

MBA
USP
ESALQ

Análisis Espacial III

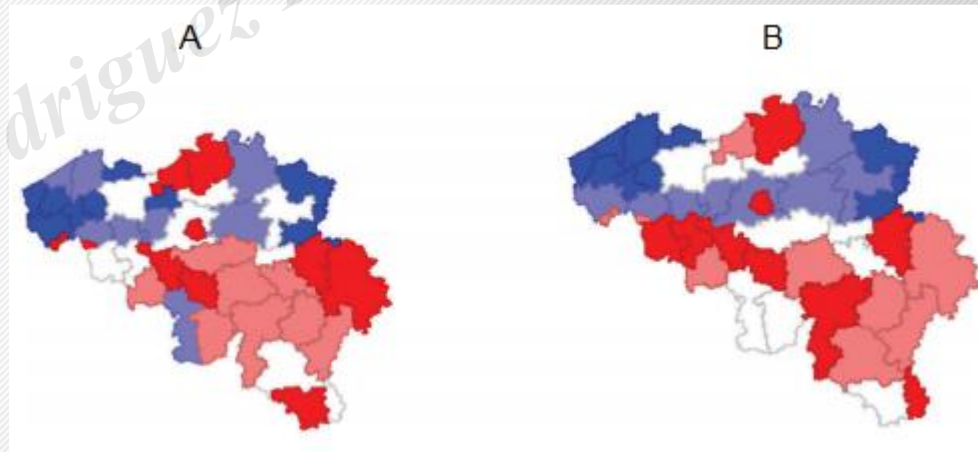
Rafael de Freitas Souza

A dark gray 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 gray rectangle is centered over the map, and a yellow rectangular highlight is positioned on the right side of the slide.

Introducción al Análisis Exploratorio de Datos Espaciales

Análisis Exploratorio de Datos Espaciales

- Conforme discuten Piatkowska, Messner y Yang (2018), la *Exploratory Spatial Data Analysis* (ESDA) se dirige a facilitar la identificación y la visualización de estándares espaciales de un determinado fenómeno.



La evolución de la delincuencia en Bélgica del 2006 al 2012.

Fuente: Piatkowska et al. (2018), con adaptaciones.

¿Cómo fue posible evidenciar las consideraciones al respecto del mapa anterior?

La construcción de los mapas discutidos fue posible, primeramente, porque fue definido un criterio de vecindad. Después de eso, una matriz de desfases espacial W fue propuesta y, sólo ahí, fueron verificadas las autocorrelaciones espaciales del fenómeno, de forma global y local. Después de eso, modelos espaciales pueden ser estimados.

Luego:

- Paso 1: Elegir un criterio de vecindad;
- Paso 2: Construir una matriz de desfases espaciales (W);
- Paso 3: Calcular las autocorrelaciones globales y locales;
- Paso 4: Estimación de modelos.

The background of the slide is a light gray map of Brazil. Overlaid on the map is a complex network of thin gray lines connecting numerous small black dots, representing a spatial network or a vector field. The network is denser in the central and northern parts of the country.

Estableciendo Vecindad

Matrices de Desfases Espaciales

Luiz R. R.

Matrices de Vecindad

El primer paso para realizar una ESDA es el establecimiento de las vecindades entre las localidades estudiadas para que podamos, por ejemplo, verificar las autocorrelaciones espaciales, señalar algunas heterogeneidades e inclusive, detectar eventuales outliers.

Conforme Anselin y Rey (2014) discuten, el establecimiento de vecindades es realizado por una matriz de ponderación espacial W que puede asumir diversas tipologías, siendo las más comunes las matrices de contigüidad, de proximidad geográfica y las matrices de proximidad socioeconómica.

Matriz W Espacialmente Ponderada por la Contigüidad

La idea de contigüidad parte del presupuesto de la existencia de una frontera física común entre regiones. En esa línea de razonamiento, se puede decir que Brasil es contiguo a la Argentina, pero no a China, por ejemplo.

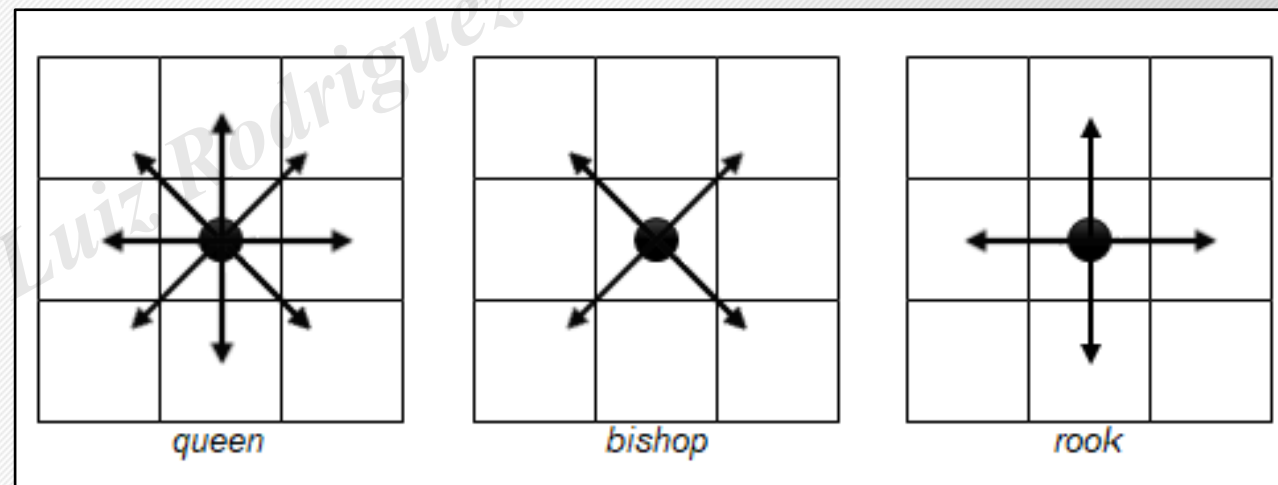
Basándose en la premisa anteriormente descrita, una Matriz de Contigüidad es una matriz W con valores binarios, en que se establece el valor 1 en la presencia de una frontera física común, y 0 para la ausencia. Matemáticamente, podemos describir sus términos w_{ij} como:

$$w_{ij} = \begin{cases} 1, & \text{caso haya contigüidad entre } i \text{ e } j; \\ 0, & \text{caso no haya contigüidad entre } i \text{ e } j. \end{cases}$$

en que $w_{ii} = 0$, por convención.

Contigüidad

La contigüidad entre elementos w_{ij} puede ser establecida por diversas convenciones. Las convenciones más comunes son la *queen* (reina), el *bishop* (Alfil) y la *rook* (torre). Los nombres no son coincidencias y se refieren a contigüidades estipuladas con base en el movimiento de las piezas del ajedrez.



Fávero et al. (2022)

Matriz W Espacialmente Ponderada por las Distancias Geográficas

- El criterio de la distancia geográfica puede ser una salida interesante para promover el equilibrio de vecinos entre observaciones. Bajo esa óptica, se debe establecer una distancia $d_i(k)$ que sea el umbral para el establecimiento de los k vecinos para determinada región i . Conforme apunta Almeida (2012), la idea de este tipo de matriz es la premisa de que regiones geográficamente más próximas poseen una mayor interacción espacial.
- Para el criterio de la distancia geográfica, los elementos de w_{ij} la matriz **W** pueden ser matemáticamente presentados por:

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

en que $w_{ii} = 0$, por convención.

Matriz W Espacialmente Ponderada por los *k-Nearest Neighbors*

Anselin y Rey (2014) discuten que la matriz W ponderada por la convención *k-nearest neighbors*, además de evitar la existencia de “islas” en el mapa, quita el peso del investigador de arbitrar una distancia óptima $d_i(k)$ entre las observaciones de su base de datos. En contrapunto, en nuestra opinión, impone que el investigador decida por el valor de k vecinos. También es necesario decir que, una vez más, $w_{ii} = 0$, por convención.

Matriz W Espacialmente Ponderada por las Distancias Sociales

Podemos, también, establecer una matriz W utilizando las distancias sociales como criterio de ponderación espacial. Podemos entender las distancias sociales como las diferencias de IDH, del Índice de Gini, de tasas de analfabetismo, de tasas de mortalidad infantil, etc. entre las observaciones.

Luiz Rodriguez
05.374.619.81

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 spatial relationships 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 horizontally across the middle of the map, containing the title text. On the right side of the map, there is a vertical rectangular area with a yellowish-brown gradient, also containing network lines.

Estandarización de Matrices Espaciales

Estandarización de Matrices Espaciales

Anselin y Rey (2014) aseguran ser rara la utilización de una matriz W en su forma binaria, sugiriendo la adopción de algún proceso de estandarización, siendo los más comunes la estandarización de la matriz W en línea, la doble estandarización y la estabilización de la varianza.

Luiz Rodriguez

Suponga la siguiente 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

Estandarización en Línea de la Matriz W

Asumiendo como ejemplo de matriz de vecindad lo expuesto por nuestra tabla, una matriz W estandarizada en línea, considera la suma de los pesos espaciales binarios en cada una de sus líneas, dividiéndolos por sus respectivos w_{ij} .
Matemáticamente:

$$w_{ij \text{ estandarizado en línea}} = \frac{w_{ij}}{\sum_j w_{ij}}$$

en que la sumatoria de los pesos espaciales de cada línea debe, obligatoriamente, ser igual a 1; mientras que la suma de todos los pesos S_0 es dada por:

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

en que n es igual al total de observaciones; si hay q observaciones sin vecinos, estas deben ser sustraídas de n , generando lo expuesto por:

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

Estandarización en Línea de la 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

Doble Estandarización de la Matriz W

La idea del procedimiento de la doble estandarización de la matriz de pesos espaciales W es la de transformarla en una matriz estocástica, cuya sumatoria de todos sus pesos S_0 sea igual a 1:

$$w_{ij \text{ doble estandarizado}} = \frac{w_{ij}}{\sum_i \sum_j w_{ij}}$$

	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

Estandarización de la Matriz W por la Estabilización de la Varianza

La estandarización de la matriz de pesos espaciales W por la estabilización de la varianza fue propuesta por Tiefelsdorf, Griffith y Boots (1999). El valor estandarizado de los pesos espaciales es obtenido, básicamente, en dos pasos.

Primeramente, cada peso original de la línea i debe ser dividido por la raíz cuadrada de las sumatorias de los pesos al cuadrado de su respectiva línea i , dando origen a un nuevo peso llamado w_{ij}^* :

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

Puesto eso, cada peso w_{ij}^* debe ser multiplicado por el factor presente en:

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

en que n es el total de observaciones; q es el total de observaciones sin vecinos; Q es dado por $\sum_i \sum_j w_{ij}^*$.

Estandarización de la Matriz W por la Estabilización de la Varianza

	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 dark gray background. Overlaid on the map is a complex network of thin, light gray lines connecting various points across the globe, representing spatial relationships or data flow. The density of connections is higher in the Northern Hemisphere, particularly in Europe and North America. A semi-transparent dark gray rectangle is centered over the map, containing the title text. On the right side of the map, there is a vertical rectangular area with a yellowish-brown tint, also featuring the network overlay.

Autocorrelación Espacial

Autocorrelación Espacial

Establecidas las vecindades, y sus respectivas matrices W de desfase espacial, podemos dirigirnos a verificar si los datos observados se distribuyen de manera aleatoria o se forman en un estándar espacial, o sea, si existe la autocorrelación espacial que involucra el fenómeno estudiado.

Griffith (2003) puntúa que la autocorrelación espacial puede ser entendida como la medida de la correlación existente entre los valores de una única variable de interés de forma geográfica.

- Las métricas de autocorrelación espacial global se dirigen a mensurar el grado de relación espacial de un fenómeno en relación con todos los valores observados en la base de datos.
- Ya las métricas de autocorrelación espacial local, miden las autocorrelaciones de las observaciones, una a una, en relación con su vecindad establecida por la matriz de desfase espacial W .

Autocorrelación Global - la Estadística I de Moran

La estadística I fue propuesta primero por Moran (1948) y, años después, Cliff y Ord (1973, 1981) presentaron un trabajo más robusto sobre las ideas originales de Moran, asentando la siguiente fórmula:

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

en que n representa el número de observaciones; z apunta los valores estandarizados de la variable dependiente Y por el procedimiento *zscores*; w_{ij} representan los pesos espaciales de la matriz \mathbf{W} de una determinada observación existente en la línea i y columna j ; y S_0 es la sumatoria de todos los pesos espaciales w_{ij} .

$H_0: -\left[\frac{1}{n-1}\right]$, que indica que los valores Y_i son independientes de los valores de las observaciones vecinas, es decir, que no hay autocorrelación espacial para dado nivel de significancia.

El Diagrama de la Estadística I de Moran

El Diagrama de Moran es un método de visualización de las autocorrelaciones globales apuntadas por la estadística I de Moran.

La técnica visual consiste en un gráfico bidimensional de dispersión con cuatro cuadrantes, a saber: Alto-Alto (AA), Alto-Bajo (AB), Bajo-Bajo (BB) y Bajo-Alto (BA):



Autocorrelación Local - la Estadística Moran Local

Anselin (1995) propuso una manera de mensurar las autocorrelaciones locales llamada *Local Indicators of Spatial Association* (LISA). La técnica LISA se dirige a identificar estándares locales de asociación espacial (Anselin, 1995, p. 93).

Conforme discuten Lansley y Cheshire (2016), la técnica LISA investiga las relaciones espaciales entre los datos considerando las vecindades establecidas. Entre los tipos de LISA propuestos por Anselin (1995) - e.g. Gamma Local, Geary Local, etc. - será presentado el tipo más comúnmente referenciado: el Moran Local.

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

en que, de forma semejante a la Estadística I de Moran, z_i y z_j representan los valores estandarizados de la variable dependiente; la sumatoria considerada incluye, sólo, cada vecino j perteneciente a la vecindad J_i establecida por la matriz de ponderación espacial W ; y los pesos espaciales w_{ij} sean, preferentemente, estandarizados en línea con el objetivo de facilitar la interpretación, sin olvidar que, por convención, $w_{ii} = 0$.

Autocorrelación Local - La Estadística G de Getis y Ord

Getis y Ord (1992) propusieron otra manera de estudiar la asociación espacial de las observaciones de una determinada base de datos, asentada en la concentración espacial.

Según Almeida (2012) la estadística G , para cada observación i , consigue apuntar una métrica que determina en qué medida los individuos de la base de datos están rodeados por observaciones con valores altos - los llamados *hot spots*; o rodeados por observaciones con valores bajos - los llamados *cool spots*.

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

en que w_{ij} representa los pesos binarios de una matriz espacialmente ponderada por las distancias y, por convención, $w_{ii} = 0$; el numerador representa la suma de todos los valores vecinos Y_j dentro de la vecindad establecida por la distancia d de i , sin incluir Y_i ; y el denominador, representa la suma de todos los valores vecinos Y_j sin incluir Y_i .

Referencias

- 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