrito Federal. As áreas que pertencem as regiões administrativas de Ceilândia, Taguatinga, Samambaia, Sobradinho e Gama apresentam comportamento diferenciado para os anos em estudo. Em 2010, os valores da função de *Kernel* para as regiões de Sobradinho e Taguatinga apresentam maior intensidade. Em 2011, o Gama apresenta um crescimento do valor da função de intensidade.

Ainda na parte pontual da análise, verificamos que através do método do vizinho mais próximo e envelopes de simulação, há existência de agrupamentos entre as ocorrências para distâncias pequenas entre os eventos. Além disso, a distribuição identificada nos mapas das Figuras 5.3, 5.4 e 5.7, gerados pela função de intensidade de *Kernel*, coincidiu com a frequência vista nas Tabelas 5.1, 5.2, 5.3 e 5.4 presentes na parte descritiva da análise. Os produtos resultantes da análise descritiva, os mapas de localização e as tabelas de frequência, confirmaram o resultado da análise da função de intensidade de *Kernel*. Os dois crimes estudados apresentaram pouca variação na mudança das regiões de maior ocorrência, com o passar do tempo.

Devido a natureza pontual dos dados e a alta heterogeneidade dos mesmos quando agrupados por setor censitário, não foi possível obter resultados confiáveis para as técnicas de área e krigagem.

Na etapa da análise de áreas, concluímos através dos resultados obtidos pelo índice de *Moran I* e os gráficos de espalhamento, a não rejeição da hipótese nula de independência espacial entre as ocorrências. Realizando este tipo de análise nos deparamos com um problema básico para os dados agregados por área, ou seja, para uma mesma população estudada, a definição espacial das fronteiras das áreas

afeta os resultados obtidos. Deve-se observar que o problema da escala é um efeito inerente aos dados agregados por áreas.

Utilizando a krigagem ordinária, não obtivemos boas estimativas para superfície interpolada utilizando a técnica da krigagem ordinária. Os valores negativos para as estimativas se deram devido a alta heterogeneidade dos dados, que por sua vez, violou a hipótese de média constante da krigagem ordinária. Uma possível solução para este problema, poderá ser a utilização da técnica de krigagem universal.

Referências Bibliográficas

- Bailey, T. C. & Gatrell, A. C. (1995). *Interactive Spatial Data Analysis*. Prentice Hall.
- Camargo, E. C. G., Câmara, G., & Fucks, S. D. (2004). *Análise Espacial de Dados Geográficos*, chapter 3 - Análise Espacial e Superfícies. EMBRAPA.
- Carvalho, M. S. & Câmara, G. (2004). *Análise Espacial de Dados Geográficos*, chapter 2 - Análise Espacial de Eventos. EMBRAPA.
- Câmara, G., Fuks, S. D., Carvalho, M. S., Monteiro, A. M., Felgueira, C. A., Assad, E., Camargo, E. C. G., Cruz, O. G., Correa, V., & Souza, W. (2004). *Análise Espacial de dados Geográficos*. EMBRAPA.
- Cressie, N. A. C. (1991). *Statistics for Spatial Data*. Wiley.
- Cruz, O. G., Câmara, G., Carvalho, M. S., & Correa, V. (2004). *Análise Espacial de Dados Geográficos*, chapter 5 - Análise Espacial de Áreas. EMBRAPA.
- GDF (2012). Administrações regionais. Technical report, Governo do Distrito Federal. Acesso em 16 nov. 2012.
- Monteiro, A. M., Câmara, G., Fucks, S. D., & Carvalho, M. S. (2004). *Análise Espacial de Dados Geográficos*, chapter 1 - Análise Espacial e Geoprocessamento. EMBRAPA.
- Monteiro, J. B. (2009). Indicador de criminalidade geral baseado em métodos multivariados e estatística espacial para controle na segurança pública do estado. Technical report, Universidade Federal do Rio Grande do Sul.
- SSP/DF (2012). Histórico sobre a secretaria. Technical report, Secretaria de Segurança Pública do Distrito Federal. Acesso em 16 nov. 2012.
- Stein, A. & Corsten, L. C. A. (1991). Universal kriging and cokriging as a regression procedure. *Biometrics*, 47(2):575–587.

Tobler, W. (1970). A computer movie simulating urban growth in the detroit region.
Economic Geography, (46):234-240 p.