Introducción al análisis espacial de datos socioeconómicos usando R

Módulo 2

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

Laboratorio de Economía Aplicada (LEA)

Pontificia Universidad Javeriana de Cali

2025-04-30

Análisis exploratorio de datos espaciales

Análisis exploratorio de datos espaciales

El análisis de datos espaciales conduce a la estadística espacial. Estos son los tipos de datos espaciales

- Datos de eventos (datos de puntos).
- Datos espacialmente continuos (datos geoestadísticos).
- Datos de regionalizados.
- Datos de interacción espacial (datos de flujo).

Estos tipos de datos no son categorías estrictas ni mutuamente exclusivas.

- Cartografía y visualización - Histogramas, mapas de valores atípicos, diagramas de caja.
- Índices de asociación espacial / autocorrelación espacial.
- Análisis de patrones puntuales.
- El objetivo es buscar una buena comprensión y descripción de los datos, sugiriendo así hipótesis para explorar.
- Busque especialmente pistas sobre la "heterogeneidad espacial" o la "dependencia espacial".
- Pocas suposiciones a priori sobre los datos.
- Nuestro objetivo es modelar y extraer inferencias adecuadas sobre un proceso generador de datos aleatorio y no observado.

Modelos más frecuentes para la modelación espacial

Modelos más frecuentes para la modelación espacial

¿El objetivo es . . . ?

- Generalmente deseamos comprender (modelar) los aspectos estructurados de nuestros datos, dejando atrás un vector de ruido aleatorio.
- Datos = Estructura + Error
- Datos = Cond. estructura media + estructura *Var*/*Cov*

Punto de Partida: Modelo Mínimos Cuadrados Ordinarios (MCO)

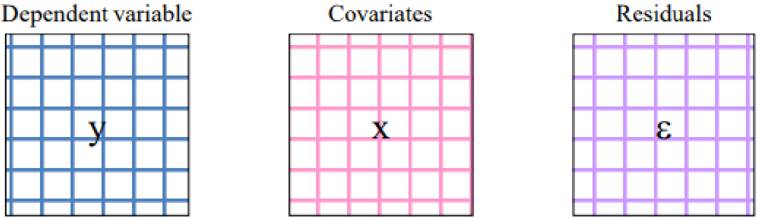


Figure 1: Regresión estándar MCO

MCO (cont.)

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_k x_{ki} + \varepsilon_i$$

En notación matricial

$$\boldsymbol{y} = \boldsymbol{X}\beta + \varepsilon$$

Donde

$$n \cdot 1 = n \cdot (k+1) * (k+1) \cdot 1 + n \cdot 1$$

Supuestos del modelo de regresión

El teorema de Gauss-Markov afirma que β es un “mejor estimador lineal insesgado” siempre que:

- **Supuestos:**
  - Independencia media: $E[\varepsilon_i|x_i] = 0$ para todo i
  - Homocedasticidad: $Var[\varepsilon_i|x_i] = \sigma^2$ para todo i
  - Perturbaciones no correlacionadas: $Cov[\varepsilon_i, \varepsilon_j] = 0$ para todo $i \neq j$
  - X es no estocástico: $Cov[X', \varepsilon] = 0$ para todo X
- **Otros que suelen incluirse:**
  - y normalmente distribuida
  - Linealidad
  - Sin multicolinealidad
  - ε_i son normales (perturbación normal)

Si se incumple el supuesto de errores Independientes e Identicamente Distribuidos (**iid**) en MCO, la inferencia estadística no es válida.

Consecuencias de MCO

Sin considerar la estructura espacial

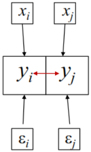
- Los coeficientes MCO de regresión estimados están sesgados e inconsistentes.
- Los coeficientes MCO de regresión estimados son ineficientes.
- El R^2 es exagerado.
- Las inferencias no son correctas.

Modelo espaciales

Modelo espacial autorregresivo (SAR)

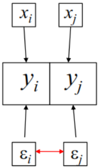
Incorpora dependencia espacial agregando un “rezago espacial” a la variable dependiente en el lado derecho de la ecuación de regresión y trata la correlación espacial como un proceso o efecto de interés.

- Los valores de y en un área están directamente influenciados por los valores de y encontrados en áreas vecinas.
- Depende de cómo definimos el vecindario.



Modelo de error espacial (SEM)

- Examina la autocorrelación espacial entre los residuos de áreas adyacentes.
- Trata la correlación espacial principalmente como un problema.
  - Desestima la idea de que la correlación espacial pueda reflejar algún proceso significativo.
- El error espacial positivo puede reflejar:
  - Un modelo mal especificado (particularmente una variable omitida que está espacialmente desordenada).
- Unidad espacial de agregación incorrecta.



SEM (cont.)

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{u}$$

$$\boldsymbol{u} = \lambda \boldsymbol{W}\boldsymbol{u} + \varepsilon$$

$$\boldsymbol{u} = [\boldsymbol{I} - \boldsymbol{W}\lambda]^{-1}\varepsilon$$

Reemplazo en estructural,

$$\boldsymbol{y} = \boldsymbol{X}\beta + [\boldsymbol{I} - \boldsymbol{W}\lambda]^{-1}\varepsilon$$

Efectos marginales son:

$$\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{X}} = \boldsymbol{I}\beta$$

Se estima por el método de Máxima Verosimilitud.

Test de identificación del tipo de modelación espacial

Test de identificación del tipo de modelación espacial

Elegir entre modelos: AIC y SC

- AIC = -2L + 2k
  donde L es la log-verosimilitud y k es el número de parámetros del modelo.
- SC = -2L + kln(N) (también conocido como BIC)
  **donde N es el número de observaciones.

Comentarios técnicos

Pasos para determinar el grado de autocorrelación espacial de los datos y ejecutar una regresión espacial:

1. Elija un criterio de vecindad ¿Qué zonas están relacionadas?
2. Asigne ponderaciones a las zonas que están vinculadas
3. Crear una matriz de ponderaciones espaciales
4. Ejecutar una prueba estadística para examinar la autocorrelación espacial
5. Ejecutar una regresión MCO
6. Determinar qué tipo de regresión espacial ejecutar
7. Ejecutar una regresión espacial
8. Aplicar la matriz de pesos

Otros modelos: SLX y SAC

Otros modelos: SLX y SAC

Modelo de rezago Espacial de las X's (SLX)

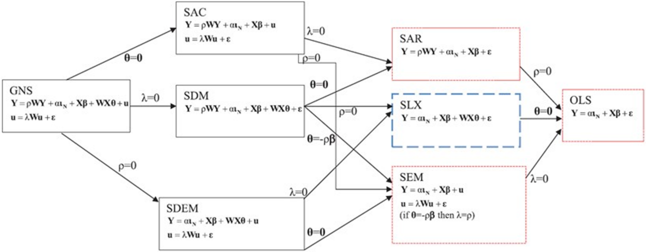
$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{WX}\theta + \varepsilon$$

Modelo Espacial Autorregresión combinado (SAC)

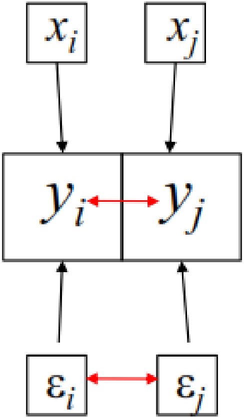
$$\boldsymbol{y} = \rho \boldsymbol{Wy} + \boldsymbol{X}\beta + \boldsymbol{u}$$

$$\boldsymbol{u} = \lambda \boldsymbol{Wu} + \varepsilon$$

Hay más


SAC
$Y = \rho WY + \alpha\iota_N + X\beta + u$
$u = \lambda Wu + \varepsilon$
GNS
$Y = \rho WY + \alpha\iota_N + X\beta + WX\theta + u$
$u = \lambda Wu + \varepsilon$
SDM
$Y = \rho WY + \alpha\iota_N + X\beta + WX\theta + \varepsilon$
SDEM
$Y = \alpha\iota_N + X\beta + WX\theta + u$
$u = \lambda Wu + \varepsilon$
SAR
$Y = \rho WY + \alpha\iota_N + X\beta + \varepsilon$
SLX
$Y = \alpha\iota_N + X\beta + WX\theta + \varepsilon$
SEM
$Y = \alpha\iota_N + X\beta + u$
$u = \lambda Wu + \varepsilon$
(if θ=-$\rho\beta$ then λ=ρ)
OLS
$Y = \alpha\iota_N + X\beta + \varepsilon$
θ=0
λ=0
ρ=0
θ=-$\rho\beta$

Hay más



Tipo de archivos necesarios:

Archivo Shape, extensión .shp. Contiene la información espacial más los atributos. Significa que formarán parte otros ficheros como shx, .dbf, y .prj. Desarrollado por ESRI

Referencias

- Root, E. Class Notes. (2015). Introduction to Spatial Regression Analysis. ICPSR Summer Workshop 2015.
- Dubé, J. , & Legros, D. (2014). Spatial econometrics using microdata.
- LeSage, J. P. , & Pace, R. K. (2009). Introduction to spatial econometrics. Boca Raton: CRC Press.
- Bivand, R. , E.J. Pebesma and V. Gomez-Rubio. (2013). Applied Spatial Data Analysis with R. New York: Springer.