CURSO 4

Aplicação de Métodos Estatísticos Espaciais para Análise de Dados Geográficos.

AULA 2 - Analise de Padrões de Áreas.

Dr. Eduardo Camargo
eduardo.camargo@inpe.br

Dr. Carlos Felgueiras
carlos.felgueiras@inpe.br

Sumário

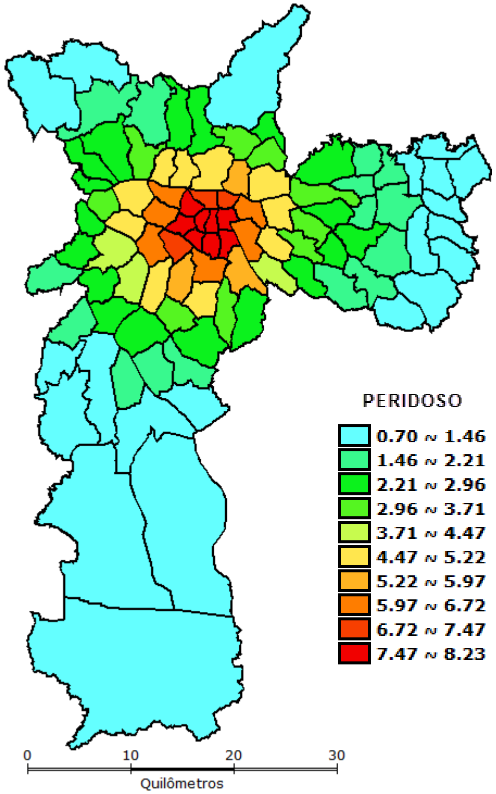
- Introdução
- Técnicas de ESDA
- Matrizes de Proximidade Espacial
- Média Espacial Móvel (μ_i)
- Indicadores Globais de Autocorrelação Espacial
  - Índices Globais de Moran (I), Geary (C) e Getis e Ord (G)
- Indicadores Locais de Associação Espacial (LISA)
  - Índice Local de Moran (I_i)
  - Índices Locais de Getis e Ord (G_i e G_i^*)
- Exemplo Prático com o Programa TerraView

Introdução

Análise de padrão de áreas: distribuição de eventos está associada a áreas.

Objetivo: determinar se há um padrão espacial nos valores agregados por áreas.

Forma usual de apresentação: através de mapas temáticos.



Exemplo: disparidade social

Percentual de idosos na cidade de São Paulo.

Existe algum padrão espacial?

Que fatores explicam essa distribuição?

ESDA: "*Coleção de técnicas para descrever e visualizar distribuições espaciais, identificar situações atípicas, descobrir padrões de associação espacial, clusters e sugerir regimes espaciais ou formas de heterogeneidade espacial*" (Anselin).

1 Visualização de padrão de áreas

técnicas convencionais de visualização cartográfica, estatísticas não-espaciais.

2 Indicadores Globais de Autocorrelação

explorar a dependência espacial, mostrando como os valores estão correlacionados no espaço;

o conceito utilizado é o de autocorrelação espacial;

exemplos de indicadores globais: Moran (I), Geary (C), *Getis e Ord (G).*

3- Indicadores Locais de Associação Espacial (LISA)

Possibilita a identificação de:

clusters: objetos com valores de atributos semelhantes;

outliers: objetos anômalos;

a presença de mais de um regime espacial.

exemplos de indicadores locais: Moran (I_i), Getis e Ord (G_i e G_i^*).

Agrupamento de atributos

- intervalos iguais
- quantis
- estatísticos

componente espacial

MG

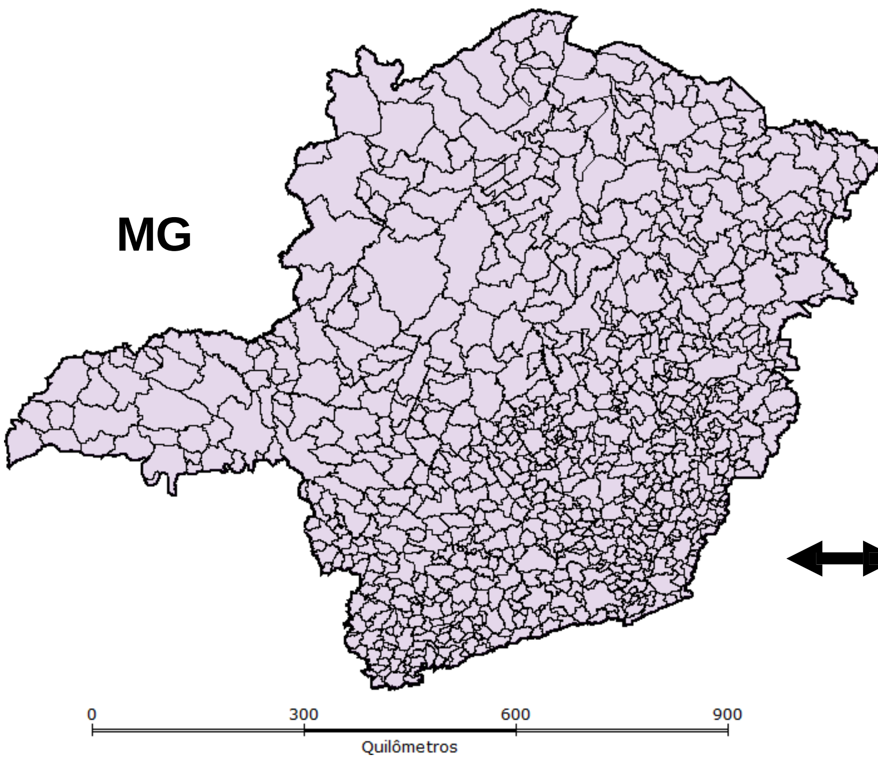


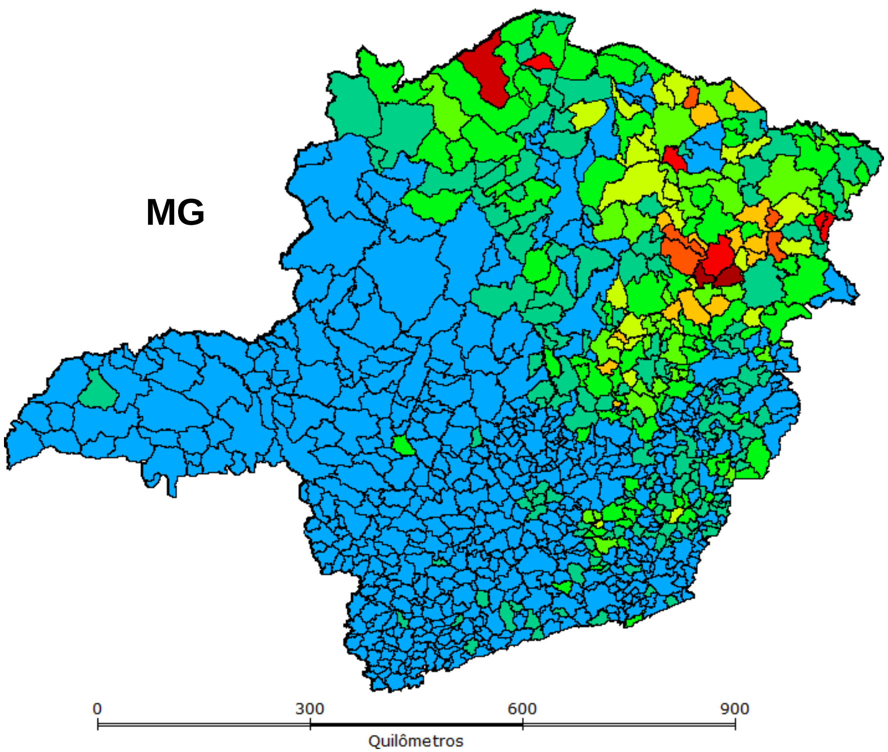
tabela de atributos

	ID	CD_GEOCODM	NM_MUNICIP	PC05
1	697	3107802	BOM JESUS DO GALHO	5.9
2	706	3108552	BRASILÂNDIA DE MINAS	2.1
3	662	3104403	ARGIRITA	4.4
4	663	3104452	ARICANDUVA	16.2
5	664	3104502	ARINOS	7.6
6	665	3104601	ASTOLFO DUTRA	0.2
7	666	3104700	ATALÉIA	12
8	667	3104809	AUGUSTO DE LIMA	6.3
9	668	3104908	BAEPENDI	3.2
10	669	3105004	BALDIM	2.4
11	670	3105103	BAMBUÍ	1.4
12	671	3105202	BANDEIRA	15.6

Agrupamento por intervalos iguais

- definidos pelos valores máximo e mínimo;
- mostram a dispersão nos dados;
- outliers podem mascarar diferenças.

Ex: proporção de crianças de 0 a 5 anos de idade residentes em domicílios particulares permanentes com responsável ou conjugue analfabeto e saneamento inadequado.



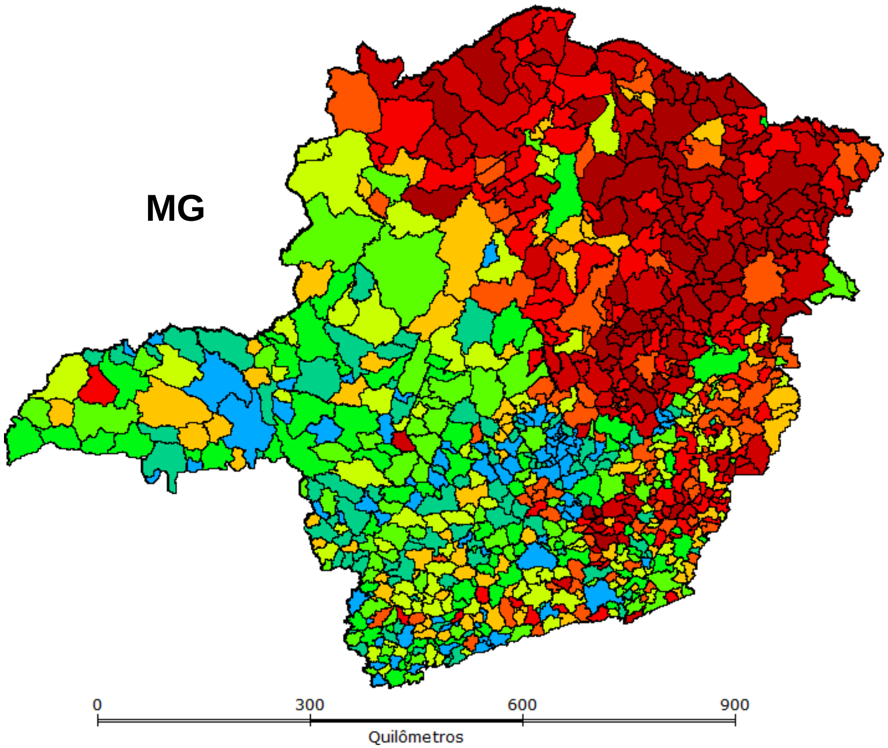
Cor	De	Para	Rótulo	Quantidade
	0.00	4.06	-0.01 ~ 4.06	532
	4.06	8.12	4.06 ~ 8.12	154
	8.12	12.18	8.12 ~ 12.18	83
	12.18	16.24	12.18 ~ 16.24	34
	16.24	20.30	16.24 ~ 20.30	24
	20.30	24.36	20.30 ~ 24.36	13
	24.36	28.42	24.36 ~ 28.42	5
	28.42	32.48	28.42 ~ 32.48	5
	32.48	36.54	32.48 ~ 36.54	1
	36.54	40.61	36.54 ~ 40.61	2

Fonte: IBGE, censo 2010.

Agrupamento por quantis

- cada agrupamento contém aproximadamente o número igual de elementos;
- conceito de ordenação;
- por exemplo, os 25% melhores e os 25% piores.

Ex: proporção de crianças de 0 a 5 anos de idade residentes em domicílios particulares permanentes com responsável ou conjugue analfabeto e saneamento inadequado.



Cor	De	Para	Rótulo	Quantidade
	0.00	0.30	-0.01 ~ 0.30	81
	0.30	0.70	0.30 ~ 0.70	89
	0.70	1.20	0.70 ~ 1.20	81
	1.20	1.70	1.20 ~ 1.70	83
	1.70	2.50	1.70 ~ 2.50	86
	2.50	3.60	2.50 ~ 3.60	86
	3.60	5.40	3.60 ~ 5.40	88
	5.40	8.00	5.40 ~ 8.00	86
	8.00	12.00	8.00 ~ 12.00	86
	12.00	40.61	12.00 ~ 40.61	87

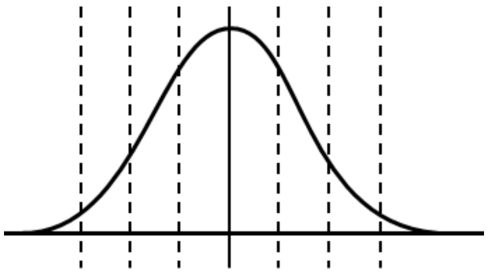
Fonte: IBGE, censo 2010.

Agrupamento estatístico por desvios padrões

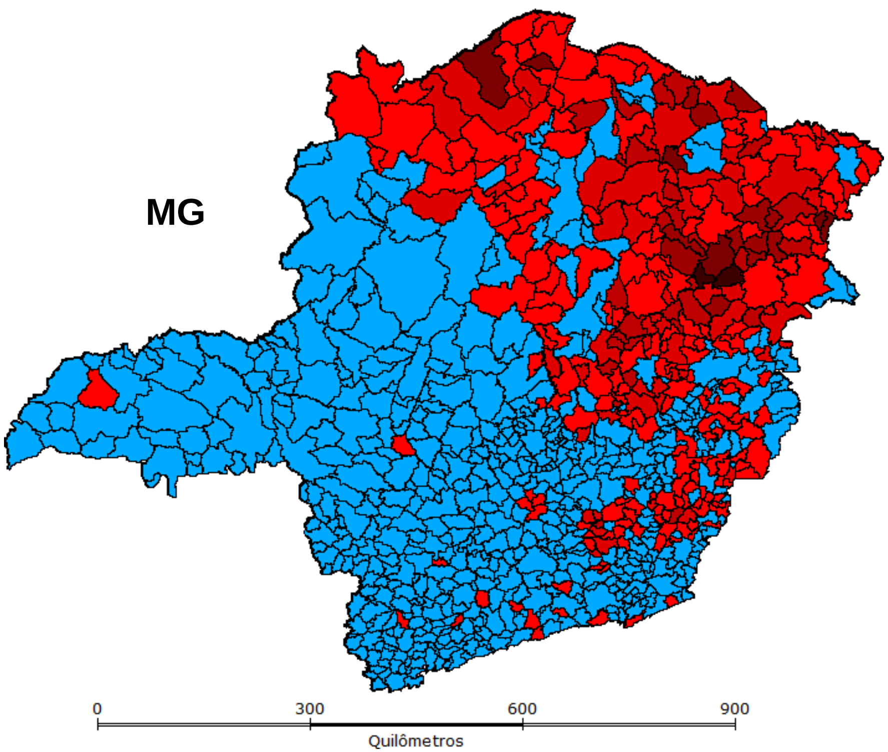
exibe a dispersão em torno da média;

quebras: 1 dp, ½ dp, ¼ dp;

caracteriza o comportamento da variável.



Ex: proporção de crianças de 0 a 5 anos de idade residentes em domicílios particulares permanentes com responsável ou conjugue analfabeto e saneamento inadequado.



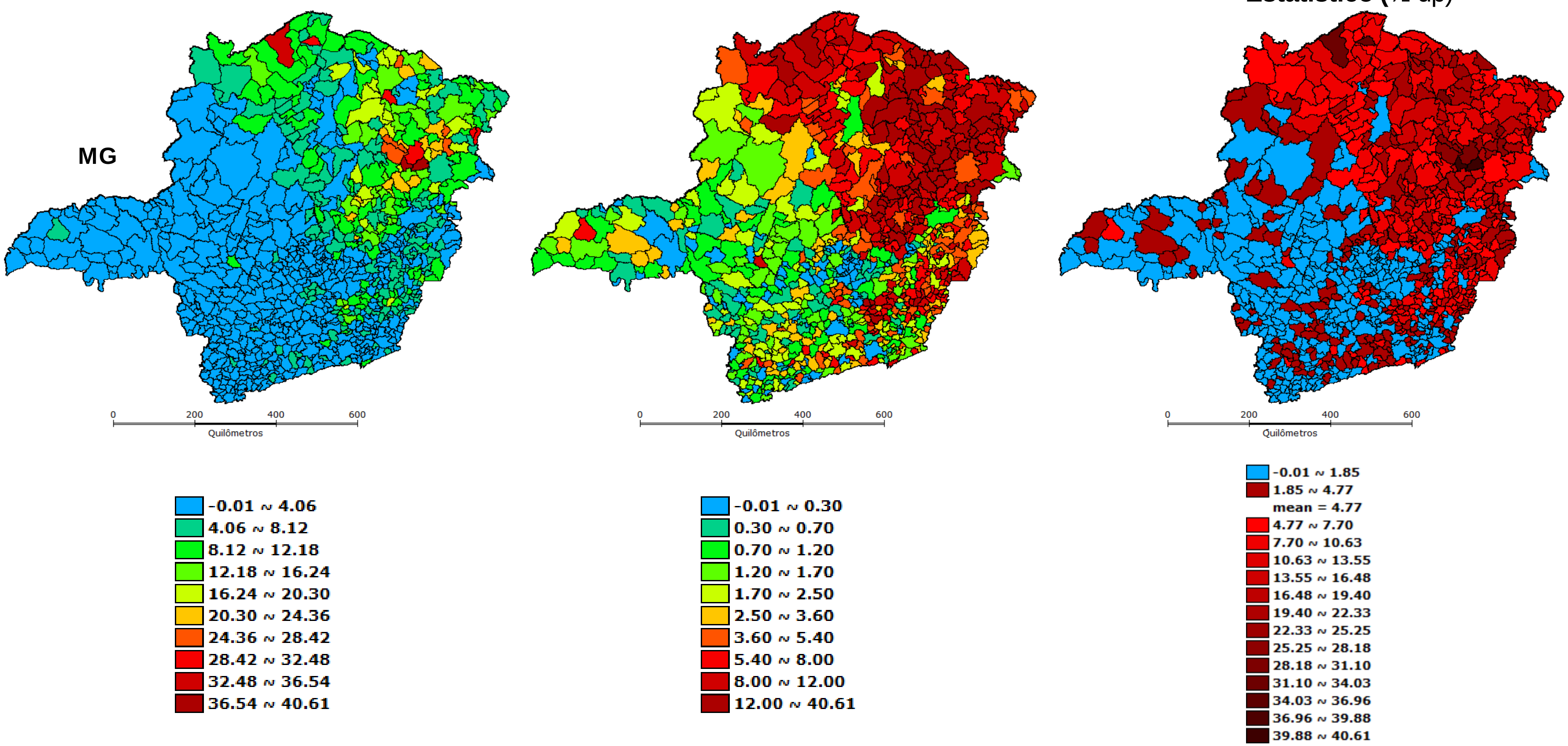
Cor	De	Para	Rótulo	Quantidade
	0.00	4.77	0.00 ~ 4.77	557
	mean = 4.77		mean = 4.77	0
	4.77	10.63	4.77 ~ 10.63	195
	10.63	16.48	10.63 ~ 16.48	54
	16.48	22.33	16.48 ~ 22.33	28
	22.33	28.18	22.33 ~ 28.18	11
	28.18	34.03	28.18 ~ 34.03	6
	34.03	39.88	34.03 ~ 39.88	0
	39.88	40.61	39.88 ~ 40.61	2

Fonte: IBGE, censo 2010.

Visualização de Padrão de Áreas

Cuidados com apresentação: mapas coloridos podem esconder diferenças e levar a várias interpretações sobre os mesmos dados.

Passos Iguais | **Quantis** | **Estatístico** (½ dp)



Proporção de crianças de 0 a 5 anos de idade residentes em domicílios particulares permanentes com responsável ou conjugue analfabeto e saneamento inadequado.

Aplicar os Métodos de Agrupamento de Atributos
- intervalos iguais
- quantis
- estatísticos

Executar o Roteiro Prático

Aula 2 - Análise Padrão de Áreas <u>até a pág. 8</u>

Efeitos de Primeira Ordem

Média Espacial Móvel

Dependência Espacial Global

Efeitos de segunda ordem

Indicadores globais: Moran (I), Geary (C), Getis e Ord (G)

Dependência Espacial Local

LISA (Local Indicators of Spatial Association)

Indicadores locais: Moran Local (I_i), Getis e Ord (G_i e G_i*)

Matrizde Proximidade Espacial (W)

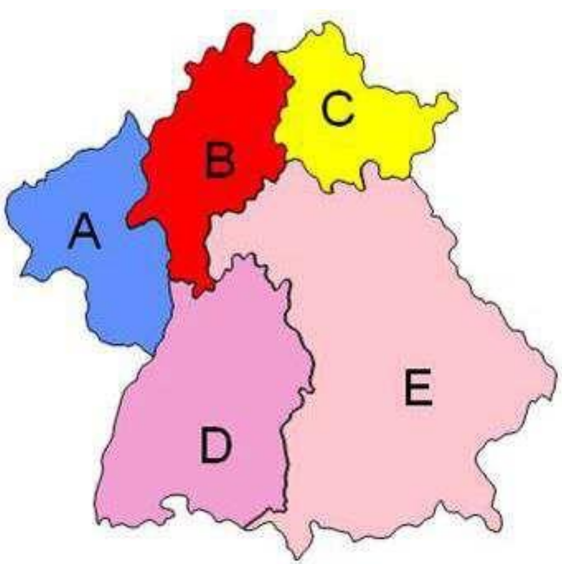
A matriz W

Tem dimensão n x n, em que n é o número de áreas (objetos). Os elementos w_{ij} representam uma medida de proximidade entre objetos O_i e O_j.

Critérios de construção

$w_{ij} = 1$, se O_i toca O_j

$w_{ij} = 1$, se distância entre O_i e $O_j < d$



	A	B	C	D	E
A	0	1	0	1	0
B	1	0	1	1	1
C	0	1	0	0	1
D	1	1	0	0	1
E	0	1	1	1	0

Média Espacial Móvel

A Média Espacial Móvel explora o valor médio (μ_i) do atributo na região de estudo (efeitos de primeira ordem);

Pode ser empregada para mostrar padrões e tendências espaciais;

Seu estimador é definido como:

$$\hat{\mu}_i = \frac{\sum_{j=1}^{n} w_{ij}\, y_i}{\sum_{j=1}^{n} w_{ij}} \quad i = 1, 2, \ldots, n$$

em que:

w_{ij} : são pesos atribuídos conforme a conexão entre as áreas i e j;

y_i : é o valor do atributo em cada área;

n : é o número de áreas (polígonos).

Média Espacial Móvel

Exemplo

$$\hat{\mu}_i = \frac{\sum_{j=1}^{n} w_{ij} y_i}{\sum_{j=1}^{n} w_{ij}} \quad i = 1, 2, \ldots, n$$

$$\begin{bmatrix} 19{,}66 \\ 16{,}00 \\ 16{,}00 \\ 14{,}66 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 20 \\ 15 \\ 24 \\ 5 \end{bmatrix}$$

$$\hat{\mu}_A = (20x1/3) + (15x1/3) + (24x1/3) + (5x0) = 19{,}66$$

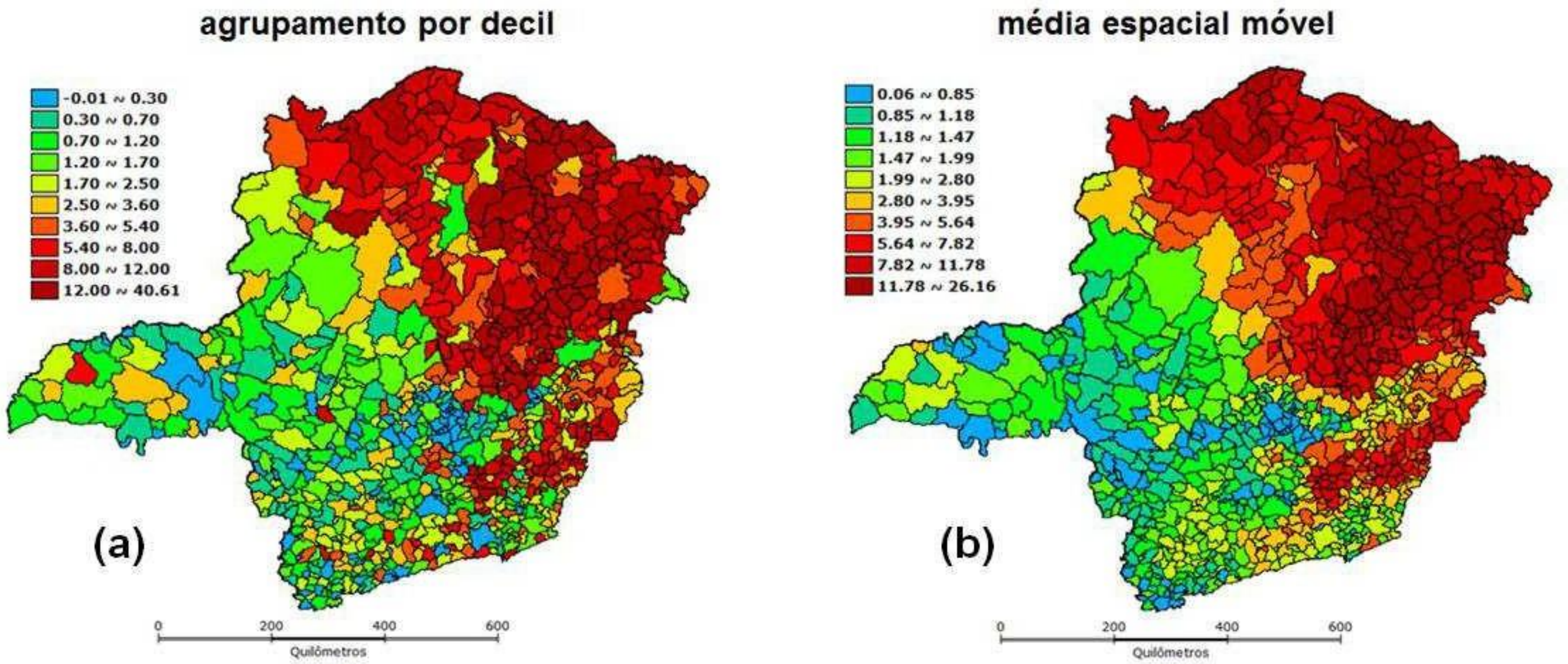
$$\hat{\mu}_B = (20x1/4) + (15x1/4) + (24x1/4) + (5x1/4) = 16{,}0$$

$$\hat{\mu}_C = (20x1/4) + (15x1/4) + (24x1/4) + (5x1/4) = 16{,}0$$

$$\hat{\mu}_D = (20x0) + (15x1/3) + (24x1/3) + (5x1/3) = 14{,}66$$

Média Espacial Móvel

A figura (**b**) ilustra um exemplo do uso do estimador de Média Espacial Móvel para proporção de crianças de 0 a 5 anos de idade residentes em domicílios particulares permanentes com responsável ou conjugue analfabeto e saneamento inadequado (Fonte: IBGE, censo 2010).



Indicadores Globais de Autocorrelação Espacial

Como visto a média espacial móvel é útil quando se deseja investigar padrões e tendências espaciais.

Para muitos tipos de dados é **importante explorar a dependência espacial**, mostrando como os valores estão relacionados no espaço.

O conceito mais utilizado é o de **autocorrelação espacial**: mede o quanto o valor observado de um atributo numa região é independente dos valores desta mesma variável nas localizações vizinhas.

Uma das formas de detecção de similaridade entre áreas é através do índice global de Moran (I).

Indicadores Globais de Autocorrelação Espacial

O índice global de Moran (I) é definido como:

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

em que:

n corresponde ao número de áreas;

y_i é o valor do atributo considerado na área i;

$\bar{y}$ representa o valor médio do atributo na região de estudo;

w_{ij} são os pesos atribuídos conforme a conexão entre as áreas i e j.

Indicadores Globais de Autocorrelação Espacial

O índice global de Moran (I): O que é necessário entender?

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

i. Qual o significado do valor do índice global de Moran (I)?

ii. Como interpretar a equação acima ?

iii. Qual sua significância ou validade estatística? Como avaliar?

Indicadores Globais de Autocorrelação Espacial

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

É análogo ao coeficiente de correlação convencional;

(I) varia de -1 a +1, quantifica o grau de autocorrelação espacial existente:

-1 : significa autocorrelação espacial negativa ou inversa;

0 : significa aleatoriedade;

+1 : significa autocorrelação espacial positiva ou direta.

Interpretação da equação do índice global de Moran (I)

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

A equação de (I) pode ser simplificada:

i) normalizando os atributos;

ii) alterando a matriz de proximidade (W), de forma que a soma dos elementos de cada linha seja igual a 1.

$$I = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\, z_i\, z_j}{\sum_{i=1}^{n} z_i^{\,2}}$$

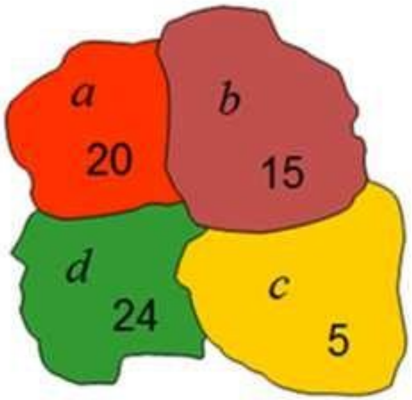
Indicadores Globais de Autocorrelação Espacial

Interpretação da equação do índice global de Moran (I)

$$I = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} z_i z_j}{\sum_{i=1}^{n} z_i^2}$$

Considere o exemplo abaixo.

Passo 1: normalizar os atributos



$$\longrightarrow \quad z_i = \frac{(y_i - \bar{y})}{\sigma} \quad \longrightarrow$$

$z_a =$	0,5628
$z_b =$	-0,1407
$z_c =$	-1,1257
$z_d =$	1,5479

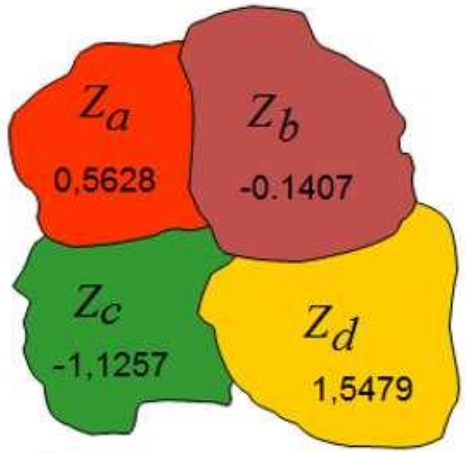
$$\textit{Média} \quad \bar{y} = \frac{20+15+24+5}{4} = 16$$

$$\textit{Variância} \quad \sigma^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n} = \frac{(20-16)^2 + (15-16)^2 + (24-16)^2 + (5-16)^2}{4} = 50{,}5$$

$$\textit{Desvio Padrão} \quad \sigma = \sqrt{\sigma^2} = \sqrt{50{,}5} = 7{,}1063$$

Interpretação da equação do índice global de Moran (I)

Passo 2: alterar a matriz de proximidade (W), de forma que a soma dos elementos de cada linha seja igual a 1.



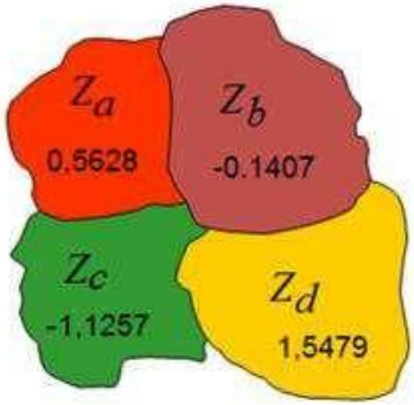
	a	*b*	*c*	*d*
a	0	1	1	0
b	1	0	1	1
c	1	1	0	1
d	0	1	1	0

	a	*b*	*c*	*d*
a	0	1/2	1/2	0
b	1/3	0	1/3	1/3
c	1/3	1/3	0	1/3
d	0	1/2	1/2	0

Indicadores Globais de Autocorrelação Espacial

Interpretação da equação do índice global de Moran (I)

Passo 3: efetuar o cálculo.



	a	b	c	d
a	0	1/2	1/2	0
b	1/3	0	1/3	1/3
c	1/3	1/3	0	1/3
d	0	1/2	1/2	0

$$I = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} z_i z_j}{\sum_{i=1}^{n} z_i^2} = \frac{\begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} * \begin{bmatrix} 0{,}3167 & -0{,}0792 & 0{,}6335 & -0{,}8711 \\ -0{,}0792 & 0{,}0197 & -0{,}1583 & 0{,}2177 \\ 0{,}6335 & -0{,}1583 & 1{,}2672 & -1{,}7424 \\ -0{,}8711 & 0{,}2177 & -1{,}7424 & 2{,}3959 \end{bmatrix}}{4} = \frac{\begin{bmatrix} 0 & -0{,}0396 & 0{,}3167 & 0 \\ -0{,}0264 & 0 & -0{,}0527 & 0{,}0725 \\ 0{,}2111 & -0{,}0527 & 0 & -0{,}5808 \\ 0 & 0{,}1088 & -0{,}8712 & 0 \end{bmatrix}}{4} = \frac{-0{,}9143}{4} = -0{,}228$$

A significância do índice de Moran (*I*). Como avaliar?

Um dos aspectos mais relevantes com relação ao índice de Moran (I) é estabelecer sua validade estatística; isto é:

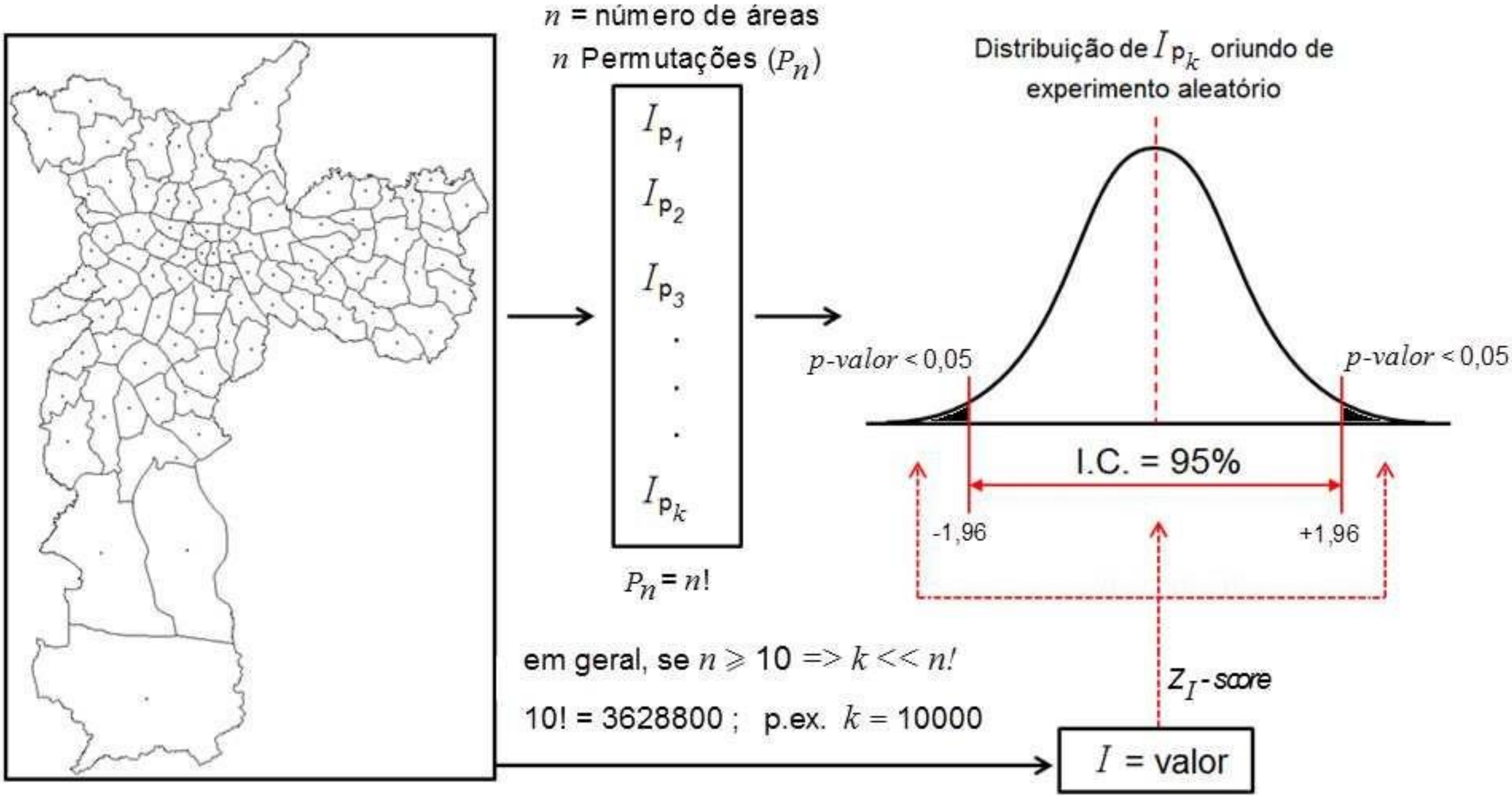
Será que os valores medidos representam correlação espacial significativa?

Para estimar a significância de I é preciso associar a este uma distribuição estatística. Duas abordagens são possíveis:

i. simulação da distribuição por permutação aleatória nos valores dos atributos;

ii. distribuição aproximada (hipótese da normalidade).

Teste de pseudo-significância



A validade estatística do índice de Moran sob a distribuição aproximada.

Para um número suficiente de regiões o índice Moran (I) tem uma distribuição amostral que é aproximadamente normal, dada por:

$$\boxed{E(I) = \frac{1}{(n-1)}} \qquad \boxed{\sigma^2 = \frac{n^2(n-1)S_1 - n(n-1)S_2 - 2S_o^2}{(n+1)(n-1)^2 S_o^2}} \qquad \boxed{I_N = \frac{I - E(I)}{\sigma}} = Z_I\text{-}score$$

em que:

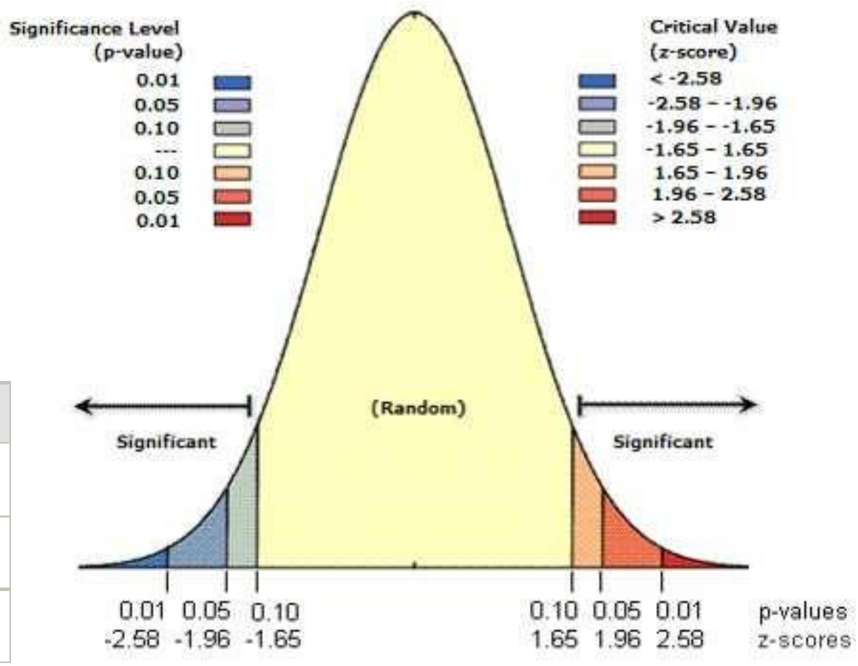
n: é o número de regiões;

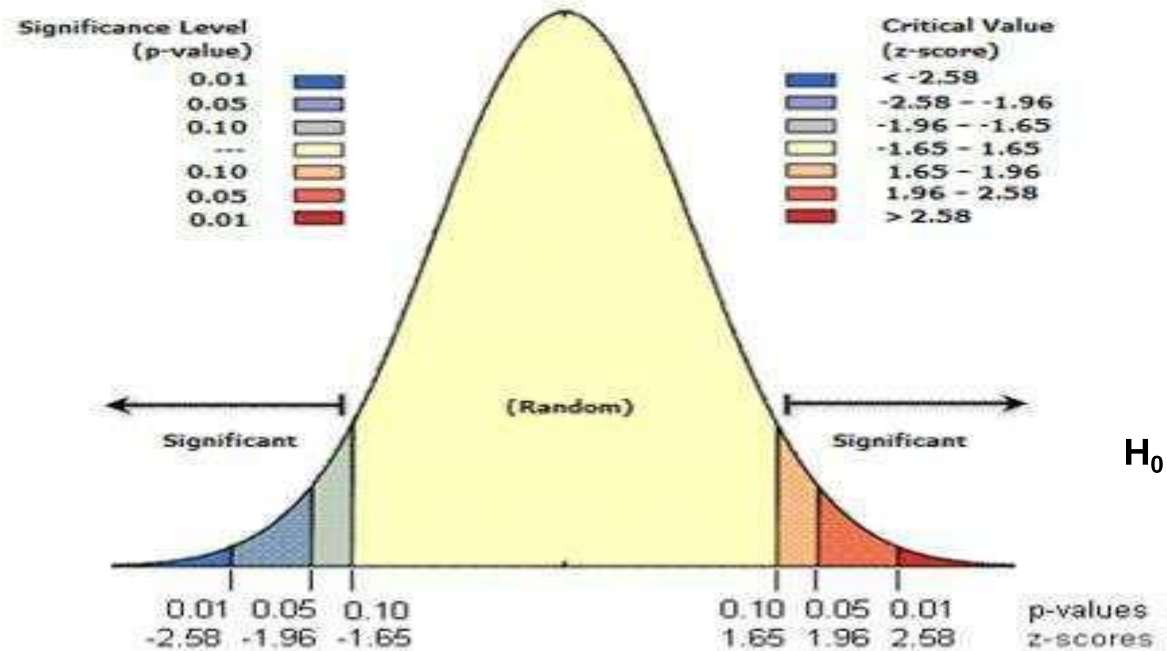
$$S_o = \sum\sum w_{ij} \quad para \quad i \neq j$$

$$S_1 = \sum\sum\left(w_{ij} + w_{ij}\right)^2 \quad para \quad i \neq j$$

$$S_2 = \sum\left(\sum w_{ij} + \sum w_{ij}\right)^2 \quad para \quad i \neq j$$

z-score (Standard Deviations)	p-value (Probability)	Confidence level
< -1.65 or > +1.65	< 0.10	90%
< -1.96 or > +1.96	< 0.05	95%
< -2.58 or > +2.58	< 0.01	99%





Hipótese Nula
H_0: processo é aleatório

p-valor é estatisticamente significante e *z-score* é negativo: **Rejeita H_0**

altos e baixos valores mais espacialmente dispersos

altos e baixos valores espacialmente aleatórios

altos e baixos valores mais espacialmente agrupados

p-valor é estatisticamente significante e *z-score* é positivo: **Rejeita H_0**

p-valor não é estatisticamente significante: **Aceita H_0**

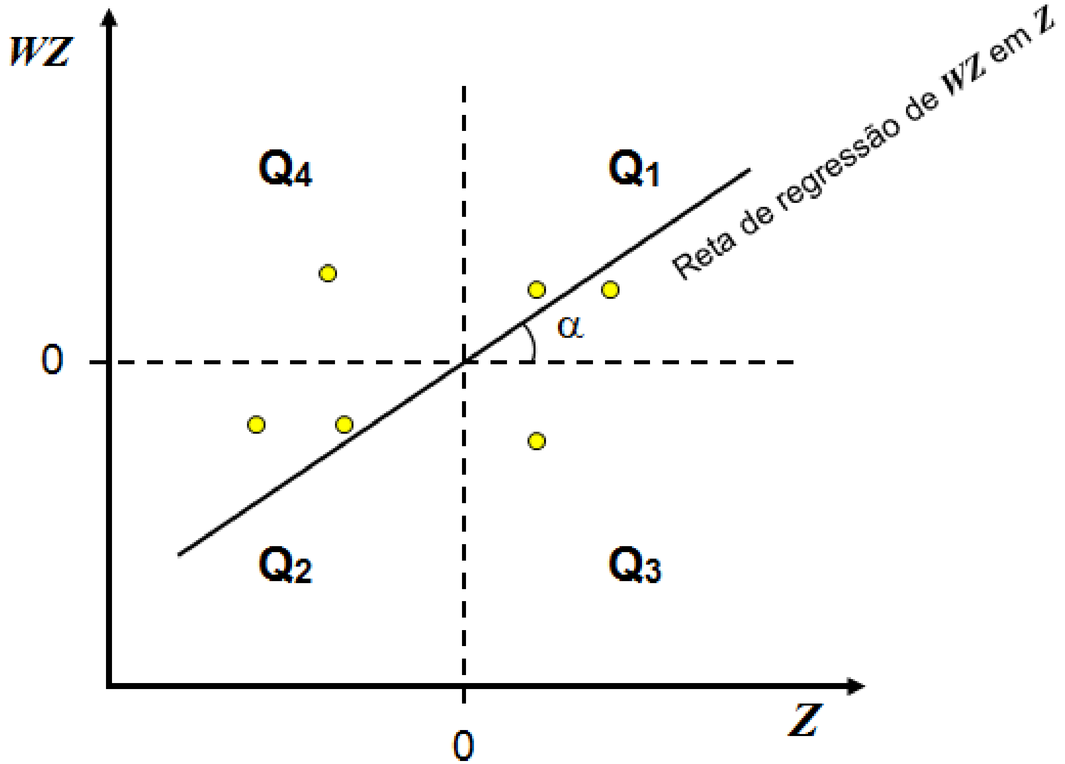
Indicadores Globais de Autocorrelação Espacial

Uma maneira adicional de visualizar o índice de Moran (I) proposta por Anselin (1996), é através do **Diagrama de Espalhamento de Moran**.

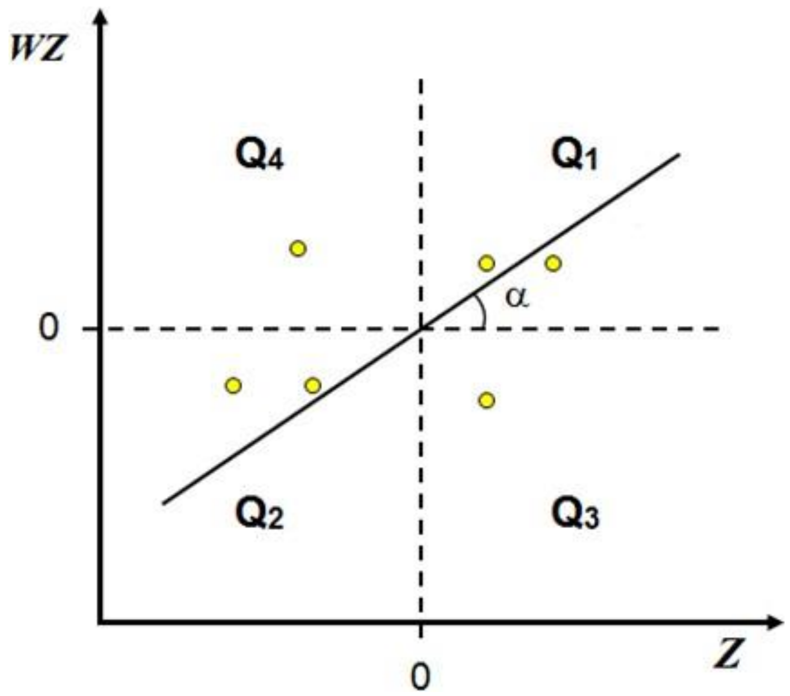
Este diagrama relata espacialmente o relacionamento entre os valores do vetor de desvios $\boldsymbol{Z}$ (Z_i - $\bar{Z}$) e os valores das médias locais $\boldsymbol{WZ}$, indicando diferentes regimes espaciais presentes nos dados.

$$I = \frac{Z^t WZ}{Z^t Z}$$

Nesta formulação, I equivale ao coeficiente de regressão linear (a inclinação da reta de regressão).



Indicadores Globais de Autocorrelação Espacial



Q$_1$ (Z [+], WZ [+]) e **Q**$_2$ (Z [-], WZ [-])

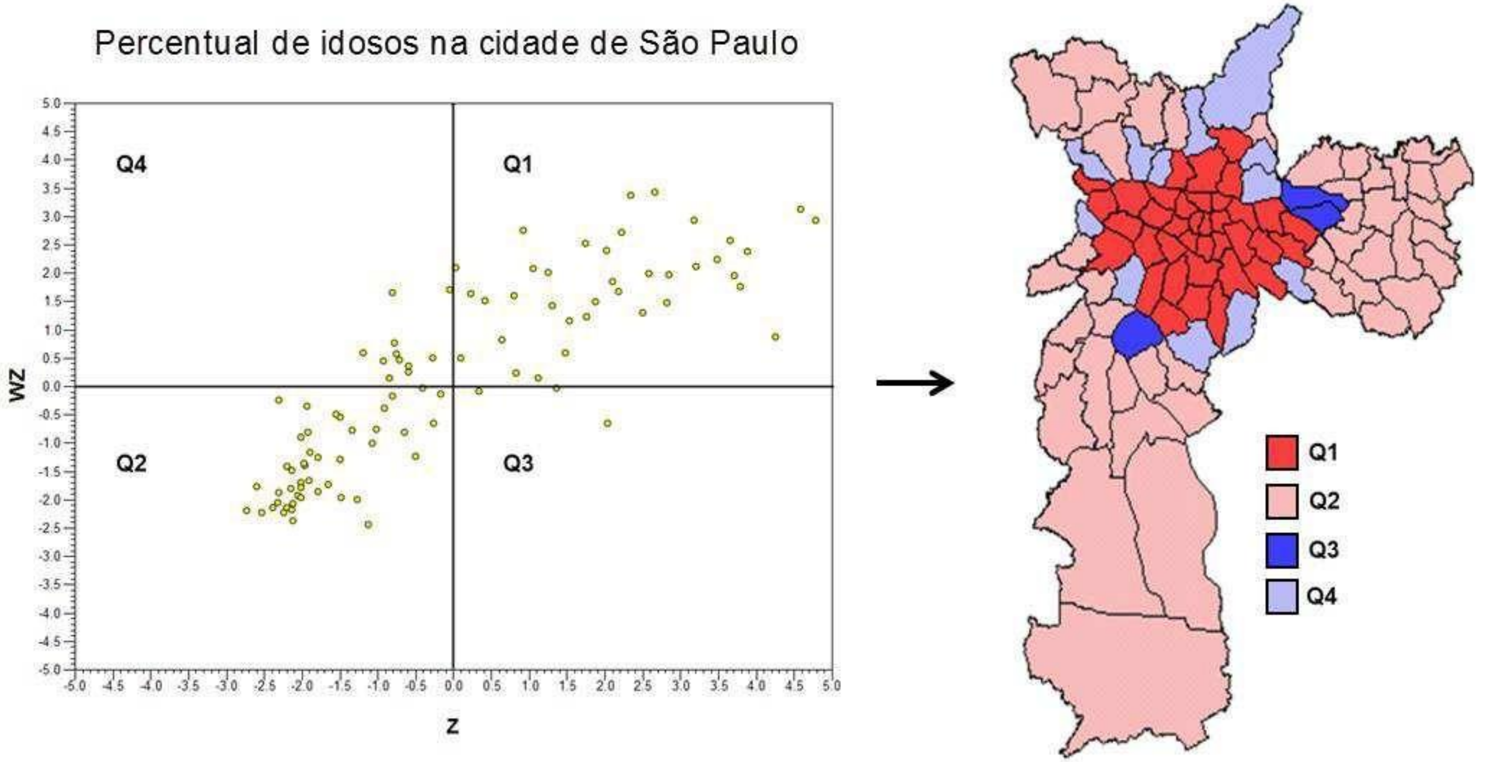
Indicam pontos de associação espacial positiva, no sentido que uma localização possui vizinhos com valores semelhantes.

Q$_3$ (Z [+], WZ [-]) e **Q**$_4$ (Z [-], WZ [+])

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

O Diagrama de Espalhamento de Moran na forma de mapa temático (Box Map).

Cada polígono é apresentado indicando seu quadrante no diagrama de espalhamento.



Pontos localizados em **Q_3** e **Q_4** podem ser vistos como extremos, por indicar regiões que não seguem o mesmo processo de dependência espacial das demais observações. Estes pontos marcam regiões de transição entre regimes espaciais distintos.

Um outro indicador global de autocorrelação espacial é o de Geary (C), definido como:

$$C = \frac{(n-1)}{2\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(Y_i - Y_j)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

Os termos da equação acima seguem as mesmas definições de Moran (I).

O indicador Geary (C) <u>normalmente</u> assume valores entre 0 a 2:

$0 \leq c < 1$: indica autocorrelação espacial positiva ou direta;

$1 < c \leq 2$: indica autocorrelação espacial negativa ou inversa.

$c \cong 1$: indício de independência espacial entre as áreas;

Indicadores Globais de Autocorrelação Espacial

Interpretação do indicador Geary (C)

$$C = \frac{(n-1)}{2\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(Y_i - Y_j)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

$$\Downarrow$$

$$C = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(Y_i - Y_j)^2}{2\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \Bigg/ \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{(n-1)}$$

variabilidade da vizinhança

variabilidade geral
(variância dos dados)

Indicadores Globais de Autocorrelação Espacial

Interpretação do indicador Geary (C)

$$C = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(Y_i - Y_j)^2}{2\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \Bigg/ \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{(n-1)}$$

variabilidade da vizinhança

variabilidade geral
(variância dos dados)

- Se a variabilidade da vizinhança for próxima da variabilidade geral $\Rightarrow C \cong 1 \Rightarrow$ indício de independência espacial entre as áreas;

- Se áreas vizinhas tem a tendência de serem similares, a variabilidade da vizinhança será menor que a variabilidade geral $\Rightarrow 0 < C < 1 \Rightarrow$ autocorrelação espacial positiva;

- Se áreas vizinhas não são similares, a variabilidade entre vizinhos será maior que a variabilidade geral $\Rightarrow 1 < C < 2 \Rightarrow$ autocorrelação espacial negativa;

Indicadores Globais de Autocorrelação Espacial

Moran (I) x Geary (C)

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

$$C = \frac{(n-1)}{2\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(Y_i - Y_j)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

O índice de Moran difere do índice de Geary por utilizar a diferença entre cada área (y_i e y_j) e a média global ($\bar{y}$), e não a diferença entre pares de áreas (y_i e y_j), como é o caso do índice de Geary.

Assim, o índice de Moran é mais sensível a valores extremos, enquanto o índice de Geary é uma medida sensível às diferenças em pequenas distâncias entre áreas.

Indicadores Globais de Autocorrelação Espacial

Indicador global de Getis e Ord (G), definido como:

$$G = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j}\, y_i y_j}{\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j}, \forall j \neq i$$

y_i e y_j : valores de atributos associados às i-ésima e j-ésima áreas;

$w_{i,j}$: peso espacial entre as áreas i e j;

n : número de áreas;

$\forall j \neq i$: indica que as áreas não podem ser as mesmas.

Indicadores Globais de Autocorrelação Espacial

Indicador global de Getis e Ord (G), definido como:

$$G = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j}\, y_i y_j}{\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j}, \forall j \neq i$$

Um esquema de ponderação binário é recomendável para esta estatística (G);

Note que a unica diferença entre o numerador e o denominador é a ponderação $w_{i,j}$;

G trabalha somente com valores positivos. Consequentemente se os pesos $w_{i,j}$ são binários (0/1) ou são sempre < 1, o range de valores para G será entre 0 e 1;

O valor de G é uma medida relativa, sendo de difícil interpretação;

G deve ser interpretado em relação E[G] e Var[G] => *no cálculo do* z_G*-score.*

Indicador global de Getis e Ord (G)

$$G = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j}\, y_i y_j}{\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j}, \forall j \neq i$$

$$\boxed{z_G = \frac{G - \mathrm{E}[G]}{\sqrt{\mathrm{V}[G]}}}$$

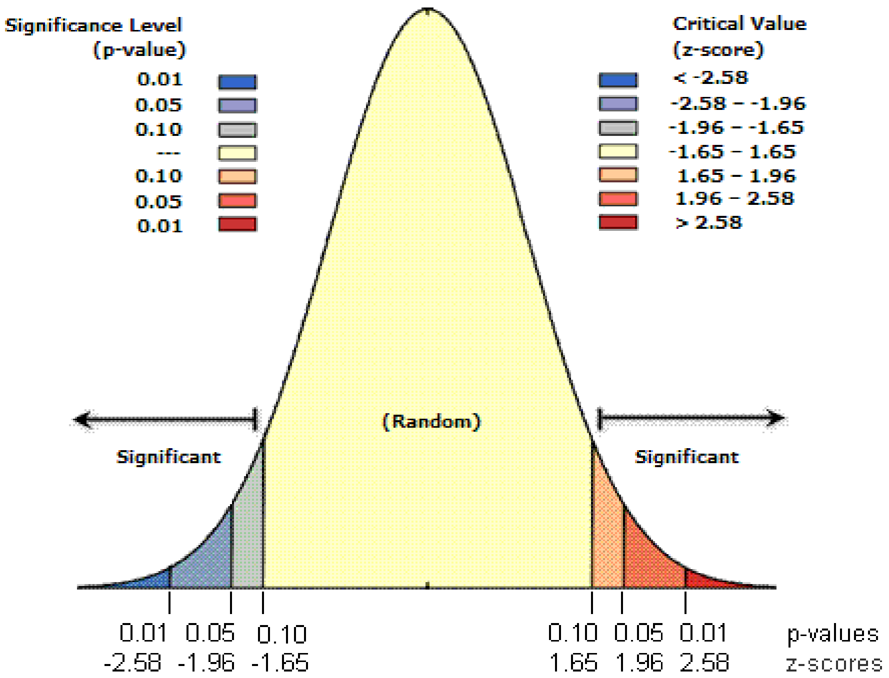
z_G-*score*

$$\mathrm{E}[G] = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j}}{n(n-1)}, \ \forall j \neq i$$

$$\mathrm{V}[G] = \mathrm{E}[G^2] - \mathrm{E}[G]^2$$

Interpretação do z_G-*score*

$$z_G = \frac{G - \mathrm{E}[G]}{\sqrt{\mathrm{V}[G]}}$$



Quando o *p-valor* for pequeno e significante estatísticamente a hipótese nula (aleatória) é rejeitada e o valor de z_G-*score* torna-se importante:

Se z_G-*score* é positivo => $G > \mathrm{E}[G]$ => presença de clusters de valores altos; (***hot spot***)

Se z_G-*score* é negativo => $G < \mathrm{E}[G]$ => presença de clusters de valores baixos (***cold spot***)

Indicadores Locais de Autocorrelação Espacial

Conforme visto os indicadores globais de Moran e Geary fornecem um valor único como medida da associação espacial;

Muitas vezes é necessário examinar padrões numa escala local;

Neste caso, é preciso utilizar indicadores locais de associação espacial que possam ser associados a diferentes localizações de uma variável distribuída espacialmente.

A utilização dos indicadores locais em conjunto com os indicadores globais refinam nosso conhecimento sobre o processos que dão origem a dependência espacial.

Indicadores Locais de Autocorrelação Espacial

Os indicadores locais de associação espacial produzem um valor específico para cada objeto (área).

Isto acarreta a identificação de:

i. clusters: objetos com valores de atributos semelhantes;

ii. outliers: objetos anômalos;

iii. a presença de mais de um regime espacial.

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

i. permitir a identificação de padrões de associação espacial significativos;

ii. ser uma decomposição do índice global de associação espacial.

Indicador local de Moran I_i (Anselin, 1996).

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

Se $I_i > 0$ => clusters de valores similares (altos ou baixos);

Se $I_i < 0$ => clusters de valores distintos.

Normalizando as variáveis e alterando matriz de proximidade (W), de forma que a soma em cada linha seja igual a 1, o indicador local (I_i) se reduz a:

$$I_i = z_i \sum_{j=1}^{n} w_{ij} z_j$$

Indicadores Locais de Autocorrelação Espacial

De forma similiar aos indicadores globais a significância do índices locais de Moran (I_i) são também avaliadas (Anselin, 1995).

Uma vez determinada a significância estatística de (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**.

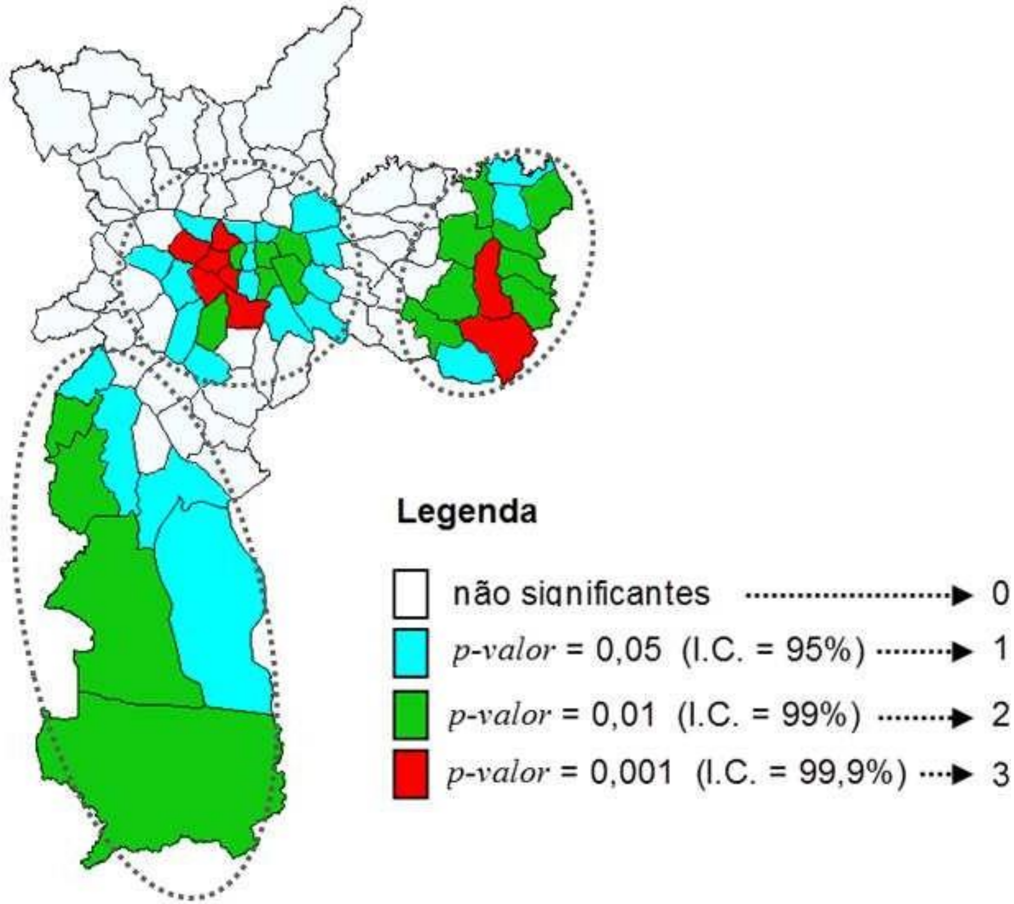
Na geração do **LISA MAP** os índices locais I_i são classificados como:

não significantes;

com confiança de: 95% (1,96 σ); 99% (2,54 σ) e 99,9% (3,2 σ).

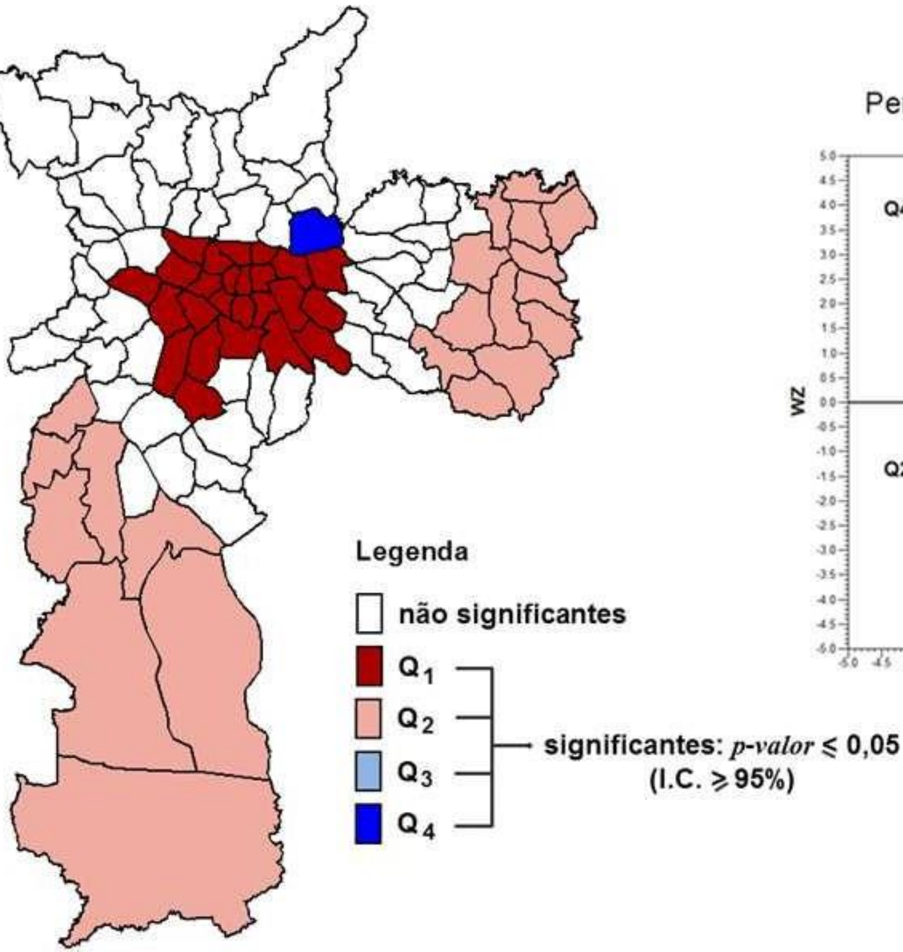
Indicadores Locais de Autocorrelação Espacial

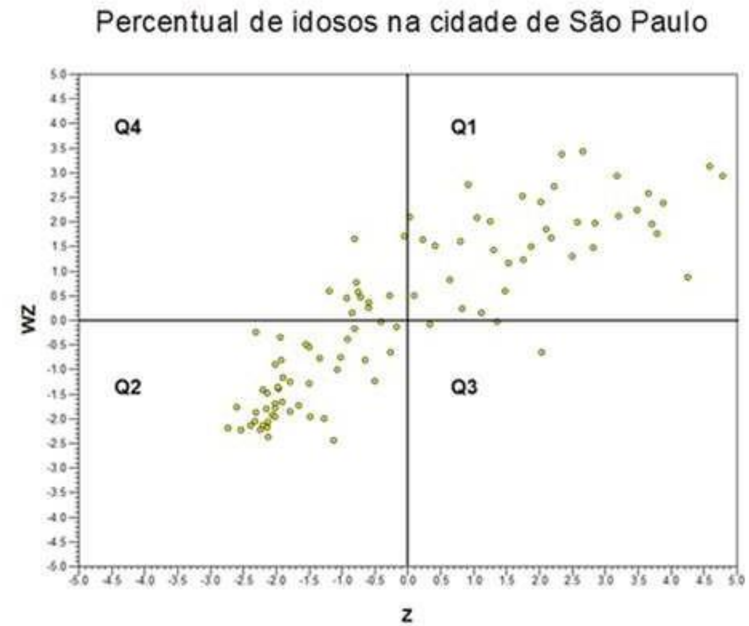
O **LISA MAP** ilustrado na figura apresenta a distribuição dos valores de correlação local para o percentual de idosos dos distritos de SP.



Este resultado indica claramente uma forte polarização centro-periferia, indicando a presença de bolsões.

Uma outra forma de análise é através do mapa denominado **Moran Map** (Anselin, 1999). Neste caso, os índices locais I_i são associados ao diagrama de Espalhamento de Moran.





Indicadores Locais de Autocorrelação Espacial

Os indicadores locais G_i e $G_i{}^*$ (Getis e Ord, 1992):

$$G_i(d) = \frac{\sum_{j=1}^{n} w_{ij}(d)x_j}{\sum_{j=1}^{n} x_j}, j \neq i \qquad G_i{}^*(d) = \frac{\sum_{j=1}^{n} w_{ij}(d)x_j}{\sum_{j=1}^{n} x_j}, j = i$$

em que:

w_{ij} : valor na matriz de proximidade para região i com a região j em função da distância (d).

x_i e x_j : são os valores dos atributos considerados nas áreas i e j;

d : é distância entre pontos (ex: entre os centróides);

n : o número de áreas (polígonos).

NOTA: a estatística G_i inclui no numerador a soma dos valores de todos vizinhos dentro de uma distância d do ponto considerado. $G_i{}^*$ difere de G_i por incluir a localização visitada.

Finalizar o Roteiro Prático

Aula 2 - Análise Padrão de Áreas, <u>a partir da pág. 9</u>