



Universidade Federal do ABC

AUTOCORRELAÇÃO ESPACIAL

Vitor Vieira Vasconcelos
Flávia da Fonseca Feitosa

BH1350 – Métodos e Técnicas de Análise da Informação para o Planejamento
Julho de 2017

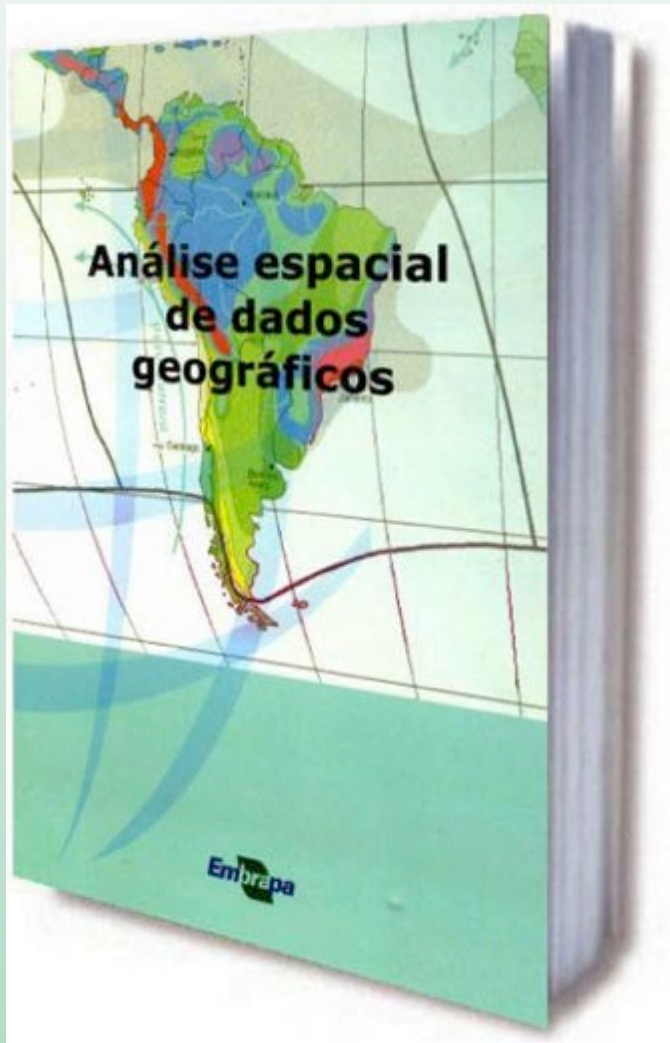
AULAS ANTERIORES...

- *A importância analítica do espaço para o Planejamento Territorial*
- *Estrutura de Dados Espaciais (Matrizes & Vetores)*
- *Revisão de Estatística*
- *Na última aula: CORRELAÇÃO ENTRE VARIÁVEIS*

**Até aqui apenas visualizamos
espacialmente os dados (mapas coropléticos)
No entanto, precisamos ir além!!!**

**DADOS ESPACIAIS SÃO
ESPECIAIS**

BIBLIOGRAFIA BÁSICA



Druck, S.; Carvalho, M.S.; Câmara, G.; Monteiro, A.V.M. (eds) "Análise Espacial de Dados Geográficos". Brasília, EMBRAPA, 2004.

Capítulo 5 – Análise Espacial de Áreas

A NATUREZA DOS DADOS ESPACIAIS

Alguns princípios:

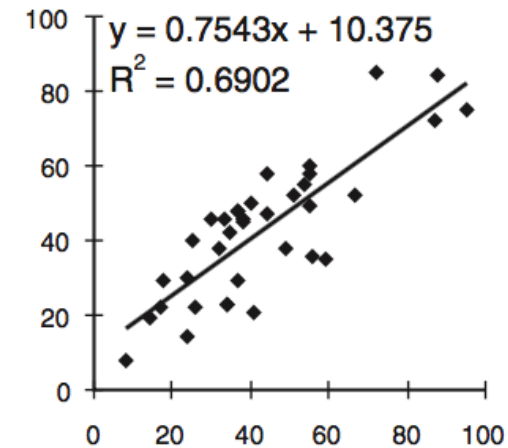
- *As questões de escala geográfica, agregação e nível de detalhamento são fundamentais para construir representações apropriadas do mundo*

O Problema da Unidade de Área Modificável

Independent variable Dependent variable

87	95	72	37	44	24
40	55	55	38	88	34
41	30	26	35	38	24
14	56	37	34	8	18
49	44	51	67	17	37
55	25	33	32	59	54

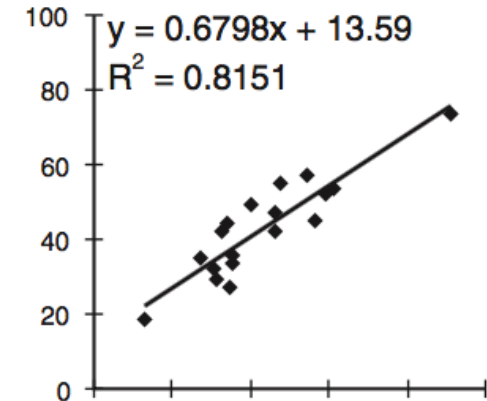
72	75	85	29	58	30
50	60	49	46	84	23
21	46	22	42	45	14
19	36	48	23	8	29
38	47	52	52	22	48
58	40	46	38	35	55



Aggregation scheme 1

91	54.5	34
47.5	46.5	61
35.5	30.5	31
35	35.5	13
46.5	59	27
40	32.5	56.5

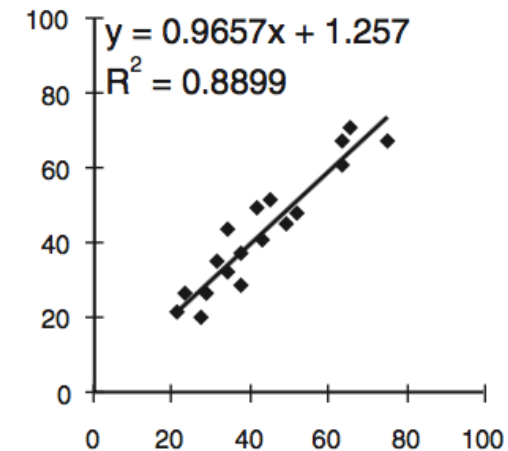
73.5	57	44
55	47.5	53.5
33.5	32	29.5
27.5	35.5	18.5
42.5	52	35
49	42	45

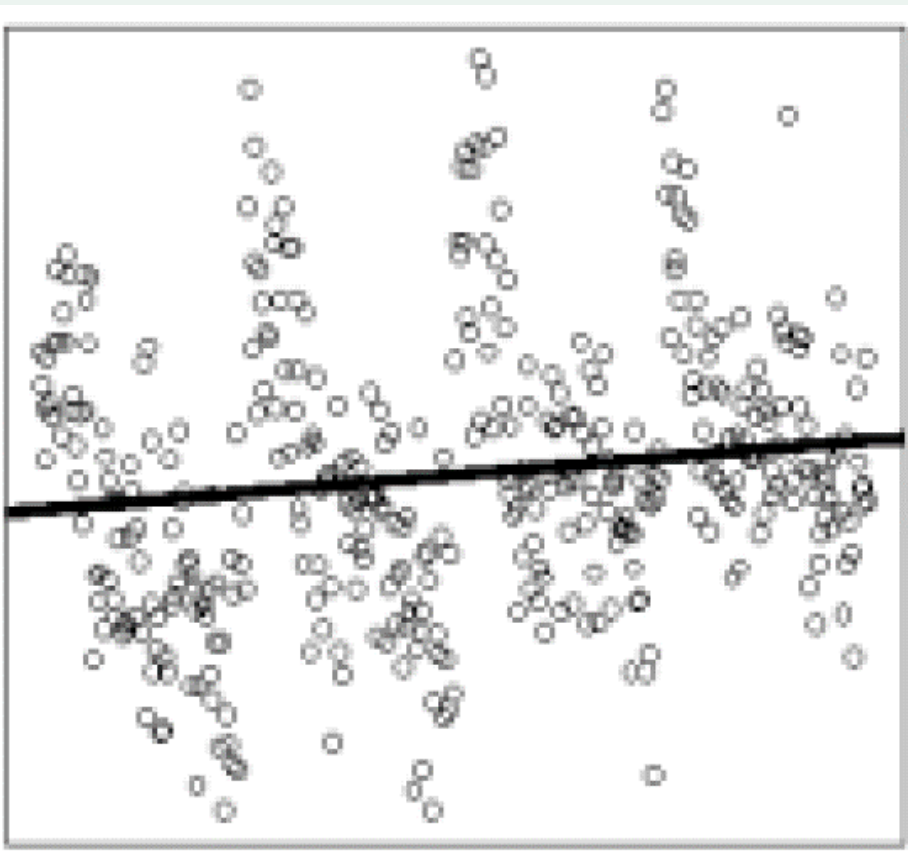


Aggregation scheme 2

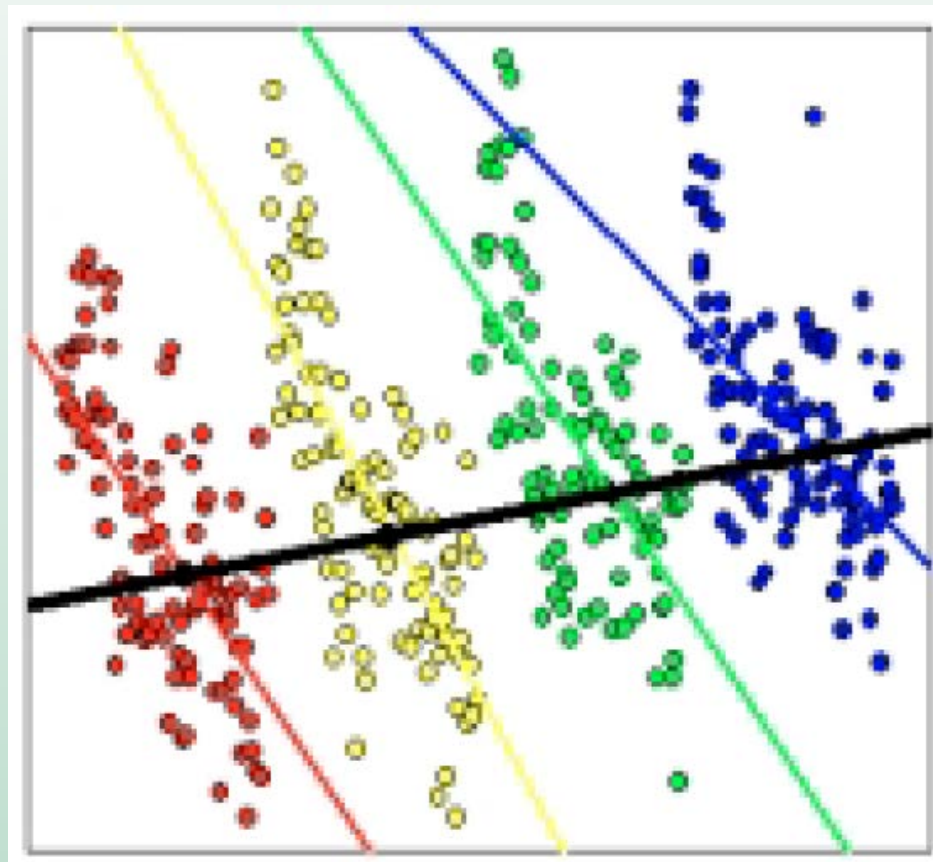
63.5	75	63.5	37.5	66	29
27.5	43	31.5	34.5	23	21
52	34.5	42	49.5	38	45.5

61	67.5	67	37.5	71	26.5
20	41	35	32.5	26.5	21.5
48	43.5	49	45	28.5	51.5





Dados agregados



Dados segregados

A NATUREZA DOS DADOS ESPACIAIS

Alguns princípios:

- *Diferentes medidas obtidas do mundo real **covariam**, e compreender a natureza espacial da covariação pode nos ajudar a entender melhor os fenômenos.*

Na representação do mundo real, é importante incorporar informações sobre como dois ou mais fatores **COVARIAM** espacialmente.

Análise dos impactos de investimentos no sistema público de transporte (ex. Monotrilho): Avaliação de como os preços dos imóveis aumentam à medida que aumenta a proximidade das estações.

Outros exemplos???

A NATUREZA DOS DADOS ESPACIAIS

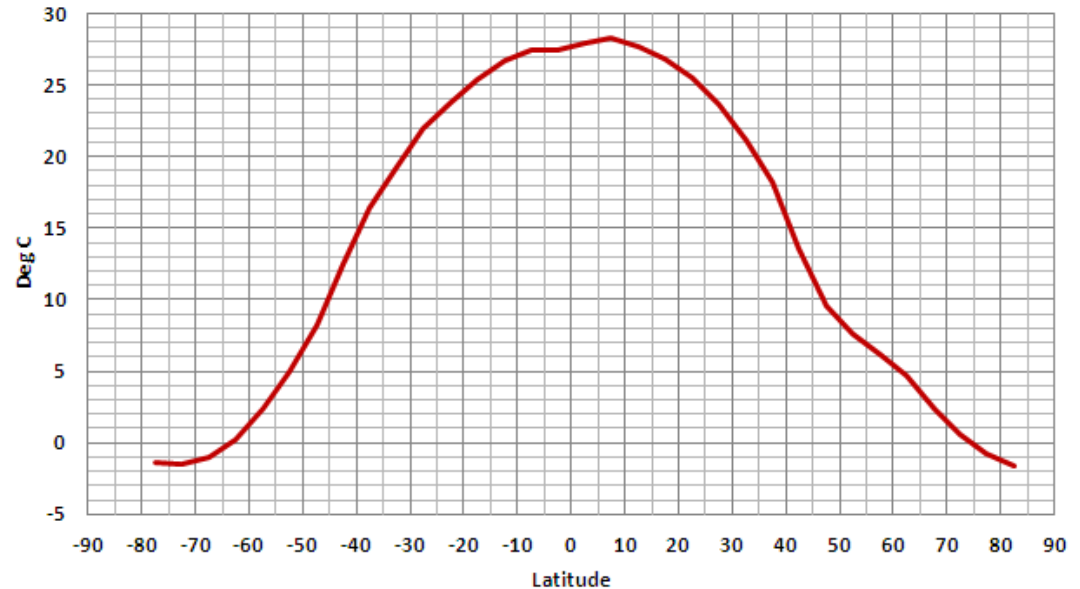
Alguns princípios:

- *Os efeitos de proximidade são fundamentais para representar e compreender a variação espacial e para reunir representações incompletas de lugares únicos.*

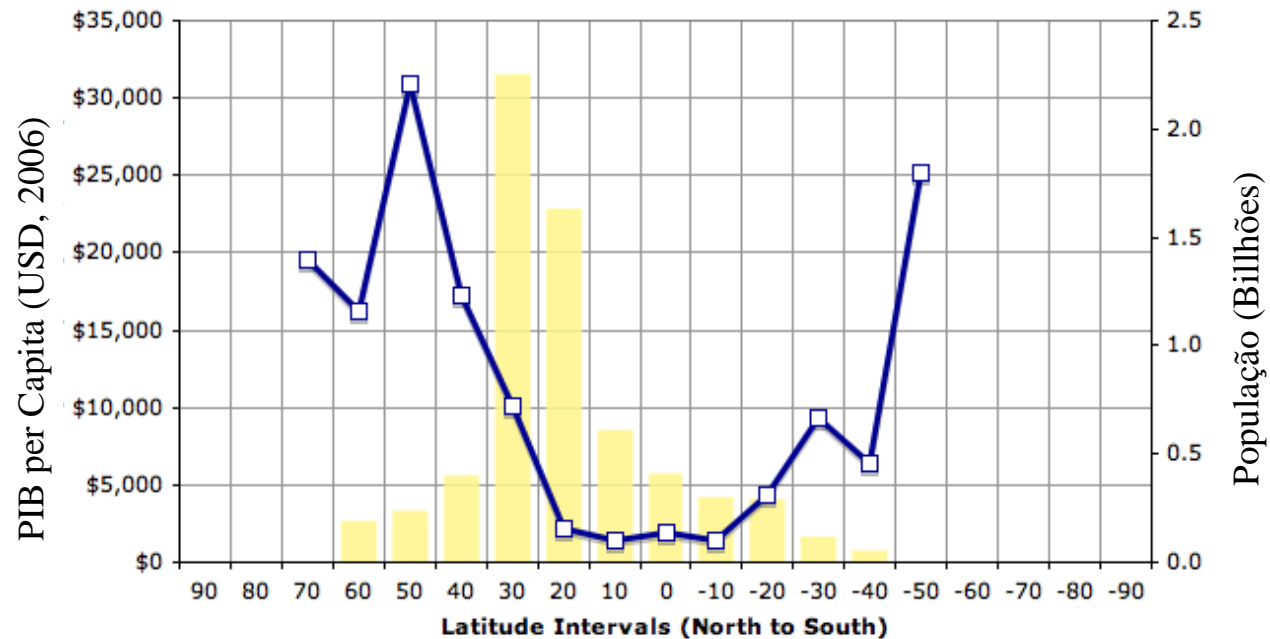
(Não dá para analisar proximidade numa planilha!!!)

Gráficos com eixos espaciais

Temperatura Média à Superfície do Mar
Média de janeiro de 1982 a Dezembro de 2011

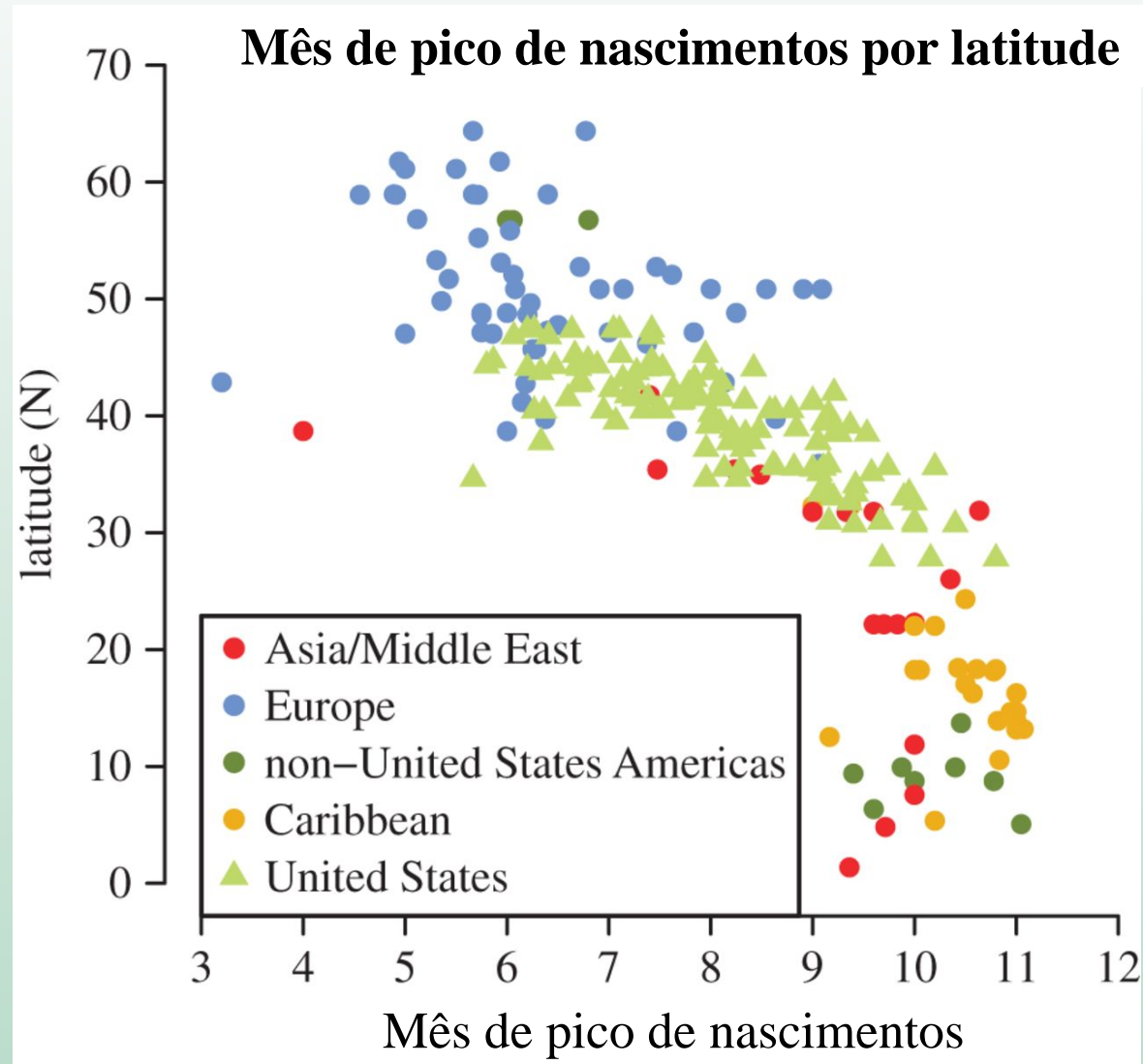


<https://medium.com/universe-factory/climate-modeling-101-4544e00a2ff2>



<https://thecivilisingmission.com/category/contemporary-history/page/3/>

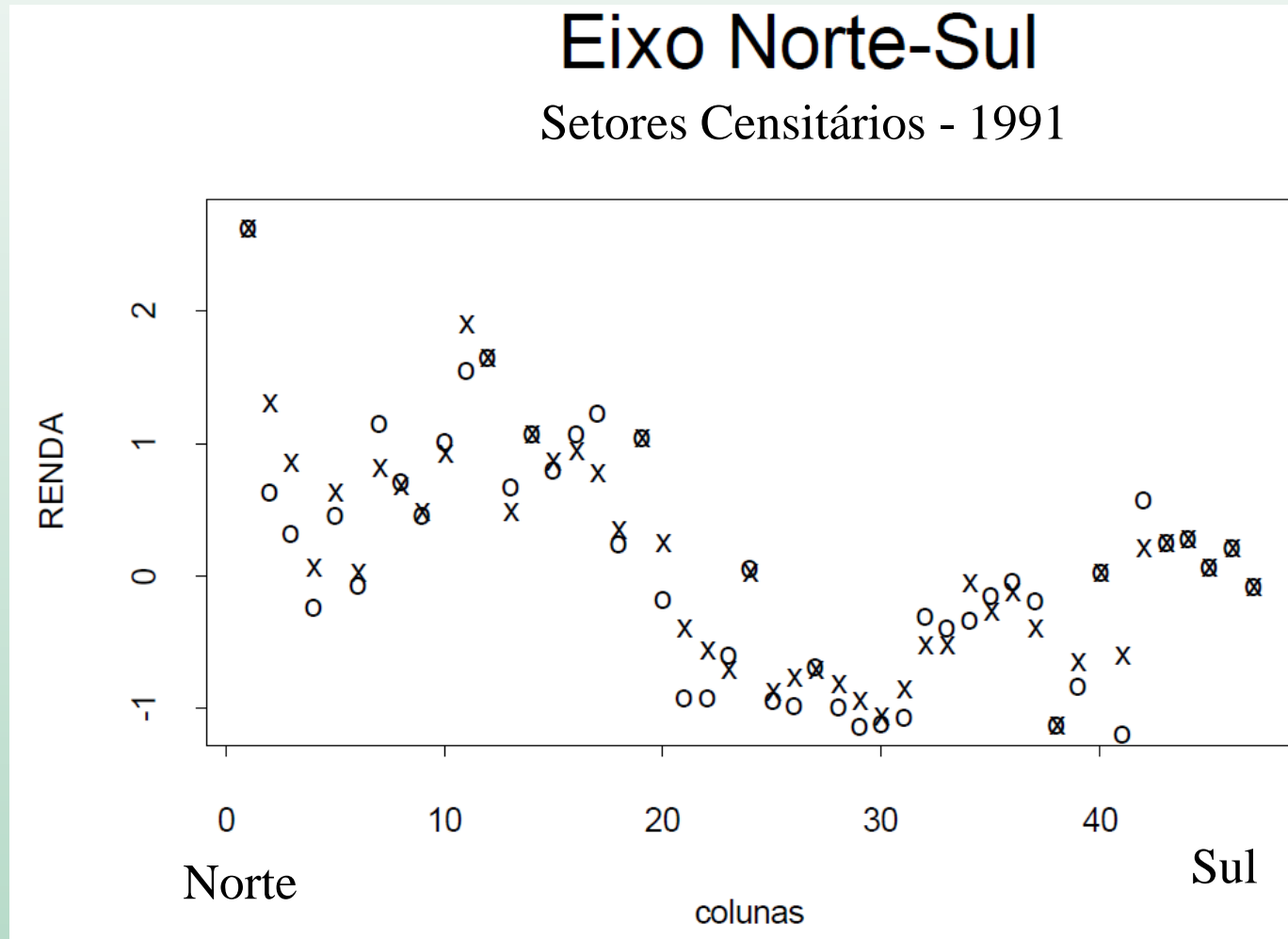
Gráficos com eixos espaciais



Martinez-Bakker, M., Bakker, K.M., King, A.A. and Rohani, P., 2014, May. Human birth seasonality: latitudinal gradient and interplay with childhood disease dynamics. In *Proc. R. Soc. B* (Vol. 281, No. 1783, p. 20132438). The Royal Society.

Gráficos com eixos espaciais

Ilha do Governador - RJ



PRIMEIRA LEI DA GEOGRAFIA

“Lei de Tobler”

*“No mundo, todas as coisas se parecem,
mas coisas mais próximas são mais
parecidas que aquelas mais distantes”*
(Waldo Tobler, 1970)

Tobler, W. R. 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography* 46: 234–40.

*A relação entre eventos próximos no espaço
pode ser formalizada no conceito de*
AUTOCORRELAÇÃO ESPACIAL

FOCO DESTA AULA

*A relação entre eventos consecutivos no tempo
pode ser formalizada no conceito de*
AUTOCORRELAÇÃO TEMPORAL

PRIMEIRA LEI DA GEOGRAFIA

“Lei de Tobler”

*“No mundo, todas as coisas se parecem,
mas coisas mais próximas são mais
parecidas que aquelas mais distantes”*

(Waldo Tobler, 1979)

DEFINIÇÃO SUCINTA DE
AUTOCORRELAÇÃO ESPACIAL
(positiva)

HETEROGENEIDADE ESPACIAL

- Tendência de lugares de serem diferentes uns dos outros.
- Heterogeneidade maior com o aumento da distância
- Alguns fenômenos geográficos **variam de maneira gradual através do espaço (estacionariedade)**, enquanto outros podem apresentar **extrema irregularidade, violando a Lei de Tobler**

Morumbi e Paraísopolis: Diferenças Abruptas Lado a Lado (Violação da Lei de Tobler)



Foto: Tuca Vieira

Autocorrelação Espacial

As medidas de autocorrelação espacial procuram lidar simultaneamente com **similaridades** na **localização dos objetos espaciais** e de **seus atributos**

Autocorrelação Positiva (Lei de Tobler): Feições similares em localização também são similares em atributos

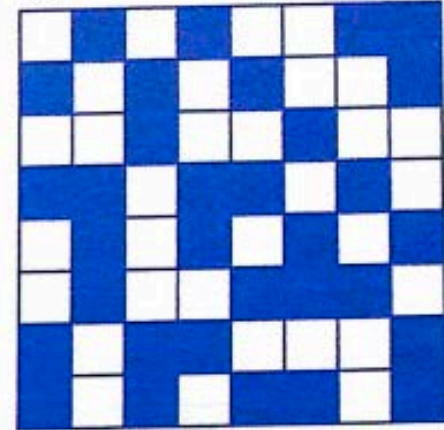
Autocorrelação Negativa (oposição à Lei de Tobler): Feições similares em localização tendem a ter atributos menos similares do que feições mais distantes

Ausência de Autocorrelação: Quando atributos são independentes da localização

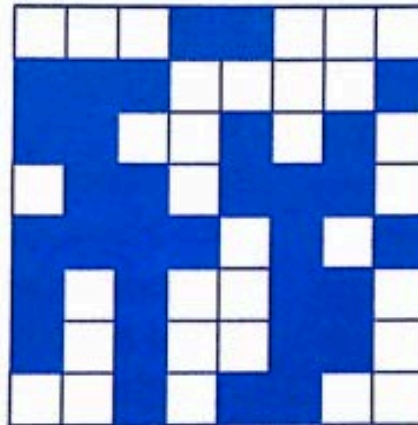
**EXTREMA
AUTOCORRELAÇÃO
NEGATIVA**



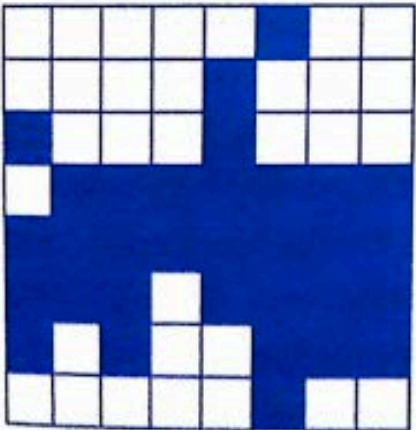
**ARRANJO
DISPERSO**



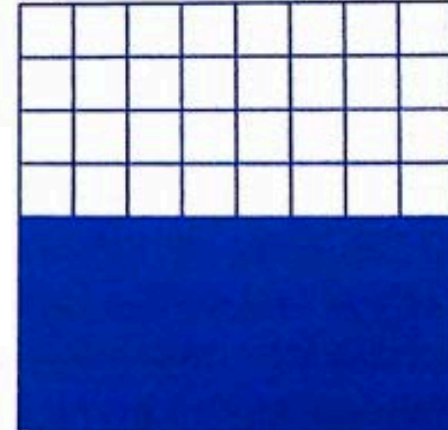
**INDEPENDÊNCIA
ESPACIAL**



**AGRUPAMENTO
ESPACIAL**



**EXTREMA
AUTOCORRELAÇÃO
ESPACIAL POSITIVA**



A Natureza dos Dados Espaciais

- **As medidas de autocorrelação espacial** nos informam sobre as interrelações dos fenômenos através do espaço, **um atributo por vez.**
- Outra faceta importante da natureza dos dados geográficos é a **tendência de existirem relações entre diversos fenômenos no mesmo local (2 ou mais variáveis).**

A interrelação das várias propriedades de um local é um aspecto importante da natureza dos dados geográficas e é fundamental para o entendimento sobre o mundo real.

NO ENTANTO, ELA TAMBÉM É UMA PROPRIEDADE QUE DESAFIA A ANÁLISE DA ESTATÍSTICA CONVENCIONAL, POIS A MAIORIA DOS MÉTODOS ASSUME INDEPENDÊNCIA DAS OBSERVAÇÕES (**CORRELAÇÃO ESPACIAL = ZERO**)

MATRIZ DE VIZINHANÇA

Para comparar a similaridade de atributos geográficos em feições vizinhas é importante estabelecer um **CRITÉRIO DE VIZINHANÇA**.

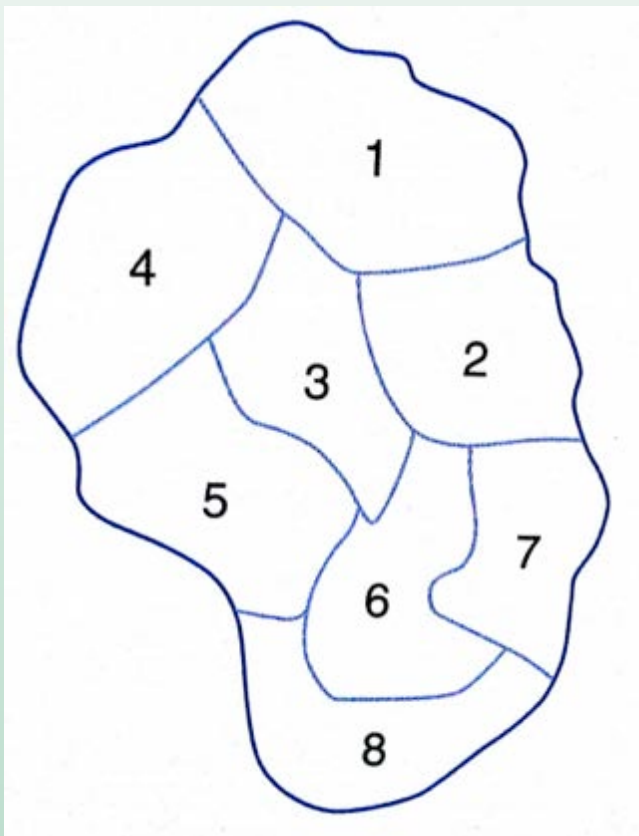
Baseado neste critério, precisamos definir uma matriz de pesos **W** na qual cada elemento w_{ij} mede a proximidade/vizinhança entre i e j (i identificando a linha; j identificando a coluna da matriz)

Um critério de vizinhança comumente adotado: **CONTIGUIDADE**

$w_{ij} = 1$, se as regiões i e j são contíguas (ou seja, são vizinhas).

$w_{ij} = 0$, caso contrário.

MATRIZ DE VIZINHANÇA



	1	2	3	4	5	6	7	8
1	0	1	1	1	0	0	0	0
2	1	0	1	0	0	1	1	0
3	1	1	0	1	1	1	0	0
4	1	0	1	0	1	0	0	0
5	0	0	1	1	0	1	0	1
6	0	1	1	0	1	0	1	1
7	0	1	0	0	0	1	0	1
8	0	0	0	0	1	1	1	0

MATRIZ DE PESOS

A matriz pode ser binária (1 ou 0, como no exemplo) ou geral (pesos contínuos, dependendo da distância – mais próximo com peso maior, é “mais vizinho”).

A matriz também pode ser normalizada.

MATRIZ DE VIZINHANÇA

MATRIZ NORMALIZADA

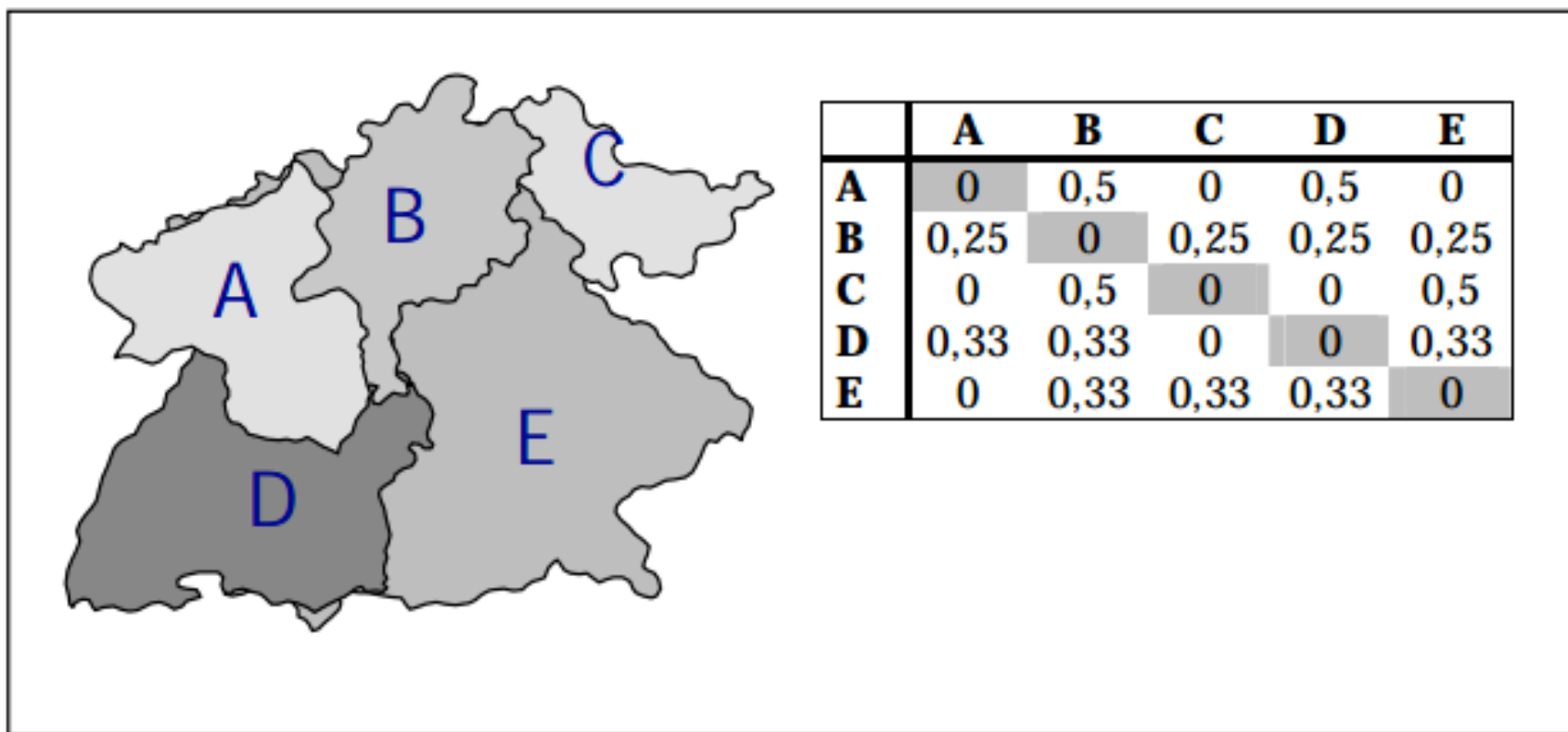
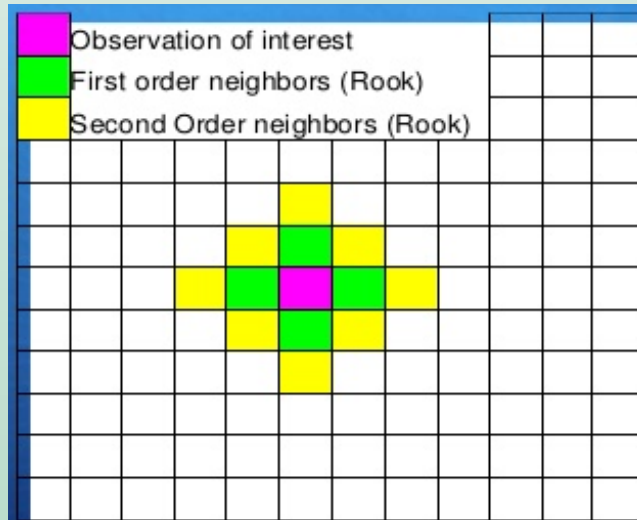


Figura 5-8- Matriz de proximidade espacial de primeira ordem, normalizada pelas linhas.

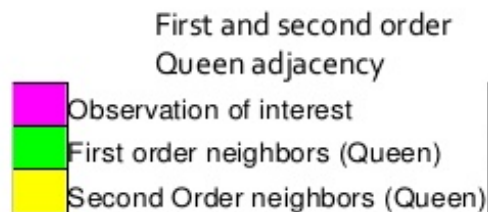
CRITÉRIOS DE VIZINHANÇA

▪ **CONTIGUIDADE: Rook (torre) e Queen (rainha)**
(para dados de “área”)

▪ **Ordem de contiguidade**

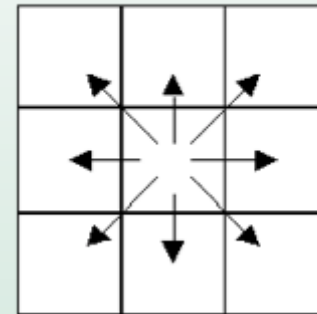


First and second order
Rook adjacency

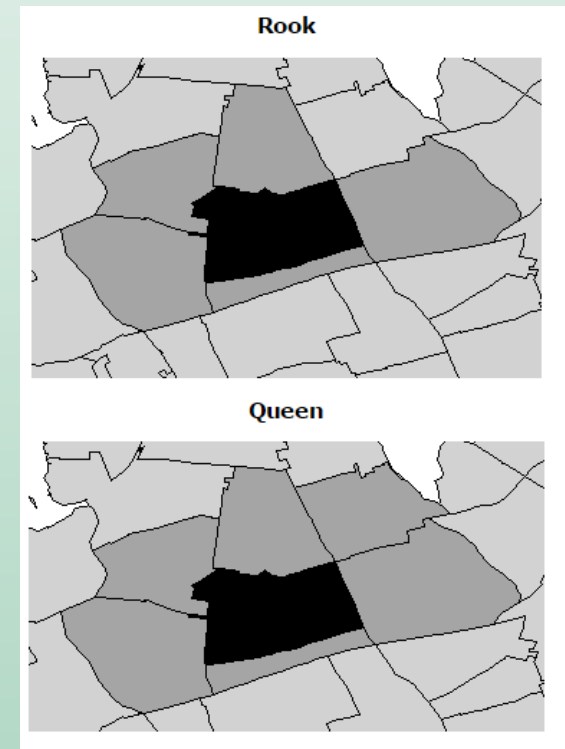
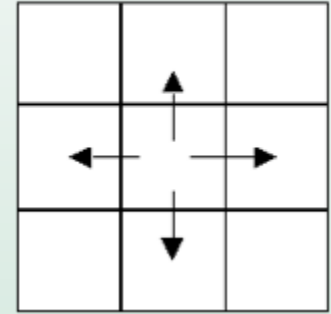


First and second order
Queen adjacency

Queen's (Kings) Case



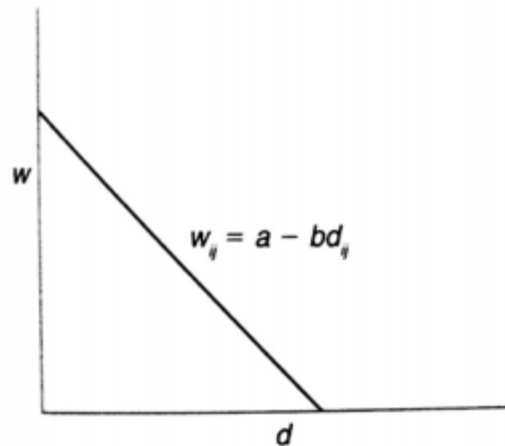
Rooks Case



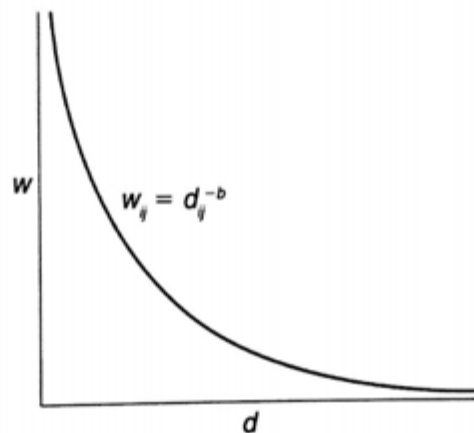
CRITÉRIOS DE VIZINHANÇA

■ DISTÂNCIA:

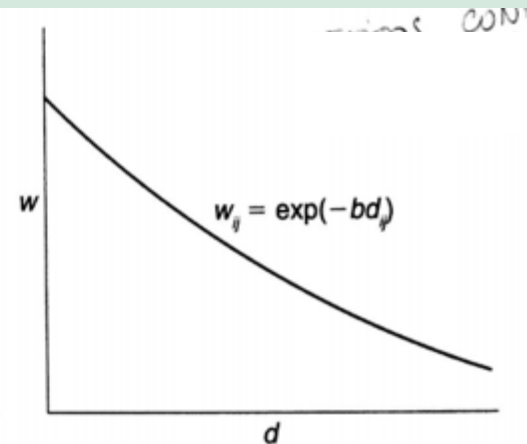
- “n” vizinhos mais próximos;
- distância máxima;
- funções de distância (w_{ij} com valores contínuos)



DECAIMENTO LINEAR



DECAIMENTO
EXPONENCIAL NEGATIVO
DA DISTÂNCIA



DECAIMENTO NEGATIVO
EXPONENCIAL DA
DISTÂNCIA

MATRIZ DE VIZINHANÇA

O critério adotado para construir uma matriz de vizinhança revela a **estrutura espacial** considerada na análise, e pode alterar os resultados das medidas de autocorrelação espacial.

Portanto, vale a pena **testar alguns critérios alternativos para a definição de “vizinhança”** (contiguidade, distância...) e, conseqüentemente, para a definição da **matriz de vizinhança** - também conhecida como *matriz de pesos* ou *matriz de proximidade espacial*.

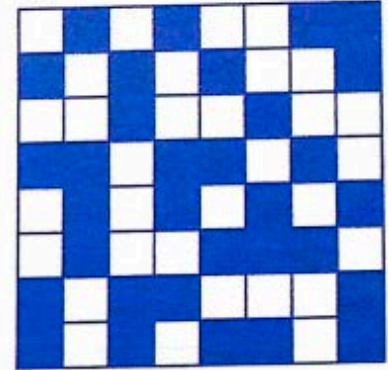
ÍNDICES DE AUTOCORRELAÇÃO ESPACIAL

(A) **EXTREMA AUTOCORRELAÇÃO NEGATIVA**



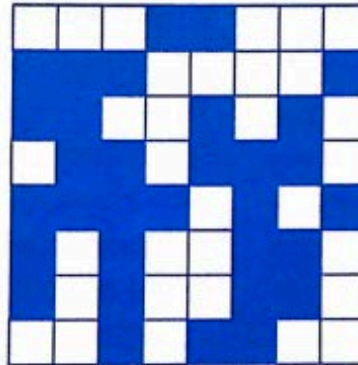
$$I = -1,000$$

(B) **ARRANJO DISPERSO**



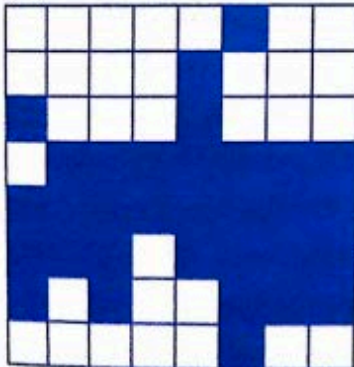
$$I = -0,393$$

(C) **INDEPENDÊNCIA ESPACIAL**



$$I = 0,000$$

(D) **AGRUPAMENTO ESPACIAL**



$$I = +0,393$$

(E) **EXTREMA AUTOCORRELAÇÃO ESPACIAL POSITIVA**



$$I = +0,857$$

ÍNDICES DE AUTOCORRELAÇÃO ESPACIAL

1. Índices Globais de Associação Espacial

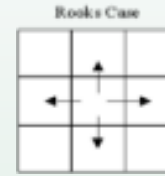
- Apresenta uma medida única para toda a área analisada.
- Índice de Moran (I), Índice de Geary

2. Índices Locais de Associação Espacial (LISA)

- Decomposições dos índices globais, podem ser visualizados na forma de mapas. Permite a identificação de diferentes regimes de associação espacial
- Índice local de Moran (I_i)

ÍNDICES DE AUTOCORRELAÇÃO ESPACIAL

VIZINHANÇA DO
TIPO “ROOK” (torre)

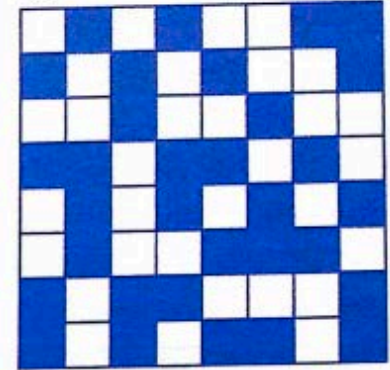


(A) EXTREMA AUTOCORRELAÇÃO NEGATIVA



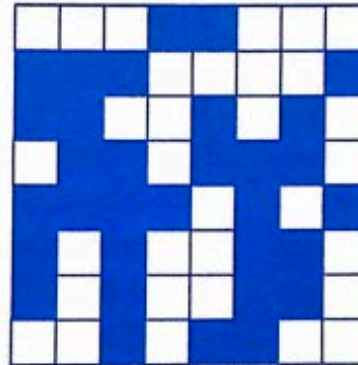
$$\begin{aligned}n_{BW} &= 112 \\n_{BB} &= 0 \\n_{WW} &= 0\end{aligned}$$

(B) ARRANJO DISPERSO



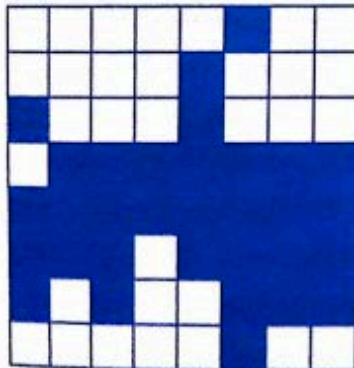
$$\begin{aligned}n_{BW} &= 78 \\n_{BB} &= 16 \\n_{WW} &= 18\end{aligned}$$

(C) INDEPENDÊNCIA ESPACIAL



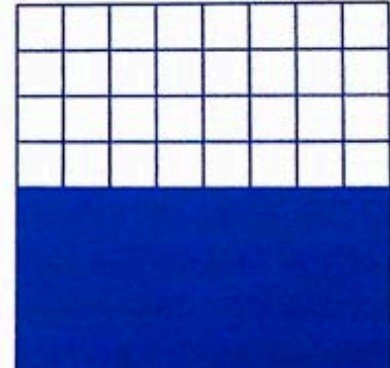
$$\begin{aligned}n_{BW} &= 56 \\n_{BB} &= 30 \\n_{WW} &= 26\end{aligned}$$

(D) AGRUPAMENTO ESPACIAL



$$\begin{aligned}n_{BW} &= 34 \\n_{BB} &= 42 \\n_{WW} &= 36\end{aligned}$$

EXTREMA
AUTOCORRELAÇÃO
ESPACIAL POSITIVA
(E)



$$\begin{aligned}n_{BW} &= 8 \\n_{BB} &= 52 \\n_{WW} &= 52\end{aligned}$$

ÍNDICE GLOBAL DE MORAN

É um índice global de autocorrelação espacial, que varia entre -1 e 1

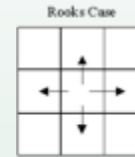
$I = 1 \rightarrow$ Extrema Autocorrelação Positiva (Lei de Tobler):
Feições similares em localização também são similares em atributos

$I = -1 \rightarrow$ Extrema Autocorrelação Negativa (oposição à Lei de Tobler): Feições similares em localização tendem a ter atributos menos similares do que feições mais distantes

$I = 0 \rightarrow$ Ausência de Autocorrelação: Quando atributos são independentes da localização

ÍNDICES DE AUTOCORRELAÇÃO ESPACIAL

VIZINHANÇA DO
TIPO “ROOK” (torre)

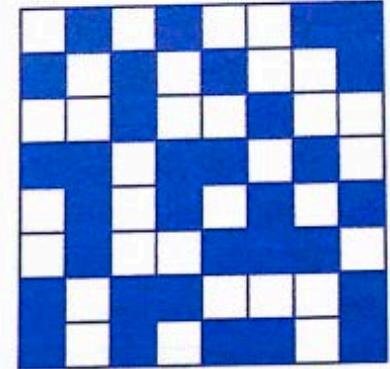


(A) EXTREMA AUTOCORRELAÇÃO NEGATIVA



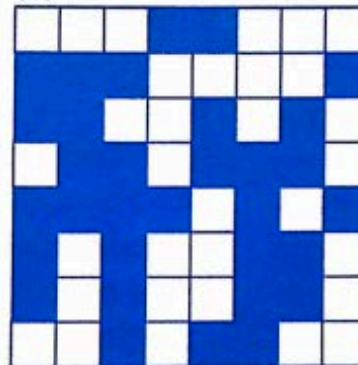
$$I = -1,000$$
$$n_{BW} = 112$$
$$n_{BB} = 0$$
$$n_{WW} = 0$$

(B) ARRANJO DISPERSO



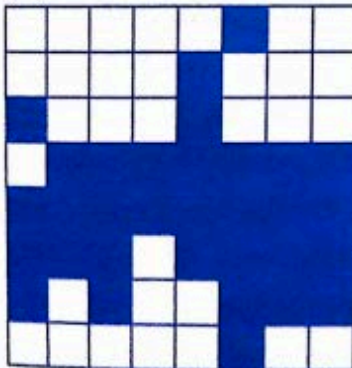
$$I = -0,393$$
$$n_{BW} = 78$$
$$n_{BB} = 16$$
$$n_{WW} = 18$$

(C) INDEPENDÊNCIA ESPACIAL



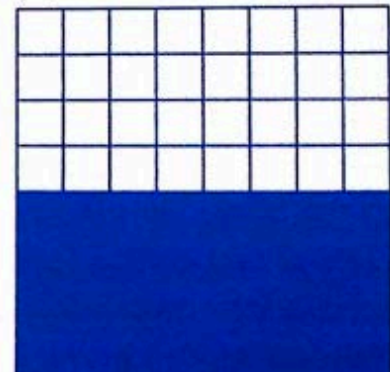
$$I = 0,000$$
$$n_{BW} = 56$$
$$n_{BB} = 30$$
$$n_{WW} = 26$$

(D) AGRUPAMENTO ESPACIAL



$$I = +0,393$$
$$n_{BW} = 34$$
$$n_{BB} = 42$$
$$n_{WW} = 36$$

EXTREMA
AUTOCORRELAÇÃO
ESPACIAL POSITIVA
(E)



$$I = +0,857$$
$$n_{BW} = 8$$
$$n_{BB} = 52$$
$$n_{WW} = 52$$

ÍNDICE GLOBAL DE MORAN

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \cdot \frac{\sum_i \sum_j w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_i (y_i - \bar{y})^2}$$

w_{ij} é o peso, $w_{ij}=1$ se observações i e j são vizinhas.
 y_i and \bar{y} representam o valor da variável na localização i
e a média da variável, respectivamente.

n é o número total de observações

$I \rightarrow$ Índice de Moran: métrica usada para testar a hipótese sobre autocorrelação espacial

Estruturado de maneira semelhante ao coeficiente de correlação de Pearson:
uma medida padronizada de covariância

Índice
Global de
Moran

$$I = \frac{\frac{n}{\sum_i \sum_j w_{ij}} \cdot \sum_i \sum_j w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_i (y_i - \bar{y})^2}$$

Coeficiente
de Correlação
de Pearson

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{(N - 1) s_x s_y}$$

ÍNDICE GLOBAL DE MORAN

De forma geral, o Índice de Moran presta-se a um teste cuja hipótese nula é de independência espacial. Neste caso, seu valor seria ZERO.

Valores positivos (entre 0 e +1) indicam autocorrelação positiva
Valores negativos (entre 0 e -1) indicam autocorrelação negativa.

No exemplo da variável “consumo de água”, obtemos um $I = 0,59$ (vizinhança por contiguidade – queen). Será que este valor representa uma correlação espacial significativa estatisticamente???

Teste de Pseudo Significância

Para estimar a significância do índice, seria preciso associar a este uma distribuição estatística, sendo mais usual relacionar a estatística teste à distribuição normal.

Porém, para evitar pressupostos em relação à distribuição, a abordagem mais comum é um **TESTE DE PSEUDO-SIGNIFICÂNCIA**

Teste de Pseudo Significância

No TESTE DE PSEUDO-SIGNIFICÂNCIA são geradas diferentes permutações dos valores de atributos associados às regiões.

Cada permutação produz 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.

Teste de Pseudo Significância

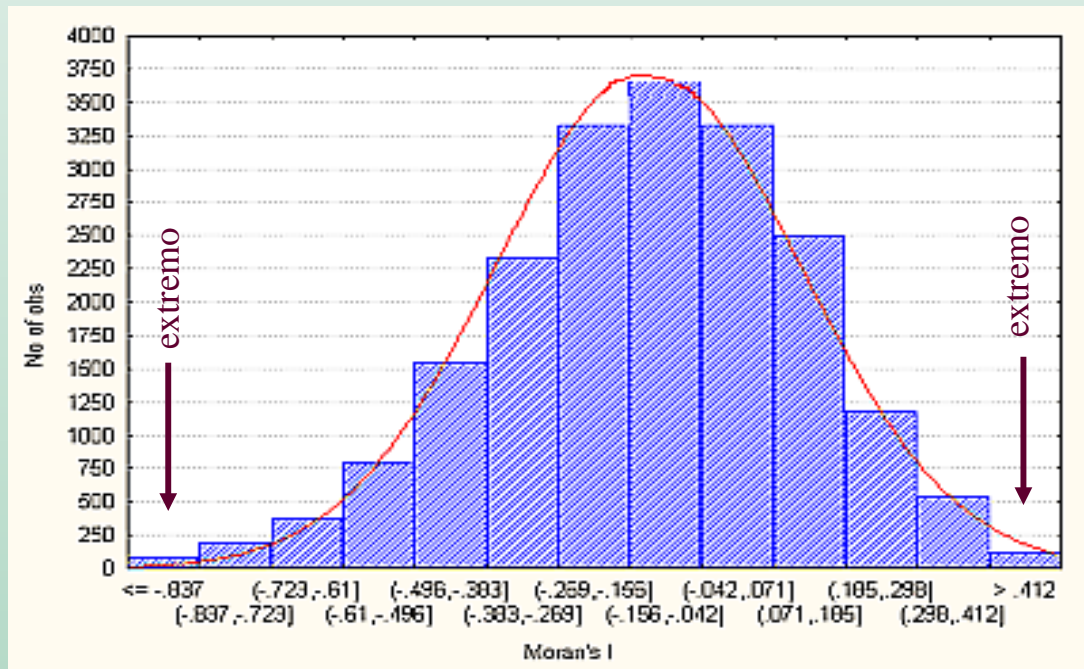
No nosso exemplo, seria como se fizéssemos inúmeras permutações com os valores de consumo de água, atribuindo-os aleatoriamente para os diversos municípios e calculando o Índice de Moran para cada uma destas permutações aleatórias.

Faríamos isso, por exemplo 999 vezes! E obteríamos, portanto, 999 valores para o Índice de Moran + o valor do índice de Moran construído sobre os dados observados. Com estes 1000 valores de I , poderíamos contruir uma distribuição do índice.

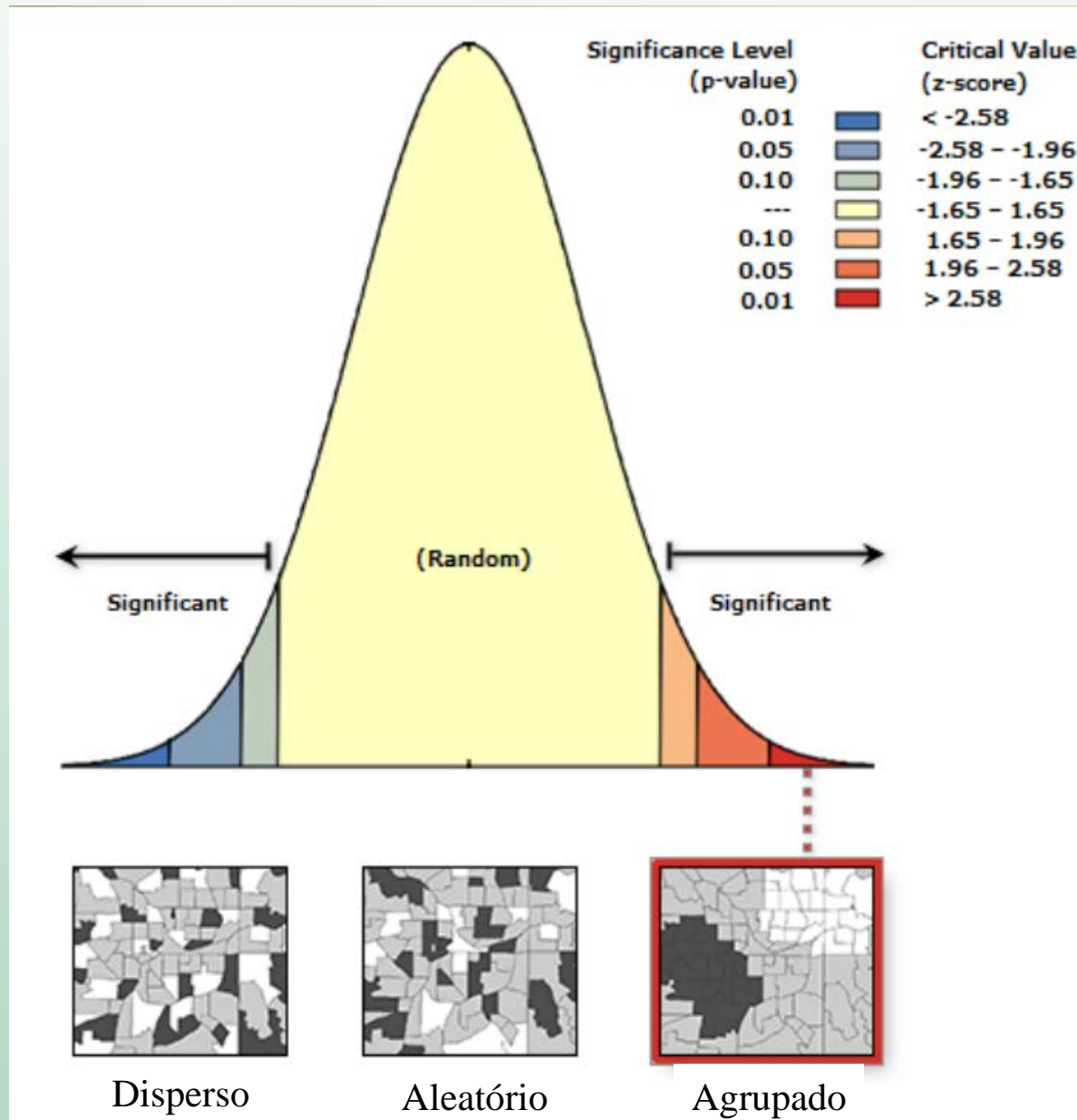
Teste de Pseudo Significância

Se o valor do índice I medido originalmente corresponder a um “extremo” da distribuição simulada, então trata-se de valor com significância estatística.

Distribuição
simulada

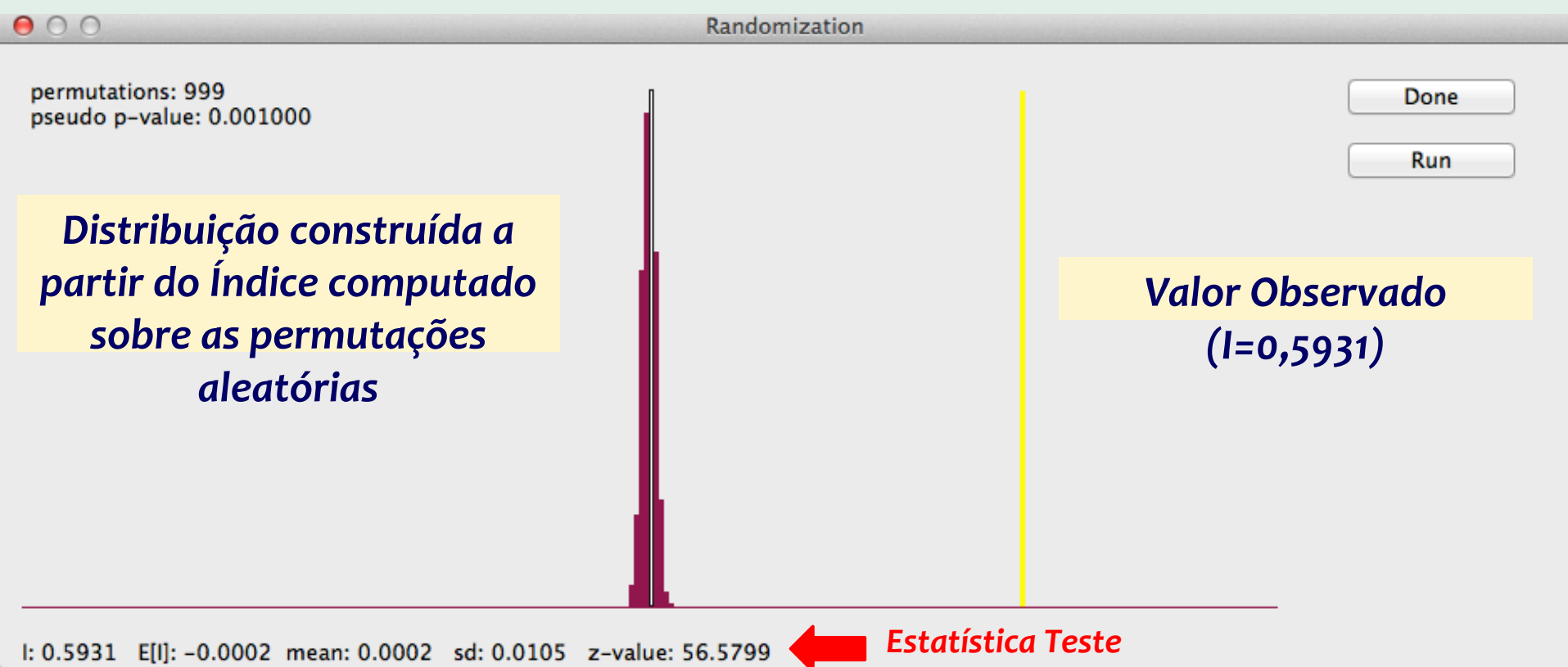


Teste de Pseudo Significância



Teste de Pseudo Significância

No exemplo do consumo de água, obtemos a seguinte distribuição simulada



Como o valor observado está na extremidade ($p\text{-valor} < 0,001$), rejeitamos a hipótese nula de que não há autocorrelação espacial

Diagrama de Espalhamento de Moran

- Maneira adicional de visualizar a dependência espacial
 - Construído com base nos valores padronizados (escore-z)
 - A ideia é comparar os valores padronizados do atributo numa área com a média dos seus vizinhos, construindo um gráfico bidimensional de z (valores padronizados) por wz (média dos vizinhos)

escores-z

$$z = \frac{X - \bar{X}}{s}$$

Diagrama de Espalhamento de Moran

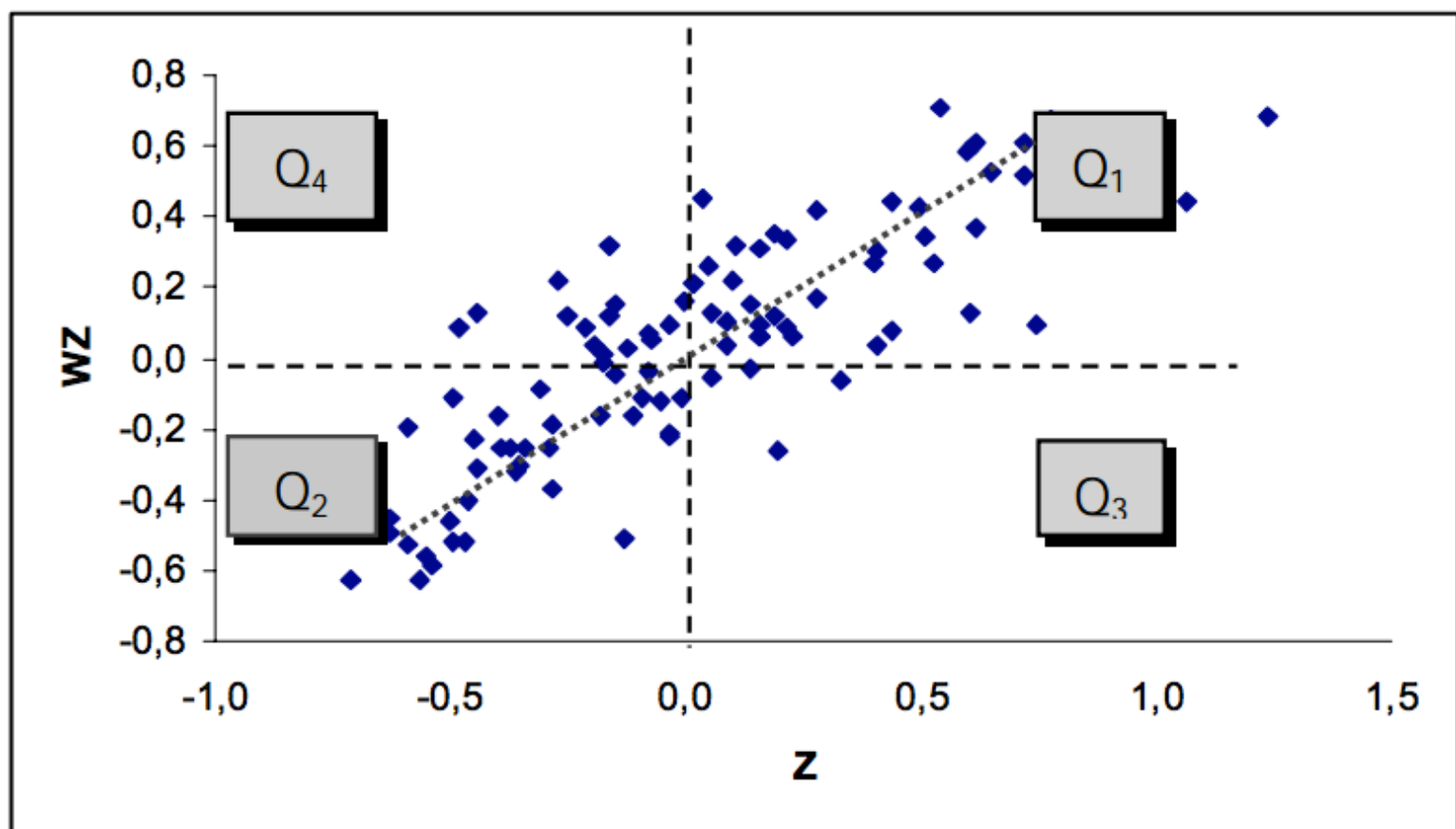
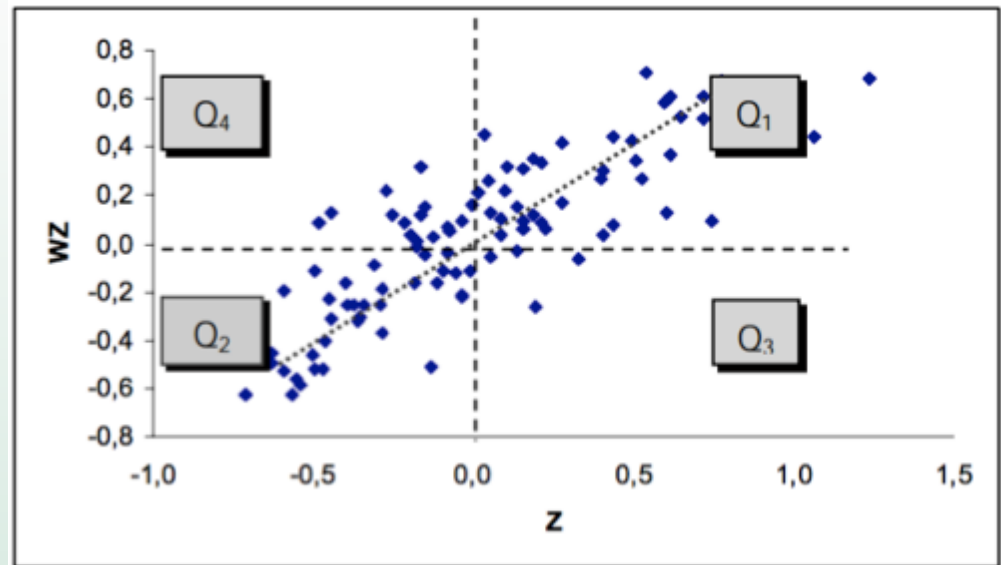


Figura 5-15 Diagrama de Espalhamento de Moran para o índice de exclusão/inclusão social de São Paulo, censo de 1991.

Diagrama de Espalhamento de Moran



É dividido em 4 quadrantes:

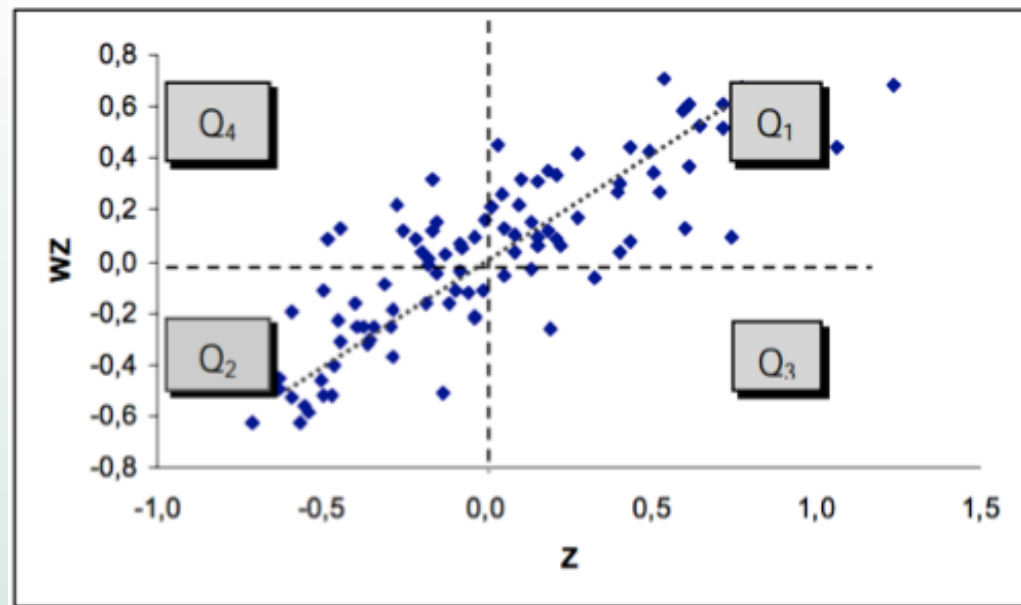
- **Q_1** (valores positivos, médias positivas) e **Q_2** (valores negativos, médias negativas):

indicam pontos de associação espacial positiva, no sentido que uma localização possui vizinhos com valores semelhantes.

- **Q_3** (valores positivos, médias negativas) e **Q_4** (valores negativos, médias positivas):

indicam pontos de associação espacial negativa, no sentido que uma localização possui vizinhos com valores distintos.

Diagrama de Espalhamento de Moran

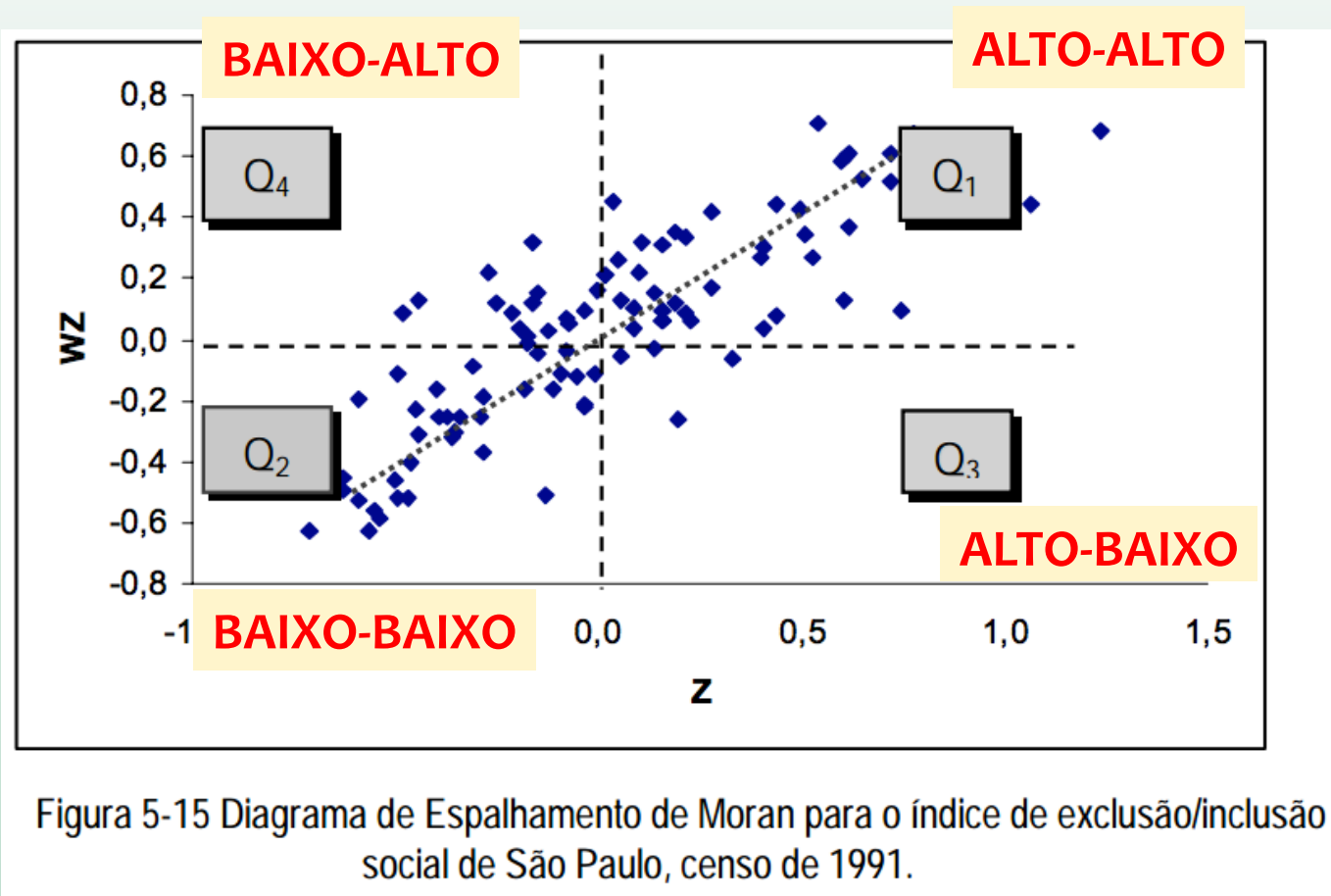


Os pontos em Q_3 e Q_4 podem ser vistos como extremos:

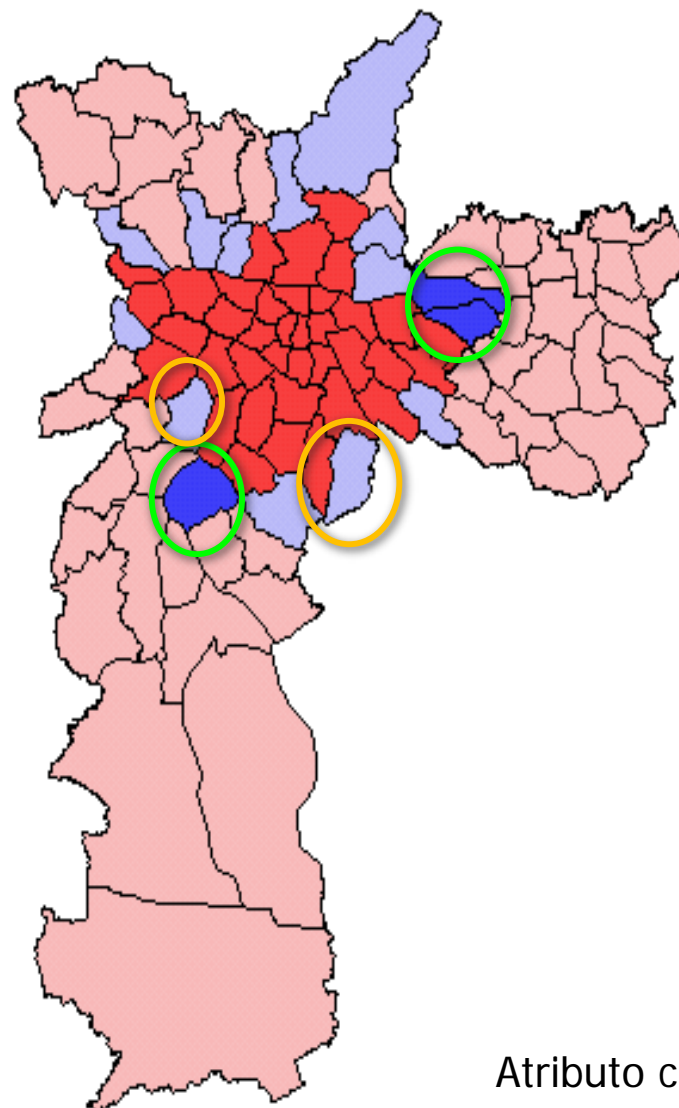
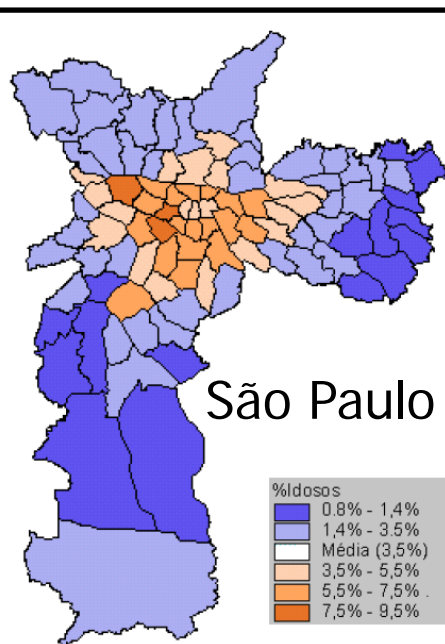
- afastados da reta de regressão linear
- regiões que não seguem o mesmo processo de dependência espacial das demais observações.

Marcam **regiões de transição** entre regimes espaciais distintos.

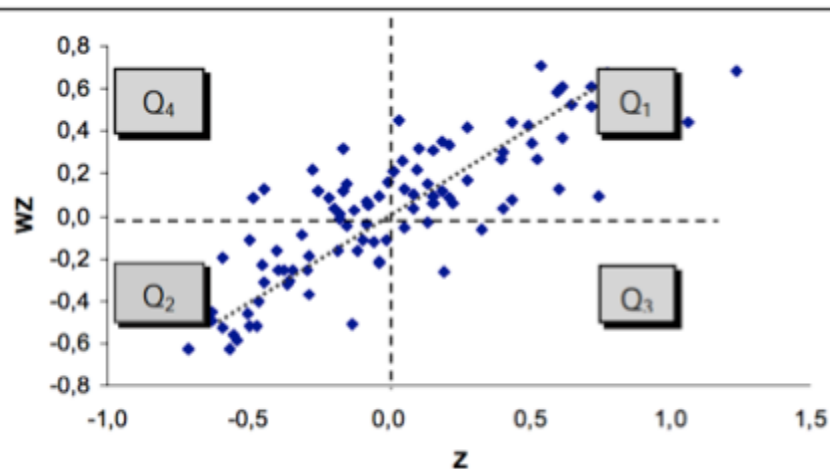
Diagrama de Espalhamento de Moran



O Diagrama também pode ser apresentado na forma de mapa temático, no qual cada polígono é apresentado indicando-se seu quadrante no diagrama de espalhamento



Atributo considerado
percentagem de idosos



Indicadores Locais de Associação Espacial (LISA)

- O estimador de autocorrelação espacial, Moran (I), fornece um **valor único** como medida da associação espacial. Por isso o chamamos de **INDICADOR GLOBAL**.
- No entanto, muitas vezes é necessário examinar padrões numa escala maior (mais detalhada).
 - Neste caso, utilizamos **INDICADORES LOCAIS DE ASSOCIAÇÃO ESPACIAL (LISA)** que possam ser associados a diferentes localizações de uma variável distribuída espacialmente.
 - A utilização destes indicadores, em conjunto com os indicadores globais, refinam nosso conhecimento sobre o processos que dão origem a dependência espacial.

Indicadores Locais de Associação Espacial (LISA)

- Os indicadores locais de associação especial produzem um valor específico para cada objeto.
- Isto acarreta a identificação de:
 - “Clusters”: objetos com valores de atributos semelhantes,
 - “Outliers”: objetos anômalos,
 - A presença de mais de um regime espacial.

Indicadores Locais de Associação Espacial (LISA)

Segundo Anselin (1995), um indicador local de associação espacial (LISA) tem que atender a dois objetivos:

- Permitir a identificação de padrões de associação espacial significativos;
- Ser uma decomposição do índice global de associação espacial.

Indicadores Locais de Associação Espacial (LISA)

Indicadores locais I_i de Moran (Anselin, 1996):

$$I_i = \frac{(y_i - \bar{y}) \sum_{j=1}^n w_{ij} (y_j - \bar{y})}{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}}$$

$I_i > 0$ “clusters” de valores similares (altos ou baixos).

$I_i < 0$ “clusters” de valores distintos (Ex: uma localização com valores altos rodeada por uma vizinhança de valores baixos).

Indicadores Locais de Associação Espacial (LISA)

De forma similar aos indicadores globais, a significância do índice local de Moran (I_i) deve ser avaliada utilizando hipótese de normalidade ou simulação de distribuição por permutação aleatória nos valores dos atributos (Anselin, 1995).

A significância estatística de Moran (I_i) é muito útil gerar um mapa indicando as regiões que apresentam correlação local significativamente diferente do resto dos dados.

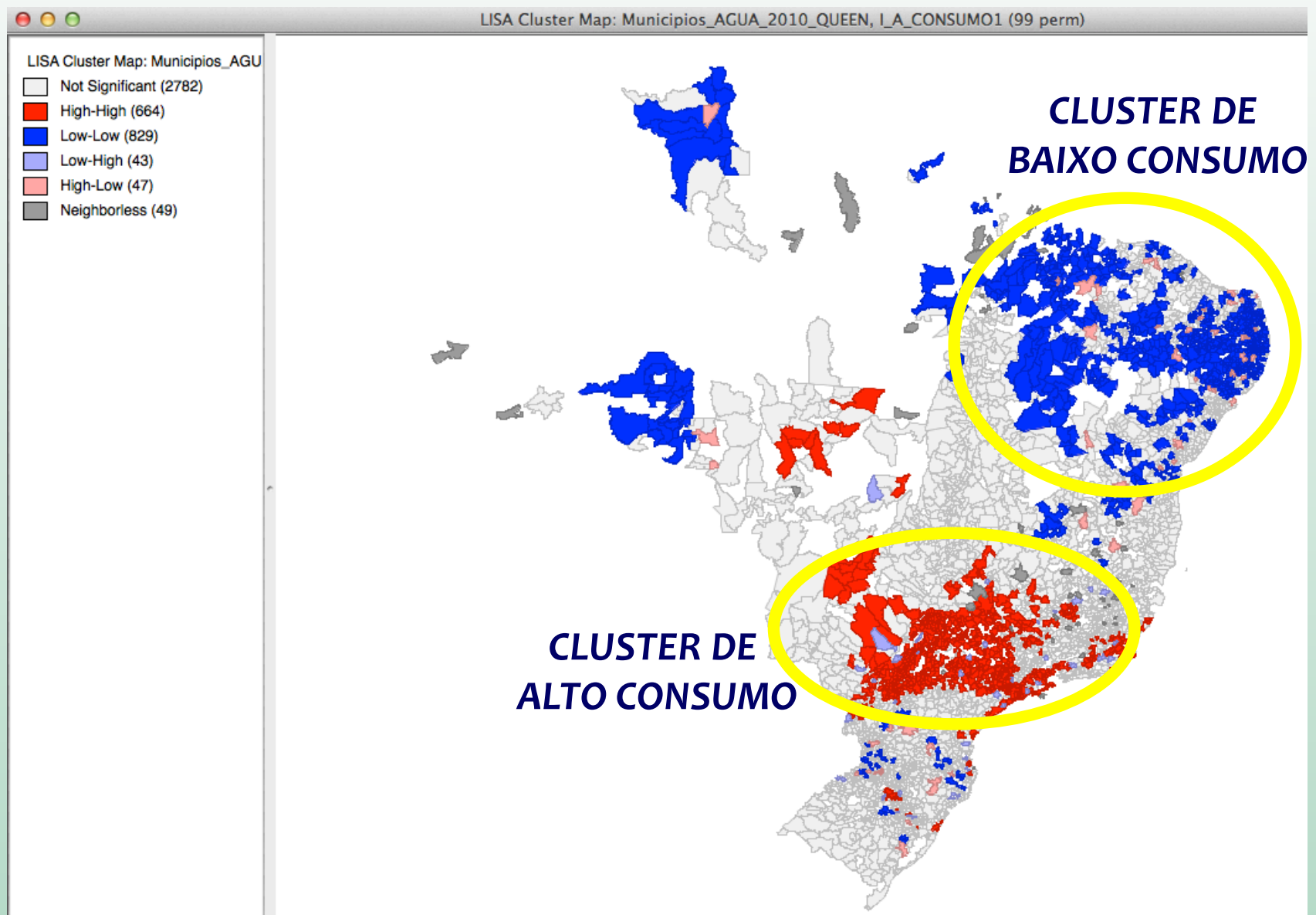
Este mapa é denominado por Anselin (1995) de **“LISA MAP”**.

Indicadores Locais de Associação Espacial (LISA)

No **LISA MAP**, os índices locais I_i são classificados como:

- Não significativos
- Significativos com confiança de 95% ($1,96\sigma$), 99% ($2,58\sigma$) e 99,9% ($3,29\sigma$).

LISA MAP



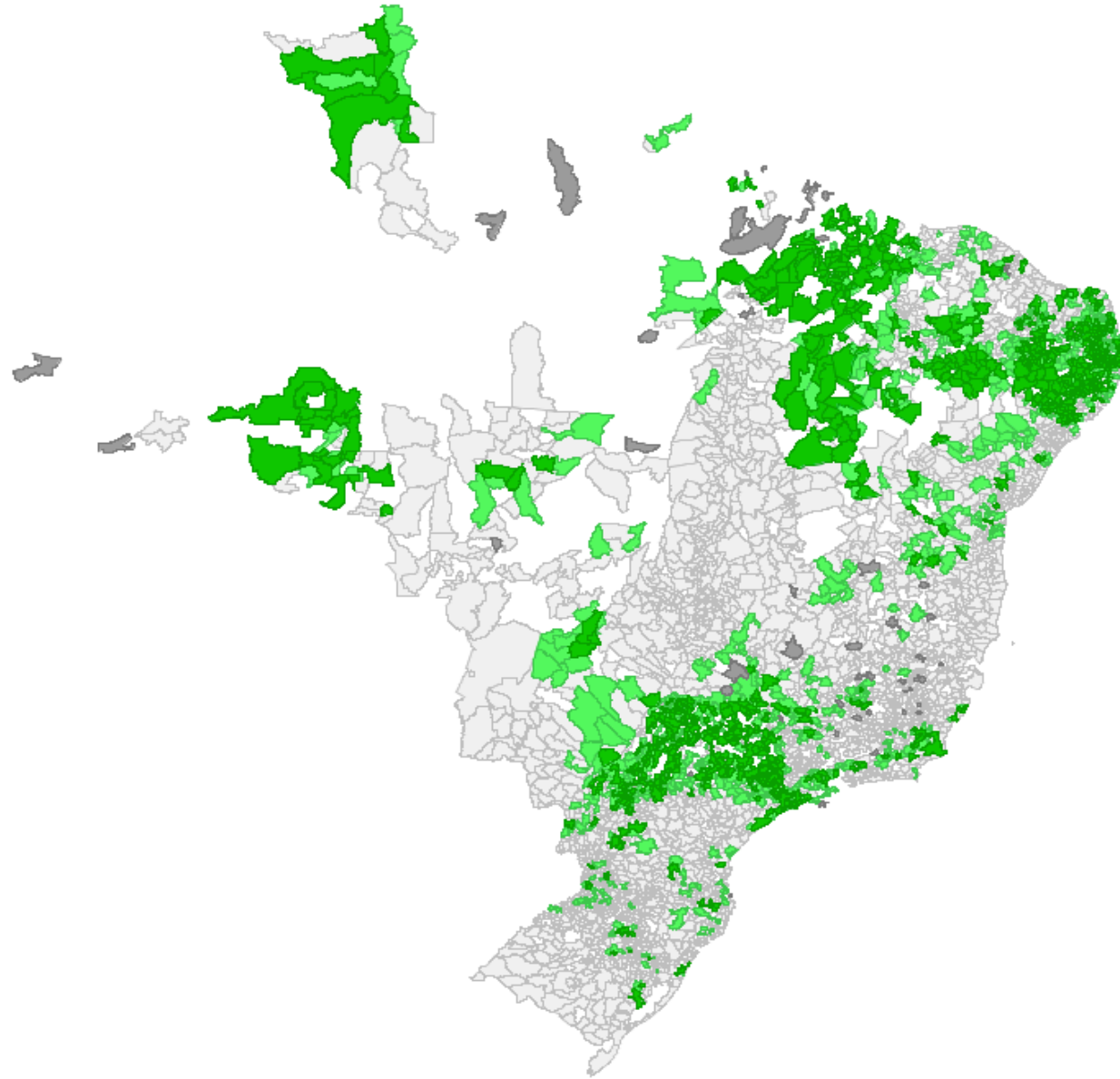
No GeoDa, apenas valores com significância maior que 95% são apresentados

MAPA DE SIGNIFICÂNCIA

LISA Significance Map: Municipios_AGUA_2010_QUEEN, I_A_CONSUMO1 (99 perm)

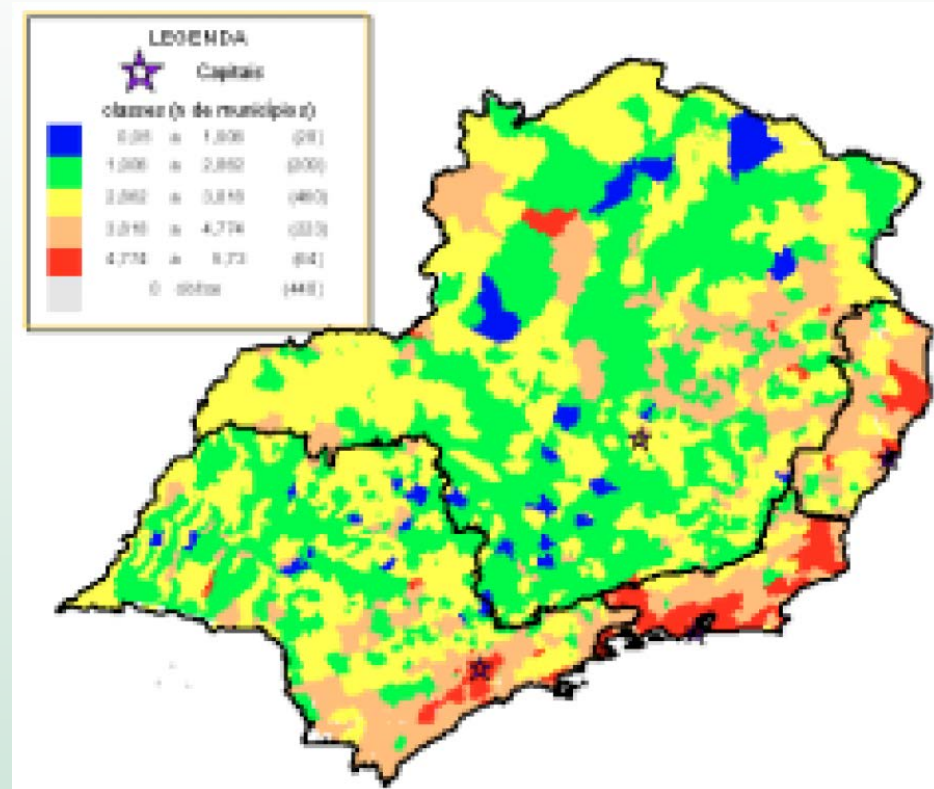
LISA Significance Map: Municipios_AGUA_201

- Not Significant (2782)
- $p = 0.05$ (799)
- $p = 0.01$ (784)
- $p = 0.001$ (0)
- $p = 0.0001$ (0)
- Neighborless (49)

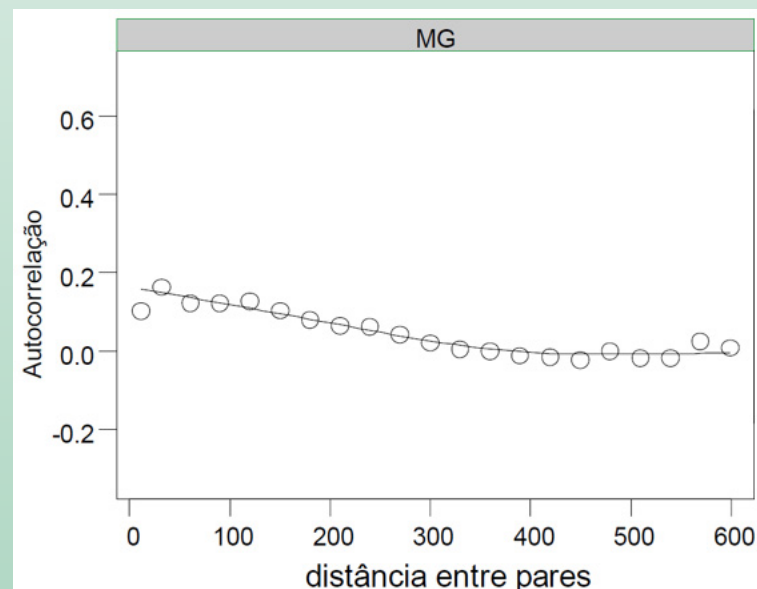
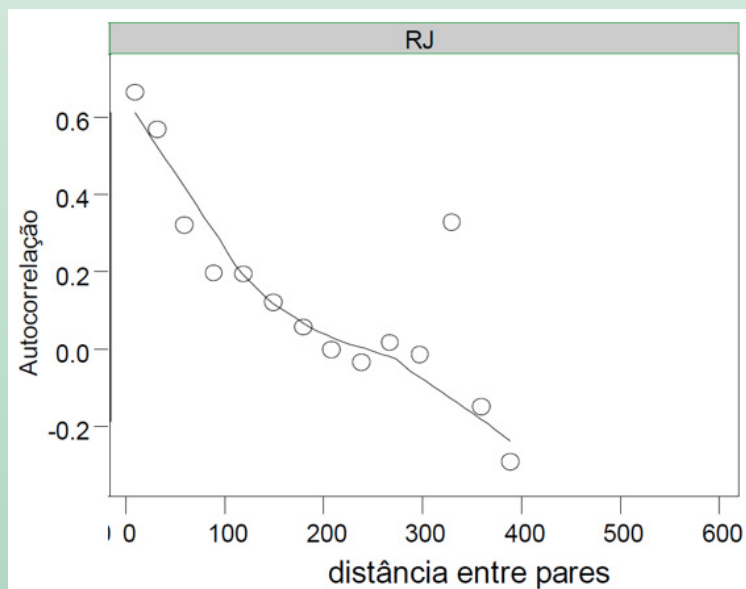


Correlograma

Índice de Mortes por Homicídios

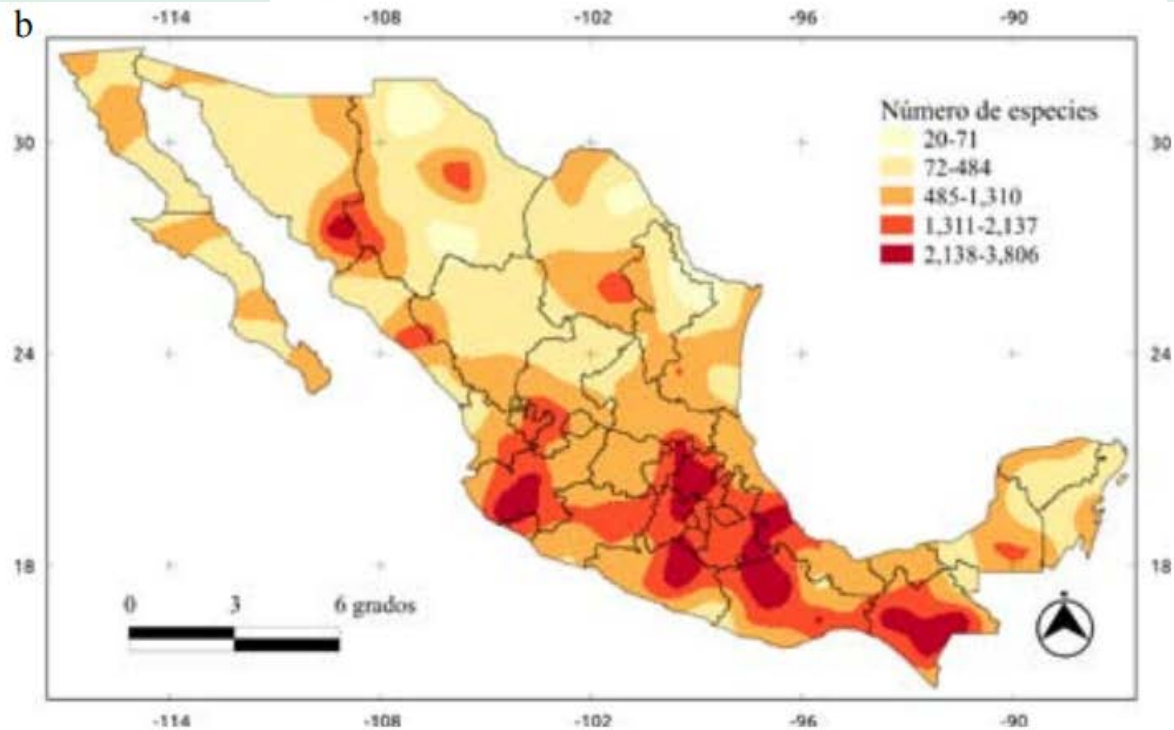
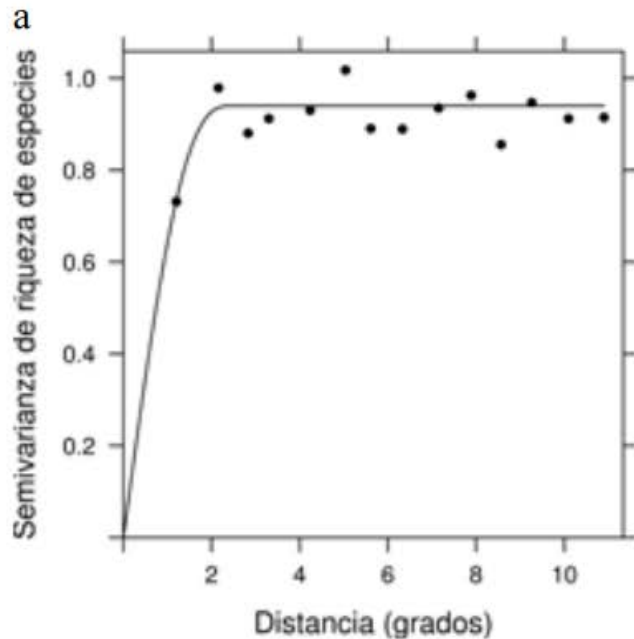
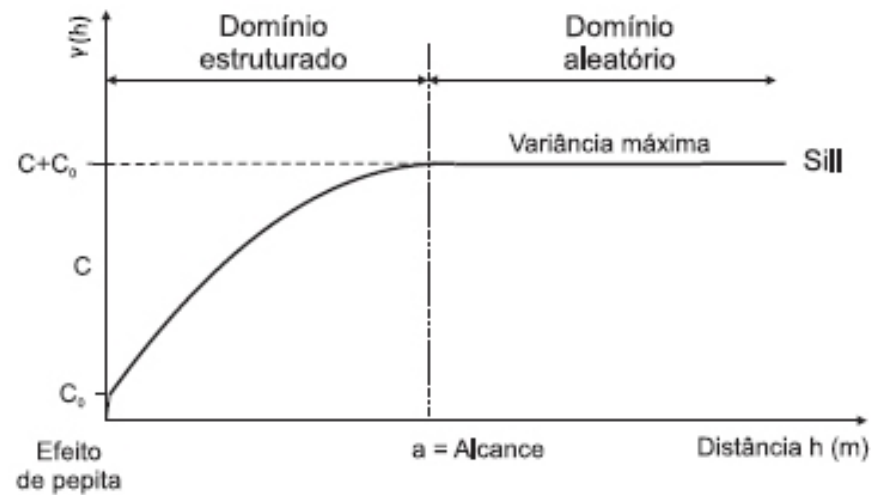


Correlograma



Variograma

C = Variância
Co = Efeito Pepita
C+Co = Patamar
A = Alcance

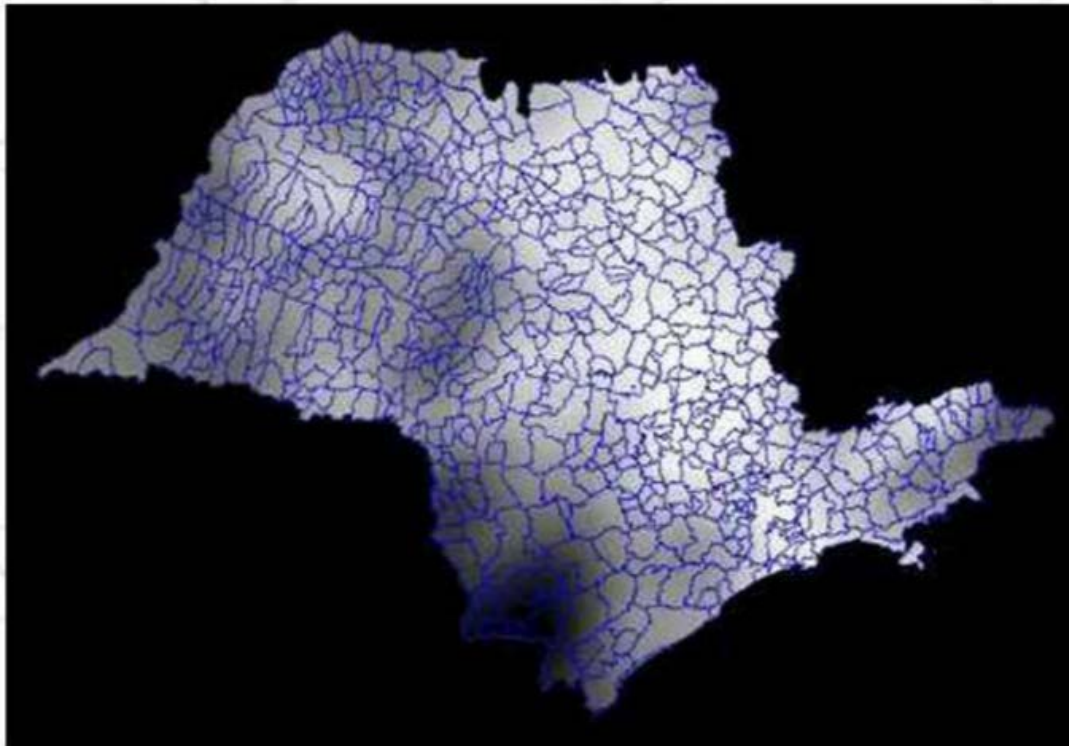


SANTOS, Carlos Eduardo dos y BIONDI, João Carlos. **Utilização de elipsoide de anisotropia variográfica como indicador cinemático em maciços rochosos fragmentados por falhas: o exemplo do depósito de asbestos crisotila cana brava (Minaçu, GO).** *Geol. USP*, 2011, vol.11, n.3, pp. 65-77.

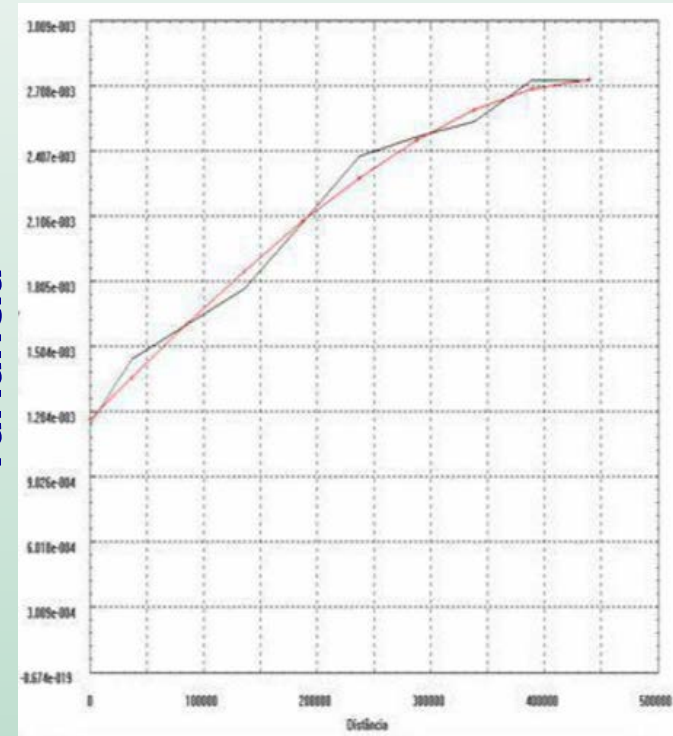
CRUZ-CARDENAS, Gustavo et al . Distribución espacial de la riqueza de especies de plantas vasculares en México. *Rev. Mex. Biodiv.*, México , v. 84, n. 4, p. 1189-1199, 2013 .

Variograma

IDH no Estado de São Paulo



Variância



Distância