

MBA
USP
ESALQ

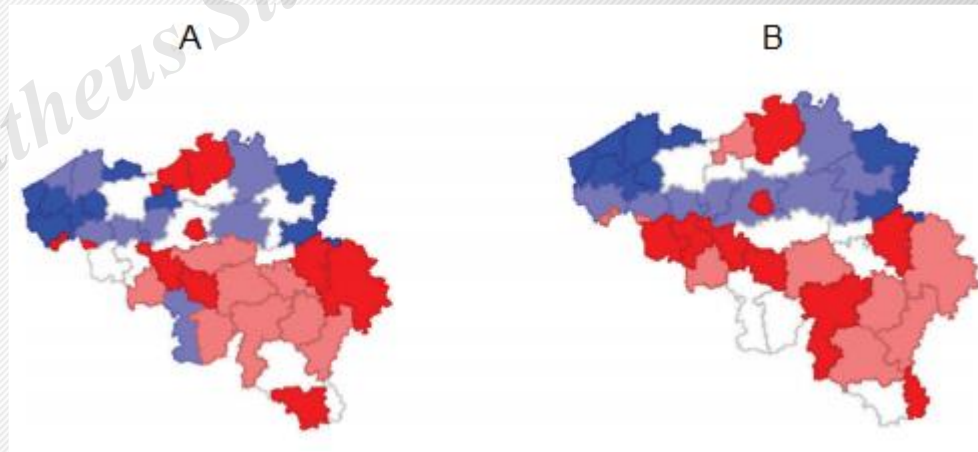
Análise Espacial III
Rafael de Freitas Souza

A dark-themed world map with a complex network of white lines connecting various points across the continents, representing spatial data or network connections. A semi-transparent dark horizontal band is centered across the map, and a semi-transparent olive-green rectangular box is positioned on the right side, partially overlapping the map and the dark band.

Introdução à Análise Exploratória de Dados Espaciais

Análise Exploratória de Dados Espaciais

- Conforme discutem Piatkowska, Messner e Yang (2018), a *Exploratory Spatial Data Analysis* (ESDA) direciona-se a facilitar a identificação e a visualização de padrões espaciais de um dado fenômeno.



A evolução da criminalidade na Bélgica entre 2006 a 2012.

Fonte: Piatkowska et al. (2018), com adaptações.

Como foi possível evidenciar as considerações a respeito do mapa anterior?

A construção dos mapas discutidos foi possível, primeiramente, porque foi definido um critério de vizinhança. Após isso, uma matriz de defasagem espacial W foi proposta e, só aí, foram verificadas as autocorrelações espaciais do fenômeno, de forma global e local. Após isso, modelos espaciais podem ser estimados.

Logo:

- Passo 1: Escolher um critério de vizinhança;
- Passo 2: Construir uma matriz de defasagens espaciais (W);
- Passo 3: Calcular as autocorrelações globais e locais;
- Passo 4: Estimação de modelos.

A map of Brazil is shown in a light gray tone. Overlaid on the map is a complex network of thin black lines connecting numerous points, representing spatial relationships or a network structure. The network is denser in the central and eastern parts of the country.

Estabelecendo Vizinhanças

Matrizes de Defasagens Espaciais

Matrizes de Vizinhanças

O primeiro passo para efetuarmos uma ESDA é o estabelecimento das vizinhanças entre as localidades estudadas para que, então possamos, por exemplo, verificar as autocorrelações espaciais, apontar algumas heterogeneidades e, inclusive, detectar eventuais outliers.

Conforme Anselin e Rey (2014) discutem, o estabelecimento de vizinhanças é feito por uma matriz de ponderação espacial W que pode assumir diversas tipologias, sendo as mais comuns as matrizes de contiguidade, de proximidade geográfica e as matrizes de proximidade socioeconômica.

Matriz W Espacialmente Ponderada pela Contiguidade

A ideia de contiguidade parte do pressuposto da existência de uma fronteira física comum entre regiões. Nessa linha de raciocínio, pode-se dizer que o Brasil é contíguo à Argentina, mas não à China, por exemplo.

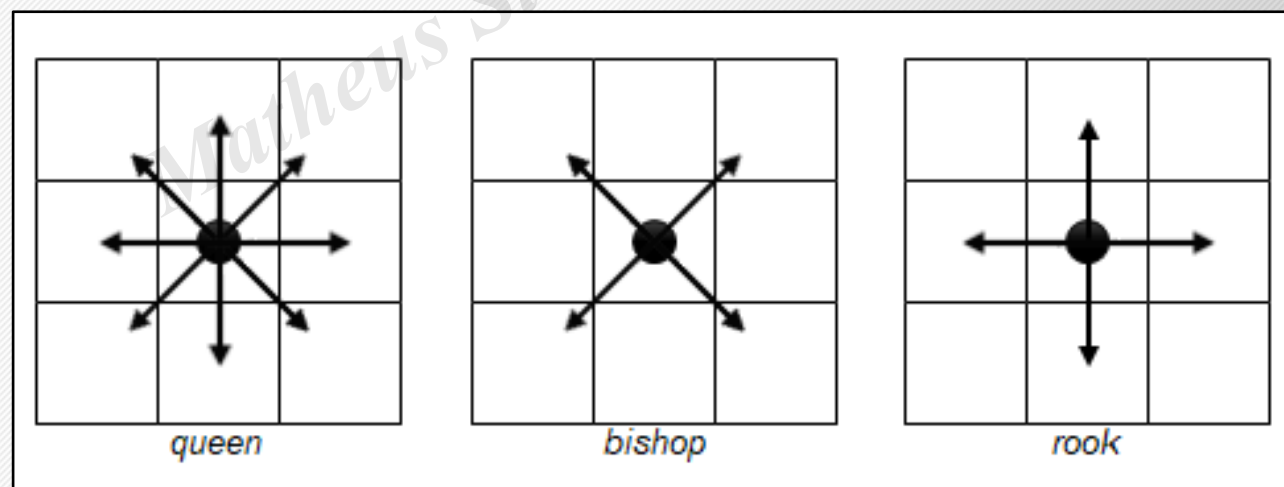
Embasando-se na premissa anteriormente descrita, uma Matriz de Contiguidade é uma matriz W com valores binários, em que se estipula o valor 1 na presença de uma fronteira física comum, e 0 para a ausência. Matematicamente, podemos descrever os seus termos w_{ij} como:

$$w_{ij} = \begin{cases} 1, & \text{caso haja contiguidade entre } i \text{ e } j; \\ 0, & \text{caso não haja contiguidade entre } i \text{ e } j. \end{cases}$$

em que $w_{ii} = 0$, por convenção.

Contiguidade

A contiguidade entre elementos w_{ij} pode ser estabelecida por diversas convenções. As convenções mais comuns são a *queen* (rainha), a *bishop* (bispo) e a *rook* (torre). Os nomes não são coincidências e referem-se a contiguidades estipuladas com base no movimento das peças do xadrez.



Fávero et al. (2022)

Matriz W Espacialmente Ponderada pelas Distâncias Geográficas

- O critério da distância geográfica pode ser uma saída interessante para promover o balanceamento de vizinhos entre observações. Sob essa ótica, deve-se estabelecer uma distância $d_i(k)$ que seja o limiar para o estabelecimento dos k vizinhos para dada região i . Conforme aponta Almeida (2012), a ideia desse tipo de matriz é a premissa de que regiões geograficamente mais próximas possuem uma maior interação espacial.
- Para o critério da distância geográfica, os elementos w_{ij} da matriz W podem ser matematicamente apresentados por:

$$w_i(k) = \begin{cases} 1, & \text{para os casos em que } d_{ij} \leq d_i(k); \\ 0, & \text{para os casos em que } d_{ij} > d_i(k). \end{cases}$$

em que $w_{ii} = 0$, por convenção.

Matriz W Espacialmente Ponderada pelos k -*Nearest Neighbors*

Anselin e Rey (2014) discutem que a matriz W ponderada pela convenção k -*nearest neighbors*, além de evitar a existência de “ilhas” no mapa, tira o peso do pesquisador de arbitrar uma distância ótima $d_i(k)$ entre as observações de sua base de dados. Em contraponto, em nossa opinião, impõe que o pesquisador decida pelo valor de k vizinhos. Também é preciso dizer que, mais uma vez, $w_{ii} = 0$, por convenção.

Matriz W Espacialmente Ponderada pelas Distâncias Sociais

Podemos, também, estabelecer uma matriz W utilizando as distâncias sociais como critério de ponderação espacial. Podemos entender as distâncias sociais como as diferenças de IDH, do Índice de Gini, de taxas analfabetismo, de taxas de mortalidade infantil, etc. entre as observações.

A world map with a dark gray background. Overlaid on the map is a complex network of white lines connecting various points across the globe, representing a spatial network or data flow. The lines are most dense in the central and eastern regions of the map. A semi-transparent dark gray rectangle is positioned behind the title text, and a semi-transparent olive green rectangle is positioned behind the logo in the bottom right corner.

Padronização de Matrizes Espaciais

Padronização de Matrizes Espaciais

Anselin e Rey (2014) afirmam ser rara a utilização de uma matriz W em sua forma binária, sugerindo a adoção de algum um processo de padronização, sendo os mais comuns a padronização da matriz W em linha, a dupla padronização e a estabilização da variância.

Suponha a seguinte matriz W

	ARG	BOL	BRA	CHL	COL	ECU	GUY	GUF	PRY	PER	SUR	URY	VEN
ARG	0	1	1	1	0	0	0	0	1	0	0	1	0
BOL	1	0	1	1	0	0	0	0	1	1	0	0	0
BRA	1	1	0	0	1	0	1	1	1	1	0	0	0
CHL	1	1	0	0	0	0	0	0	0	1	0	0	0
COL	0	0	1	0	0	1	0	0	0	1	0	0	1
ECU	0	0	0	0	1	0	0	0	0	1	0	0	0
GUY	0	0	1	0	0	0	0	0	0	0	1	0	1
GUF	0	0	1	0	0	0	0	0	0	0	1	0	0
PRY	1	1	1	0	0	0	0	0	0	0	0	0	0
PER	0	1	1	1	1	1	0	0	0	0	0	0	0
SUR	0	0	1	0	0	0	1	1	0	0	0	0	0
URY	1	0	1	0	0	0	0	0	0	0	0	0	0
VEN	0	0	1	0	1	0	1	0	0	0	0	0	0

Padronização em Linha da Matriz W

Assumindo-se como exemplo de matriz de vizinhança o exposto pela nossa tabela, uma matriz W padronizada em linha, considera o somatório dos pesos espaciais binários em cada uma de suas linhas, dividindo-os pelos seus respectivos w_{ij} . Matematicamente:

$$w_{ij\text{padronizado em linha}} = \frac{w_{ij}}{\sum_j w_{ij}}$$

em que o somatório dos pesos espaciais de cada linha deve, obrigatoriamente, ser igual a 1; enquanto que a soma de todos os pesos S_0 é dada por:

$$S_0 = \sum_i \sum_j w_{ij} = n$$

em que n é igual ao total de observações; caso haja q observações sem vizinhos, estas devem ser subtraídas de n , gerando o exposto por:

$$S_0 = \sum_i \sum_j w_{ij} = n - q$$

Padronização em Linha da Matriz W

	ARG	BOL	BRA	CHL	COL	ECU	GUY	GUF	PRY	PER	SUR	URY	VEN
ARG	0.00	0.20	0.20	0.20	0.00	0.00	0.00	0.00	0.20	0.00	0.00	0.20	0.00
BOL	0.20	0.00	0.20	0.20	0.00	0.00	0.00	0.00	0.20	0.20	0.00	0.00	0.00
BRA	0.14	0.14	0.00	0.00	0.14	0.00	0.14	0.14	0.14	0.14	0.00	0.00	0.00
CHL	0.33	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.00	0.00
COL	0.00	0.00	0.25	0.00	0.00	0.25	0.00	0.00	0.00	0.25	0.00	0.00	0.25
ECU	0.00	0.00	0.00	0.00	0.50	0.00	0.00	0.00	0.00	0.50	0.00	0.00	0.00
GUY	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.33
GUF	0.00	0.00	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.50	0.00	0.00
PRY	0.33	0.33	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PER	0.00	0.20	0.20	0.20	0.20	0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SUR	0.00	0.00	0.33	0.00	0.00	0.00	0.33	0.33	0.00	0.00	0.00	0.00	0.00
URY	0.50	0.00	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
VEN	0.00	0.00	0.33	0.00	0.33	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.00

Dupla Padronização da Matriz W

A ideia do procedimento da dupla padronização da matriz de pesos espaciais W é a de transformá-la em uma matriz estocástica, cujo somatório de todos os seus pesos S_0 seja igual a 1:

$$w_{ij \text{ duplamente padronizado}} = \frac{w_{ij}}{\sum_i \sum_j w_{ij}}$$

	ARG	BOL	BRA	CHL	COL	ECU	GUY	GUF	PRY	PER	SUR	URY	VEN
ARG	0.00	0.02	0.02	0.02	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.02	0.00
BOL	0.02	0.00	0.02	0.02	0.00	0.00	0.00	0.00	0.02	0.02	0.00	0.00	0.00
BRA	0.02	0.02	0.00	0.00	0.02	0.00	0.02	0.02	0.02	0.02	0.00	0.00	0.00
CHL	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00
COL	0.00	0.00	0.02	0.00	0.00	0.02	0.00	0.00	0.00	0.02	0.00	0.00	0.02
ECU	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00
GUY	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.02
GUF	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00
PRY	0.02	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PER	0.00	0.02	0.02	0.02	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SUR	0.00	0.00	0.02	0.00	0.00	0.00	0.02	0.02	0.00	0.00	0.00	0.00	0.00
URY	0.02	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
VEN	0.00	0.00	0.02	0.00	0.02	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00

Padronização da Matriz W pela Estabilização da Variância

A padronização da matriz de pesos espaciais W pela estabilização da variância foi proposta por Tiefelsdorf, Griffith e Boots (1999). O valor padronizado dos pesos espaciais é obtido, basicamente, em dois passos.

Primeiramente, cada peso original da linha i deve ser dividido pela raiz quadrada dos somatórios dos pesos ao quadrado de sua respectiva linha i , dando origem a um novo peso chamado w_{ij}^* :

$$w_{ij}^* = \frac{w_{ij}}{\sqrt{\sum_j w_{ij}^2}}$$

Posto isso, cada peso w_{ij}^* deve ser multiplicado pelo fator presente em:

$$\frac{n-q}{Q}$$

em que n é o total de observações; q é o total de observações sem vizinhos; Q é dado por $\sum_i \sum_j w_{ij}^*$.

Padronização da Matriz W pela Estabilização da Variância

	ARG	BOL	BRA	CHL	COL	ECU	GUY	GUF	PRY	PER	SUR	URY	VEN
ARG	0.00	0.24	0.24	0.24	0.00	0.00	0.00	0.00	0.24	0.00	0.00	0.24	0.00
BOL	0.24	0.00	0.24	0.24	0.00	0.00	0.00	0.00	0.24	0.24	0.00	0.00	0.00
BRA	0.20	0.20	0.00	0.00	0.20	0.00	0.20	0.20	0.20	0.20	0.00	0.00	0.00
CHL	0.31	0.31	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.31	0.00	0.00	0.00
COL	0.00	0.00	0.27	0.00	0.00	0.27	0.00	0.00	0.00	0.27	0.00	0.00	0.27
ECU	0.00	0.00	0.00	0.00	0.38	0.00	0.00	0.00	0.00	0.38	0.00	0.00	0.00
GUY	0.00	0.00	0.31	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.31	0.00	0.31
GUF	0.00	0.00	0.38	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.38	0.00	0.00
PRY	0.31	0.31	0.31	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PER	0.00	0.24	0.24	0.24	0.24	0.24	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SUR	0.00	0.00	0.31	0.00	0.00	0.00	0.31	0.31	0.00	0.00	0.00	0.00	0.00
URY	0.38	0.00	0.38	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
VEN	0.00	0.00	0.31	0.00	0.31	0.00	0.31	0.00	0.00	0.00	0.00	0.00	0.00

A world map with a network of white lines connecting various points across the continents, representing spatial relationships. A semi-transparent dark gray rectangle is centered over the map, containing the title text. A yellow rectangular area is visible on the right side of the map, partially obscured by the gray rectangle.

Autocorrelação Espacial

Autocorrelação Espacial

Estabelecidas as vizinhanças, e as suas respectivas matrizes W de defasagem espacial, podemos nos direcionar a verificar se os dados observados se distribuem de forma aleatória ou se formam um padrão espacial, ou seja, se existe a autocorrelação espacial envolvendo o fenômeno estudado.

Griffith (2003) pontua que a autocorrelação espacial pode ser entendida como a medida da correlação existente entre os valores de uma única variável de interesse de forma geográfica.

- As métricas de autocorrelação espacial global direcionam-se a mensurar o grau de relação espacial de um fenômeno em relação a todos os valores observados na base de dados.
- Já as métricas de autocorrelação espacial local, medem as autocorrelações das observações, uma a uma, em relação a sua vizinhança estabelecida pela matriz de defasagem espacial W .

Autocorrelação Global - a Estatística I de Moran

A estatística I foi proposta primeiramente por Moran (1948) e, anos depois, Cliff e Ord (1973, 1981) apresentaram um trabalho mais robusto sobre as ideias originais de Moran, assentando a seguinte fórmula:

$$I = \frac{n}{S_0} \times \frac{\sum_i \sum_j w_{ij} z_i z_j}{\sum_{i=1}^n z_i^2}$$

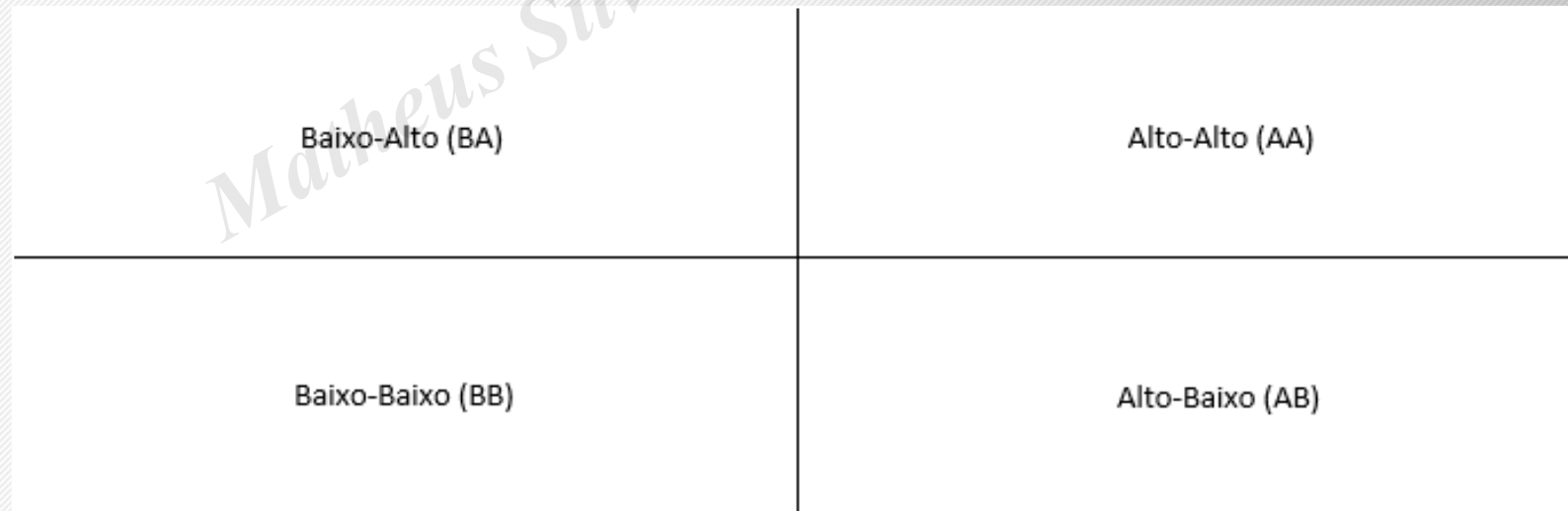
em que n representa o número de observações; z aponta os valores padronizados da variável dependente Y pelo procedimento *zscores*; w_{ij} representam os pesos espaciais da matriz W de uma dada observação existente na linha i e coluna j ; e S_0 é o somatório de todos os pesos espaciais w_{ij} .

$H_0: -\left[\frac{1}{n-1}\right]$, que indica que os valores Y_i são independentes dos valores das observações vizinhas, isto é, que não há autocorrelação espacial para dado nível de significância.

O Diagrama da Estatística I de Moran

O Diagrama de Moran é um método de visualização das autocorrelações globais apontadas pela estatística I de Moran.

A técnica visual consiste em um gráfico bidimensional de dispersão com quatro quadrantes, a saber: Alto-Alto (AA), Alto-Baixo (AB), Baixo-Baixo (BB) e Baixo-Alto (BA):



Autocorrelação Local - a Estatística Moran Local

Anselin (1995) propôs uma maneira de mensurar as autocorrelações locais chamada *Local Indicators of Spatial Association* (LISA). A técnica LISA direciona-se a identificar padrões locais de associação espacial (Anselin, 1995, p. 93).

Conforme discutem Lansley e Cheshire (2016), a técnica LISA investiga as relações espaciais entre os dados considerando as vizinhanças estabelecidas. Dentre os tipos de LISA propostos por Anselin (1995) - e.g. Gamma Local, Geary Local, etc. - será apresentado o tipo mais comumente referenciado: o Moran Local.

$$I_i = z_i \sum_j w_{ij} z_j$$

em que, de forma semelhante à Estatística I de Moran, z_i e z_j representam os valores padronizados da variável dependente; o somatório considerado inclui, apenas, cada vizinho j pertencente à vizinhança J_i estabelecida pela matriz de ponderação espacial W ; e os pesos espaciais w_{ij} sejam, preferencialmente, padronizados em linha visando facilitar a interpretação, sem olvidar que, por convenção, $w_{ii} = 0$.

Autocorrelação Local - A Estatística G de Getis e Ord

Getis e Ord (1992) propuseram uma outra maneira de estudar a associação espacial das observações de uma dada base de dados, assentada na concentração espacial.

Segundo Almeida (2012) a estatística G , para cada observação i , consegue apontar uma métrica que determina em qual medida os indivíduos da base de dados estão rodeados por observações com valores altos - os chamados *hot spots*; ou rodeados por observações com valores baixos - os chamados *cool spots*.

$$G_i(d) = \frac{\sum_{j=1}^n w_{ij}(d)Y_j}{\sum_{j=1}^n Y_j}, \text{ com } j \neq i$$

em que w_{ij} representa os pesos binários de uma matriz espacialmente ponderada pelas distâncias e, por convenção, $w_{ii} = 0$; o numerador representa a soma de todos os valores vizinhos Y_j dentro da vizinhança estabelecida pela distância d de i , sem incluir Y_i ; e o denominador, representa a soma de todos os valores vizinhos Y_j sem incluir Y_i .

Referências

Almeida, E. (2012). *Econometria Espacial Aplicada*. Campinas: Alínea.

Anselin, L. (1995). Local Indicators of Spatial Association. *Geographical Analysis*, 27(2), 93-115. doi:10.1111/j.1538-4632.1995.tb00338.x

Anselin, L., & Rey, S. J. (2014). *Modern Spatial Econometrics in Practice*. Chicago: GeoDa Press.

Cliff, A. D., & Ord, J. K. (1973). Classics in Human Geography Revisited. *Progress in Human Geography*, 19(2), 245-249. doi:10.1177/030913259501900205

Cliff, A. D., & Ord, J. K. (1981). *Spatial Processes Models and Applications*. London: Pion.

Fávero, L. P., Belfiore, P., & Freitas Souza, R. (2022). *Data Science, Analytics and Machine Learning with R*. Cambridge: Academic Press.

Getis, A., & Ord, J. K. (1992). The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis*, 24(3), 189-206. doi:10.1111/j.1538-4632.1992.tb00261.x

Griffith, D. A. (2003). *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding Through Theory and Scientific Visualization (Advances in Spatial Science)*. London: Springer.

Piatkowska, S. J., Messner, S. F., & Yang, T. -C. (2018). Xenophobic and racially motivated crime in Belgium: exploratory spatial data analysis and spatial regressions of structural covariates. *Deviant Behavior*, 39(11), 1398-1418. doi:10.1080/01639625.2018.1479917

Tiefelsdorf, M., Griffith, D. A., & Boots, B. (1999). A Variance-Stabilizing Coding Scheme for Spatial Link Matrices. *Environment and Planning A*, 31(1), 165-180. doi:10.1068/a310165