PUCP

DEPARTAMENTO DE ECONOMÍA

LABORATORIO DE ECONOMETRÍA: STATA

1ECO31

Sesión 10
Análisis Exploratorio de Datos Espaciales (Clusters)

Docente: Juan Palomino

Índice

1. Matrices Espaciales

Matrices Espaciales

La disponibilidad de polígonos permite la construcción de matrices de pesos espaciales basados en:

- Contiguidad
- Vecinos más cercanos
- Distancias

Estas matrices espaciales permiten crear indicadores de autocorrelación espacial para la identificación de clusters de fenómenos socioeconómicos a nivel territorial.

1.1 Matriz Espacial basada en Contiguidad

Matriz Espacial basado en Contiguidad

La especificación de la relación de contigüidad en la matriz de peso espacial es la siguiente:

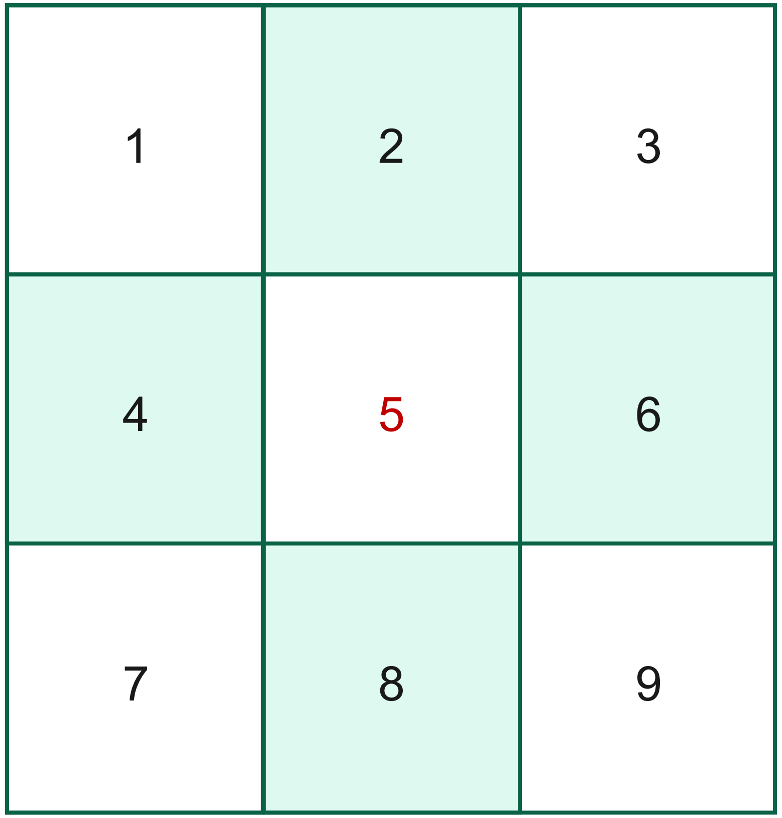
$$w_{ij} = \begin{cases} 1 \ si & región\ i\ y\ región\ j\ son\ contigüas \\ 0 \ si & región\ i\ y\ región\ j\ no\ son\ contigüas \end{cases}$$

Existen diferentes criterios binarios:

- Criterio de roca (Borde común)
- Criterio alfil (Vértice común)
- Criterio de reina (Borde y Vértice común)

Matriz Espacial basado en Contiguidad Roca

¿Cuáles son los vecinos de la región 5?



$$
W = \begin{array}{c} \\ R1 \\ R2 \\ R3 \\ R4 \\ R5 \\ R6 \\ R7 \\ R8 \\ R9 \end{array}
\begin{array}{c}
\begin{array}{ccccccccc} R1 & R2 & R3 & R4 & R5 & R6 & R7 & R8 & R9 \end{array} \\
\begin{bmatrix}
0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0
\end{bmatrix}
\end{array}
$$

El borde común son las regiones 2, 4, 6 y 8.

Matriz Espacial basado en Contiguidad Alfil

¿Cuáles son los vecinos de la región 5?

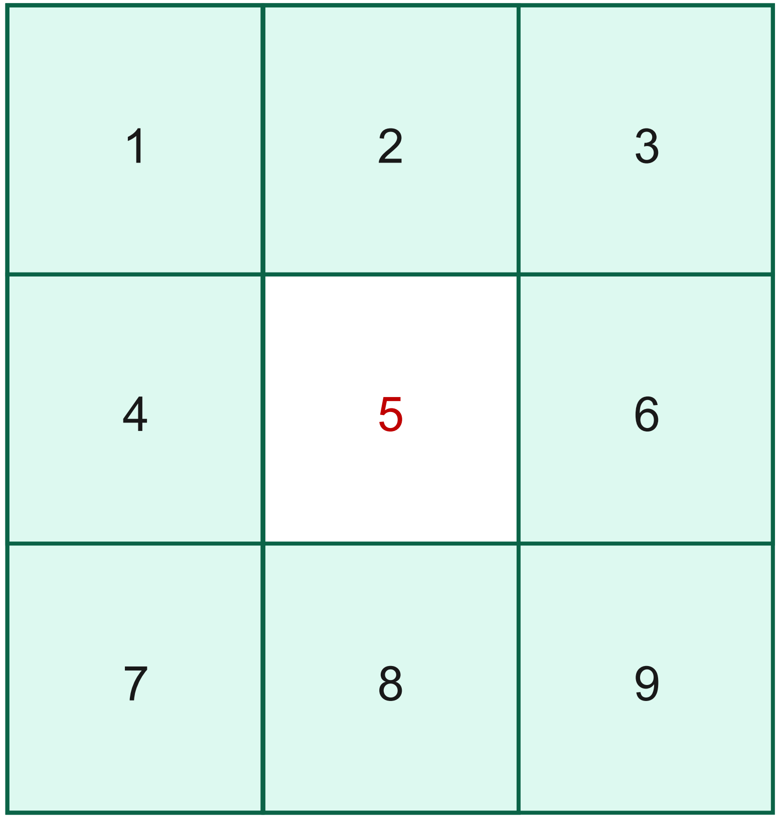
1	2	3
4	5	6
7	8	9

$$W = \begin{array}{c} \\ R1 \\ R2 \\ R3 \\ R4 \\ R5 \\ R6 \\ R7 \\ R8 \\ R9 \end{array} \begin{array}{c} \begin{array}{ccccccccc} R1 & R2 & R3 & R4 & R5 & R6 & R7 & R8 & R9 \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

El vértice común son las regiones 1, 3, 7 y 9.

Matriz Espacial basado en Contiguidad Reina

¿Cuáles son los vecinos de la región 5?



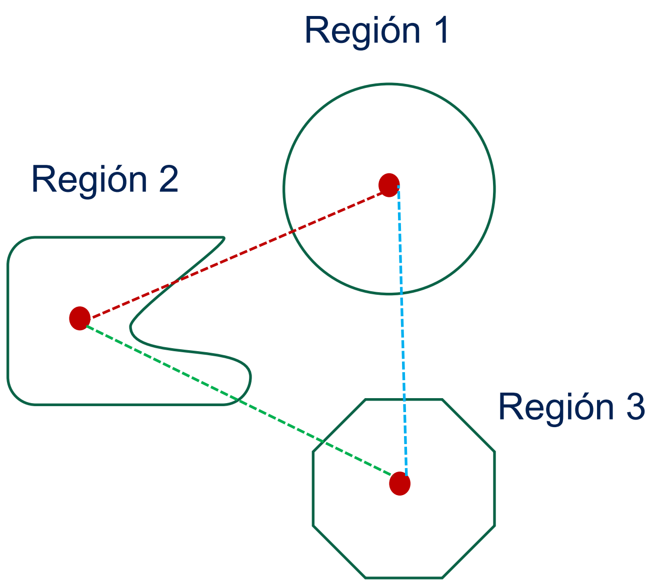
$$
W = \begin{array}{c} \\ R1 \\ R2 \\ R3 \\ R4 \\ \color{red}{R5} \\ R6 \\ R7 \\ R8 \\ R9 \end{array}
\begin{array}{c}
\begin{array}{ccccccccc} R1 & R2 & R3 & R4 & R5 & R6 & R7 & R8 & R9 \end{array} \\
\begin{bmatrix}
0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\
\color{red}{1} & \color{red}{1} & \color{red}{1} & \color{red}{1} & \color{red}{0} & \color{red}{1} & \color{red}{1} & \color{red}{1} & \color{red}{1} \\
0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0
\end{bmatrix}
\end{array}
$$

El vértice y borde común son las regiones 1, 2, 3, 4, 6, 7, 8, y 9.

1.2 Matriz Espacial basada en Distancia

Criterio de Distancia

- Los pesos también pueden definirse como una función de la distancia entre las regiones i y j, d_{ij}.
- d_{ij} se suele calcular como la distancia entre sus centroides (u otra unidad importante).
- Sean x_i y x_j la longitud e y_i y y_j las coordenadas de latitud para las regiones i y j, respectivamente.



Métricas de Distancia

Métrica Minkowski

Sean dos puntos i y j, con coordenadas respectivas (x_i, y_i) y (x_j, y_j):

$$d_{ij}^{p} = (|x_i - x_j|)^{\rho} + (|y_i - y_j|)^{\rho}$$

Métrica Euclidiana

Estableciendo $\rho = 2$ en la métrica Minkowski, tenemos:

$$d_{ij}^{p} = \sqrt{(|x_i - x_j|)^2 + (|y_i - y_j|)^2}$$

Métrica Manhattan

Estableciendo $\rho = 1$ en la métrica Minkowski, tenemos:

$$d_{ij}^{p} = (|x_i - x_j|) + (|y_i - y_j|)$$

Métricas de Distancia

Distancia Gran Circulo

La distancia euclidiana no es necesariamente la distancia más corta si se tiene en cuenta la curvatura de la tierra.

Sean dos puntos i y j, con coordenadas respectivas (x_i, y_i) y (x_j, y_j):

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

donde r es el radio de la Tierra.

La distancia arco es obtenida en millas con $r = 3{,}959$ y en kilómetros con $r = 6{,}371$

Matriz Espacial basada en Distancia

Distancia Euclidiana

$$w_{ij} = \begin{cases} d_{ij}\ si & i \neq j \\ 0\ \ si & i = j \end{cases}$$

Distancia Inversa

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

Típicamente, $\alpha = 1$ o $\alpha = 2$.

Distancia Exponencial Negativo

$$w_{ij} = \begin{cases} \exp(-\dfrac{d_{ij}}{\alpha}) & si \quad i = j \\ 0 & si \quad i \neq j \end{cases}$$

Típicamente, $\alpha = 1$ o $\alpha = 2$.

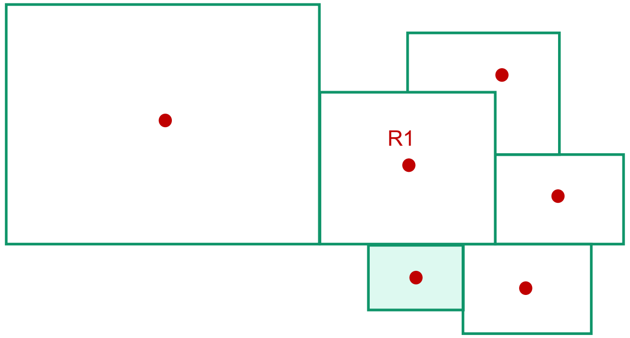
1.3 Matriz Espacial basada en k-nearest neighbor

Matriz Espacial basada en k-nearest neighbor

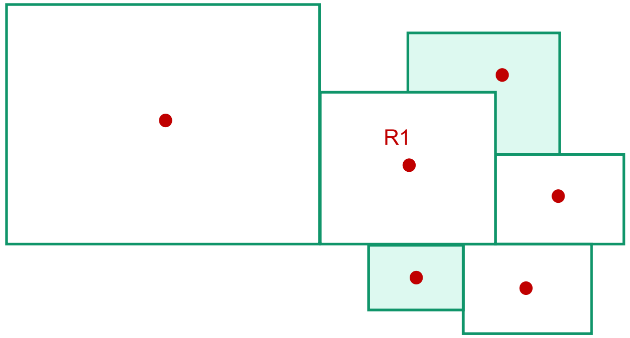
Vecinos k más cercanos: nosotros explicitamos el límite de número de vecinos.

$$w_{ij} = \begin{cases} 1 \;\; si & centroide\ de\ j\ es\ uno\ de\ los\ k\ centroides\ más\ cercanos\ a\ i \\ 0 \;\; si & otra\ manera \end{cases}$$

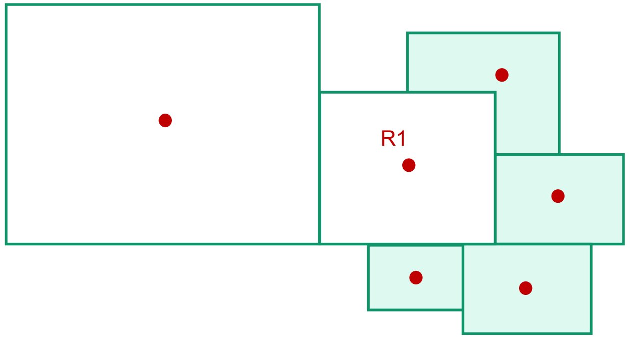
$k = 1$



$k = 2$



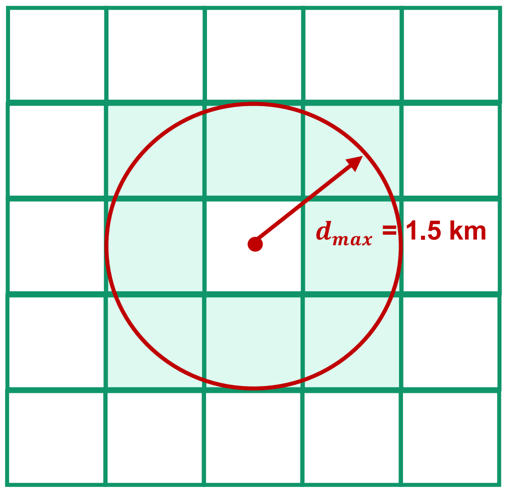
$k = 4$

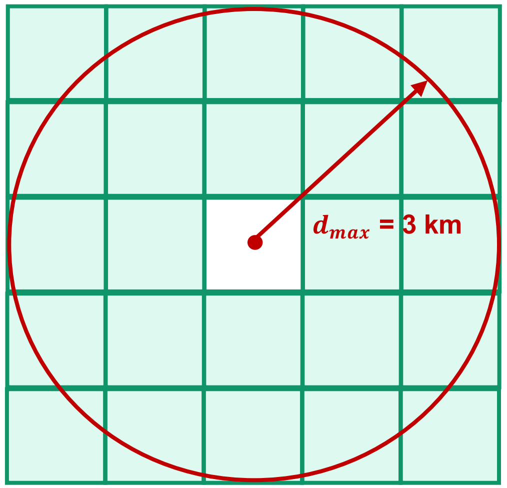


Matriz Espacial basada en umbrales de distancia

Umbral de distancia: especifica que una región i es vecina de j si la distancia entre ellos es menor que una distancia máxima especificada.

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





2. Rezago Espacial

Matriz Estandarizada

- Las matrices W se utilizan para calcular los promedios ponderados en los que se coloca más peso en observaciones cercanas que en observaciones distantes.
- Los elementos de una matriz de pesos estandarizados por filas son iguales a:

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

- 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.
- Bajo la estandarización de filas:
  - La suma de elementos de cada fila es igual a 1.
  - Las matrices ya no son simétricas

Matriz no estandarizada

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

Matriz estandarizada

$$W = \begin{bmatrix} 0 & 0.33 & 0 & 0.33 & 0.33 & 0 & 0 & 0 & 0 \\ 0.2 & 0 & 0.2 & 0.2 & 0.2 & 0.2 & 0 & 0 & 0 \\ 0 & 0.33 & 0 & 0 & 0.33 & 0.33 & 0 & 0 & 0 \\ 0.2 & 0.2 & 0 & 0 & 0.2 & 0 & 0.2 & 0.2 & 0 \\ 0.125 & 0.125 & 0.125 & 0.125 & 0 & 0.125 & 0.125 & 0.125 & 0.125 \\ 0 & 0.2 & 0.2 & 0 & 0.2 & 0 & 0 & 0.2 & 0.2 \\ 0 & 0 & 0 & 0.33 & 0.33 & 0 & 0 & 0.33 & 0 \\ 0 & 0 & 0 & 0.2 & 0.2 & 0.2 & 0.2 & 0 & 0.2 \\ 0 & 0 & 0 & 0 & 0.33 & 0.33 & 0 & 0.33 & 0 \end{bmatrix}$$

Suma

=1
=1
=1
=1
=1
=1
=1
=1
=1

Rezago Espacial

- El operador del rezago espacial toma la forma $y_L = Wy$ con dimensión $n \times 1$, donde cada elemento viene dado por $yL_i = \sum_j w_{ij} y_j$, es decir, un promedio ponderado de los valores "y" en el vecino de i.

- Por ejemplo:

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

- Usando una matriz de pesos estandarizado por fila:

$$Wy = \begin{pmatrix} 0 & 1 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 10 \\ 50 \\ 30 \end{pmatrix} = \begin{pmatrix} 50 \\ 5 + 15 \\ 50 \end{pmatrix}$$

- Asimismo, cuando W es estandarizado, cada elemento $(Wy)_i$ es interpretado como un promedio ponderado de los valores "y" para los vecinos i.

3. Indicadores de Autocorrelación Espacial

Indicadores de Autocorrelación Espacial

Uno de los elementos centrales del AEDE se enfoca en los indicadores de autocorrelación espacial. Anselin (1996) ha generado dos taxonomías de autocorrelación espacial:

Autocorrelación Espacial Global

Es una medida de clustering general en los datos. Asume que solo un estadístico resume el área de estudio (Homogeneidad):

- Moran's I
- Gery's C
- Getis and Ord's G(d)

Autocorrelación Espacial Local

Una medida de autocorrelación espacial para cada locación individual.

- Indices Locales para el Análisis Espacial (LISA)

3.1 Autocorrelación Espacial Global

Autocorrelación Espacial Global

El principal estadístico empleado para autocorrelación espacial global es el I de Moran.

$$I = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \, (x_i - \bar{x})(x_j - \bar{x})}{S_0 \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Donde $S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}$ y w_{ij} es un elemento de la matriz de pesos espacial que mide la distancia espacial o conectividad entre las regiones i y j. En forma matricial:

$$I = \frac{n \, z'Wz}{S_0 z'z}$$

Donde $z = x - \bar{x}$. Si la matriz W es estandarizada por fila, entonces:

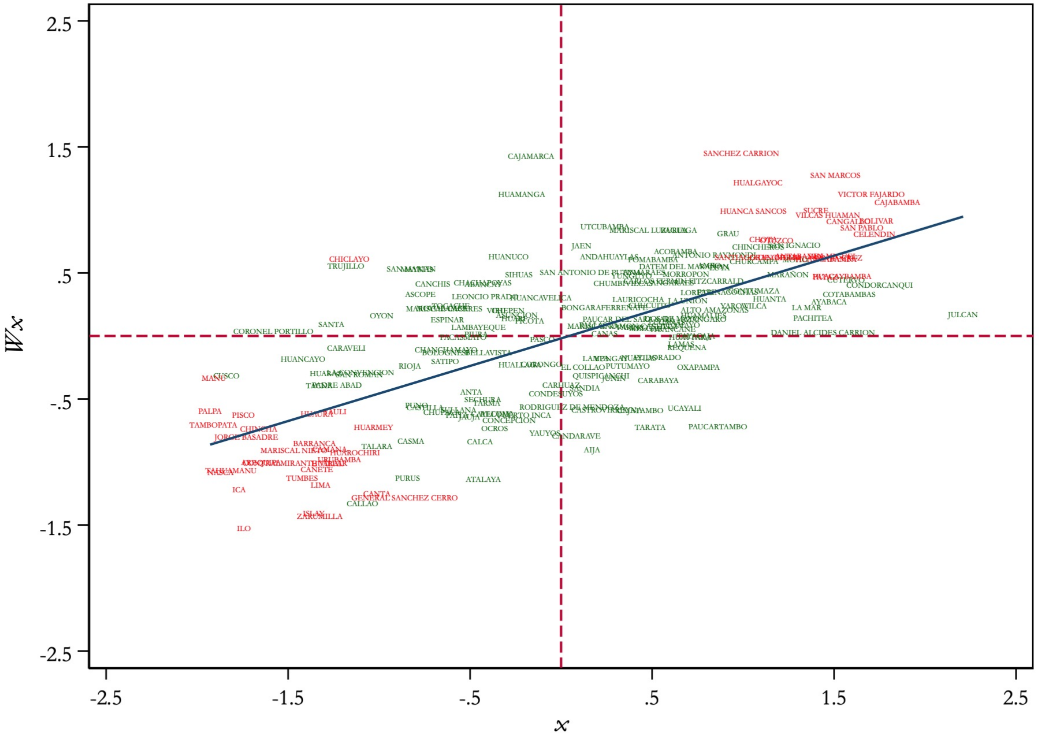
$$I = \frac{z'Wz}{z'z}$$

Porque $S_0 = n$. El rango de valores va desde -1 (dispersión perfecta) a +1 (correlación perfecta). Un valor de cero indica un patrón espacial aleatorio.

Diagrama de Dispersión de Moran

- El Scatter plot de Moran es una representación de la distribución de la variable frente a su rezago espacial, en el eje x es la variable normalizada X, y en el eje "y" el rezago espacial normalizado.
- Se divide en cuatro cuadrantes:

Cuadrante II (negativo X, positivo WX)



Cuadrante I (positivo X, positivo WX)

Cuadrante III (negativo X, negativo WX)

Cuadrante IV (positivo X, negativo WX)

3.2 Autocorrelación Espacial Local

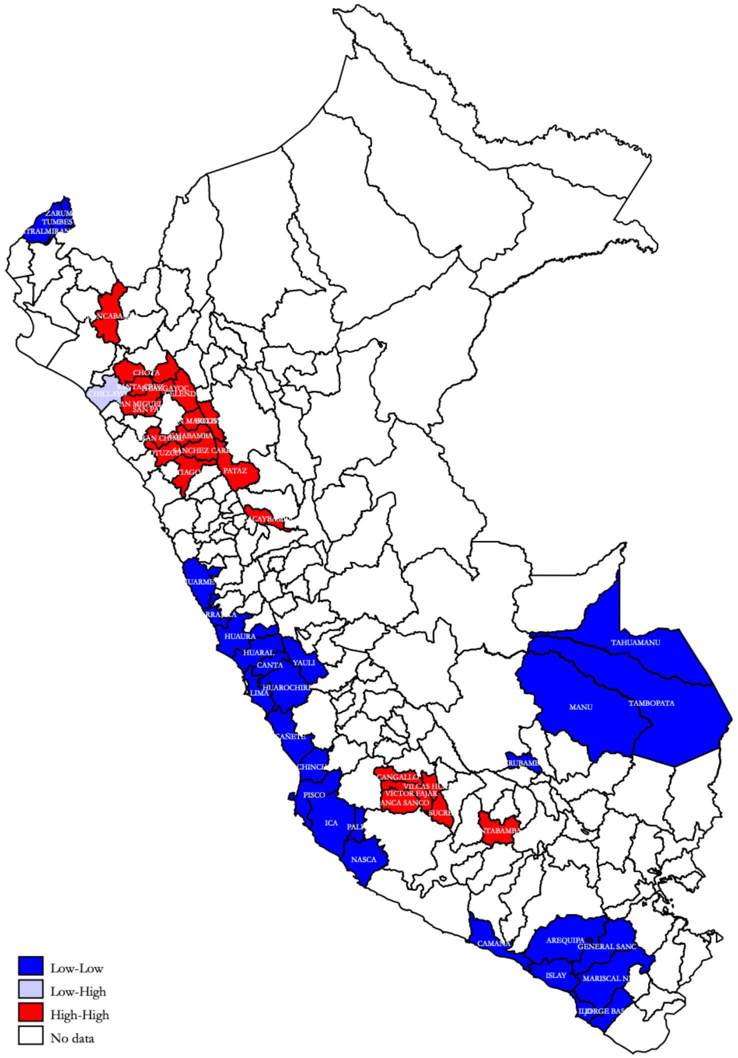
Autocorrelación Espacial Local

- Los indicadores globales resumen la información dado un conjunto de observaciones espaciales, pero no son capaces de identificar cluster.
- La identificación de los cluster es crucial para poder dar cuenta de estructuras espaciales en la distribución de las observaciones en el espacio.
- Anselin (1995) formaliza los indicadores locales, son bastante similares a los globales pero la matriz W solo es considerada para el vecindaria i.

Autocorrelación Espacial Local

- Detecta la contribución de cada localización a la autocorrelación espacial global.
- Los indicadores de autocorrelación espacial local son usados para identificar puntos calientes / frios (hot/cold spot): concentraciones espacial de altos/bajos valores u outliers espaciales.
- I de Moran local:

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



PUCP