Universidade de Brasília

Instituto de Ciências Exatas

Departamento de Estatística

Relação entre o Índice I de Moran e a Quantidade de Observações

por

Daniel Fernandes de Oliveira 09/0110030

Brasília

2013

Daniel Fernandes de Oliveira 09/0110030

Relação entre o Índice I de Moran e a Quantidade de Observações

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

Dedico este trabalho aos meus pais

Daniel Fernandes de Oliveira

Agradecimentos

Aos meus pais, irmãos e a toda minha família que, com muito carinho e apoio, não mediram esforços para que eu chegasse até esta etapa de minha vida.

Ao professor Alan. pela paciência na orientação e incentivo que tornaram possível a conclusão desta monografia.

A todos os professores do curso, que foram tão importantes na minha vida acadêmica e no desenvolvimento desta monografia.

Aos amigos e colegas, pelo incentivo e pelo apoio constantes.

Resumo

Desde o trabalho clássico de Moran (1950) sobre correlação espacial, muitos estudos foram desenvolvidos. O coeficiente I idealizado pelo autor varia de -1 a $+1$, onde -1 indica máxima autocorrelação espacial negativa, 0 indica ausência de autocorrelação espacial e $+1$ indica máxima autocorrelação espacial positiva. Entretanto, esses valores máximos geralmente não ocorrem, mas não por conta do fenômeno, e sim pela quantidade de dados. Assim, o objetivo deste trabalho é mostrar a relação entre o índice I de Moran e a quantidade de dados. Após simular valores em polígonos regulares e irregulares viu-se que o valor do índice I de Moran é de fato influenciado pela quantidade de dados e uma correção deste pode ser estabelecida dividindo-se o valor obtido do índice I pelo seu valor máximo para a quantidade de dados utilizada. Concluiu-se também que, para os dados simulados a alocação do maior valor da variável nas regiões extremas resultou em maiores valores do índice além de que para grandes quantidade de dados o índice tende a 1 para dados obtidos da forma simulada.

Lista de Tabelas

4.1 Quadrado 7×7 Matriz de Distância e Vizinhança(Todos os valores) . 20

4.2 Quadrado 7×7 Matriz de Distância e Vizinhança 21

4.3 Quadrado 9×9 Matriz de Distância e Vizinhança 21

4.4 Quadrado 11×11 Matriz de Distância e Vizinhança 21

4.5 Quadrado 13×13 Matriz de Distância e Vizinhança 21

4.6 Quadrado 15×15 Matriz de Distância e Vizinhança 22

4.7 Quadrado 17×17 Matriz de Distância e Vizinhança 22

4.8 Quadrado 19×19 Matriz de Distância e Vizinhança 22

4.9 Quadrado 21×21 Matriz de Distância e Vizinhança 22

4.10 Quadrado 23×23 Matriz de Distância e Vizinhança 23

4.11 Quadrado 25×25 Matriz de Distância e Vizinhança 23

4.12 Losângo 7×7 Matriz de Distância e Vizinhança 35

4.13 Losângo 9×9 Matriz de Distância e Vizinhança 35

4.14 Losângo 11×11 Matriz de Distância e Vizinhança 35

4.15 Losângo 13×13 Matriz de Distância e Vizinhança 36

4.16 Losângo 15×15 Matriz de Distância e Vizinhança 36

4.17 Losângo 17×17 Matriz de Distância e Vizinhança 36

4.18 Losângo 19 × 19 Matriz de Distância e Vizinhança 36

4.19 Losângo 21 × 21 Matriz de Distância e Vizinhança 37

4.20 Losângo 23 × 23 Matriz de Distância e Vizinhança 37

4.21 Valores I de Moran: Quadrados 49

4.22 Valores I de Moran: Losângos 51

4.23 São Paulo: Valores do índice I de Moran 55

4.24 Pernambuco: Valores do índice I de Moran 60

4.25 Tocantins valores do I de Moran 63

4.26 Columbus valores do I de Moran 66

4.27 Valores máximos do I de Moran por Quantidade de observações . . . 70

Lista de Figuras

1.1 Exemplo pico 25 4

2.1 Exemplos da mapas coropléticos 6

2.2 Exemplo da matriz de proximidades 8

2.3 Exemplo do Diagrama de Espalhamento de Moran 12

2.4 Exemplo de Box *map* 13

2.5 Exemplo de Moran *map* 14

3.1 Exemplos de estados brasileiros (Tocantins e São Paulo) 16

3.2 Matriz quadrada de dimensão 7 em 2D e 3D, respectivamente 17

3.3 Diferença entre o uso da matriz de proximidade por distância e por vizinhança, respectivamente 17

4.1 Mapas coropléticos 7×7, matriz de distâncias pico em 2,6,8 e 14 respectivamente 24

4.2 Mapas coropléticos 7×7, matriz de vizinhanças: Pico em 4,22,3 e 5 respectivamente 24

4.3 Mapas coropléticos 9×9, matriz de distâncias pico em 2,8,10 e 18 respectivamente 25

4.4 Mapas coropléticos 9×9, matriz de vizinhanças: Pico em 37,5,4 e 6 respectivamente 25
4.5 Mapas coropléticos 11×11, matriz de distâncias pico em 13,21,2 e 10 respectivamente 26
4.6 Mapas coropléticos 11×11, matriz de vizinhanças: Pico em 6,56,5 e 7 respectivamente 26
4.7 Mapas coropléticos 13×13, matriz de distâncias pico em 33,11,27 e 39 respectivamente 27
4.8 Mapas coropléticos 13×13, matriz de vizinhanças: Pico em 7,79,6 e 8 respectivamente 28
4.9 Mapas coropléticos 15×15, matriz de distâncias pico em 18,28,32 e 44 respectivamente 28
4.10 Mapas coropléticos 15×15, matriz de vizinhanças pico em 106,8,105 e 91 respectivamente 29
4.11 Mapas coropléticos 17×17, matriz de distâncias pico em 4,14,52 e 68 respectivamente 29
4.12 Mapas coropléticos 17×17, matriz de vizinhanças pico em 9,137,8 e 10 respectivamente 30
4.13 Mapas coropléticos 19×19, matriz de distâncias pico em 40,55,4 e 16 respectivamente 30
4.14 Mapas coropléticos 19×19, matriz de vizinhanças pico em 10,137,9 e 11 respectivamente 31

4.15 Mapas coropléticos 21 × 21, matriz de distâncias pico em 23,41,45 e 61 respectivamente 31
4.16 Mapas coropléticos 21 × 21, matriz de vizinhanças pico em 11,211,10 e 12 respectivamente 32
4.17 Mapas coropléticos 23 × 23, matriz de distâncias pico em 73,89,50 e 66 respectivamente 32
4.18 Mapas coropléticos 23 × 23, matriz de vizinhanças pico em 12,254,11 e 13 respectivamente 33
4.19 Mapas coropléticos 25 × 25, matriz de distâncias pico em 28,48,54 e 72 respectivamente 34
4.20 Mapas coropléticos 25 × 25, matriz de vizinhanças pico em 12,13,14 e 301 respectivamente 34
4.21 Mapas coropléticos 7 × 7, matriz de distâncias pico em 11,23,4 e 22 respectivamente 38
4.22 Mapas coropléticos 7 × 7, matriz de vizinhanças: Pico em 4,22,11 e 23 respectivamente 39
4.23 Mapas coropléticos 9 × 9, matriz de distâncias pico em 14,38,5 e 37 respectivamente 39
4.24 Mapas coropléticos 9 × 9, matriz de vizinhanças: Pico em 5,37,13 e 41 respectivamente 40
4.25 Mapas coropléticos 11 × 11, matriz de distâncias pico em 17,57,28 e 58 respectivamente 40

4.26 Mapas coropléticos 11 × 11, matriz de vizinhanças: Pico em 17,57,6 e 61 respectivamente 41
4.27 Mapas coropléticos 13 × 13, matriz de distâncias pico em 20,80,7 e 79 respectivamente 41
4.28 Mapas coropléticos 13 × 13, matriz de vizinhanças: Pico em 20,80,7 e 79 respectivamente 42
4.29 Mapas coropléticos 15 × 15, matriz de distâncias pico em 38,108,23 e 107 respectivamente 42
4.30 Mapas coropléticos 15 × 15, matriz de vizinhanças pico em 23,107,38 e 108 respectivamente 43
4.31 Mapas coropléticos 17 × 17, matriz de distâncias pico em 26,138,42 e 44 respectivamente 43
4.32 Mapas coropléticos 17 × 17, matriz de vizinhanças pico em 26,138,43 e 139 respectivamente 44
4.33 Mapas coropléticos 19 × 19, matriz de distâncias pico em 48,174,28 e 30 respectivamente 44
4.34 Mapas coropléticos 19 × 19, matriz de vizinhanças pico em 29,173,48 e 174 respectivamente 45
4.35 Mapas coropléticos 21 × 21, matriz de distâncias pico em 53,213,32 e 212 respectivamente 45
4.36 Mapas coropléticos 21 × 21, matriz de vizinhanças pico em 53,213,32 e 212 respectivamente 46

4.37 Mapas coropléticos 23 × 23, matriz de distâncias pico em 81,58,35 e 57 respectivamente 46
4.38 Mapas coropléticos 23 × 23, matriz de vizinhanças pico em 81,58,35 e 57 respectivamente 47
4.39 Gráficos do I de Moran em relação a quantidade de dados, quadrado distância 51
4.40 Gráficos do I de Moran em relação a quantidade de dados, quadrado vizinho 52
4.41 Gráficos do I de Moran em relação a quantidade de dados, losângo distância 52
4.42 Gráficos do I de Moran em relação a quantidade de dados, losângo vizinho 53
4.43 Regiões que geraram maior valor do índice (SP1) 56
4.44 Regiões que geraram maior valor do índice (SP2) 57
4.45 Regiões que geraram maior valor do índice (SP3) 58
4.46 Regiões que geraram maior valor do índice (SP4) 59
4.47 Regiões que geraram maior valor do índice (PE1) 61
4.48 Regiões que geraram maior valor do índice (PE2) 61
4.49 Regiões que geraram maior valor do índice (PE3) 62
4.50 Regiões que geraram maior valor do índice (PE4) 62
4.51 Regiões que geraram maior valor do índice (TO1) 64
4.52 Regiões que geraram maior valor do índice (TO2) 64

4.53 Regiões que geraram maior valor do índice (TO3) 65

4.54 Regiões que geraram maior valor do índice (TO4) 65

4.55 Regiões que geraram maior valor do índice (COL1) 66

4.56 Regiões que geraram maior valor do índice (COL2) 67

4.57 Regiões que geraram maior valor do índice (COL3) 67

4.58 Regiões que geraram maior valor do índice (COL4) 68

Sumário

RESUMO **iv**

1 INTRODUÇÃO **1**

1.1 OBJETIVOS 3

1.2 Metodologia 3

2 ANÁLISE DE ÁREA **5**

2.1 INTRODUÇÃO 5

2.2 Matriz de proximidade espacial 6

2.3 Índice I de Moran 9

2.3.1 Índices Globais de Associação Espacial 10

2.4 Análise Gráfica 11

2.4.1 Diagrama de Espalhamento de Moran 11

2.4.2 Box map, LISA map e Moran map 13

3 SIMULAÇÃO DE POLÍGONOS E MATRIZES **15**

3.1 INTRODUÇÃO 15

3.2 Polígonos e Matrizes 15

3.3 Simulação 16

4 RESULTADOS **19**

4.1 Polígonos Regulares 19

4.1.1 Tabelas Quadrado 19
4.1.2 Figuras Quadrado 23
4.1.3 Tabelas Losângo 35
4.1.4 Figuras Losângo 38
4.1.5 Análise dos Resultados para Polígonos Regulares 48
4.2 Polígonos Irregulares 54
4.2.1 São Paulo 55
4.2.2 Pernambuco 60
4.2.3 Tocantins 63
4.2.4 Columbus 66
5 CONCLUSÃO **71**
REFERÊNCIAS **72**

Capítulo 1

INTRODUÇÃO

O índice I de Moran, construído pelo estatístico Moran (1950), é a estatística mais antiga e mais utilizada para identificar autocorrelações espaciais. Foi desenvolvido utilizando a mesma idéia do coeficiente de correlação linear de Pearson, mas difere deste por mensurar a correlação entre o par y_i e y_j e por atribuir um peso à proximidade espacial. Para realizar a atribuição deste peso, pode-se utilizar inicialmente uma matriz binária e simétrica W, denominada de matriz de proximidade espacial, isto é, w_{ij} poderia ser tanto 0 quanto 1 e $w_{ij} = w_{ji.}$. Em alguns casos, esta matriz é padronizada nas linhas, ou seja, cada linha soma 1. Como o coeficiente linear de Pearson, o índice I de Moran considera o desvio padrão $\frac{(y_i - \overline{y})}{s_y}$, porém o índice I se relaciona com o desvio padrão da mesma variável y em uma vizinhança j vizinha a i, isto é:

$$\left(\frac{y_i - \overline{y}}{s_y}\right)\left(\frac{y_j - \overline{y}}{s_y}\right) \tag{1.1}$$

A idéia é verificar se, quando o desvio de uma área é maior, os devios dos valores vizinhos tendem a ser maiores (a mesma idéia é válida para desvios pequenos). Obviamente, o importante é o comportamento médio destes pares. Assim, calculando

a média da Equação (1.1) e após algumas operações matemáticas, obtemos :

$$I = \frac{1}{\sum_{i \neq j} w_{ij}} \sum_{i \neq j} w_{ij} \left(\frac{y_i - \overline{y}}{s_y} \right) \left(\frac{y_j - \overline{y}}{s_y} \right) \tag{1.2}$$

onde:

y_i = valor da variável na região i;

y_j = valor da variável na região j;

$\overline{y}$ = média de y;

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

s_y = desvio padrao de y;

Assim como o Coeficiente de Pearson, o índice I de Moran varia de -1 a 1, em que -1 indica a autocorrelação espacial negativa máxima, 0 indica a ausência de autocorrelação espacial e +1 indica a autocorrelação positiva máxima. Podemos escrever o índice I de Moran em notação matricial, obtendo:

$$I = \frac{z^t.Wz}{z^t.z} \text{ ou } I = (z^t z)^{-1} z^t W z \tag{1.3}$$

onde:

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

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

Wz = matriz $n \times n$ dos desvios de Wz;

O problema consiste nos valores máximos raramente serem encontrados, mas não por conta do fenômeno, e sim por conta da quantidade de obsevações, sendo que em alguns casos (principalmente com poucos dados) a interpretação do índice não seria a mais correta.

Sendo assim, o objetivo desse trabalho é simular diversos tipos de áreas e quantidade de observações na tentativa de encontrar uma forma de ajustar o índice de Moran, considerando a quantidade de dados.

1.1 OBJETIVOS

O objetivo geral do trabalho é encontrar uma relação entre o índice I de Moran máximo e a quantidade de dados observados.

Os objetivos específicos são:

- Utilizar diferentes formas de polígonos;
- Utilizar matrizes de proximidades binária e por distância euclidiana;
- Realizar as análises utilizando o *software* SAS 9.2 com a PROC IML.

1.2 Metodologia

Utilzando o SAS, foram gerados polígonos regulares e irregulares, uma variável hipotética foi criada e cada região dos polígonos recebeu um valor desta variável. Os valores foram destribuidos de forma a forçar com que o índice I de Moran atingisse o valor de correlação espacial máximo $+1$, isto foi feito selecionando uma região para receber o valor máximo da variável, essa região foi chamada de pico e a partir deste pico as regiões em volta recebiam o valor do pico $-10, -20, -30, ..., -10 * n$, sendo n a distância da região ao pico, até todas as regiões possuirem um valor para a variável. Para um melhor entendimento do índice, foram feitas simulações de tal

forma que todas as regiões de cada polígono recebessem o valor máximo ou pico uma vez. Este processo é demonstrado, para um exemplo de polígono regular quadrático 7×7 com valor máximo na região central 25, na figura a seguir:

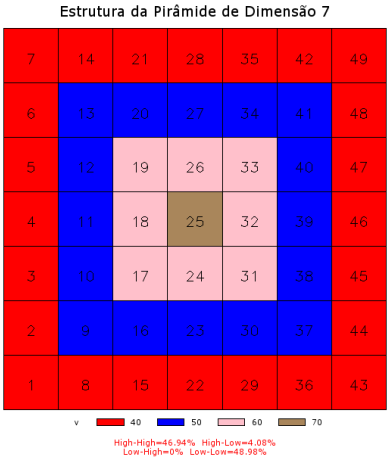


Figura 1.1: Exemplo pico 25

O processo descrito foi realizado para quadrados e losângos de dimensões $(7 \times 7, 9 \times 9, 11 \times 11, ..., 25 \times 25)$ e para as regiões que delimitam os estados de São Paulo, Pernambuco, Tocantis e para a região da cidade de Columbus EUA.

Capítulo 2

ANÁLISE DE ÁREA

2.1 INTRODUÇÃO

A estatística espacial está sendo desenvolvida desde o século passado. Esse tipo de estatística é utilizada, principalmente, quando se estuda eventos agregados por municípios, bairros ou setores censitários, para os quais não temos informações pontuais dos dados, mas um valor geral da área. Esses valores podem ser chamados de indicadores. Alguns exemplos desses indicadores são (Druck et al., 2004):

- Contagem, por exemplo número de chefes de família;
- Proporções, como percentual de adultos analfabetos;
- Médias, como renda média de um chefe de família por município;
- Mediana, como mediana etária dos homens.

Os dados agregados por área são, usualmente, apresentados utilizando-se do uso de mapas coloridos com o padrão especial do fenômeno, onde cores diferentes indicam valores diferentes conhecidos por mapas coropléticos. Pode-se observar esta forma de disposição de dados na Figura 2.1.

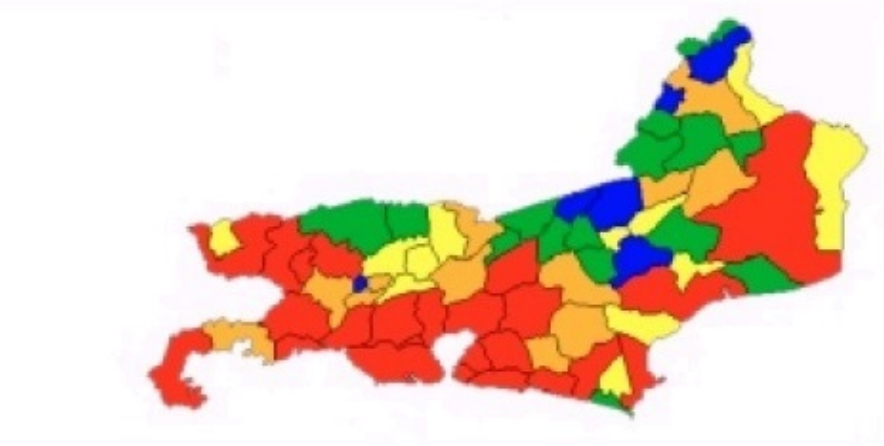


Figura 2.1: Exemplos da mapas coropléticos

Apesar da visualização dos mapas levar a algumas conclusões intuitivas, o estudo de dados espaciais é muito abrangente. Pode-se propor modelos para explicar a distribuição dos dados, analisar a relação área-indivíduo, utilizar diversas técnicas de visualização dos dados, além da visualização intuitiva de observar os valores extremos no mapa, estimar a autocorrelação espacial com o uso de índices, dentre outros.

Neste trabalho será estudado um índice de autocorrelação espacial específico, o índice I de Moran. Porém, para se entender um índice deste tipo é necessário compreender a matriz de proximidade espacial.

2.2 Matriz de proximidade espacial

Segundo Druck et al. (2004) a matriz de proximidade espacial, também conhecida como matriz de vizinhança, é uma ferramenta utilizada para indicar a relação de proximidade entre diferentes áreas. Dado um conjunto de n áreas $(A_1, ..., A_n)$, constrói-se a matriz $W^{(1)}(n \times n)$, em que cada um dos elementos w_{ij} representa uma

medida de proximidade entre A_i e A_j, w_{ij} pode ser calculada utilizando os seguintes critérios (Assunção, 2003):

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

- $w_{ij} = 1$, se A_i compartilha um lado comum com A_j, caso contrário $w_{ij} = 0$;

- $w_{ij} = 1$, se o centróide da área j é um dos k centróides mais próximos daquela área i e $w_{ij} = 0$, caso contrário.k pode assumir qualquer valor inteiro positivo, dependendo do caso. Note que nesse caso a matriz W não é simétrica, pois não necessariamente os k vizinhos mais próximos de i serão os mesmos para j, a não ser para $k = 1$.

- $w_{ij} = \frac{l_{ij}}{l_i}$, onde l_{ij} é o comprimento da fronteira entre A_i e A_j e l_i é o perímetro de A_i.

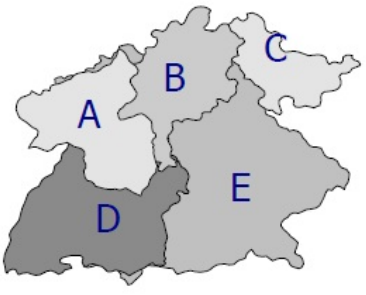
- Seja d_{ij} a distância entre os centróides de i e j. Fazendo $w_{ij} = \frac{1}{(1+d_{ij})}$ as áreas extremamente próximas, com $d_{ij} \approx 0$, teriam assim $w_{ij} \approx 1$, e à medida que estas áreas se afastam, teriam $w_{ij} \approx 0$ (lê-se: aproximadamente 0), dando a mesma idéia dos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

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

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

Uma forma de operacionalizar isso é fazer w_{ij} da mesma forma que a escolha anterior, apenas trocando d_{ij} pelo tempo necessário t_{ij} para ir de i até j em um determinado meio de transporte. Outra forma seria fazer w_{ij} como o valor ou a quantidade de trocas comerciais entre as áreas i e j. Com isso, áreas distantes geograficamente mas com trocas comerciais significativas teriam mais peso na matriz W do que áreas próximas geograficamente mas com poucas trocas comerciais.

É muito útil normalizar a matriz de proximidade em suas linhas, ou seja, fazer com que a soma de cada linha seja igual a 1. Esta normalização facilita a utilização da matriz em cálculos, principalmente nos cálculos dos índices de autocorrelação espacial. A Figura 2.2 ilustra uma matriz de proximidade espacial normalizada, onde os valores da matriz refletem a adjacência entre as áreas.



	A	B	C	D	E
A	0	0,5	0	0,5	0
B	0,25	0	0,25	0,25	0,25
C	0	0,5	0	0	0,5
D	0,33	0,33	0	0	0,33
E	0	0,33	0,33	0,33	0

Figura 2.2: Exemplo da matriz de proximidades

A idéia da matriz de proximidade espacial pode ser generalizada para vizinhos de maior ordem (vizinhos dos vizinhos). Com critério análogo ao adotado para a matriz de vizinhança de primeira ordem, pode-se construir as matrizes $(W^{(2)}, ..., W^{(n)})$ onde $W^{(n)}$ representa a matriz de proximidade de ordem n. Por exemplo, na Figura 2.2, as áreas A e C são vizinhas na matriz de proximidade espacial de ordem 2. Por simplicidade, os coeficientes da matriz de primeira ordem são designados simplesmente por w_{ij}, e os das matrizes de ordem k por $w_{ij}^{(k)}$ (Druck et al., 2004).

2.3 Índice I de Moran

Para uma melhor compreensão do índice I de Moran, é útil entender como outros índices de correlação, espacial ou não, funcionam. Um índice de correlação bastante conhecido é o coeficiente de correlação linear de Pearson, utilizado para medir a correlação entre duas variáveis, x e y (Bussab and Morettin, 2004) e é definido por:

$$r = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{y_i - \overline{y}}{s_i} \right) \left(\frac{x_i - \overline{x}}{s_x} \right) \tag{2.1}$$

onde:

$\overline{y}$ = média de y;

$\overline{x}$ = média de x;

s_y = desvio padrão de y;

s_x = desvio padrão de x;

Este índice varia de -1 a +1, com -1 indicando correlação negativa máxima e +1 correlação positiva máxima, não havendo correlação quando o valor do índice é zero.

Os índices de autocorrelação espacial possuem o mesmo objetivo do coeficiente

de correlação de Pearson, mas diferem deste por considerar, em sua estrutura, a informação referente ao fator espacial.

A dependência espacial pode ser medida de duas formas: através de índices globais que mostram essa dependência em um fator, e através de índices locais que verificam a dependência espacial em cada área estudada (Silva, 2006). O índice I de Moran é um índice global, por isso, nesse trabalho, só serão discutidos índices deste tipo.

2.3.1 Índices Globais de Associação Espacial

Além do índice I de Moran, outro índice de autocorrelação espacial global seria o índice C de Geary, que possui uma estrutura semelhante à do índice I de Moran. Entretanto no númerador calcula-se a diferença entre a medida na área i e a medida na área j. O índice C varia entre 0 e 2, onde valores próximos de 1 indicam ausência de autocorrelação espacial, próximos de 0 indicam autocorrelação positiva e próximos de 2 autocorrelação negativa. O coeficiente de Geary é mais indicado quando existe uma pequena quantidade de vizinhanças (Lembo, 2005).

$$C = \frac{n-1}{2} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(y_i - y_j)^2}{(\sum_{i=1}^{n}(y_i - \overline{y})^2)(\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij})} \tag{2.2}$$

onde:

y_i = valor da variável y na região i;

y_j = valor da variável y na região j;

$\overline{y}$ = média de y;

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

n = número de observações;

2.4 Análise Gráfica

A depêndencia espacial pode ser visualizada através de gráficos usualmente denominados "Diagrama de Espalhamento de Moran". Para um melhor entendimento da dependência espacial, é interessante indicar no mapa as regiões referentes a esses índices. Esta apresentação é amigável e facilita a localização de aglomerações (Silva, 2006).

2.4.1 Diagrama de Espalhamento de Moran

O índice de Moran tem uma representação gráfica elaborada por Anselin (1988), denominada *Moran Scatterplot* ou "Diagrama de Espalhamento de Moran". Na forma matricial, o índice de Moran é dado por: 1.3

Neste caso, o índice I de Moran é equivalente ao coeficiente angular β de uma regressão linear de Wz em z. β indica a inclinação da reta de regressão, sendo assim, podemos analisar a associação espacial de z_i com a média dos vizinhos Wz dispondo esses pontos ao redor da reta. O diagrama é dividido em quatro quadrantes (Q1, Q2, Q3 e Q4). Os pontos que estão em Q1 são chamados alto-alto ($high - high$) e indicam que para valores altos de z, na média existem valores altos de Wz. Já os pontos que estão em Q3 são chamados baixo-baixo ($low - low$) indicando que para valores baixos de z, na média existem baixos valores de Wz. Os pontos que estão em Q2 e Q4 são denominados baixo-alto ($low - high$) e alto-baixo ($high - low$) respecitvamente, indicando que para baixos(ou altos) valores de z, na média existem altos(ou baixos) valores de Wz. A Figura 2.3, representa esse tipo de apresentação gráfica (Silva, 2006).

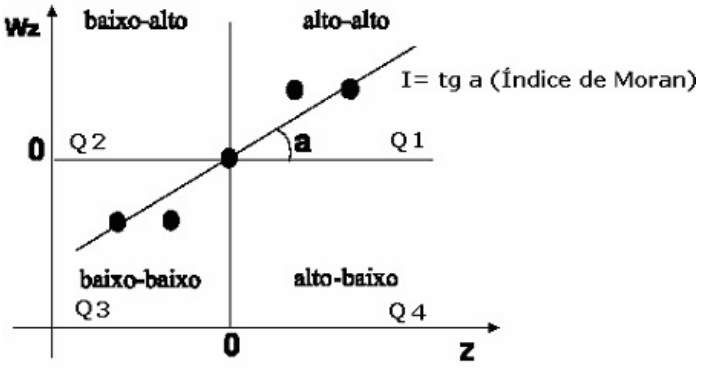


Figura 2.3: Exemplo do Diagrama de Espalhamento de Moran
Fonte: Silva (2006).

A depêndencia espacial é maior à medida que mais pontos se aglomeram no primeiro e terceiro quadrantes, pontos no segundo e quarto quadrantes descaracterizam este fato. A ausência de depêndencia espacial sugere que o fator "espaço" não influencia na váriavel y. Sendo assim, os dados podem ser caracterizados como eventos aleatórios (Silva, 2006).

2.4.2 Box map, LISA map e Moran map

Pode-se também vizualizar o Diagrama de Espalhamento de Moran através de um mapa(caso exista a malha geográfica correta), sendo este tipo de visualização denominada *Box Map* (Figura 2.4). Os quatro quadrantes ($high - high$, $high - low$, $low - high$, $low - low$) são representados por diferentes cores para facilitar a identificação.

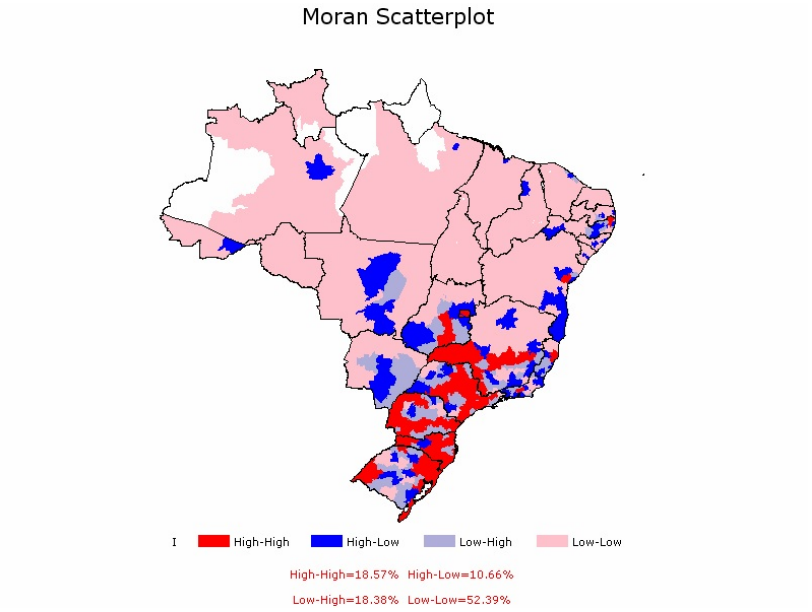


Figura 2.4: Exemplo de Box *map*
Fonte: Silva (2006).

Utiliza-se o LISA *map* para indicar, no mapa, os pontos estatisticamente significativos calculados pelos índices locais. Geralmente usam-se os níveis de 99,9%, 99% e 95% de significância. O nível de significância mais utilizado é o de 95%, e combinando o *Box map* com o LISA *map*, obtem-se um novo tipo de mapa denominado *Moran map* (Figura 2.5). Este gráfico permite identificar quais das classificações definidas no *Moran Scatterplot* são verdadeiramente significativas (Silva, 2006).

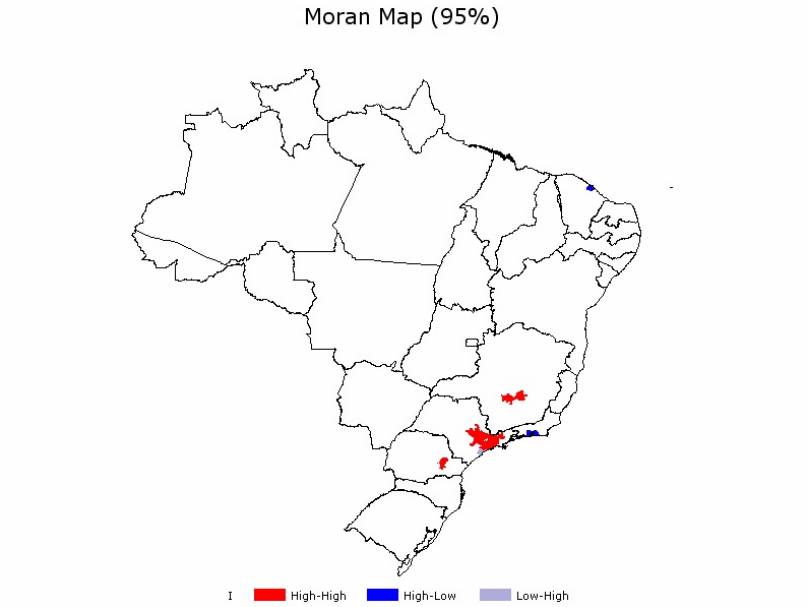


Figura 2.5: Exemplo de Moran *map*
Fonte: Silva (2006).

Quando se verifica a autocorrelação espacial com as ferramentas descritas anteriormente, o próximo passo é utilizar modelos de regressão que incorporem os efeitos espaciais. Estes modelos podem ser globais ou locais (Silva, 2006). Como o foco desse trabalho não consiste na análise da correlação espacial em si, mas na análise do índice I de Moran, a regressão espacial não será discutida, porém pode-se encontrar este tópico nos trabalhos de Silva (2006); Anselin (1988); Lesage (1999).

Capítulo 3

SIMULAÇÃO DE POLÍGONOS E MATRIZES

3.1 INTRODUÇÃO

Para mensurar os valores máximos do índice I de Moran serão necessárias diversas simulações, levando em conta diferentes polígonos e matrizes de proximidade, assim como a localização de pontos de pico dos dados. O objetivo é verificar se os valores máximos do índice variam ou não quando simuladas diferentes situações espaciais.

3.2 Polígonos e Matrizes

Os polígonos consistem, em estatística espacial, na forma geométrica das regiões estudadas, estas podendo possuir as formas mais básicas como quadrados e triângulos, ou formas mais complexas e não regulares. Deseja-se que a diferença entre os polígonos não gere diferenças nos resultados do máximo do índice I de Moran para uma mesma quantidade de observações. Serão simulados valores em polígonos referentes a alguns estados e municípios brasileiros assim como para a cidade de Columbus, Ohio, pela existência de uma grande variabilidade nas formas

que os delimitam, como nos exemplos da Figura 3.1.

Figura 3.1: Exemplos de estados brasileiros (Tocantins e São Paulo)

Em cada polígono pode-se utilizar diversas formas da matriz de proximidade, conforme visto na Seção 2.2.

3.3 Simulação

A fim de avaliar se os valores máximos do índice I de Moran se alteram em diferentes situações serão realizadas simulações no SAS 9.2 com a PROC IML, de forma a gerar polígonos regulares e irregulares com diversas quantidades de observações. Como ponto de partida foram utilizados os polígonos regulares de forma quadricular e forma de losango. Simularam-se valores de uma variável hipotética de tal sorte que cada região do polígono possuisse o valor máximo, ou pico dos dados, conforme mostra a Figura 3.2 que exemplifica uma forma quadricular com dimensão 7×7 e pico na região 25, sendo as cores os valores da variável. As dimensões variaram em 7×7, 9×9, 11×11, $\ldots$, 25×25 tanto para os quadrados quanto para os losangos. Essa variedade de dimensões serve para analisarmos os valores do índice para diversas quantidades de dados. Nesses polígonos, para o cálculo do índice I de

Moran, aplicaram-se duas matrizes de proximidade, uma de distâncias, que consiste na caracterização de vizinhança levando em conta a distância entre os centros das regiões, e uma de fronteiras, que considera vizinhas as regiões que possuem mais de dois pontos em comum ou que fazem fronteira uma com a outra. A Figura 3.3 demonstra a diferença destas matrizes, com o uso de mapas coropléticos de quadrados de dimensões 19x19 com os picos dos dados localizados nos centros dos polígonos, pode-se ver a diferença do comportamento dos vizinhos em torno do pico.

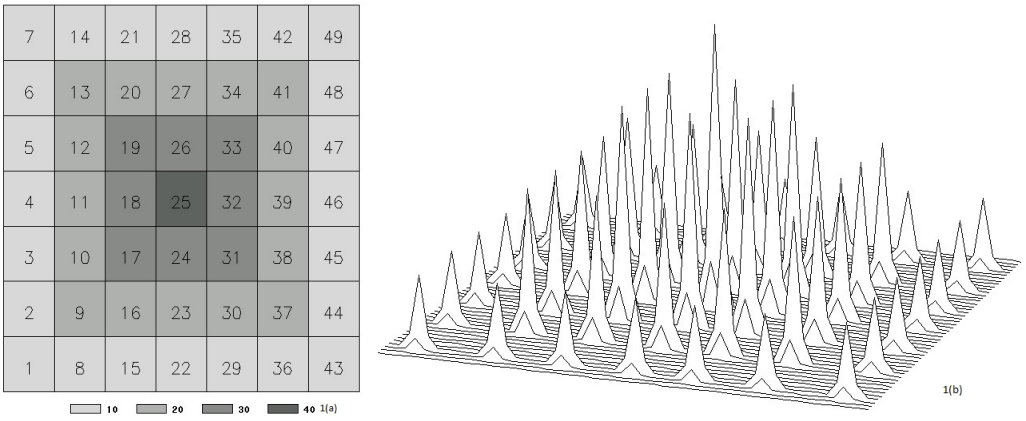


Figura 3.2: Matriz quadrada de dimensão 7 em 2D e 3D, respectivamente

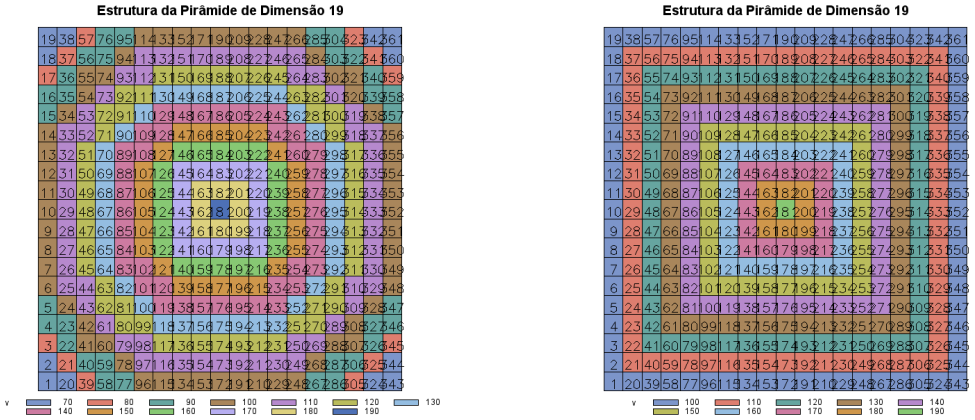


Figura 3.3: Diferença entre o uso da matriz de proximidade por distância e por vizinhança, respectivamente

Após simulados os valores da forma descrita na metodologia e obtido o conhecimento sobre o comportamento do índice nas diversas regiões dos polígono regulares, serão calculados, a partir de dados simulados, os valores do I de Moran para polígonos irregulares, com os polígonos que delimitam os estados de São Paulo, Pernambuco,Tocantis e a cidade de Columbus utilizados como exemplo.

Capítulo 4

RESULTADOS

Como o interesse deste trabalho é utilizar os valores máximos do índice I de Moran, para a correção do índice, após as simulações obteram-se os $I's$ utilizando como pico dos dados todas as regiões do polígono, uma de cada vez. Afim de compreender que regiões geram os maiores valores do índice serão computados os quatro maiores valores deste para as diversas dimensões de polígonos regulares, utilizando duas formas de matriz de proximidades, por vizinhança e distância. Com a finalidade de compreender a variabilidade do índice serão obtidos também os menores valores.

4.1 Polígonos Regulares

Seguem as tabelas com as quatro regiões que produziram os maiores valor do índice assim como o menor para os polígonos regulares (quadrado e losângo) com dimensões 7×7, 9×9, 11×11, $\ldots$, 23×23 para os losângos e até 25×25 nos quadrados.

4.1.1 Tabelas Quadrado

Para todas as dimensões foram obtidas tabelas com a localidade do pico e o valor do Índice de Moran gerado.Os valores máximos das variáveis foram alocados

nas regiões da metade dos polígonos visto que os polígonos utilizados são simétricos. Como exemplo seguem os valores e regiões de pico da metade do quadrado de dimensão 7×7, e em seguida as tabelas com os quatro maiores valores e o menor valor do índice para todas as dimensões. Nas tabelas a seguir a coluna pico indica o quadrado, ou região do polígono que quando usado como pico, resultou nos valores do índice descritos.

Tabela 4.1: Quadrado 7×7 Matriz de Distância e Vizinhança(Todos os valores)

Distância		Vizinhança	
***I* de Moran**	**Pico**	***I* de Moran**	**Pico**
0.85165	1	0.83034	1
0.85858	2	0.84058	2
0.83881	3	0.84525	3
0.83512	4	0.84645	4
0.83881	5	0.84525	5
0.85858	6	0.84058	6
0.85165	7	0.83034	7
0.85858	8	0.84058	8
0.84462	9	0.83309	9
0.81883	10	0.80557	10
0.79345	11	0.79193	11
0.81883	12	0.80557	12
0.84462	13	0.83309	13
0.85858	14	0.84058	14
0.83881	15	0.84525	15
0.81883	16	0.80557	16
0.75094	17	0.72653	17
0.64311	18	0.65104	18
0.75094	19	0.72653	19
0.81883	20	0.80557	20
0.83881	21	0.84525	21
0.83512	22	0.84645	22
0.79345	23	0.79193	23
0.64311	24	0.65104	24
0.46772	25	0.51935	25

Tabela 4.2: Quadrado 7×7 Matriz de Distância e Vizinhança

Dimensão 7	
I **de Moran**	**Pico**
0.85858	2
0.85858	6
0.85858	8
0.85858	14
0.46772	25

Dimensão 7	
I **de Moran**	**Pico**
0.84645	4
0.84645	22
0.84525	3
0.84525	5
0.51935	25

Tabela 4.3: Quadrado 9×9 Matriz de Distância e Vizinhança

Dimensão 9	
I **de Moran**	**Pico**
0.91276	2
0.91276	8
0.91276	10
0.91276	18
0.71462	41

Dimensão 9	
I **de Moran**	**Pico**
0.90638	37
0.90638	5
0.90572	4
0.90572	6
0.71462	41

Tabela 4.4: Quadrado 11×11 Matriz de Distância e Vizinhança

Dimensão 11	
I **de Moran**	**Pico**
0.94327	13
0.94327	21
0.94129	2
0.94129	10
0.79148	61

Dimensão 11	
I **de Moran**	**Pico**
0.93708	6
0.93708	56
0.93671	5
0.93671	7
0.76818	61

Tabela 4.5: Quadrado 13×13 Matriz de Distância e Vizinhança

Dimensão 13	
I **de Moran**	**Pico**
0.95786	3
0.95786	11
0.95786	27
0.95786	39
0.84932	85

Dimensão 13	
I **de Moran**	**Pico**
0.95483	7
0.95483	79
0.95463	6
0.95463	8
0.8261	85

Tabela 4.6: Quadrado 15×15 Matriz de Distância e Vizinhança

Dimensão 15	
I **de Moran**	**Pico**
0.96895	18
0.96895	28
0.96895	32
0.96895	44
0.88269	113

Dimensão 15	
I **de Moran**	**Pico**
0.96602	106
0.96602	8
0.96589	105
0.96589	91
0.86493	113

Tabela 4.7: Quadrado 17×17 Matriz de Distância e Vizinhança

Dimensão 17	
I **de Moran**	**Pico**
0.97548	4
0.97548	14
0.97548	52
0.97548	68
0.90466	145

Dimensão 17	
I **de Moran**	**Pico**
0.97351	9
0.97351	137
0.97343	8
0.97343	10
0.89209	145

Tabela 4.8: Quadrado 19×19 Matriz de Distância e Vizinhança

Dimensão 19	
I **de Moran**	**Pico**
0.98068	41
0.98068	55
0.98024	4
0.98024	16
0.92810	181

Dimensão 19	
I **de Moran**	**Pico**
0.97877	10
0.97877	172
0.97872	9
0.97872	11
0.91184	181

Tabela 4.9: Quadrado 21×21 Matriz de Distância e Vizinhança

Dimensão 21	
I **de Moran**	**Pico**
0.98419	23
0.98419	41
0.98417	45
0.98417	61
0.93702	221

Dimensão 21	
I **de Moran**	**Pico**
0.98261	11
0.98261	211
0.98257	10
0.98257	12
0.93702	221

Tabela 4.10: Quadrado 23 × 23 Matriz de Distância e Vizinhança

Dimensão 23	
I **de Moran**	**Pico**
0.98693	73
0.98693	89
0.98684	50
0.98684	66
0.95949	196

Dimensão 23	
I **de Moran**	**Pico**
0.98549	12
0.98549	254
0.98547	11
0.98547	13
0.93800	265

Tabela 4.11: Quadrado 25 × 25 Matriz de Distância e Vizinhança

Dimensão 25	
I **de Moran**	**Pico**
0.98868	54
0.98884	89
0.98684	50
0.98684	66
0.95627	313

Dimensão 25	
I **de Moran**	**Pico**
0.98772	12
0.98772	254
0.98698	11
0.98698	13
0.94694	313

4.1.2 Figuras Quadrado

Seguem os mapas coropléticos para os picos representados nas tabelas acima:

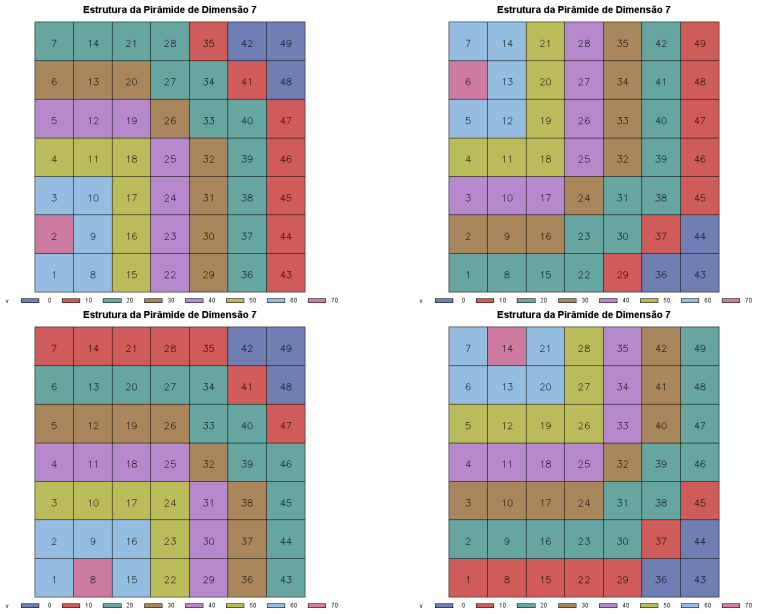


Figura 4.1: Mapas coropléticos 7×7, matriz de distâncias pico em 2,6,8 e 14 respectivamente

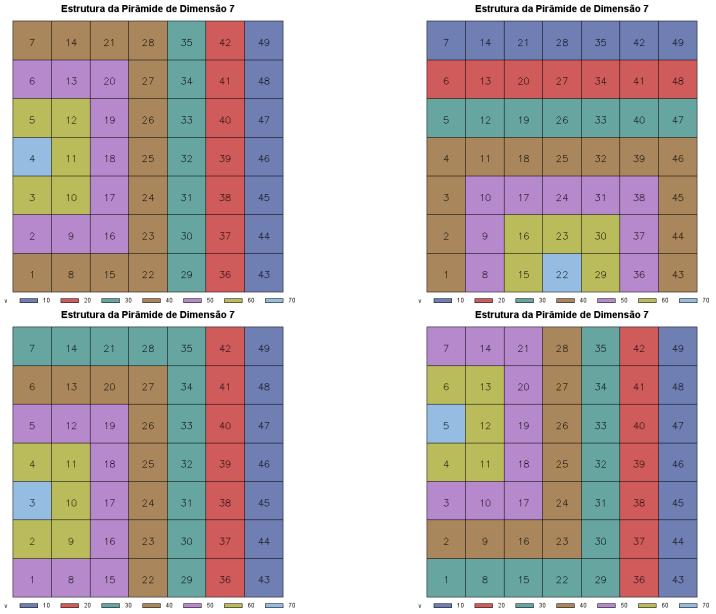


Figura 4.2: Mapas coropléticos 7×7, matriz de vizinhanças: Pico em 4,22,3 e 5 respectivamente

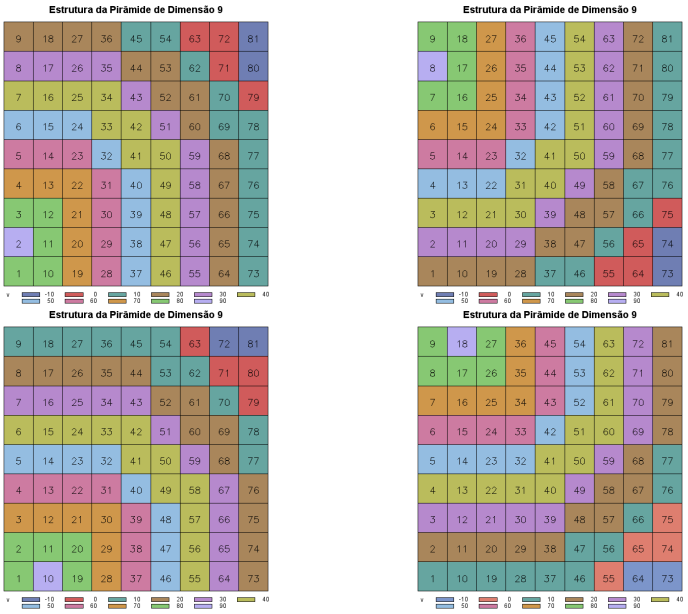


Figura 4.3: Mapas coropléticos 9×9, matriz de distâncias pico em 2,8,10 e 18 respectivamente

Figura 4.4: Mapas coropléticos 9×9, matriz de vizinhanças: Pico em 37,5,4 e 6 respectivamente

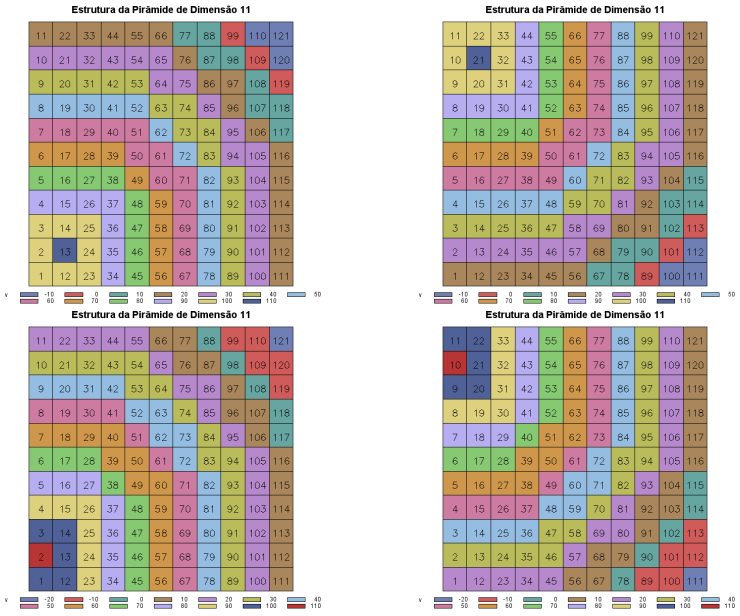


Figura 4.5: Mapas coropléticos 11 × 11, matriz de distâncias pico em 13,21,2 e 10 respectivamente

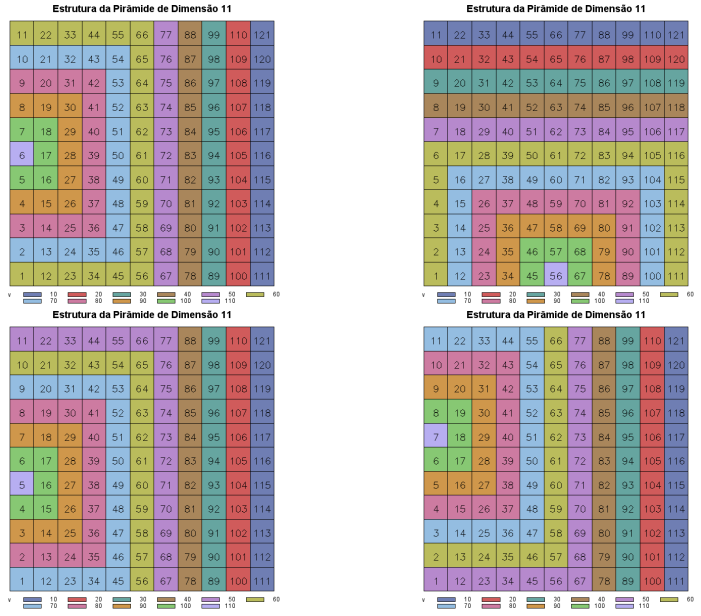


Figura 4.6: Mapas coropléticos 11 × 11, matriz de vizinhanças: Pico em 6,56,5 e 7 respectivamente

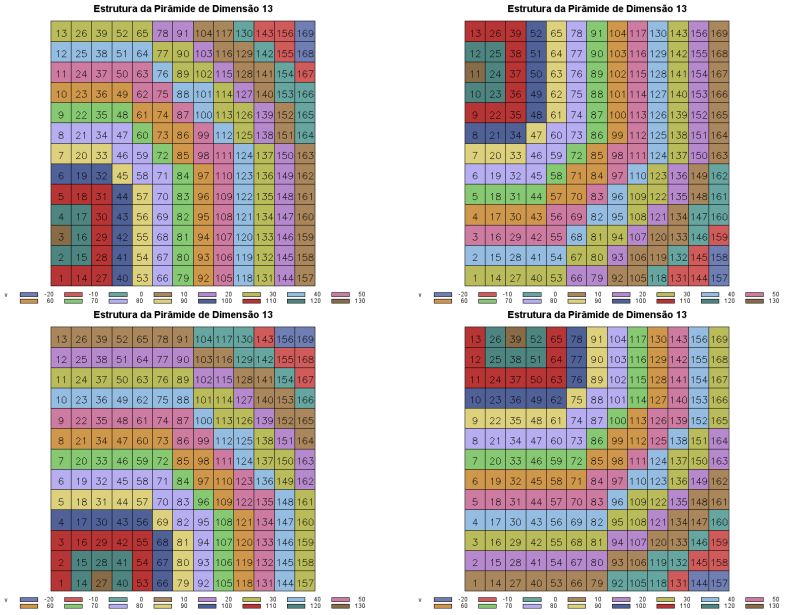


Figura 4.7: Mapas coropléticos 13 × 13, matriz de distâncias pico em 33,11,27 e 39 respectivamente

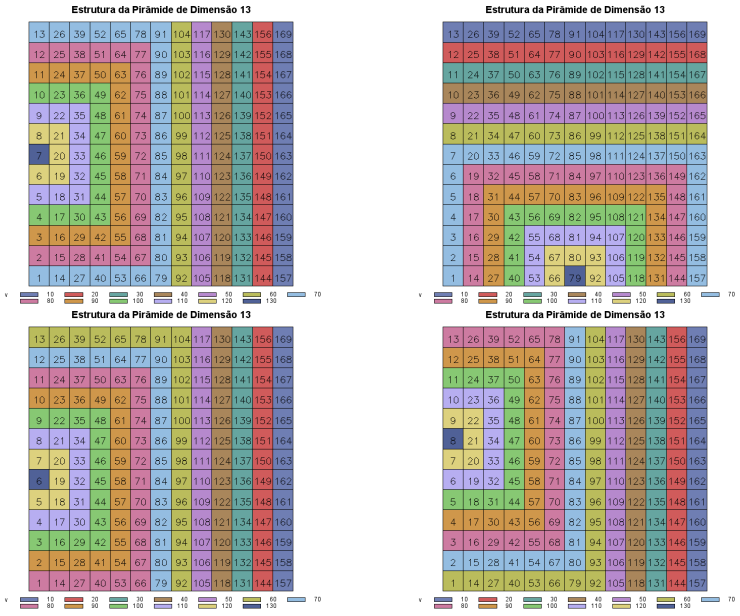


Figura 4.8: Mapas coropléticos 13 × 13, matriz de vizinhanças: Pico em 7,79,6 e 8 respectivamente

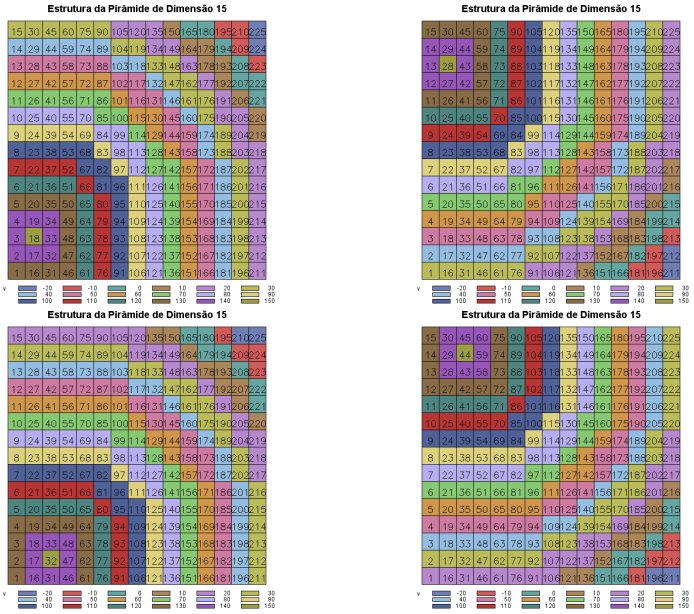


Figura 4.9: Mapas coropléticos 15 × 15, matriz de distâncias pico em 18,28,32 e 44 respectivamente

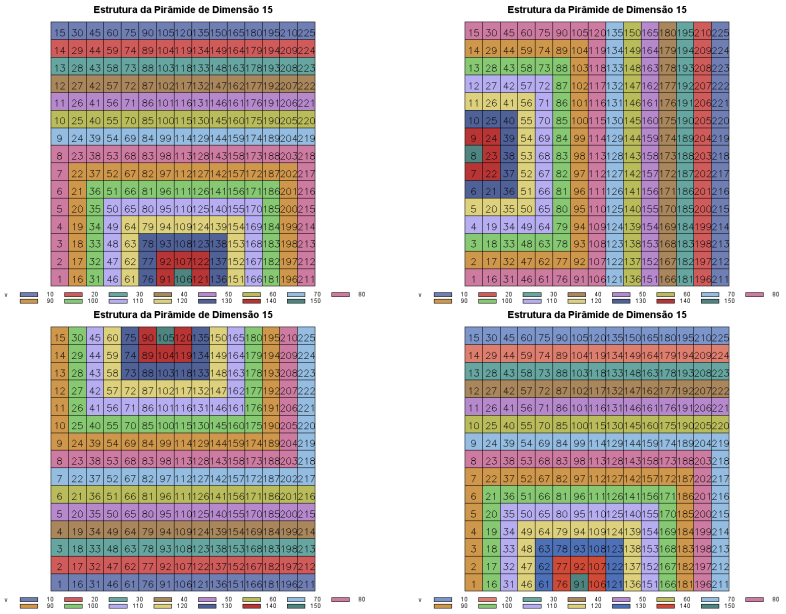


Figura 4.10: Mapas coropléticos 15×15, matriz de vizinhanças pico em 106,8,105 e 91 respectivamente

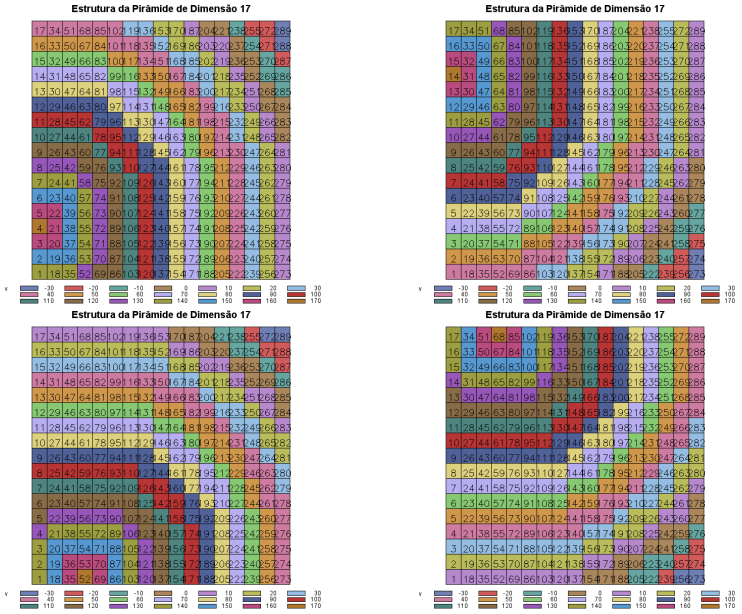


Figura 4.11: Mapas coropléticos 17×17, matriz de distâncias pico em 4,14,52 e 68 respectivamente

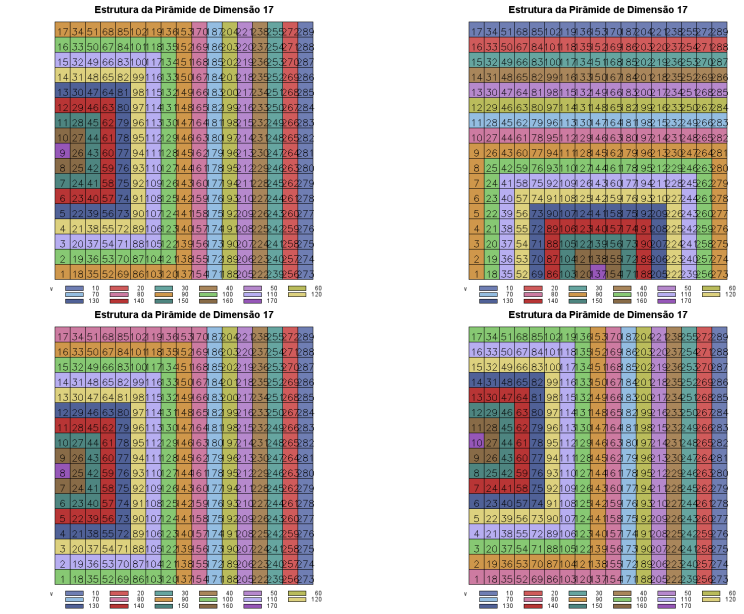


Figura 4.12: Mapas coropléticos 17×17, matriz de vizinhanças pico em 9,137,8 e 10 respectivamente

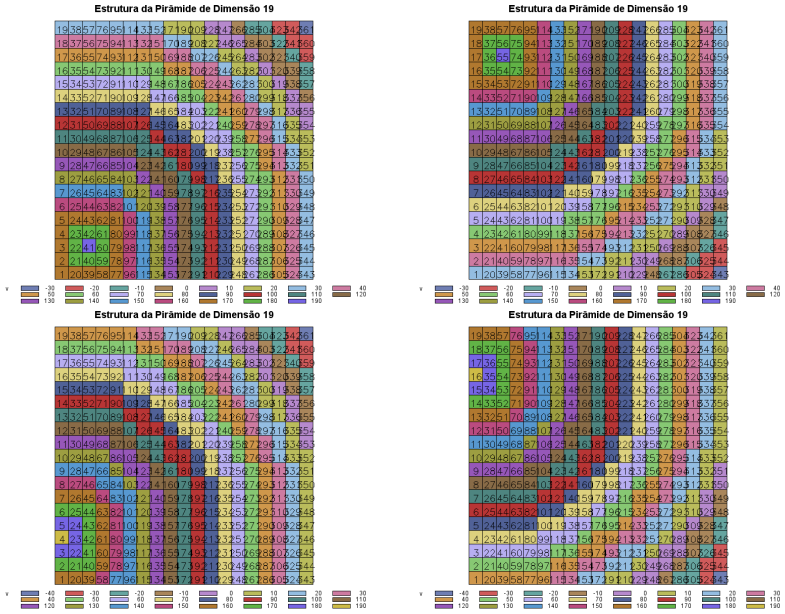


Figura 4.13: Mapas coropléticos 19×19, matriz de distâncias pico em 40,55,4 e 16 respectivamente

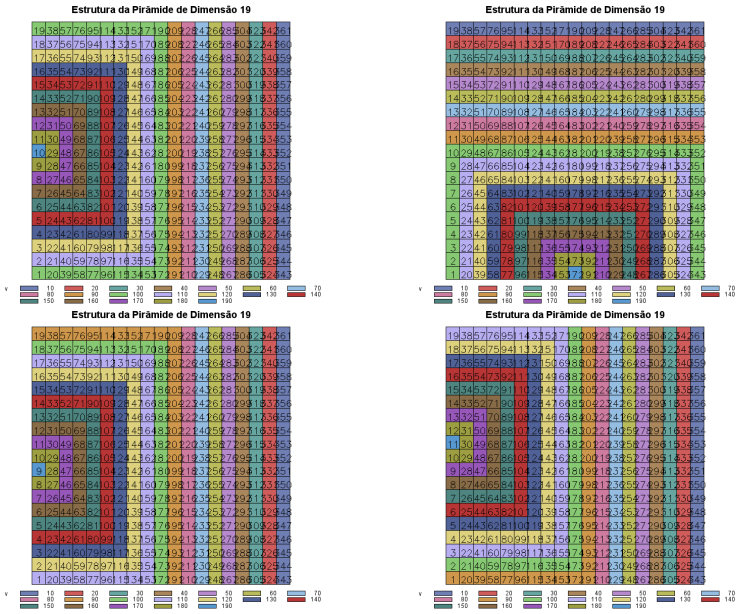


Figura 4.14: Mapas coropléticos 19×19, matriz de vizinhanças pico em 10,137,9 e 11 respectivamente

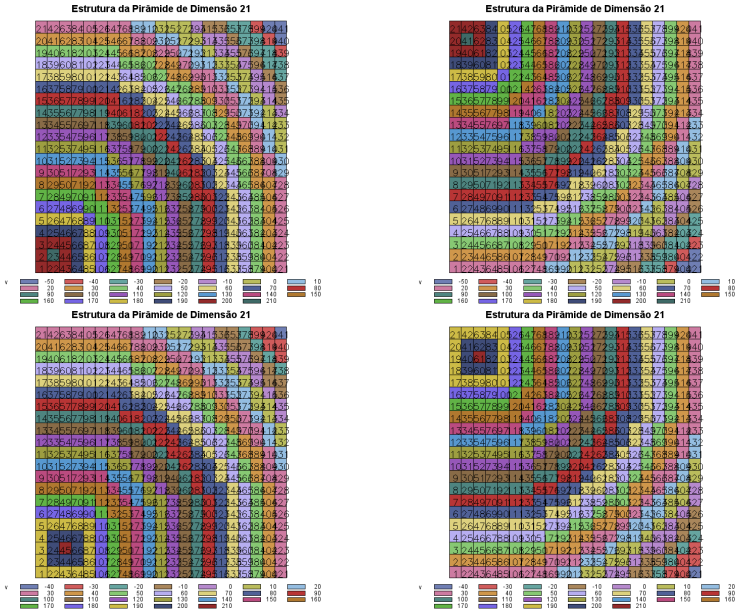


Figura 4.15: Mapas coropléticos 21×21, matriz de distâncias pico em 23,41,45 e 61 respectivamente

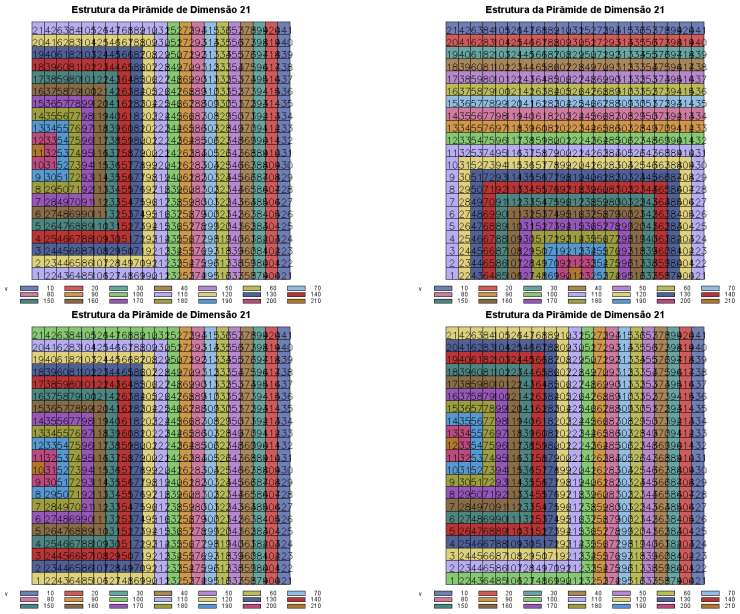


Figura 4.16: Mapas coropléticos 21×21, matriz de vizinhanças pico em 11,211,10 e 12 respectivamente

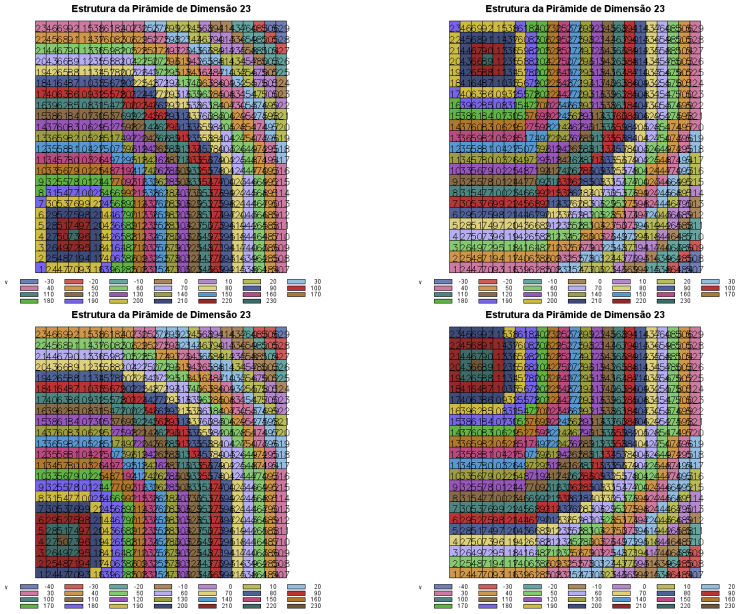


Figura 4.17: Mapas coropléticos 23×23, matriz de distâncias pico em 73,89,50 e 66 respectivamente

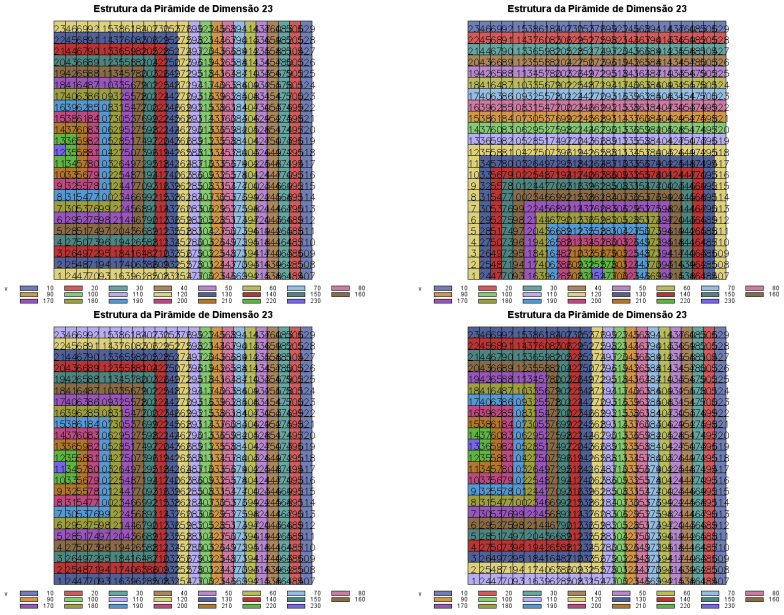


Figura 4.18: Mapas coropléticos 23×23, matriz de vizinhanças pico em 12,254,11 e 13 respectivamente

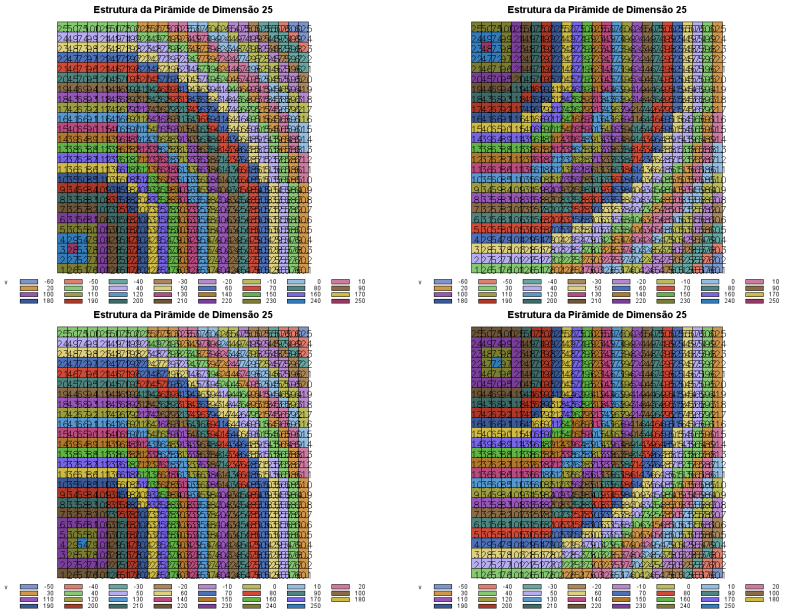


Figura 4.19: Mapas coropléticos 25×25, matriz de distâncias pico em 28,48,54 e 72 respectivamente

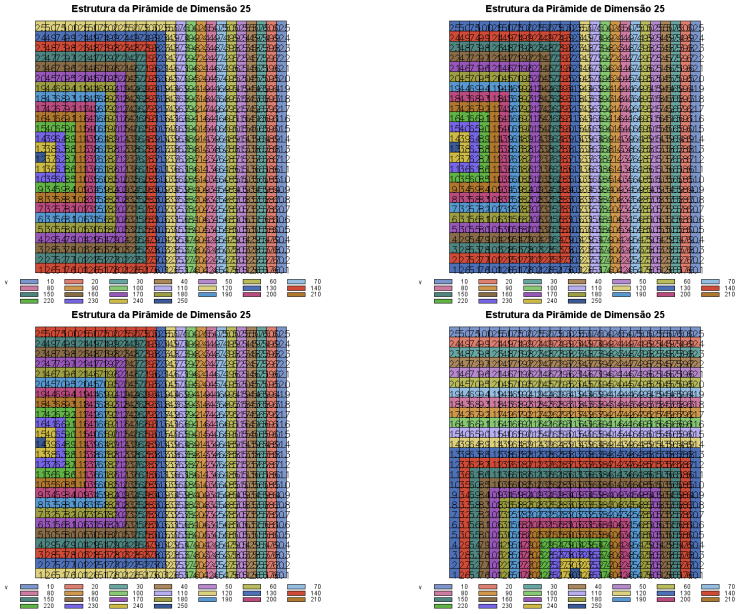


Figura 4.20: Mapas coropléticos 25×25, matriz de vizinhanças pico em 12,13,14 e 301 respectivamente

4.1.3 Tabelas Losângo

Em seguida encontram-se as tabelas com os valores obtidos para os losângos simulados:

Tabela 4.12: Losângo 7 × 7 Matriz de Distância e Vizinhança

Dimensão 7	
***I* de Moran**	**Pico**
0.73297	11
0.73297	23
0.72817	4
0.72817	22
0.34843	25

Dimensão 7	
***I* de Moran**	**Pico**
0.75568	4
0.75568	22
0.73297	11
0.73297	23
0.34843	25

Tabela 4.13: Losângo 9 × 9 Matriz de Distância e Vizinhança

Dimensão 9	
***I* de Moran**	**Pico**
0.85124	14
0.85124	38
0.84776	5
0.84776	37
0.55745	41

Dimensão 9	
***I* de Moran**	**Pico**
0.83910	5
0.83910	37
0.83175	13
0.83175	15
0.55745	41

Tabela 4.14: Losângo 11 × 11 Matriz de Distância e Vizinhança

Dimensão 11	
***I* de Moran**	**Pico**
0.89644	17
0.89644	57
0.88575	28
0.88575	58
0.68800	61

Dimensão 11	
***I* de Moran**	**Pico**
0.90480	17
0.90480	57
0.89647	6
0.89647	56
0.68800	61

Tabela 4.15: Losângo 13 × 13 Matriz de Distância e Vizinhança

Dimensão 13	
I **de Moran**	**Pico**
0.92534	20
0.92534	80
0.91990	7
0.91990	79
0.76026	85

Dimensão 13	
I **de Moran**	**Pico**
0.93361	20
0.93361	80
0.92517	7
0.92517	79
0.76795	85

Tabela 4.16: Losângo 15 × 15 Matriz de Distância e Vizinhança

Dimensão 15	
I **de Moran**	**Pico**
0.94350	38
0.94350	108
0.94230	23
0.94230	107
0.82207	113

Dimensão 15	
I **de Moran**	**Pico**
0.95094	23
0.95094	107
0.94675	38
0.94675	108
0.82207	113

Tabela 4.17: Losângo 17 × 17 Matriz de Distância e Vizinhança

Dimensão 17	
I **de Moran**	**Pico**
0.95606	26
0.95606	138
0.95375	42
0.95375	44
0.85536	145

Dimensão 17	
I **de Moran**	**Pico**
0.96220	26
0.96220	138
0.96062	43
0.96062	139
0.85900	145

Tabela 4.18: Losângo 19 × 19 Matriz de Distância e Vizinhança

Dimensão 19	
I **de Moran**	**Pico**
0.96377	48
0.96377	174
0.96279	28
0.96279	30
0.88593	181

Dimensão 19	
I **de Moran**	**Pico**
0.96995	29
0.96995	173
0.96963	48
0.96963	174
0.88099	181

Tabela 4.19: Losângo 21 × 21 Matriz de Distância e Vizinhança

Dimensão 21	
I **de Moran**	**Pico**
0.97177	53
0.97177	213
0.97148	32
0.97148	212
0.89625	221

Dimensão 21	
I **de Moran**	**Pico**
0.97583	53
0.97583	213
0.97553	32
0.97553	212
0.90569	221

Tabela 4.20: Losângo 23 × 23 Matriz de Distância e Vizinhança

Dimensão 23	
I **de Moran**	**Pico**
0.97555	81
0.97546	58
0.97540	35
0.97505	57
0.91296	265

Dimensão 23	
I **de Moran**	**Pico**
0.97740	81
0.97735	58
0.97698	35
0.97684	57
0.92031	265

4.1.4 Figuras Losângo

Seguem os mapas coropléticos com os picos que geraram maior valor do I de Moran.

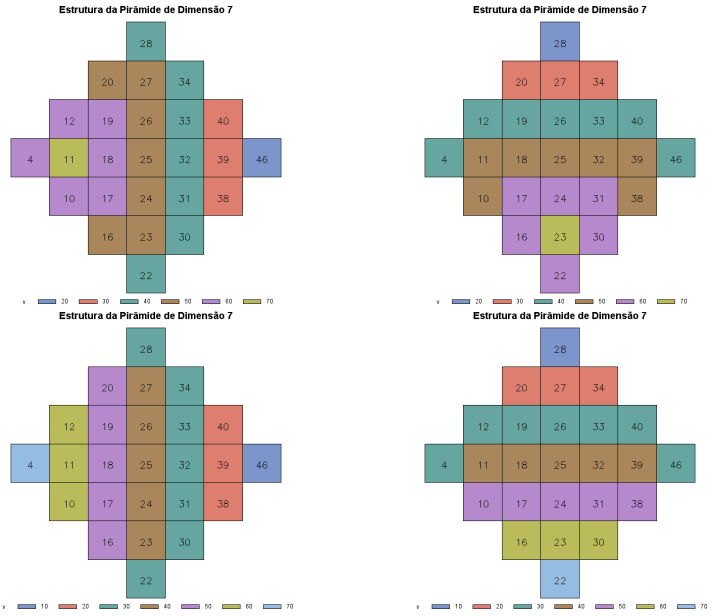


Figura 4.21: Mapas coropléticos 7×7, matriz de distâncias pico em 11,23,4 e 22 respectivamente

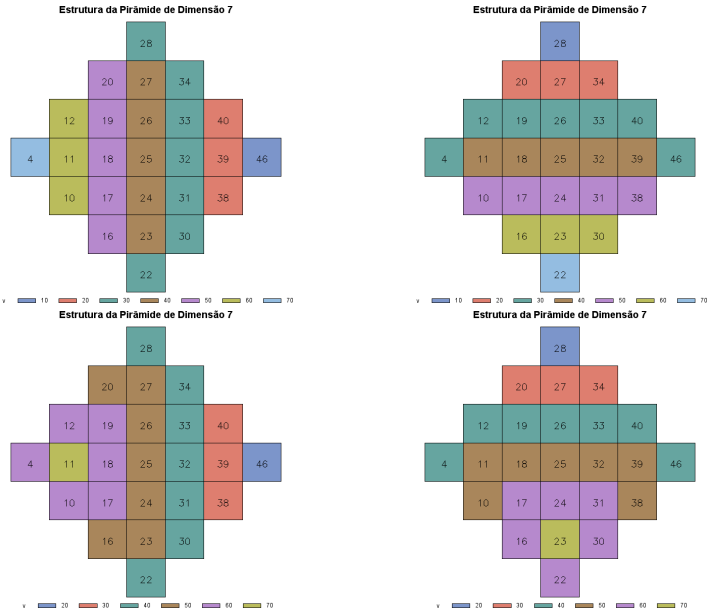


Figura 4.22: Mapas coropléticos 7×7, matriz de vizinhanças: Pico em 4,22,11 e 23 respectivamente

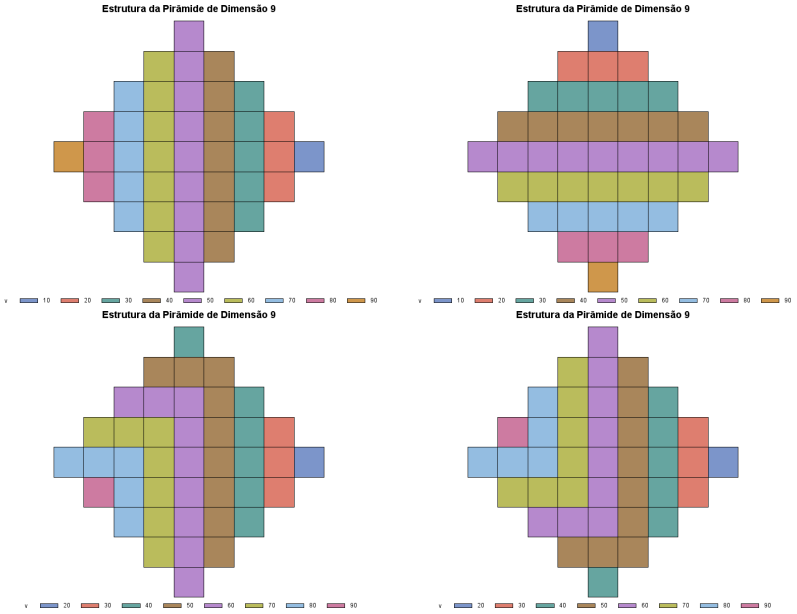


Figura 4.23: Mapas coropléticos 9×9, matriz de distâncias pico em 14,38,5 e 37 respectivamente

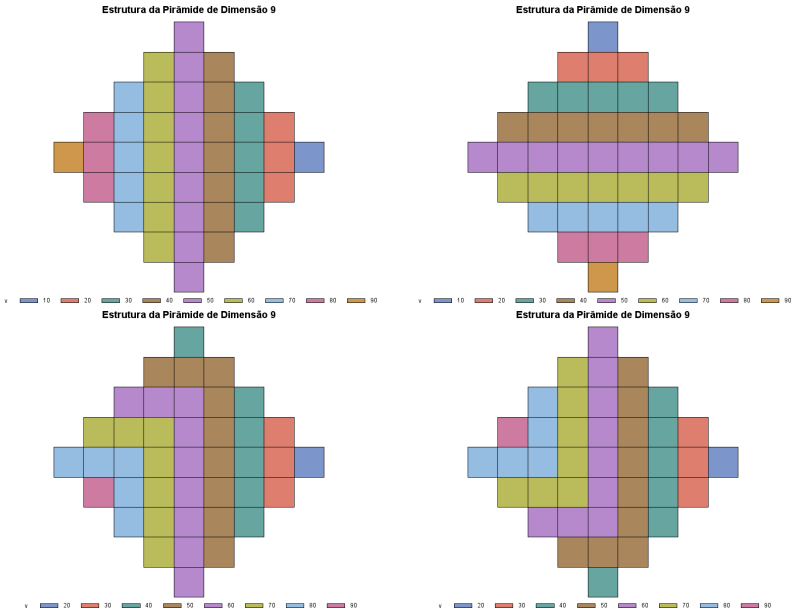


Figura 4.24: Mapas coropléticos 9×9, matriz de vizinhanças: Pico em 5,37,13 e 41 respectivamente

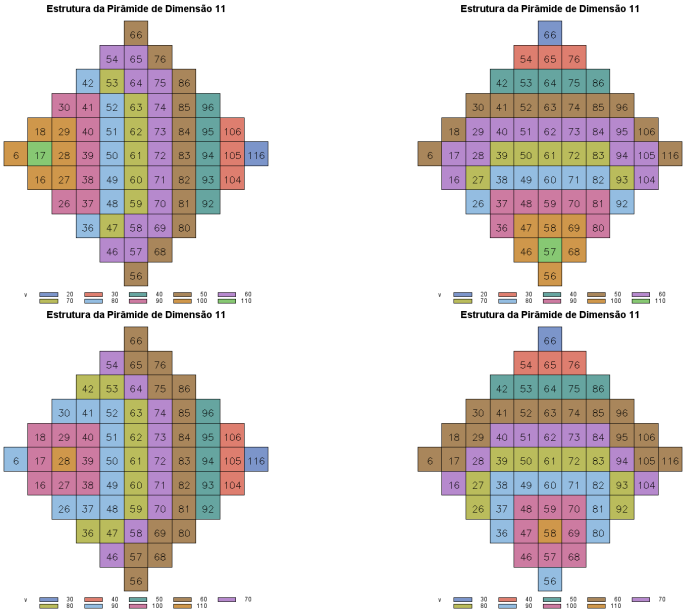


Figura 4.25: Mapas coropléticos 11×11, matriz de distâncias pico em 17,57,28 e 58 respectivamente

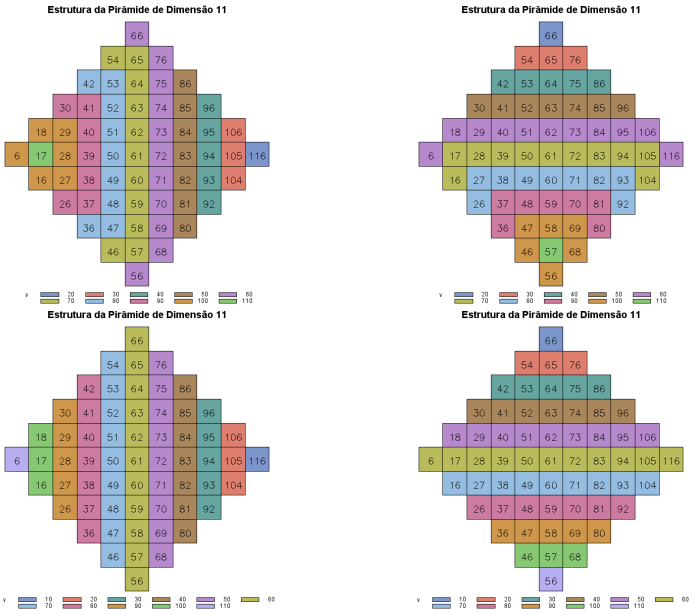


Figura 4.26: Mapas coropléticos 11 × 11, matriz de vizinhanças: Pico em 17,57,6 e 61 respectivamente

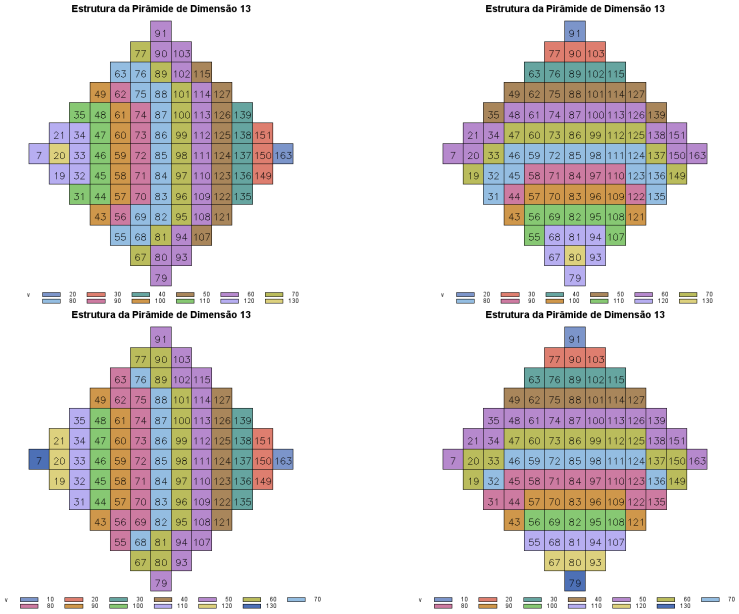


Figura 4.27: Mapas coropléticos 13 × 13, matriz de distâncias pico em 20,80,7 e 79 respectivamente

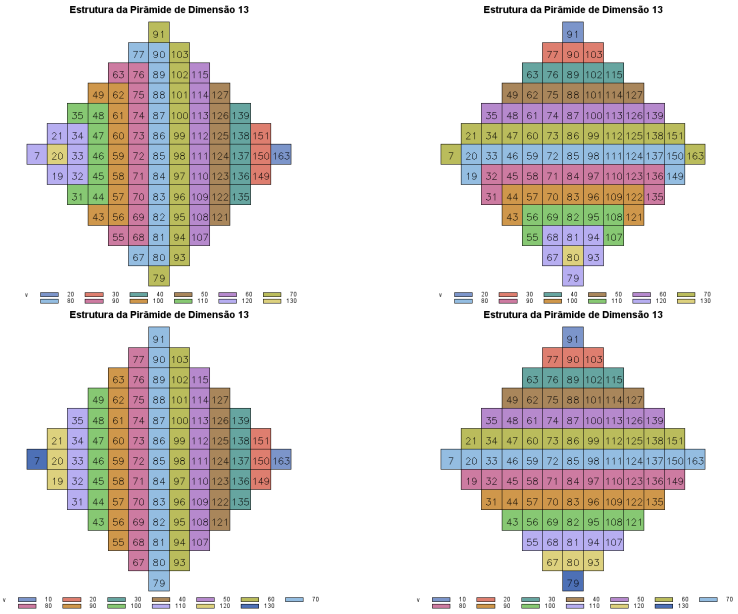


Figura 4.28: Mapas coropléticos 13 × 13, matriz de vizinhanças: Pico em 20,80,7 e 79 respectivamente

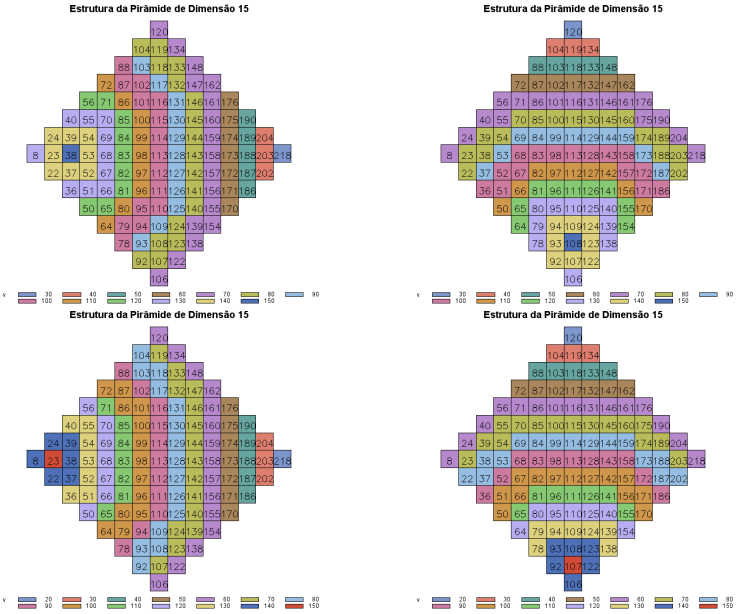


Figura 4.29: Mapas coropléticos 15 × 15, matriz de distâncias pico em 38,108,23 e 107 respectivamente

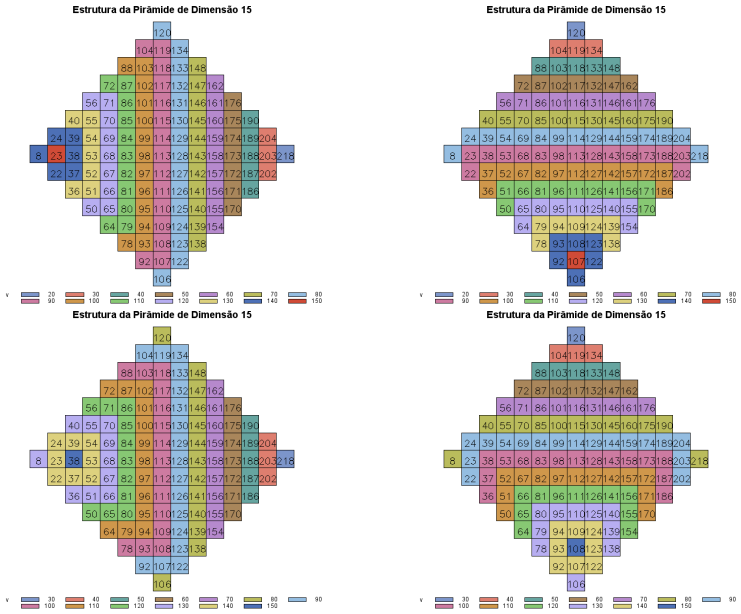


Figura 4.30: Mapas coropléticos 15×15, matriz de vizinhanças pico em 23,107,38 e 108 respectivamente

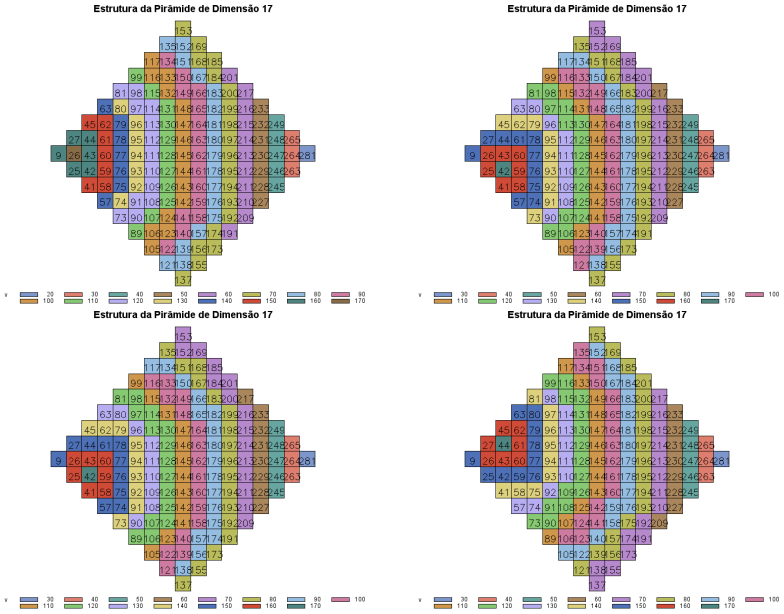


Figura 4.31: Mapas coropléticos 17×17, matriz de distâncias pico em 26,138,42 e 44 respectivamente

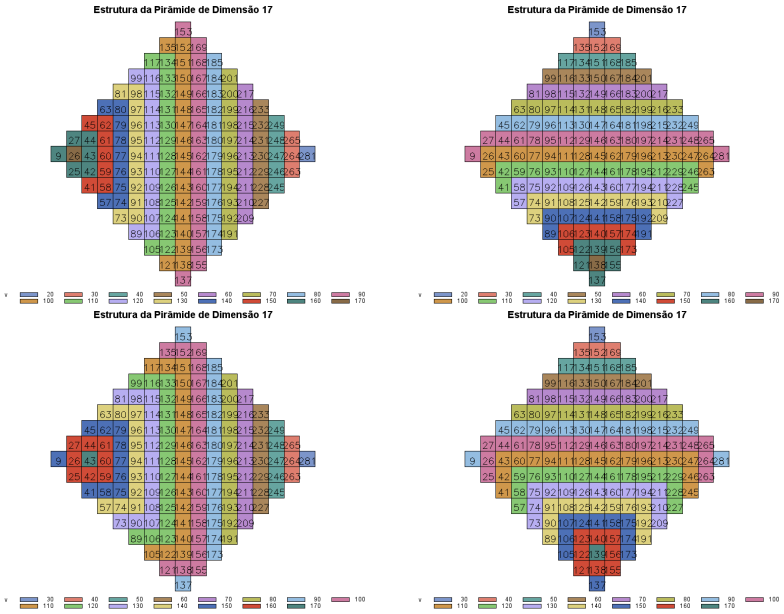


Figura 4.32: Mapas coropléticos 17×17, matriz de vizinhanças pico em 26,138,43 e 139 respectivamente

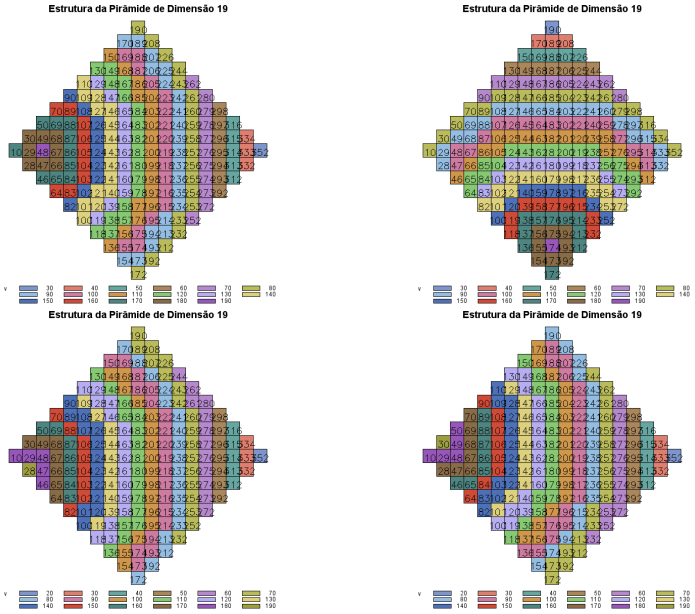


Figura 4.33: Mapas coropléticos 19×19, matriz de distâncias pico em 48,174,28 e 30 respectivamente

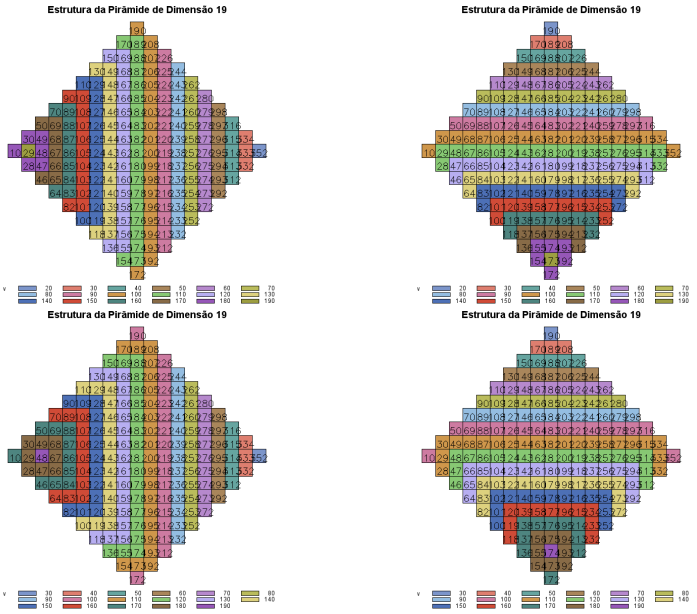


Figura 4.34: Mapas coropléticos 19×19, matriz de vizinhanças pico em 29,173,48 e 174 respectivamente

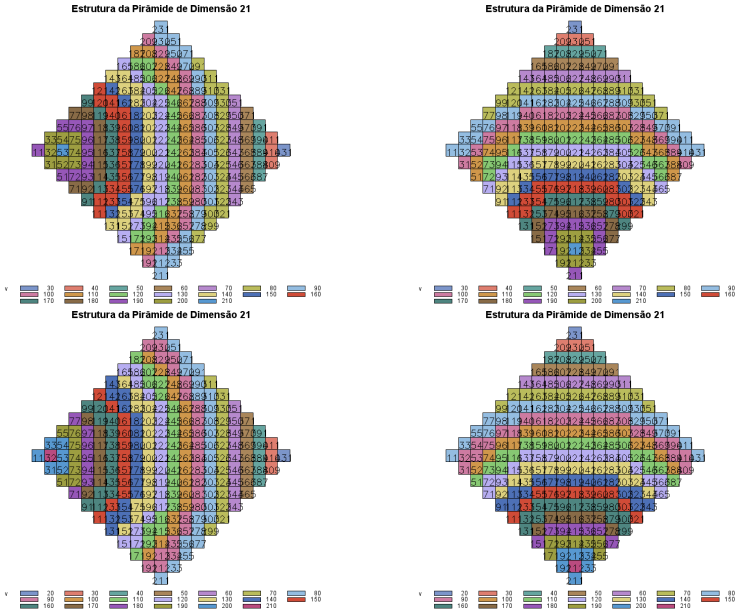


Figura 4.35: Mapas coropléticos 21×21, matriz de distâncias pico em 53,213,32 e 212 respectivamente

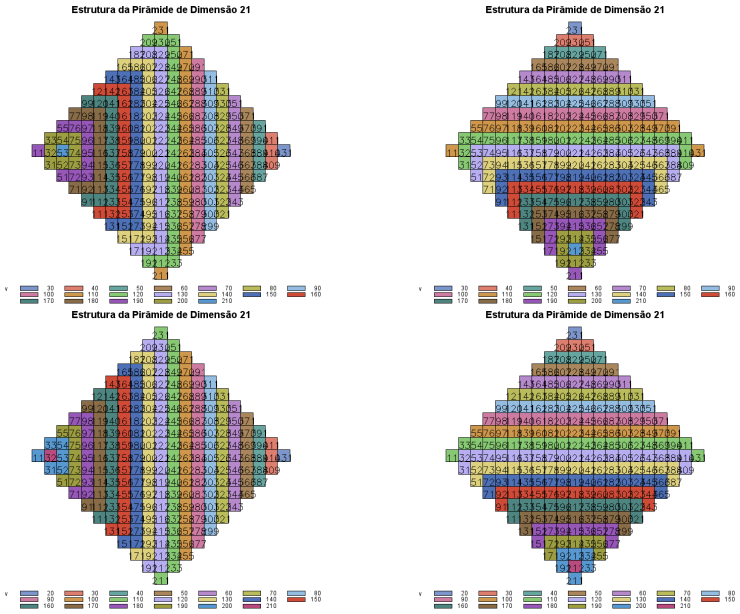


Figura 4.36: Mapas coropléticos 21×21, matriz de vizinhanças pico em 53,213,32 e 212 respectivamente

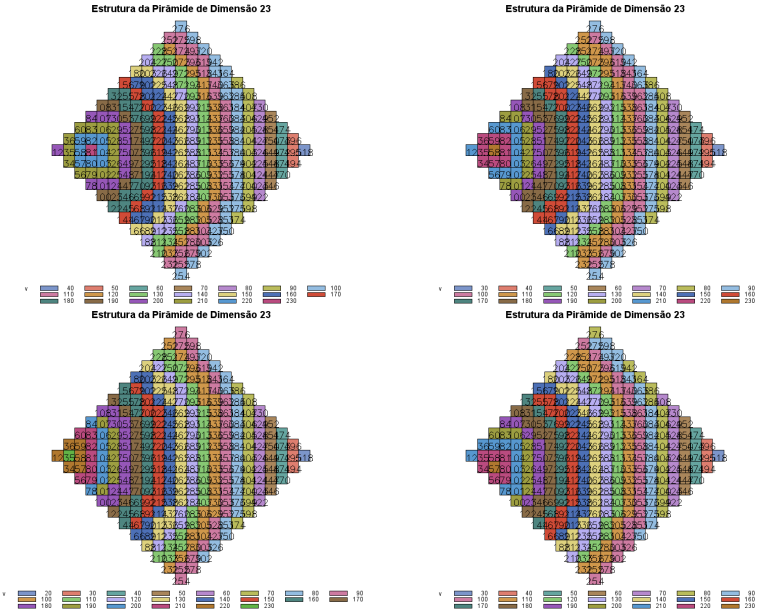


Figura 4.37: Mapas coropléticos 23×23, matriz de distâncias pico em 81,58,35 e 57 respectivamente

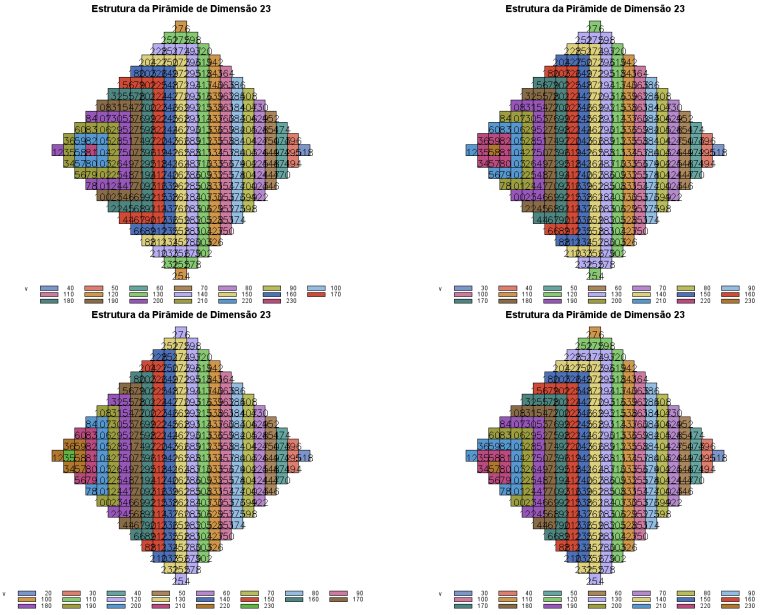


Figura 4.38: Mapas coropléticos 23 × 23, matriz de vizinhanças pico em 81,58,35 e 57 respectivamente

4.1.5 Análise dos Resultados para Polígonos Regulares

Nota-se, pelos mapas coropléticos e tabelas, que os maiores valores do índice se concentram nas partes extremas dos polígonos. Tendo em mente que os polígonos são simétricos, os maiores valores se encontrarão nas regiões das extremidades esquerda, direita, superior e inferior. Isso ocorre pois as regiões das extremidades do polígono possuem uma maior quantidade de vizinhos distantes, gerando uma maior diferença entre o valor do pico e o valor dos vizinhos nos dados simulados. Pela fórmula do índice

$$I = \frac{1}{\sum_{i \neq j} w_{ij}} \sum_{i \neq j} w_{ij} \left(\frac{y_i - \overline{y}}{s_y} \right) \left(\frac{y_j - \overline{y}}{s_y} \right) \tag{4.1}$$

essa distância no valor das variáveis acarreta em altos valores do I.

Observa-se também que quando há uma pequena quantidade de dados simulados os valores do índice aumentam rapidamente. Como exemplos, nos quadrados e losangos de dimensões 7×7 e 9×9 que possuem 49 e 81 regiões nos quadrados e 25 e 41 áreas nos losangos, o máximo do índice I de Moran saltou de $0,85858$ para $0,91276$ no quadrado com matriz de proximidades baseada nas distâncias e de $0,84645$ para $0,90638$ no mesmo polígono com matriz de proximidades baseada nos vizinhos. Já nos losangos observam-se saltos de $0,73297$ a $0,85124$ e $0,75568$ a $0,83910$ para os dois casos de matriz de proximidades. Essa diferença nos valores diminui a medida que a quantidade de regiões ou dados aumenta porém continuam sempre aumentando, indicando um possível viés do índice para uma pequena quantidade de dados. Conclui-se também que assintoticamente o índice deverá chegar ao valor de correlação máxima positiva 1, com os dados obtidos da forma simulada

nesse trabalho, como pode ser observado nas Figuras 4.39 4.40 4.41 4.42, os gráficos foram feitos com os maiores resultados do índice nos diversos polígonos e dimensões, com as Tabelas 4.21, 4.21, 4.22 e 4.22 representadas a seguir:

Tabela 4.21: Valores I de Moran: Quadrados

Dimensão	**Quantidade de observações**	**I de Moran**
7	49	0,85858
9	81	0,91276
11	121	0,94327
13	169	0,95786
15	225	0,96895
17	289	0,97548
19	361	0,98068
21	441	0,98419
23	529	0,98693
25	625	0,98868

Dimensão	**Quantidade de observações**	**I de Moran**
7	49	0,84645
9	81	0,90639
11	121	0,93708
13	169	0,95483
15	225	0,96602
17	289	0,97351
19	361	0,97877
21	441	0,98261
23	529	0,98549
25	625	0,98772

Uma regressão pode ser feita utilizando estes valores da forma:

$Y = ax + b$

onde:

Y = Valor máximo do I de Moran

a = Inclinação da reta de regressão

b = Intercepto da reta com o eixo

x = Quantidade de regiões no polígono

Temos, para os dados simulados os seguintes modelos:

Quadrado com matriz de distâncias

$Y = 0,000167241x + 0,907405393$

Quadrado com matriz de vizinhanças

$Y = 0,000181045x + 0,899564957$

Losângo com matriz de distâncias

$Y = 0,00076449x + 0,814785028$

Losângo com matriz de vizinhanças

$Y = 0,000711538x + 0,830292427$

Estimou-se que os valores mínimos do I tendem a 1 a partir de uma quantidade de 13.225 regiões, geradas por um quadrado 115×115, isso mostra que para grandes quantidades de dados a região de pico não influência no valor do índice, apenas o fenômeno em questão.

Tabela 4.22: Valores I de Moran: Losângos

Dimensão	Quantidade de observações	I de Moran
7	25	0,73297
9	41	0,8391
11	61	0,89644
13	85	0,92534
15	113	0,9435
17	145	0,95606
19	181	0,96377
21	221	0,97177
23	265	0,97334

Dimensão	Quantidade de observações	I de Moran
7	25	0,75568
9	41	0,85124
11	61	0,9048
13	85	0,93361
15	113	0,95094
17	145	0,9622
19	181	0,96995
21	221	0,97583
23	265	0,9774

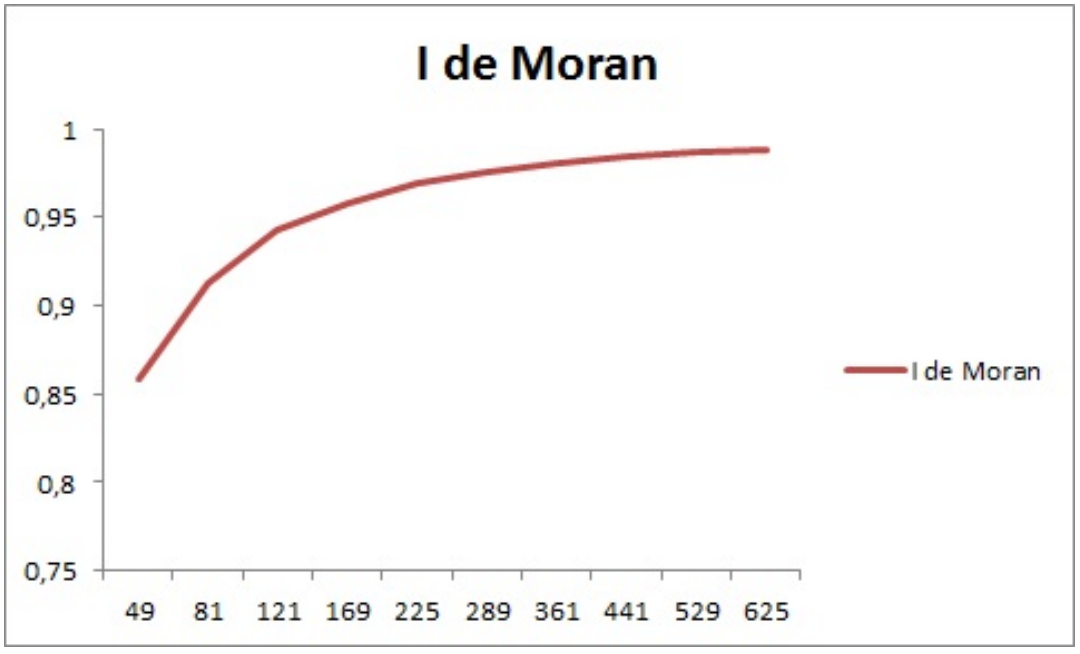


Figura 4.39: Gráficos do I de Moran em relação a quantidade de dados, quadrado distância

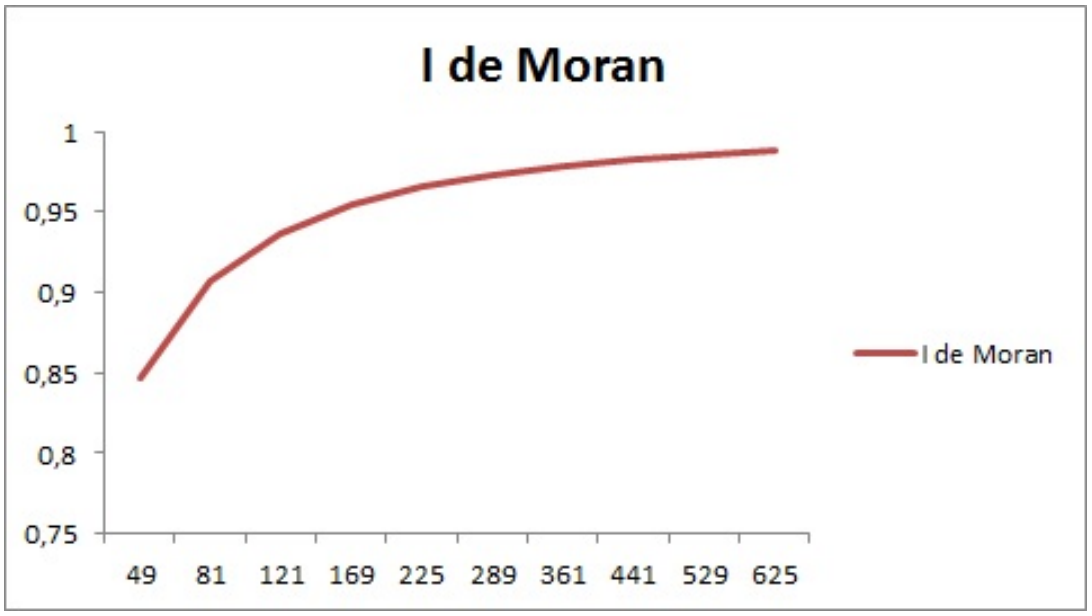


Figura 4.40: Gráficos do I de Moran em relação a quantidade de dados, quadrado vizinho

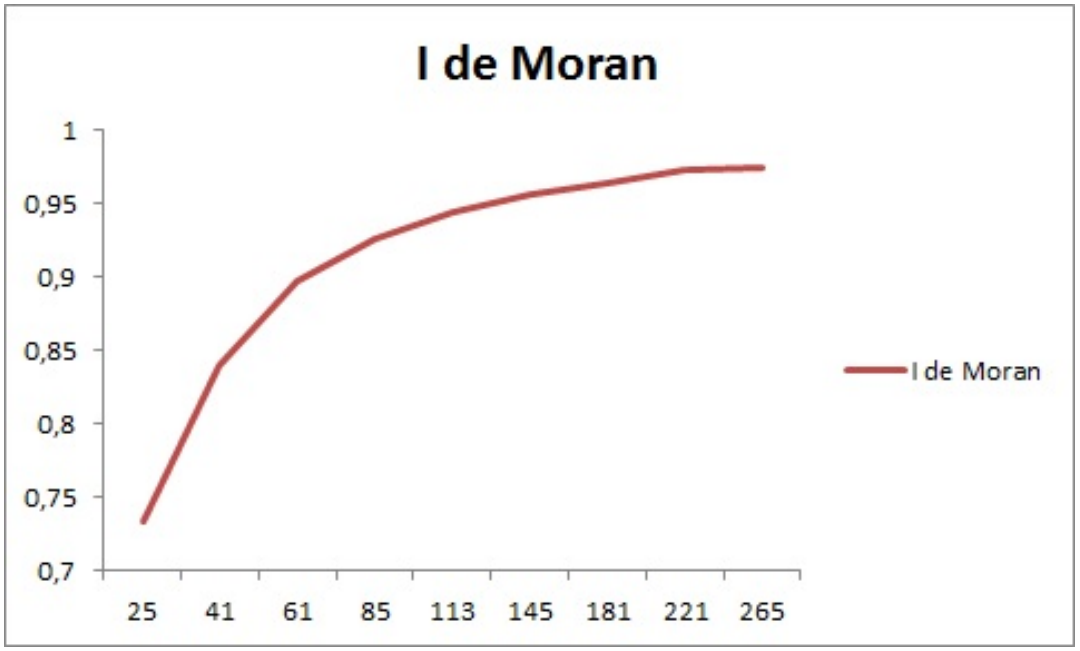


Figura 4.41: Gráficos do I de Moran em relação a quantidade de dados, losângo distância

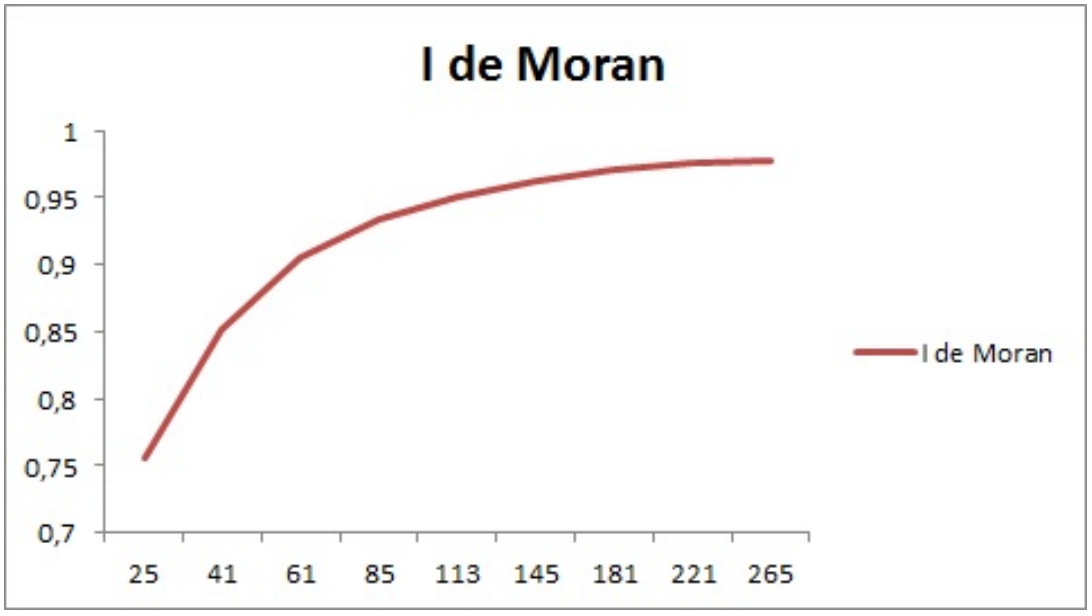


Figura 4.42: Gráficos do I de Moran em relação a quantidade de dados, losângo vizinho

Os valores mínimos também demonstram um possível viés do índice para uma pequena quantidade de dados. Observando a distância entre maiores e menores valores, percebe-se que a diferença diminui à medida que a quantidade de dados aumenta, sendo essa diferença muito maior para polígonos com pequenas quantidade de dados. Como exemplo, no quadrado utilizando matriz de distâncias 7×7 que possui um mínimo de $0,46772$ e máximo de $0,85858$, e no quadrado 9×9 utilizando a mesma matriz os valores são de $0,71462$ e $0,91276$. Esse comportamento do índice novamente demonstra um possível viés para pequenas quantidades de dados já que para polígonos com grandes dimensões os valores de máximo e mínimo são muito próximos e todos tendem à correlação máxima 1 para dados obtidos da forma do trabalho.

Ao observar os valores, nota-se uma diferença do comportamento do índice entre quadrados e losângos. Enquanto nos quadrados a matriz de proximidade baseada na distância gerou os maiores valores, nos losângos a matriz baseada nas vizinhanças resultou em maiores magnitudes do I.

4.2 Polígonos Irregulares

Como dito anteriormente, os polígonos irregulares serão representados pelos formatos de três estados brasileiros,divididos em municípios, e uma cidade americana, sendo estes São Paulo(SP) com 645 municípios, Tocantins(TO) que possui 139 municípios e Pernambuco(PE) com 181 municípios além da cidade de Columbus(COL) com 49 bairros. Esses polígonos foram selecionados devido à variedade do formato das delimitações geográficas que os representam. Já a cidade de colombus foi selecio-

nada por possuir exatamente 49 regiões em suas delimitações, sendo assim facilmente comparável com os polígonos regulares 7×7. Como na análise prévia dos polígonos regulares não notaram-se diferenças significativas entre o uso das matrizes proximidade baseadas na distância e na vizinhança, optou-se pelo uso de matrizes de proximidade baseadas na vizinhança para o cálculo do índice nos polígonos irregulares. Para os polígonos irregulares serão observados apenas os quatro maiores valores do índice.

Os códigos indicados nas colunas pico das tabelas, e nos mapas coropléticos, são os códigos dos municípios, utilizados pelo IBGE.

4.2.1 São Paulo

Segue a tabela 4.23 com os valores para São Paulo, assim como os mapas coropléticos.

Tabela 4.23: São Paulo: Valores do índice I de Moran

São Paulo	
I de Moran	**Pico**
0.99722	351880
0.99710	352940
0.99709	354780

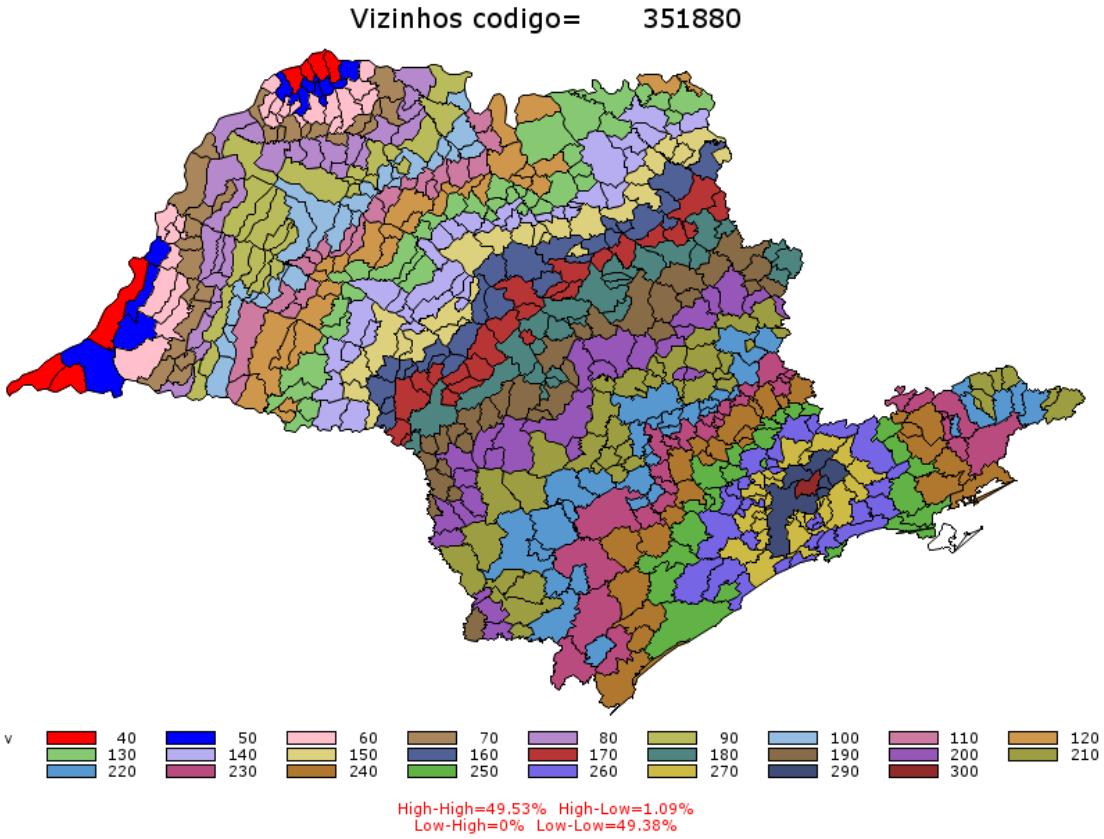


Figura 4.43: Regiões que geraram maior valor do índice (SP1)

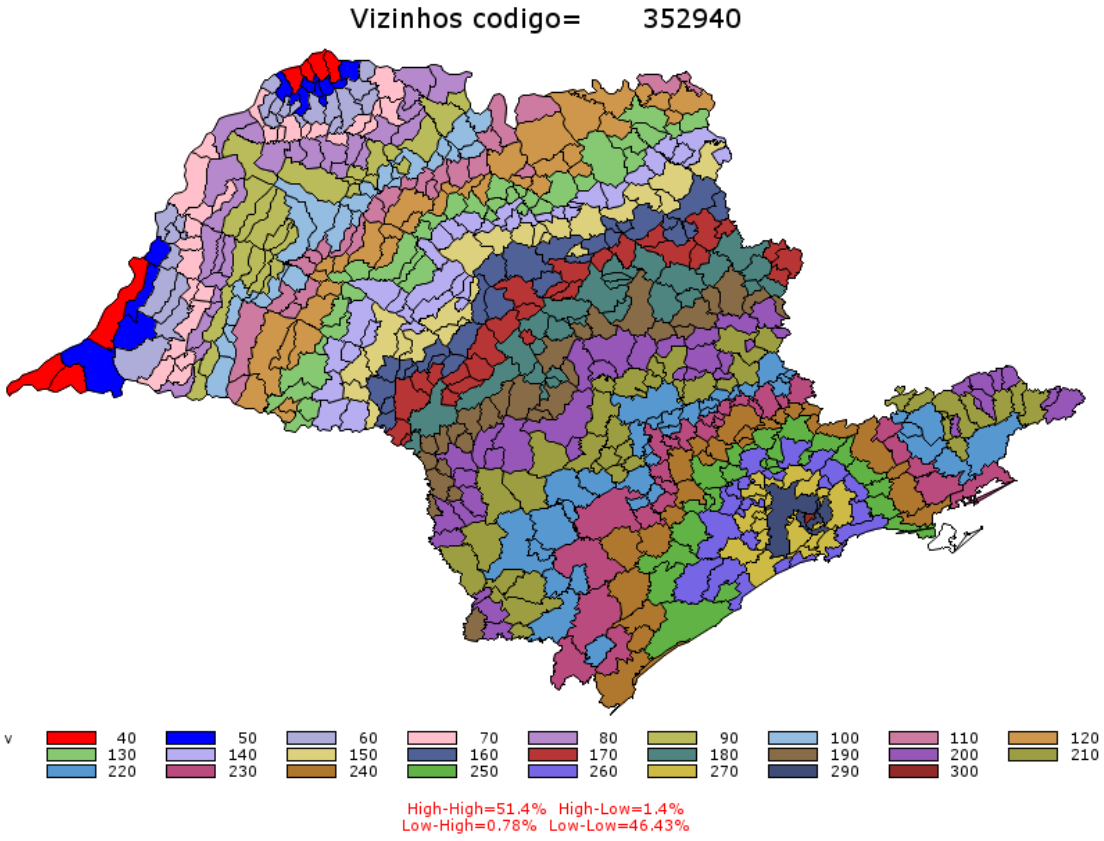


Figura 4.44: Regiões que geraram maior valor do índice (SP2)

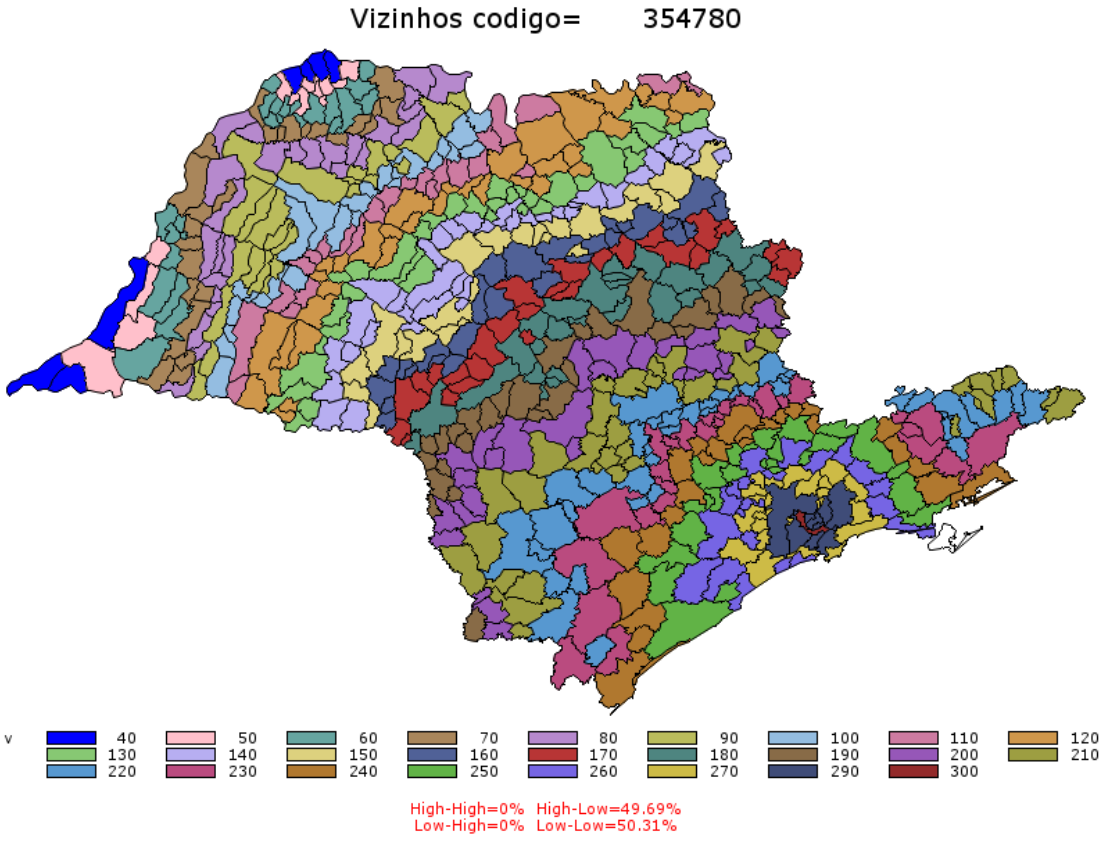


Figura 4.45: Regiões que geraram maior valor do índice (SP3)

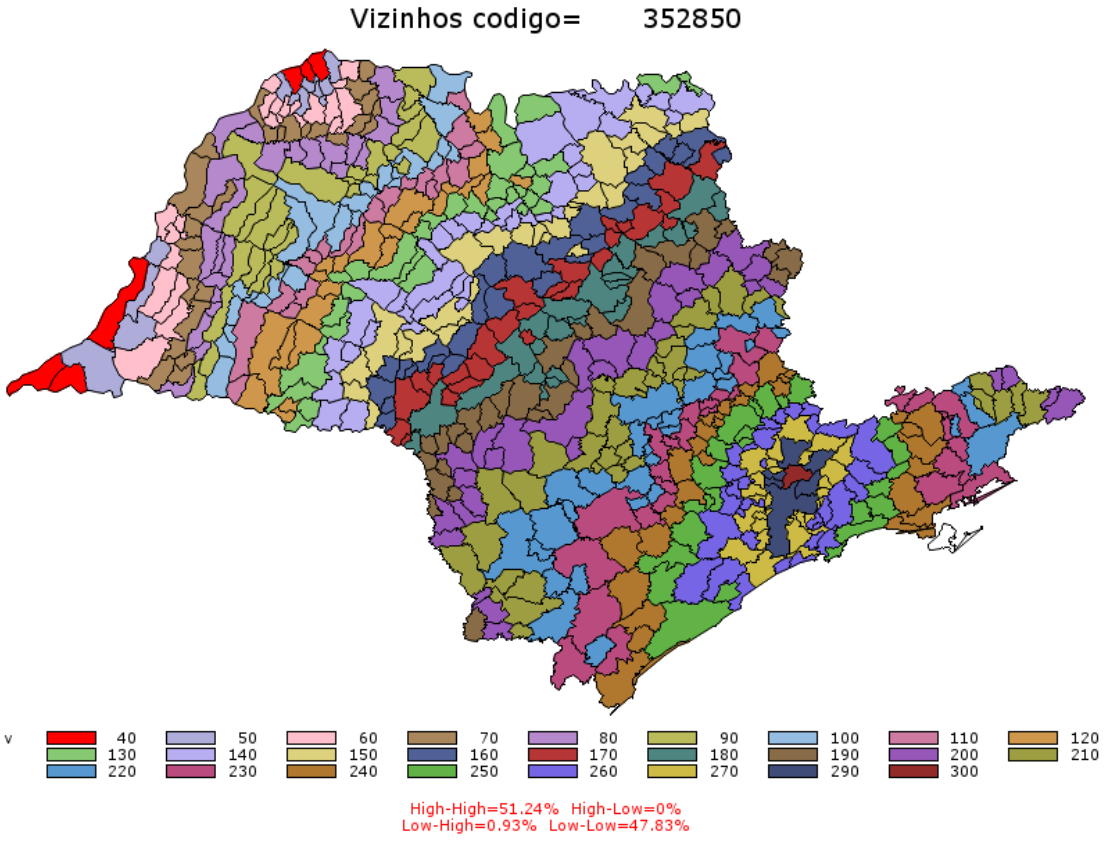


Figura 4.46: Regiões que geraram maior valor do índice (SP4)

Para uma comparação será utilizado o quadrado 25×25. Visto que são paulo possui 645 áreas e uma simulação de um quadrado 25×25 produz 625 áreas, uma comparação entre polígono regular e irregular não seria tão errônea. Observando-se os resultados, temos um máximo do índice I para os dados simulados no polígono regular de 0.98772 e no irregular obteve-se 0.99722. Nota-se que apesar de uma quantidade de dados bem próxima, o polígono irregular possui um valor máximo do índice I de moran superior ao regular. Isso se dá pelo fato da falta de simetria dos polígonos irregulares, de modo que uma extremidade possui mais vizinhos do que a outra, resultando em maiores valores do índice. Nota-se, também que o comportamento do máximo do índice referente a posição do pico dos dados é similar tanto no polígono regular quanto no irregular isto é, ambos geram valores máximos do I quando o pico dos dados está nas extremidades ou próximo delas.

4.2.2 Pernambuco

Segue a Tabela 4.24 com os valores para Pernambuco, assim como os mapas coropléticos.

Tabela 4.24: Pernambuco: Valores do índice I de Moran

Pernambuco	
I de Moran	**Pico**
0.99126	261245
0.98998	261520
0.98986	261040
0.98978	260630

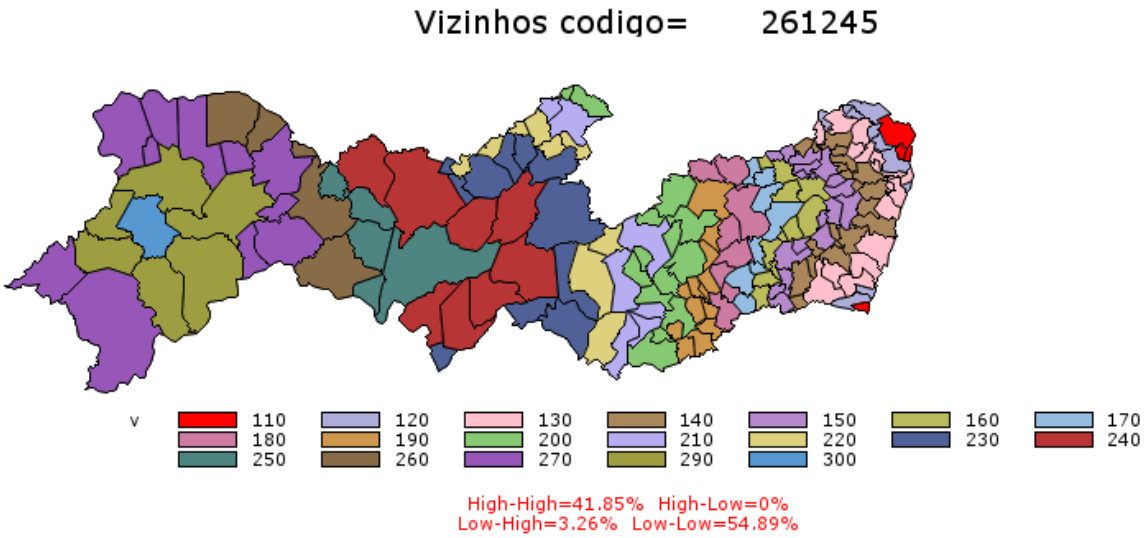


Figura 4.47: Regiões que geraram maior valor do índice (PE1)

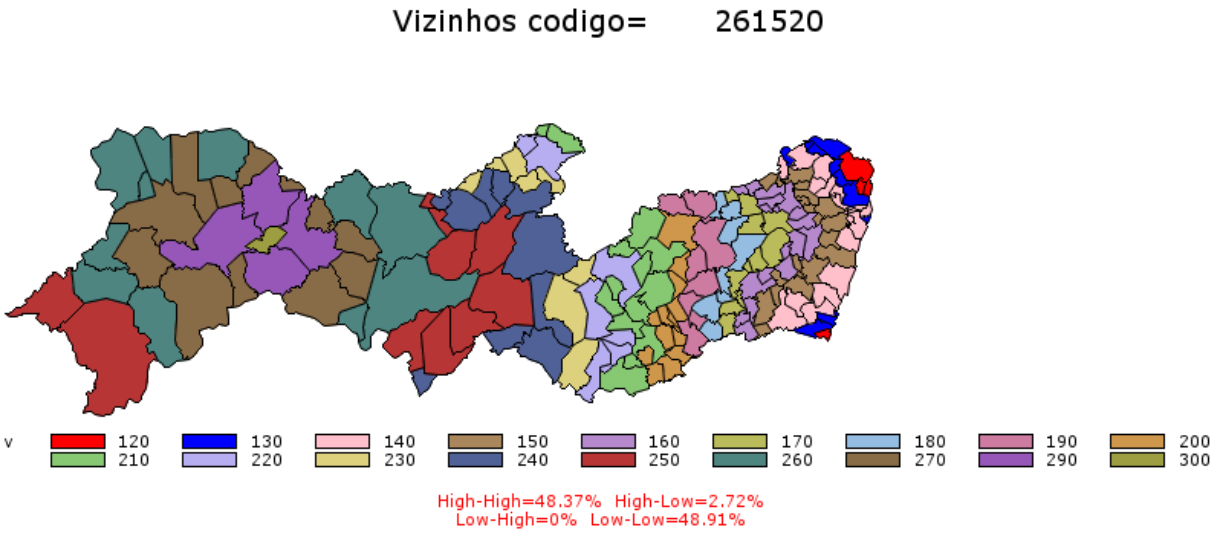


Figura 4.48: Regiões que geraram maior valor do índice (PE2)

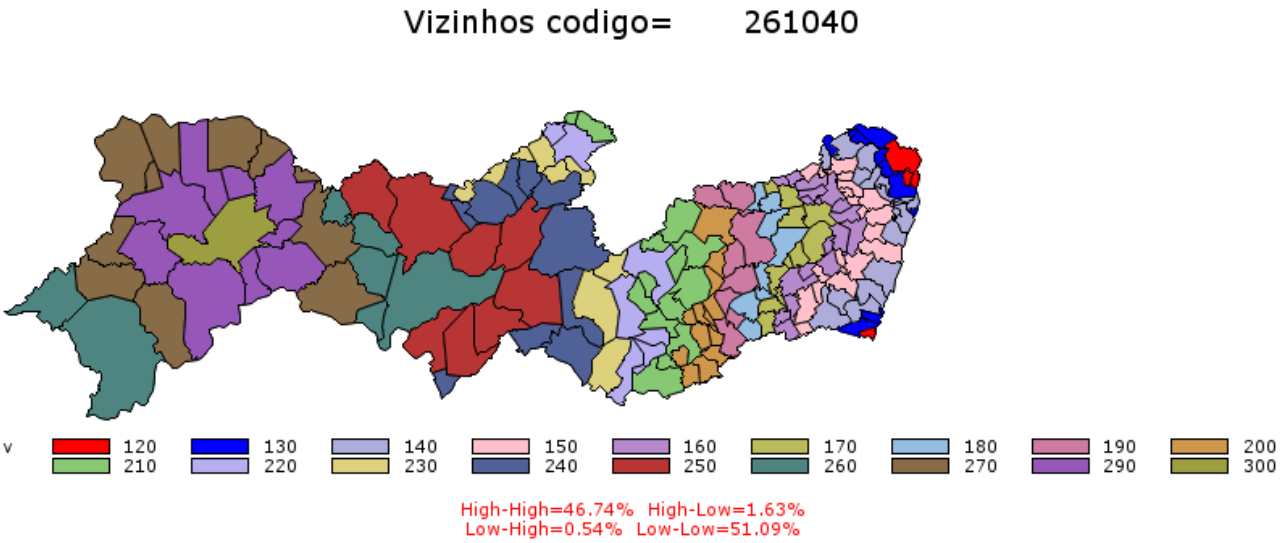


Figura 4.49: Regiões que geraram maior valor do índice (PE3)

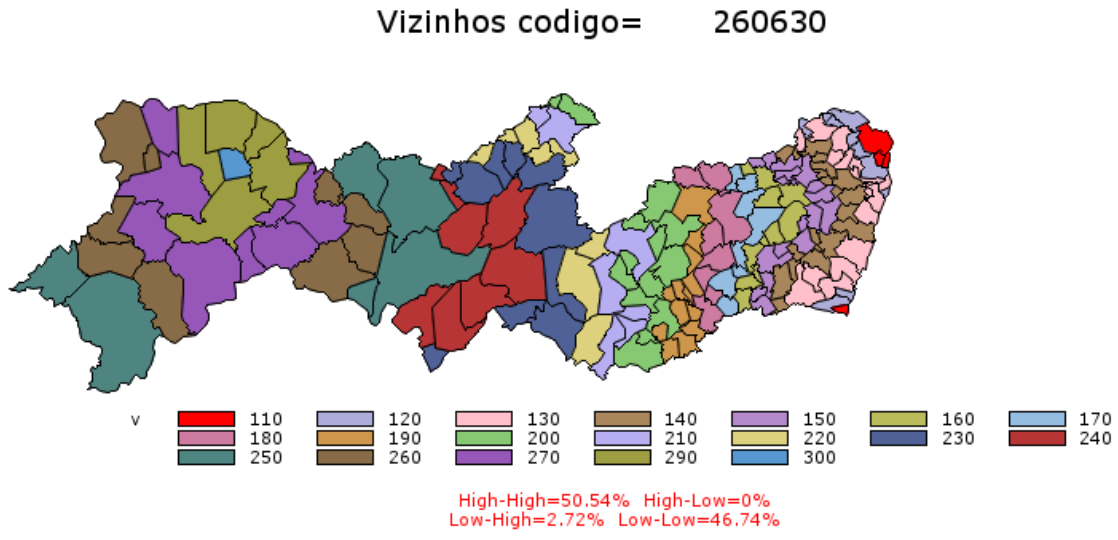


Figura 4.50: Regiões que geraram maior valor do índice (PE4)

O estado de Pernambuco(PE) possui 185 municípios, e o polígono irregular que delimita sua área pode ser comparado ao losângo gerado a partir de um quadrado 19×19 visto que este polígono regular tem 181 regiões. Novamente o valor máximo do índice I de Moran foi maior no polígono irregular sendo $0,99126$ para o polígono de Pernambuco e $0,96995$ no losango 19×19, evidenciando que a forma do polígono e o número de vizinhos que a região de pico possui influência no índice I, assim como a quantidade de regiões.

4.2.3 Tocantins

Segue a Tabela 4.25 com os valores referentes a tocantis, assim como os mapas coropléticos.

Tabela 4.25: Tocantins valores do I de Moran

Tocantins	
I de Moran	**Pico**
0.96098	170240
0.95880	170040
0.95879	170560
0.95792	171780

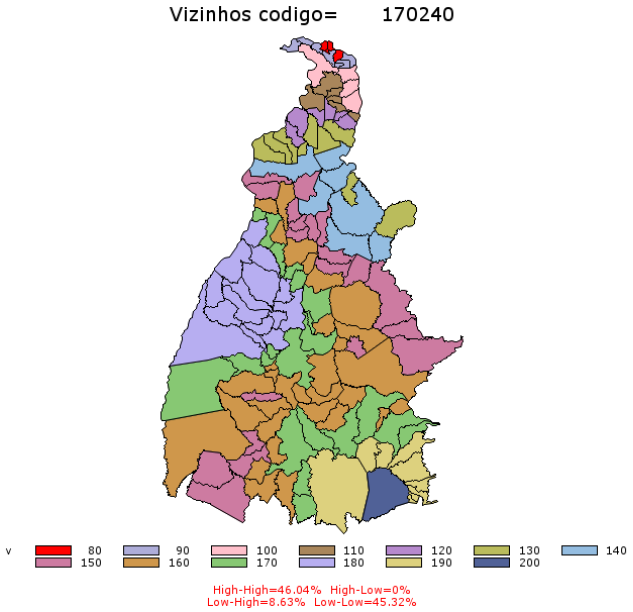


Figura 4.51: Regiões que geraram maior valor do índice (TO1)

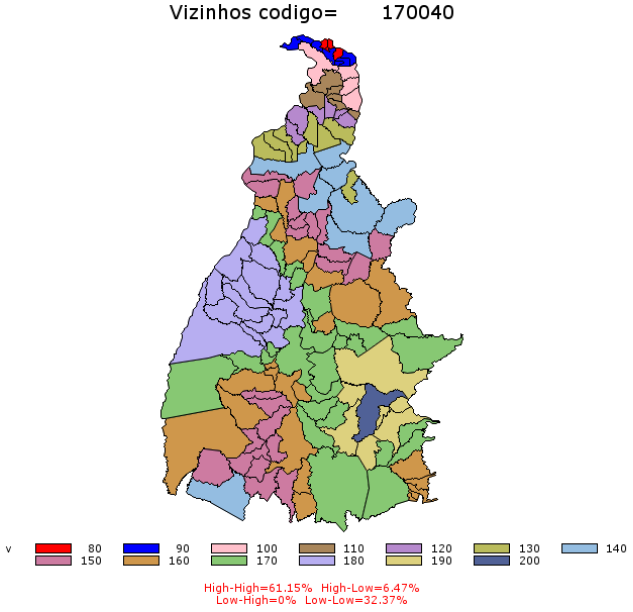


Figura 4.52: Regiões que geraram maior valor do índice (TO2)

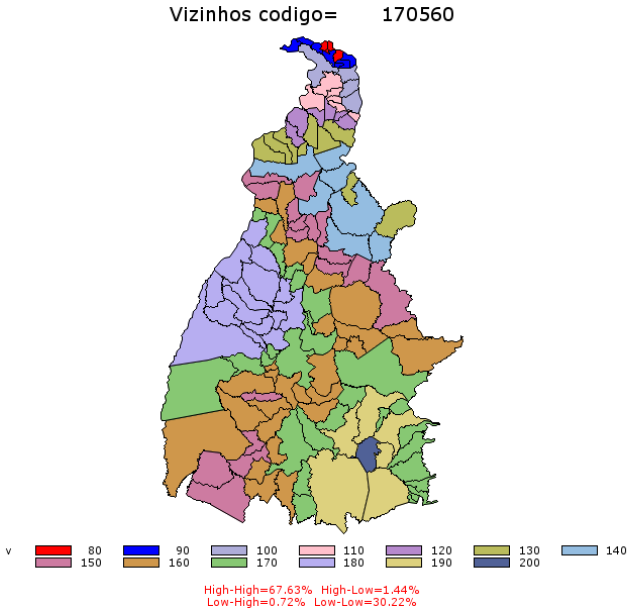


Figura 4.53: Regiões que geraram maior valor do índice (TO3)

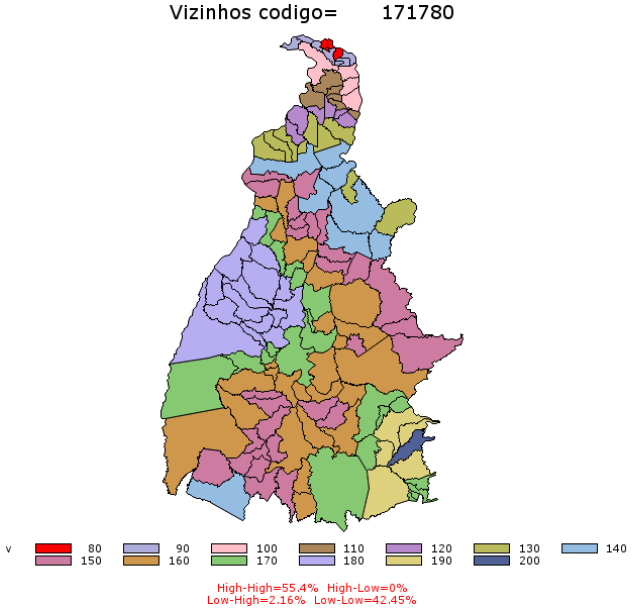


Figura 4.54: Regiões que geraram maior valor do índice (TO4)

Tocantis por sua vez tem 139 municípios em suas delimitações e pode ser comparado ao losângo criado a partir de um quadrado 17×17 já que este polígono compreende 145 regiões, uma quantidade próxima. O máximo encontrado para tocantins foi de $0,96098$ e no losângo $0,95606$, nos levando as mesmas conclusões de anteriormente.

4.2.4 Columbus

Segue a Tabela 4.26 com os valores para Columbus, assim como os mapas coropléticos.

Tabela 4.26: Columbus valores do I de Moran

Columbus	
I de Moran	**Pico**
0.93452	36
0.92325	39
0.92324	40
0.91680	41

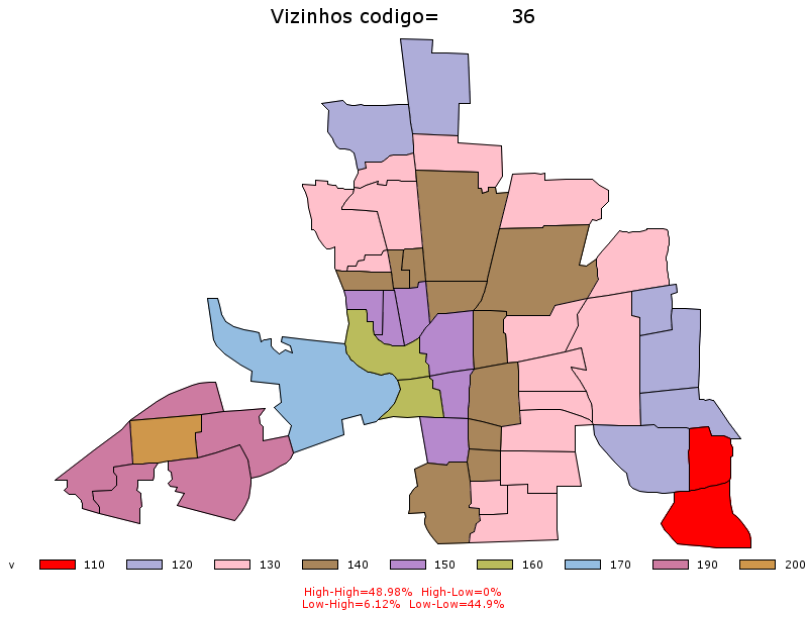


Figura 4.55: Regiões que geraram maior valor do índice (COL1)

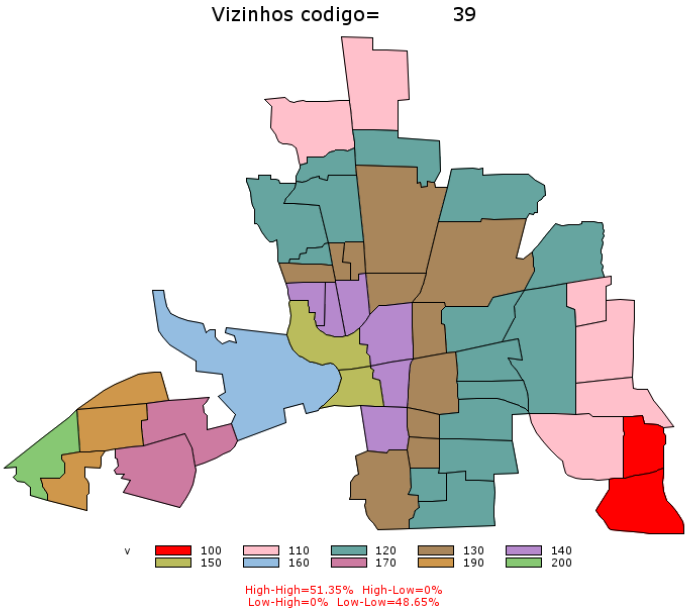


Figura 4.56: Regiões que geraram maior valor do índice (COL2)

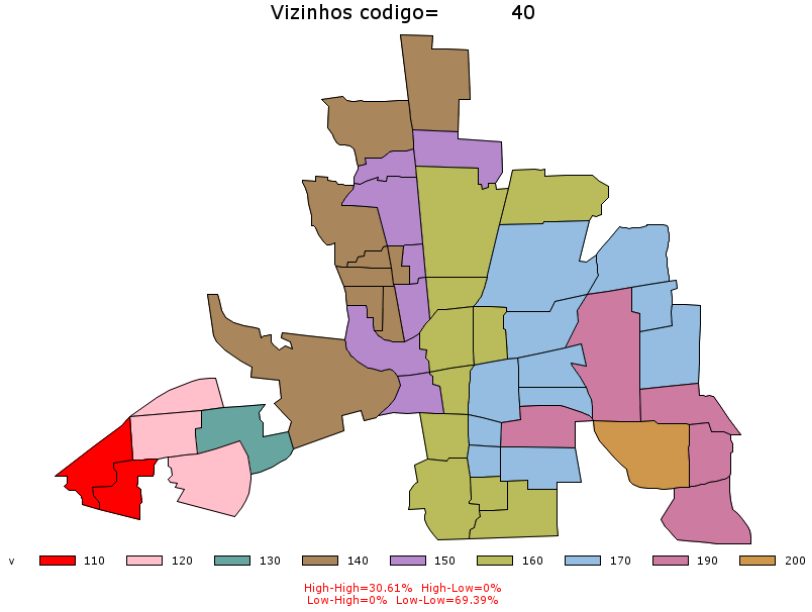


Figura 4.57: Regiões que geraram maior valor do índice (COL3)

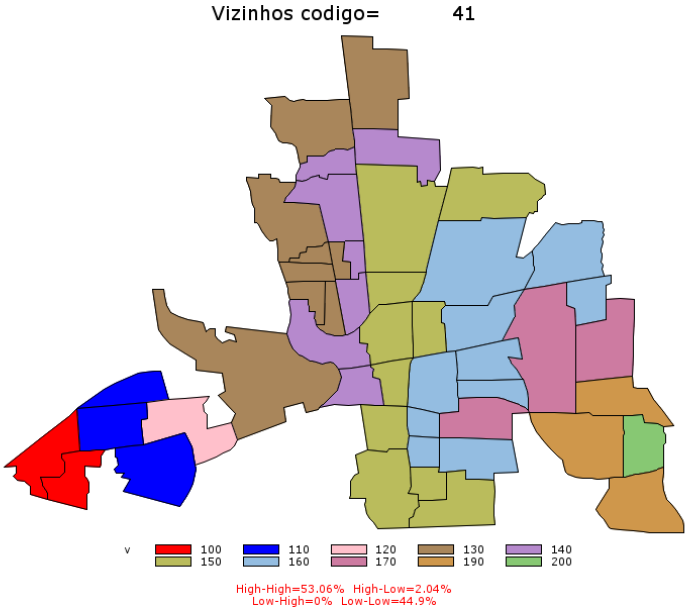


Figura 4.58: Regiões que geraram maior valor do índice (COL4)

Columbus possui exatamente 49 bairros e foi selecionado justamente para uma comparação exata com o quadrado 7×7, com os resultados para o índice de $0,93452$ para columbus e $0,73297$ no quadrado. A comparação exata foi a que gerou a maior diferença entre polígono regular e irregular, levando-se em conta que Columbus foi o menor polígono irregular utilizado conclui-se que de fato a quantidade de observações afeta drasticamente o valor do índice.

Com as informações obtidas segue a Tabela 4.27 com os valores a serem utilizados na correção do índice para pequenas quantidades de dados, sendo o índice corrigido:

$$I^* = \frac{\text{I computado para quantidade de dados n}}{\text{I máximo para quantidade de dados n}} \tag{4.2}$$

Tabela 4.27: Valores máximos do I de Moran por Quantidade de observações

Dimensão	**Quantidade de observações**	**I de Moran**
7	49	0,84645
9	81	0,90639
11	121	0,93708
13	169	0,95483
15	225	0,96602
17	289	0,97351
19	361	0,97877
21	441	0,98261
23	529	0,98549
25	625	0,98772

Para ilustrar a real dependência espacial, será utilizado um exemplo de Lesage (1999) envolvendo 49 bairros de Columbus, Ohio com informações coletadas por Anselin (1988) sobre a criminalidade da cidade. A correlação espacial computada foi de 0.5109472385, utilizando o ajuste para polígonos com 49 regiões temos o I corrigido de:

$$I^* = \frac{0.5109472385}{0,84645} = 0,603635 \tag{4.3}$$

Vê-se que o valor aumenta quando tirada a influência da quantidade de dados.

Capítulo 5

CONCLUSÃO

Após análise das informações em relação ao índice de correlação I de Moran, tanto para os polígonos regulares quanto irregulares, conclui-se que a quantidade de dados de fato influencia no resultado do índice e que a medida que o número de dados cresce os valores tanto mínimos quanto máximos do índice tendem a 1, quando obtidos de forma semelhante à simulada neste trabalho.

Conclui-se também que o valor máximo do índice I para os dados simulados depende da região de pico dos dados e viu-se que as regiões que resultavam nos maiores valores foram as regiões das extremidade em todos os polígonos tanto regulares quanto irregulares. Isso se dá por essas regiões possuírem maiores quantidades de vizinhos distantes. Viu-se também que a irregularidade do polígono também influencia nos resultados, pois a simetria tem relação com o número de vizinhos.

Além disso pode se obter uma correção para o índice utilizando os valores máximos calculados neste trabalho.

Os máximos simulados neste trabalho foram gerados considerando que a variável diminuia de 10 em 10 unidades de medida a partir do pico, caso a variável diminuisse mais rapidamente (de 50 em 50 por exemplo) os resultados poderiam ser outros.

Referências Bibliográficas

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers.

Assunção, R. M. (2003). Índices de autocorrelação espacial.

Bussab, W. O. & Morettin, P. A. (2004). *Estatística Básica*. Saraiva.

Druck, S., de Carvalho, M. S., Câmara, G., & Monteiro, A. M. V. (2004). *Análise Espacial de Dados Geográficos*. Embrapa Cerrados.

Lembo, A. J. (2005). Spatial autocorrelation. Technical report, Cornell University. Department of Crop and Soil Sciences.

Lesage, J. P. (1999). The theory and practice of spatial econometrics. Technical report, Department of Economics, University of Toledo, EUA.

Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika*, (37):17–23.

Silva, A. R. (2006). Avaliação de modelos de regressão espacial para análise de cenários do transporte rodoviário de carga. Master's thesis, Universidade de Brasília.

Tobler, W. R. (1979). Cellular geography. *Philosophy in Geography*. Edited by S. Gale and G. Olsson. Eds, Amsterdam.