

Introducción al análisis espacial de datos socioeconómicos usando R

Módulo 1

Mauricio Quiñones Domínguez
Ph.D. Política Pública
Profesor Dpto. de Economía y Finanzas

Laboratorio de Economía Aplicada (LEA)

Pontificia Universidad Javeriana de Cali

2025-04-29

- 1** Presentación del Curso
- 2** Motivación del curso
- 3** Heterogeneidad espacial vs Autocorrelación Espacial
- 4** Autocorrelación espacial: matriz de pesos
- 5** Aplicación: Caso de estudio
- 6** Análisis de modelación espacial

Presentación del Curso

Presentación del Curso

- La dimensión espacial que es objeto de análisis en otras disciplinas, afecta también las variables socioeconómicas. Al integrar la distancia geográfica con la modelación econométrica, se pueden refinar las estimaciones y capturar efectos que antes no eran tenidos en cuenta.

Objetivos del curso

- 1 Presentar por qué se considera que se debe tener en cuenta la distribución en el espacio para la modelación econométrica.
- 2 Definir y construir una matriz de pesos espaciales.
- 3 Interpretar los resultados e índices que componen el análisis de modelación espacial.
- 4 Conocer algunos de los modelos más frecuentes para la modelación espacial (Modelos: SAR, SEM y otros).
- 5 Presentar cómo se identifica el tipo de modelación espacial.

Metodología de aprendizaje

- Metodología de aprendizaje-enseñanza: “Aprender haciendo”
- Los datos analizados son de la encuesta de origen-destino (EOD) para Bogotá del 2019.
- Compartimos shape file y archivo de excel con atributos.

Sugerencia:

- Guarda los archivos y tu trabajo en 1 carpeta.

Motivación del curso

Motivación del curso

Pregunta:

¿Por qué se considera que se debe tener en cuenta la distribución en el espacio para la modelación econométrica?

Dependencia espacial

- i. Primera ley de Geografía de Tobler: “Todo está relacionado con todo lo demás, pero las cosas cercanas están más relacionadas que las distantes”. (1970: 236).

Ley de gravitación universal

$$F = G \frac{m_1 m_2}{d^2}$$

Labels in the diagram:

- fuera de atracción (pointing to F)
- constante de gravitación universal (pointing to G)
- masa del cuerpo 1 (pointing to m₁)
- masa del cuerpo 2 (pointing to m₂)
- dividido entre (pointing to the denominator d²)
- cuadrado (pointing to the 2 in d²)
- distancia entre los cuerpos (pointing to d)

Autocorrelación espacial. Las áreas vecinas son similares (Tobler, 1970)

- ii. Las áreas de observación son generalmente de diferentes tamaños.
Heterogeneidad → Heterocedasticidad.

Motivación del curso (cont.)

LOCALIDADES BOGOTÁ

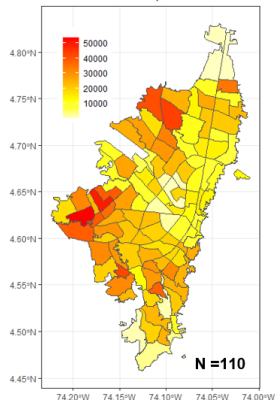


N = 110

UTAN: Unidad Territorial
de Análisis de Movilidad

Mapa indicativo de los polígonos de distintos tamaños

Número de habitantes por km2



Heterogeneidad espacial vs Autocorrelación Espacial

Heterogeneidad espacial vs Autocorrelación Espacial

- Los efectos espaciales son propiedades de los datos espaciales que dan como resultado la tendencia de que las observaciones espacialmente próximas de un atributo sean más parecidas que las observaciones más distantes (primera ley de Tobler). Esta agrupación (clustering) en el espacio puede resultar de:
 - Propiedades inherentes de algunas regiones en el área de estudio que hacen que las unidades analíticas dentro de ellas sean diferentes de las de otras regiones (*heterogeneidad espacial*) o
 - Algún tipo de interacción espacialmente modelada entre unidades vecinas (*autocorrelación espacial*) o
 - Ambos.

Heterogeneidad espacial

- Tipificados por diferenciación de la unidad geográfica o proceso espacial que se expresa a lo largo de toda la unidad geográfica bajo estudio.
- **No se asume ninguna interacción espacial** en el proceso que genera heterogeneidad espacial; se desprende de lo que hace única intrínsecamente a cada ubicación o en propiedades compartidas.
- Las observaciones no están distribuidas homogéneamente en el espacio.
- Variación a gran escala de la media en las unidades espaciales.
- Los valores de las variables están influenciados por su ubicación en el espacio.
- Resultado de la interacción de las características únicas de las unidades y su ubicación espacial.

Causas de la heterogeneidad espacial

- Patrones de interacción social que crean características únicas de las unidades espaciales.
 - Regímenes espaciales: legados de las relaciones regionales entre el centro y la periferia → diferencias entre unidades (población, PIB, etc.)
- Diferencias en las características físicas de las unidades espaciales. Tamaño y topografía diferentes.
- *Si asumimos que la heterogeneidad espacial es un proceso razonable de generación de datos entonces ...*
 - La aparente agrupación de datos **no** se debe a interacción espacial entre áreas.
 - No hay influencias vecinas a pequeña escala ni efecto de contagio.
 - Un modelo de regresión con covariables adecuadas puede eliminar la autocorrelación residual.
 - Los efectos estructurales puramente espaciales pueden incorporarse en la especificación del modelo.

Autocorrelación Espacial

- “... la existencia de una relación funcional entre lo que sucede en un punto del espacio y lo que sucede en otro lugar” (Anselin, 1988:11).
- Esto significa una falta de independencia entre las observaciones (por definición); pero la “relación funcional” es la clave; interacción entre actores sociales/económicos.
- En nuestros modelos, se expresa como un proceso espacial de corta distancia y localizado.
- Para el examen de datos en una red espacial irregular (por ejemplo, UTAM), este proceso espacial generalmente se maneja a través de la declaración exógena de un “**vecino**” (una región de influencia) definida para cada observación y se operacionaliza mediante una “**matriz de ponderaciones**”.
- Este proceso se desprende informalmente de la llamada “Primera Ley de la Geografía”; viola el supuesto de unidades de análisis independientes.

Autocorrelación Espacial (Cont.)

- *Si asumimos que la dependencia espacial es un proceso razonable de generación de datos entonces . . .*
 - Estamos diciendo que la agrupación de nuestros datos resulta de algún tipo de interacción espacial; existencia de influencias vecinas de pequeña escala; (definición de vecindario; elección de matriz de ponderaciones).
 - Creemos que podemos plantear teóricamente razones por las que la agrupación resulta de la interacción espacial o del contagio entre nuestras unidades de observación.
 - Necesitamos pensar en qué tipo de modelo de dependencia debería introducir mejor los efectos vecinos; muchas opciones.
- Covariación localizada entre las medias (u otras estadísticas) dentro de la región.
- Tendencia de las medias a “seguirse” en el espacio.
- Da lugar a grupos de valores similares.

Causas de la autocorrelación espacial

- El proceso socioeconómico subyacente ha dado lugar a una distribución agrupada de los valores de las variables. Procesos de **agrupación**.
- Agrupación de personas similares en zonas localizadas. Procesos de **interacción espacial**.
- Las personas cercanas tienen más probabilidades de interactuar, compartir. Procesos de **difusión**.
- Los vecinos aprenden unos de otros. Procesos de **dispersión**.
- La gente se desplaza, pero suelen ser distancias cortas, se llevan consigo sus conocimientos. **Jerarquías espaciales**.
- Influencias económicas que unen a las personas. Desajuste entre el proceso y las unidades espaciales, distritos frente a zonas de comercio minorista o grupos de bloques censales frente a redes de barrios.

Desafíos de la autocorrelación espacial

- No se deben interpretar correlaciones a nivel grupal como si fueran aplicables a nivel individual. Robinson, W.S. (1950). “Ecological Correlations and the Behavior of Individuals”. American Sociological Review. Robinson → **“Falacia ecológica”**.
- **Modifiable Analysis Unit Problem o MAUP (Problema de la unidad de área modificable)** - “Los usuarios habituales de correlaciones ecológicas saben que el tamaño del coeficiente depende en gran medida del número de subáreas. . . . [E]l tamaño de la correlación ecológica . . . aumentará numéricamente a medida que se produzca la consolidación de áreas más pequeñas en áreas más grandes. . . ”. (Robinson, págs. 357-8).

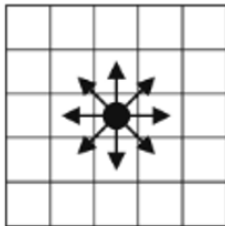
Autocorrelación espacial: matriz de pesos

Autocorrelación espacial: matriz de pesos

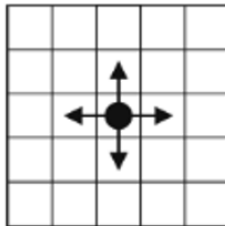
- La matriz de pesos espaciales W contiene las ponderaciones para cada unidad geográfica, con base en algún criterio de valoración. El insumo principal para su construcción es la matriz de **distancias**.
- La matriz de distancias calcula la cercanía entre las unidades espaciales analizadas. La medida de distancia más común es la distancia Euclidiana, pero no es la única.

Decisión 1: ¿Quién es vecino?

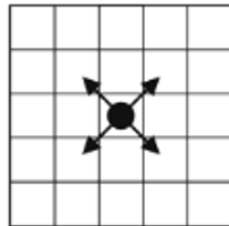
Matriz criterio Reina, torre o alfil. Estos términos se explican por sí solos y se refieren a qué tipos de celdas adyacentes elegimos incluir como “vecinas”.



Queen



Rook



Bishop

Usando la matriz reina, para una distancia definida, la matriz W se llena con ceros y unos para cumplir el criterio de cercanía.

Matriz de contigüidad

- W o “Matriz de contigüidad” – un término general que identifica vecinos con 1 y no vecinos con 0.

$$\begin{pmatrix} & R1 & R2 & R3 & R4 & R5 & R6 & R7 \\ R1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ R2 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ R3 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ R3 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ R5 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ R6 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ R7 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

- Luego se debe normalizar a la unidad por cada fila. **Matriz de ponderaciones:** reservaremos

- Por lo tanto, se obtiene una matriz W de dimensiones $N \times N$, en este caso 7×7 , que será usada para “rezagar” espacialmente las variables de interés, de acuerdo al tipo de modelación.

$$W = \begin{pmatrix} 0 & w_{12} & \cdots & w_{1j} & \cdots & w_{1N} \\ w_{21} & 0 & \cdots & w_{2j} & \cdots & w_{2N} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ w_{i1} & w_{i2} & \cdots & w_{ij} & \cdots & w_{iN} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ w_{N1} & w_{N2} & \cdots & w_{Nj} & \cdots & 0 \end{pmatrix}$$

- Se han generado los elementos w_{ij} basándose en la adyacencia.

Otras opciones:

- Hay muchas otras opciones; por ejemplo...
 - Distancia y distancia inversa.
 - k vecinos más cercanos (k nearest neighbors - knn).
 - Pesos de Cliff-Ord.
-
- La matriz de ponderaciones debe basarse en lo posible en la teoría.
 - Muchas opciones en R.
 - Puede editar W en un editor de texto general.

Aplicación: Caso de estudio

Librerías en R:

Datos:

Análisis de modelación espacial

Medición de la autocorrelación espacial

- Dos clases de pruebas de autocorrelación espacial:
- **Medidas de autocorrelación espacial global**
 - Identifica si los datos en su conjunto presentan un patrón espacial
 - ¿Existe “agrupación”?
- Medida más común: estadístico I de Moran (Alfred Pierce Moran (1950)).
- **Estadísticos de indicadores locales de asociación espacial (Local indicators of spatial association - LISA)**
 - Identifica qué unidades de observación de nuestro conjunto de datos están significativamente autocorrelacionadas espacialmente con las unidades vecinas.
- La noción clave es medir que dos propiedades varían juntas. Cuánto importa la proximidad en los datos espaciales.

I de Moran Global

Estadístico I de Moran está inspirado en la covarianza, por lo tanto, tiene similitud con **el coeficiente de correlación de Pearson**. **Rango: [-1,1]**. Sin embargo, su interpretación no es igual.

Coeficiente de correlación de Pearson. Rango: [-1,1]

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

I de Moran global

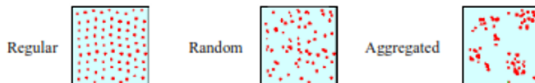
$$I = \left(\frac{n}{(\sum_{i=1}^n \sum_{j=1}^n w_{ij})} \right) \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{j=1}^n (y_i - \bar{y})^2}$$

Donde w_{ij} son los elementos de la matrix de pesos o ponderaciones W .

Autocorrelación espacial y aleatoriedad espacial

■ Aleatoriedad espacial

- Los valores observados en una ubicación no dependen de los valores observados en ubicaciones vecinas.
- El patrón espacial de valores observado es igual de probable que cualquier otro patrón espacial.
- La ubicación de los valores puede modificarse sin que ello afecte al contenido informativo de los datos.



Autocorrelación espacial y aleatoriedad espacial (cont.)

Valor esperado de la I de Moran bajo la hipótesis de ausencia de autocorrelación espacial:

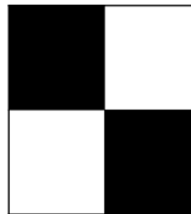
$$E(I) = -1/(n - 1)$$

$$Var(I) = \frac{n^2 S_1 - n S_2 + 3 S_0^2}{S_0^2 (n^2 - 1)} - (E(I))^2$$

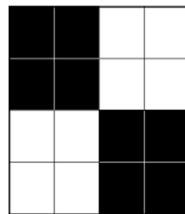
Donde S_0, S_1, S_2 son expresiones definidas que dependen de w_{ij} .

Modifiable Analysis Unit Problem o MAUP (Problema de unidad de área modificable)?.

Las medidas de autocorrelación espacial dependen de la escala



Moran's $I = -1.00$



Moran's $I = +0.33$

Inferencia

- **Hipótesis nula:** Ausencia de autocorrelación espacial

$$t = \frac{I - E(I)}{\sqrt{\text{Var}(I)}}$$

Si no puedo rechazar la hipótesis nula, entonces se debe modelar la autocorrelación espacial.

Un proceso no correlacionado tiene un I esperado = 0

- Valores > 0 indican autocorrelación espacial positiva.
- Valores < 0 indican autocorrelación espacial negativa.

Estadísticos de indicadores locales de asociación espacial (Local indicators of spatial association - LISA).

- Medidas locales
 - Valor calculado para cada observación; identificación de “puntos calientes” y “puntos fríos”.
- El estadístico local de Moran para unidad de área es:

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

- Donde z_i es una variable estandarizada

$$z_i = (x_i - \bar{x}) / SD_x$$

Diagrama de dispersión de Moran

- **Ubicaciones de asociación espacial positiva.** (“Soy parecido a mis vecinos”)
 - Q_1 (valores [+], valores cercanos [+]): **H-H**
 - Q_3 (valores [-], valores cercanos [-]): **L-L**
- **Ubicaciones de asociación espacial negativa.** (“Soy diferente a mis vecinos”)
 - Q_2 (valores [-], valores cercanos [+]): **L-H**
 - Q_4 (valores [+], valores cercanos [-]): **H-L**

