Tema 1: Introducción a la estadística espacial (Geoestadística)

Diplomado en Geoestadística y análisis espacial

Alexis Adonai Morales Alberto

2025

scidata
matemáticas para
la ciencia de datos

Contenidos

Introducción: ¿Por qué necesitamos la estadística espacial?

Un aspecto importante de cualquier estudio que involucre unidades espaciales (ciudades, regiones, países, etc.) son las posibles relaciones e interacciones entre ellas. Por ejemplo, cuando se modela la contaminación a nivel regional, es complicado analizar cada región como unidades independientes. De hecho, las regiones no pueden analizarse como entidades aisladas ya que están interrelacionadas espacialmente por interacciones ecológicas y económicas. Por lo tanto, es muy probable la existencia de externalidades ambientales: un aumento en la contaminación de la región $i's$ afectará la contaminación en las regiones vecinas, pero el impacto será menor para las regiones más lejanas.

Introducción: ¿Por qué necesitamos la estadística espacial?

Considere la Figura 1.1, donde la región 3 está altamente industrializada, mientras que las regiones 1, 2, 4 y 5 son áreas residenciales. Si la Región 3 aumenta su actividad económica, entonces la contaminación no solo aumentará en esa región, sino también en las regiones vecinas. También se espera que la contaminación aumente en la región 1 y 5 pero en magnitudes menores. Podríamos pensar que la externalidad ambiental en R3 causa degradación ambiental en otras regiones, a través de interacciones espacio-económicas (por ejemplo, transporte de entrada y salida de la región 3) e interacciones espacio-ecológicas (por ejemplo, emisiones de carbono).

Introducción: ¿Por qué necesitamos la estadsística espacial?

$$\mathrm{R1} \longleftarrow \mathrm{R2} \longleftarrow {\color{red}\mathrm{R3}} \longrightarrow \mathrm{R4} \longrightarrow \mathrm{R5}$$

Figura: 1.1.- Externalidades ambientales

Introducción: ¿Por qué necesitamos la estadsística espacial?

Del mismo modo, si estudiamos el crimen a nivel de ciudad, de alguna manera deberíamos incorporar la posibilidad de que el crimen esté localizado. Por ejemplo, identificación de concentración o grupo de mayor actividad delictiva ha surgido como un mecanismo central para orientar una respuesta de justicia penal y prevención del delito al problema de la delincuencia. Estos grupos de delincuencia se conocen comúnmente como focos: Ubicaciones geográficas de alta concentración delictiva, en relación con la distribución de la delincuencia en toda la región de interés.

Ambos ejemplos establecen implícitamente que la ubicación geográfica y la distancia son importantes. De hecho, reflejan la importancia de la primera ley de la geografía. Según Waldo Tobler: *Todo está relacionado con todo lo demás*, pero las cosas cercanas están más relacionadas que las distantes. Esta primera ley es la base de los conceptos fundamentales de **dependencia espacial** y **autocorrelación espacial**.

Dependencia espacial

La dependencia espacial refleja una situación donde los valores observados en un lugar o región, es decir observación i, depende de los valores de las observaciones vecinas en ubicaciones cercanas. Formalmente, se puede afirmar:

$$y_i = f\left(y_j\right), \quad i = 1, \cdots, n \; j \neq i \tag{1}$$

En otras palabras, lo que sucede en la región i, depende de lo que sucede en la región j para todo $j \neq i$.

Autocorrelación espacial

Otro concepto importante es la **autocorrelación espacial**. En el espacio, el término de autocorrelación ser refiere a la correlación entre el valor de la variable dentro en dos lugares diferentes. tras formas de definir el mismo concepto son:

- (1) Correlación entre el mismo atributo en dos (o más) ubicaciones diferentes.
- (2) Coincidencia de similitud de valores con similitud de ubicación.

Esencialmente, la autocorrelación espacial se ocupa de establecer si la presencia de una variable en una región de un sistema regional hace más o menos probable la presencia de esa variable en regiones vecinas. La contraparte de la autocorrelación espacial (y la dependencia espacial) es la aleatoriedad espacial.

Autocorrelación espacial

La aleatoriedad espacial significa que no podemos observar ningún patrón espacial en los datos. Es decir, el valor que observamos en alguna unidad espacial es igualmente probable que en cualquier otra unidad espacial. La aleatoriedad espacial es importante porque formará la hipótesis nula más adelante. Si se rechaza, entonces hay evidencia de estructura espacial. Como ejemplo, la Figura 1.2 traza la distribución espacial de la pobreza en la Región Metropolitana, Chile. Se puede observar que existe algún patrón espacial donde se agrupan las comunas con similar tasa de pobreza.

Autocorrelación espacial

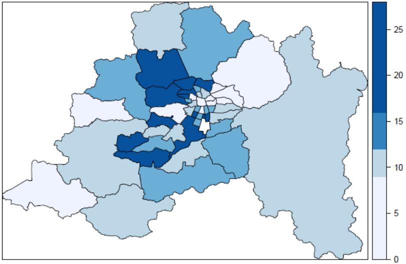

Figura: 1.2.- Distribución espacial de pobreza en la región metropolitana, Chile

Formalmente, la existencia de autocorrelación espacial puede ser expresada por el siguiente momento condicional:

$$Cov\left(y_i, y_j\right) = \mathbb{E}\left(y_i y_j\right) - \mathbb{E}\left(y_i\right)\mathbb{E}\left(y_j\right) \neq 0 \text{ para } i \neq j \tag{2}$$

Donde y_i e y_j son observaciones en la variable aleatoria de lugar o región i y j en el espacio, y i,j pueden ser puntos o unidades de área. Por lo tanto, existente una correlación espacial distinta de 0 entre los atributos de una característica definida en las ubicaciones i y j si la covarianza entre los valores de los atributos de las características en esos puntos es distinta de cero.

Si esta covarianza es positiva (es decir, si los datos con valores de atributo por encima de la media tienden a estar cerca de otros datos con valores por encima de la media), entonces decimos que hay una autocorrelación espacial positiva; si lo contrario es cierto, entonces decimos que hay autocorrelación espacial negativa. La figura 1.3 muestra un ejemplo de autocorrelación espacial positiva y negativa.

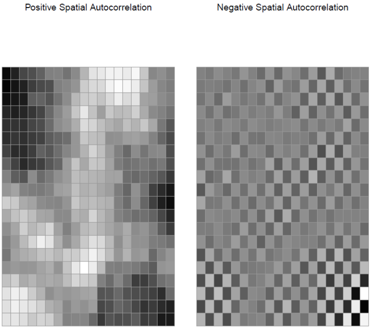

Figura: 1.3.- Autocorrelación espacial

Autocorrelación espacial

La autocorrelación positiva es mucho más común, pero existe una autocorrelación negativa, por ejemplo, en estudios de competencia de asistencia social o de subvenciones federales entre gobiernos locales (Saavedra, 2000; Boarnet y Glazer, 2002) y estudios de empleo regional (Filiztekin, 2009; Pavlyuk, 2011), las compras de lotería transfronterizas (Garrett y Marsh, 2002), la inversión extranjera directa en los países de la OCDE (Garrettsen y Peeters, 2009) y las ubicaciones de la industria manufacturera turca (Basdas, 2009). En definitiva, nos interesa estudiar patrones espaciales no aleatorios e intentar explicar esta no aleatoriedad. Las posibles causas de la falta de aleatoriedad son (Gibbons et al., 2015):

Autocorrelación espacial

1. Las empresas pueden asignarse aleatoriamente en el espacio, pero algunas características de las ubicaciones varían en el espacio e influyen en los resultados.
2. Es posible que la ubicación no tenga un efecto causal en los resultados, pero los resultados pueden estar correlacionados en el espacio porque las personas o empresas heterogéneas no se asignan al azar en el espacio.
3. Los individuos o las empresas pueden asignarse aleatoriamente en el espacio, pero interactúan de modo que las decisiones de un agente afectan los resultados de otros agentes.
4. Los individuos o las empresas pueden asignarse de forma no aleatoria a través del espacio y las características de otros cercanos influyen directamente en los resultados individuales.

Matriz de pesos espaciales

Uno de los temas cruciales en la econometría espacial es el problema de incorporar formalmente la dependencia espacial en el modelo. El principal problema es que tenemos más parámetros que observaciones. Entonces, la pregunta es: ¿Cuál sería un buen criterio para definir la cercanía en el espacio? O, en otras palabras, ¿cómo determinar qué otras unidades del sistema influyen en la que se está considerando?.

El dispositivo típicamente utilizado en análisis espacial para definir el concepto de cercanía en el espacio es la llamada *matriz de peso espacial*, o más simplemente, matriz $\mathbf{W}$. Si asumimos que hay n objetos espaciales (regiones, ciudades, países), entonces $\mathbf{W}$ será una matriz cuadrada de dimensión $n \times n$. Esta matriz impone una estructura en términos de cuáles son los vecinos de cada lugar. Asigna pesos que miden la intensidad de la relación entre pares de unidades espaciales. Así, cada elemento (i, j) de $\mathbf{W}$ que denotamos por w_{ij} expresa el grado de proximidad espacial entre el par.

Matriz de pesos espaciales

Esta matriz ($\mathbf{W}$) se puede representar de la forma:

$$\mathbf{W} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{bmatrix} \tag{3}$$

En general, asumimos que los elementos diagonales de esta matriz de *vecinos espaciales* se establecen en cero: *Las regiones no son vecinas de sí mismas*

Matriz de pesos espaciales

Una definición formal de la matriz de pesos espaciales es la siguiente:

Definición

***Matriz de pesos espaciales**.- Sea n el número de unidades espaciales. La matriz de peso espacial, $\mathbf{W}$, una matriz simétrica y no estocástica positiva $n \times n$ con el elemento w_{ij} en la ubicación i, j. Los valores de w_{ij} o los pesos para cada par de ubicaciones son asignados por algunas reglas preestablecidas que definen las relaciones espaciales entre las ubicaciones. Por convención, $w_{ij} = 0$ para los elementos diagonales.*

Matriz de pesos espaciales

No estocástico significa que el investigador toma W como conocido a priori y, por lo tanto, todos los resultados son condicionado a la especificación de $\mathbf{W}$.

Tenga en cuenta también que la definición de W requiere una regla para w_{ij}. En otras palabras, necesitamos averiguar cómo asignar un número real a w_{ij}, para $i \neq j$, que representa la fuerza de la relación espacial entre i y j. Hay varias formas de hacerlo. Pero, en general, hay dos criterios básicos. El primer tipo establece una relación basada en bordes compartidos o vértices de datos de celosía o polígono irregular (contigüidad). El segundo tipo establece una relación basada en la distancia entre ubicaciones. En términos generales, la contigüidad es más apropiada para datos geográficos expresados como polígonos (las llamadas unidades de área), mientras que la distancia es adecuada para datos puntuales, aunque en la práctica la distinción no es tan absoluta.

Pesos basados en límites

La disponibilidad de datos de polígonos o celosías permite la construcción de matrices de ponderación espacial basadas en la contigüidad. Una especificación típica de la relación de contigüidad en la matriz de peso espacial es:

$$w_{ij} = \begin{cases} 1 \text{ si } i \text{ y } j \text{ Son contiguas} \\ 0 \text{ si } i \text{ y } j \text{ No son contiguas} \end{cases} \tag{4}$$

En una cuadrícula regular, los vecinos (contigüidad) se pueden definir de varias formas. En analogía con el juego de ajedrez, se distinguen la contigüidad de la torre, la contigüidad del alfil y la contigüidad de la reina.

Pesos basados en límites: Contigüidad de la torre

En este caso, dos ubicaciones son vecinas si comparten al menos parte de un borde o lado común. En la Figura 1.4 tenemos una cuadrícula regular con 9 regiones: cada cuadrado representa una región. Si, por ejemplo, queremos definir los vecinos de la región 5 usando el criterio de la torre, entonces sus vecinos serán las regiones 2, 4, 6 y 8. Esos representan las regiones rellenas en rojo.

Pesos basados en límites: Contigüidad de la torre

Si continuamos con este razonamiento, entonces la matriz de $9 \times 9\mathbf{W}$ será:

$$\mathbf{W} = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix} \tag{5}$$

Pesos basados en límites: Contigüidad de la torre

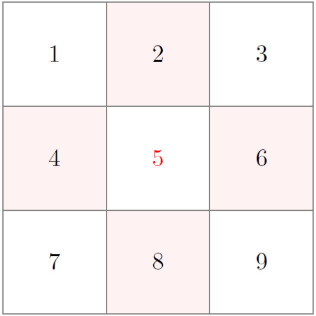

Figura: 1.4.- Contigüidad de la torre.

Pesos basados en límites: Contigüidad del alfil

En la contigüidad del alfil (que rara vez se usa en la práctica), los vecinos de la región i se ubican en sus esquinas. La figura 1.5 muestra los vecinos de la región 5 bajo este esquema. Los vecinos son las regiones 1, 3, 7 y 9. Tenga en cuenta que las regiones del interior tendrán más vecinos que las de la periferia.

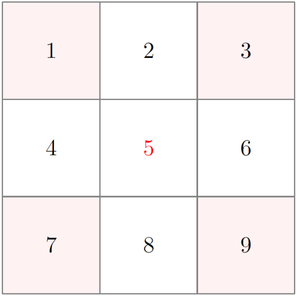

Figura: 1.5.- Contigüidad del alfil.

Pesos basados en límites: Contigüidad del alfil

La matriz $\mathbf{W}$ resultante será:

$$\mathbf{W} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{6}$$

Este criterio rara vez se utiliza en la práctica.

Pesos basados en límites: Contigüidad de la reina

En la contigüidad reina, cualquier región que toque el límite de la región i, ya sea en un lado o en un solo punto, se considera vecina. Bajo este criterio, los vecinos de 5 serán las regiones: 1, 2, 3, 4, 6, 7, 8 y 9.

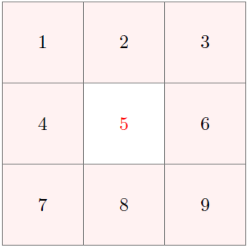

Figura: 1.5.- Contigüidad de la reina.

Pesos basados en distancias

Los pesos también se pueden definir como una función de la distancia entre la región i y j, d_{ij}. Esta distancia generalmente se calcula como la distancia entre sus centroides, pero, por supuesto, puede ser entre otros puntos relevantes para cada unidad espacial, como la capital, o la ciudad más grande, o cada región. A diferencia de los pesos basados por contigüidad, las matrices basadas en distancias sólo necesitan las coordenadas de los puntos.

Hay varias formas de calcular la distancia entre dos unidades espaciales. Sean x_i y x_j la longitud; y y_i e y_j las coordenadas de latitud para la región i y j, respectivamente. El concepto más general de distancia es la métrica de Minkowski:

$$d_{ij}^p = \left(\left| x_i - x_j \right|^p + \left| y_i - y_j \right|^p \right) \tag{7}$$

Pesos basados en distancias

Para dos puntos i y j, con respectivas coordenadas (x_i, y_i) y (x_j, y_j). y con p como el parámetro. El caso especial más familiar es la distancia euclidiana o lineal con $p = 2$:

$$d_{ij}^{e} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{7.1}$$

Otra métrica empleada es la distancia entre bloques de Manhattan. Esta medida solo considera el movimiento en las direcciones este-oeste y norte-sur, es decir, por ángulos rectos. Esto produce una medida de distancia donde $p = 1$:

$$d_{ij}^{m} = |x_i - x_j| + |y_i - y_j| \tag{7.2}$$

Pesos basados en distancias

Las tres medidas presentadas arriba son útiles si consideramos la tierra como un plano. Por ejemplo, la distancia euclidiana es la longitud de una línea recta en un mapa y no es necesariamente la distancia más corta si se tiene en cuenta la curvatura de la tierra. La distancia del gran círculo tiene en cuenta la curvatura de la Tierra. Los barcos y aviones suelen seguir la geometría del gran círculo para minimizar la distancia y ahorrar tiempo y dinero. En particular, la distancia del gran círculo se calcula como:

$$d_{ij}^{cd} = r \times \arccos^{-1}\left[\cos\left|x_i - x_j\right| \cos y_i \cos y_j + \sin y_i \sin y_j\right] \tag{8}$$

Donde r es el radio de la tierra. La distancia arc es obtenida en miles con $r = 3959$ y en kilómetros con $r = 6371$.

Pesos basados en distancias: Distancia inversa

Ahora se transforma la información acerca de la distancia entre puntos espaciales en un esquema de peso- La idea es que $w_{ijt} \rightarrow 0$ como $d_{ij} \rightarrow \infty$. En otras palabras, cuando más cerca esta j de i, mayor debe ser w_{ij} para cumplir con la primera ley de Tobler.

En el esquema de ponderación de distancia inversa, los pesos están inversamente relacionados con la distancia de separación, como se muestra a continuación:

$$w_{ij} = \begin{cases} \frac{1}{d_{ij}^{\alpha}} & \text{si } i \neq j \\ 0 & \text{si } i = j \end{cases} \tag{9}$$

Pesos basados en distancias: Distancia inversa

De la ecuación (9) donde el exponente α es un parámetro que suele establecer el investigador. En la práctica, los parámetros rara vez se estiman, pero normalmente se establecen en $\alpha = 1$ o $\alpha = 2$. Por lo tanto, los pesos vienen dados por el recíproco de la distancia: Cuanto mayor es la distancia entre las unidades espaciales, menor es el peso espacial. o la conexión espacial. Finalmente, por convención, los elementos diagonales de los pesos espaciales se establecen en cero y no se calculan. Introducir un valor de $d_{ii} = 0$ produciría una división por cero para pesos de distancia inversa.

Pesos basados en distancias: Modelo exponencial negativo

Aquí los pesos disminuyen exponencialmente con la distancia de separación:

$$w_{ij} = \exp\left(-\frac{d_{ij}}{\alpha}\right) \tag{10}$$

Donde α es un parámetro que es comúnmente elegido por el investigador. Dado que los pesos están dados por la exponencial de la distancia negativa, cuanto mayor sea la distancia entre i y j, menor será w_{ij}.

Pesos basados en distancias

Tanto la distancia inversa como la distancia exponencial negativa dependen no solo del valor del parámetro y la forma funcional, sino también de la métrica utilizada para la distancia. Dado que los pesos están inversamente relacionados con la distancia, los valores más grandes para el último producirán valores pequeños para el primero, y viceversa. Esto puede ser un problema en la práctica cuando las distancias son tan grandes que los pesos correspondientes de la distancia inversa se acercan a cero, lo que puede resultar en una matriz de peso espacial cero. Además, puede ocurrir un problema potencial cuando la métrica de la distancia es tal que las distancias toman valores menores que uno, lo que generalmente no es un resultado deseado (Anselin y Rey, 2014).

Pesos basados en distancias: Vecinos más cercanos de k

Un tipo alternativo de pesos espaciales que evita el problema de los aislamientos es seleccionar los k-vecinos más cercanos. A diferencia de la banda de distancia, esta no es una relación simétrica. Sin embargo, un problema potencial con este tipo de vecinos es la ocurrencia de lazos, es decir, cuando más de una ubicación j tiene la misma distancia de i. Existen varias soluciones para romper el empate, desde seleccionar aleatoriamente uno de los vecinos de k-ésimo orden hasta incluirlos a todos.

Distancia de umbral (pesos de banda de distancia)

En contraste con el método de los k-vecinos más cercanos, la distancia de umbral especifica que una región i es vecina de j si la distancia entre ellos es menor que una distancia máxima especificada:

$$w_{ij} = \begin{cases} 1 \text{ si } 0 \leq d_{ij} \leq d_{max} \\ 0 \text{ si } d_{ij} > d_{max} \end{cases} \tag{11}$$

Para evitar aislamientos que resultarían de una distancia crítica demasiado estricta, la distancia debe elegirse de manera que cada ubicación tenga al menos un vecino. Tal distancia se ajusta a un criterio máximo-mínimo, es decir, es la mayor de las distancias vecinas más cercanas.

Finalmente, es importante señalar que una matriz de pesos obtenida a partir de una banda de distancia siempre es simétrica, ya que la distancia es una relación simétrica.

Matriz de pesos de fila estandarizada

En la práctica, las ponderaciones espaciales rara vez se utilizan en su forma binaria (o de distancia), sino que están sujetas a una transformación o estandarización. En particular, nos gustaría calcular promedios ponderados en los que se asigna más peso a las observaciones cercanas que a las observaciones distantes. Para ello, podemos definir una matriz de pesos estandarizada por filas $\mathbf{W}^s$, cuyo elemento w_{ij}^s viene dado por:

$$w_{ij}^s = \frac{w_{ij}}{\sum_j w_{ij}} \tag{12}$$

Matriz de pesos de fila estandarizada

Esto asegura que todos los pesos estén entre 0 y 1 y facilita la interpretación de la operación con la matriz de pesos como un promedio de los valores vecinos como veremos a continuación. La matriz de ponderaciones estandarizadas por filas también garantiza que los parámetros espaciales en muchos procesos estocásticos espaciales sean comparables entre modelos (Anselin y Bera, 1998).

Otra característica importante es que, bajo la estandarización por filas, el elemento de cada fila suma a la unidad y la suma de todos los pesos, $S_0 = \sum_i \sum_j w_{ij} = n$, el número total de observaciones. Esta es una buena interpretación que exploraremos más adelante.

Otra cuestión importante es la simetría. Como hemos aprendido, algunas matrices de ponderación espacial son simétricas. Una característica importante de la matriz simétrica es que todas sus raíces características son reales. Sin embargo, después de la estandarización por filas, las matrices ya no son simétricas.

Matriz de pesos de fila estandarizada

La matriz estandarizada por filas también se conoce en la literatura como matriz estocástica por filas:

Definición

Matriz estocástica por filas: *Una matriz real* $n \times n$ $\mathbf{A}$ *es llamada matriz de Markov o matriz estocástica por filas si:*

1. $a_{ij} \geq 0$ *Para* $1 \leq i, j \leq n_i$
2. $\sum_{j=1}^{n} \alpha_{ij} = 1$ *Para* $1 \leq i \leq n$

Matriz de pesos de fila estandarizada

Una característica de la matriz estocástica por fila es relativa para sus eigen valores:

Teorema

Eigenvalores para la matriz estocástica por fila*: Cada eigenvalor w_i de una matriz estocástica por fila satisface $|w| \leq 1$*

Por lo tanto, los valores propios de la matriz de vecindad estocástica por filas (es decir, normalizada por filas, estandarizada por filas o de Markov) $\mathbf{W}^s$ están en el rango $[1, +1]$.

Finalmente, el comportamiento de $\mathbf{W}^s$ es importante para las propiedades asintóticas de los estimadores y estadísticos de prueba (Anselin y Bera, 1998, pp. 244). En particular, la matriz $\mathbf{W}$ también debe ser exógena, a menos que la endogeneidad se considere explícitamente en la especificación del modelo.

Variables espaciales rezagadas

Ahora que se ha discutido sobre la matriz de pesos espaciales, se puede crear la llamada variable espacial rezagada u operador de rezago espacial. El operador de rezago espacial toma la forma $\mathbf{y_L} = \mathbf{Wy}$ con dimensión $n \times 1$, donde cada elemento es dado por $\mathbf{y}_{Li} = \sum_j w_{ij} y_j$, $i.e$, como un peso promedio de los valores $\mathbf{y}$ en el vecino de i. Por ejemplo:

$$\mathbf{Wy} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 10 \\ 50 \\ 30 \end{bmatrix} = \begin{bmatrix} 50 \\ 10 + 30 \\ 50 \end{bmatrix}$$

Variables espaciales rezagadas

Usando una matriz de pesos estandarizada por fila:

$$\mathbf{Wy} = \begin{bmatrix} 0 & 1 & 0 \\ 0{,}5 & 0 & 0{,}5 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 10 \\ 50 \\ 30 \end{bmatrix} = \begin{bmatrix} 50 \\ 5 + 15 \\ 50 \end{bmatrix}$$

Variables espaciales rezagadas

Como resultado para la unidad espacial i, el rezago espacial de y_i referido como $\mathbf{y_{Li}}$ (la variable W_y observada para el lugar de i) es:

$$\mathbf{y_{Li}} = w_{i,1}y_i + w_{i,2}y_2 + \cdots + w_{i,n}y_i$$

o,

$$\mathbf{y_{Li}} = \sum_{j=1}^{n} w_{i,j}y_j$$

Donde los pesos w_{ij} consisten de los elementos de la i-ésima fila de la matriz $\mathbf{W}$, emparejado con los elementos correspondientes del vector de . En otras palabras, esto es una suma de pesos de los valores observados en ubicaciones vecinas, ya que los no vecinos no esta incluidos.

Espacial de orden superior

Hasta ahora hemos aprendido cómo definir el espacio geográfico por la matriz $\mathbf{W}$. Sin embargo, una pregunta interesante es cómo definir los vecinos de orden superior. Por ejemplo, nos puede interesar definir los vecinos de los vecinos de una unidad espacial. O incluso nos pueden interesar los vecinos de vecinos de vecinos de unidad espacial i. Para discutir este caso interesante, necesitamos definir matrices de peso espacial de orden superior.

Definimos la matriz de peso espacial de orden superior l como $\mathbf{W^l}$. Entonces, por ejemplo, el peso espacial de orden $l = 2$ viene dado por $\mathbf{W^2} = \mathbf{WW}$, la matriz de peso espacial de orden $l = 3$ viene dada por $\mathbf{W^3} = \mathbf{WWW}$, y así sucesivamente. ¿Cuál es el significado del elemento w_{ij} en este caso? Para pesos espaciales de orden 2, el elemento w_{ij} de la matriz de pesos es 1 si el polígono j es adyacente a los vecinos de primer orden del polígono i y es 0 en caso contrario. Así, para los pesos vecinos espaciales de orden n, el elemento w_{ij} de la matriz de pesos $\mathbf{W}$ es 1 si el polígono j es adyacente a los vecinos de orden $n1$ del polígono i, y es 0 en caso contrario.

Espacial de orden superior

Para ilustrar estos puntos, considere la siguiente estructura espacial:

$$\mathbf{W} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \tag{12}$$

Cuando $\mathbf{W}^2 = \mathbf{W}\mathbf{W}$ basado en la matriz de contigüidad de primer orden de 5×5 de $\mathbf{W}$ para (12) es:

$$\mathbf{W}^2 = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 2 & 0 & 1 & 0 \\ 1 & 0 & 2 & 0 & 1 \\ 0 & 1 & 0 & 2 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix} \tag{13}$$

Tenga en cuenta que para la región $R1$, los vecinos de segundo orden son las regiones $R1$ y $R3$. Es decir, la región $R1$ es vecina de segundo orden tanto de sí misma como de la región R3, que es vecina de la región vecina $R2$.

Ahora considere $R2$. El primer panel de la Figura 1.7 muestra los vecinos de primer orden de $R2$ dados por la matriz de ponderación espacial en (12): los vecinos de primer orden son $R1$ y $R3$. El panel B considera los vecinos de segundo orden de $R2$: Los vecinos de segundo orden son el propio $R2$ y $R4$. Para entender esto, tenga en cuenta que hay un efecto de retroalimentación del primer impacto de $R2$ proveniente de $R1$ y $R3$ (vecinos de primer orden de $R2$). Esto explica por qué el elemento $w_{22}^2 = 2$. Además, hay un efecto indirecto proveniente de $R4$ a través de $R3$ que finalmente impacta en $R2$. Esto representa el valor de 1 para el elemento w_{24}^2.

De manera similar, para la región $R3$, los vecinos de segundo orden son las regiones $R1$ (que es vecina de la región vecina $R2$), $R3$ (vecina de segundo orden consigo misma) y $R5$ (que es vecina de la región vecina $R4$).

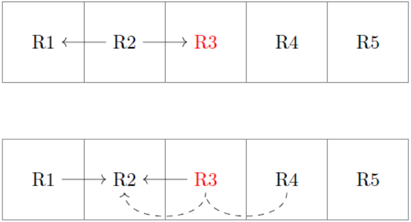

Figura: 1.6.- Vecinos de orden superior.

Los vecinos de tercer orden son:

$$\mathbf{W}^3 = \begin{bmatrix} 0 & 2 & 0 & 1 & 0 \\ 2 & 0 & 3 & 0 & 1 \\ 0 & 3 & 0 & 3 & 0 \\ 1 & 0 & 3 & 0 & 2 \\ 0 & 1 & 0 & 2 & 1 \end{bmatrix} \tag{14}$$

Ejemplos de diferentes pesos espaciales

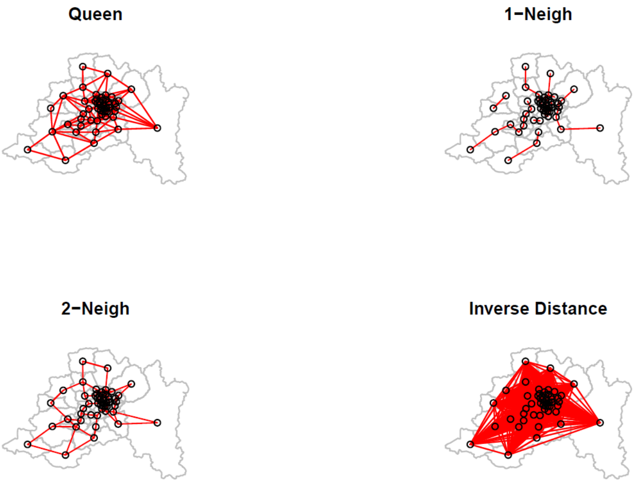

Figura: 1.7.- Tipos de pesos espaciales.

Referencias

- Arbia, G. (2014). *A Primer for Spatial Econometrics: With Applications in R* (1st ed.). Palfrave mcmillan, India.
- LeSage, P. J. and Kelley, P. R. (2009). *Introduction to Spatial Econometrics* (1st ed.). CRC Press, USA.
- Paul, E. J. (2014). *Spatial Econometrics: From Corss-Sectional Data to Spatial Panels* (1st ed.). Springer, London.
- Sarrias. M. (2020). *Notes on Spatial Econometrics*. Universidad de Talca, Chile.