

Spatial Autocorrelation in GeoDa

GeoDa is a software package developed by the Spatial Analysis Lab (Luc Anselin) at the University of Illinois. It is a collection of software tools designed for exploratory spatial data analysis (ESDA). GeoDa adheres to ESRI's shape files as the standard for storing the information. GeoDa is highly interactive based on dynamically linked windows. For example, you can select cases from any of several windows (histogram, scatter plot) and see those cases highlighted in the other active windows.

The software can be downloaded for free at <https://www.geoda.uiuc.edu/downloadin.php>
Extended documentations, tutorials, and support are also available at <https://www.geoda.uiuc.edu>.

Example 1: Child Poverty from U.S. 2000 Census

*Variables *See varlabels.xls for more details*

LOGPOV: Natural log of the odds ratio of the proportion in poverty

PFHH: Proportion of female-headed households

Example 1

cpoverty48.shp
FIPS

3,109 counties in
48 lower states

Contiguity-Based Spatial Weights: 1st order Rook, 1st order Queen

- Tools > Weights > Create

```
Input File: cpoverty48.shp
Output File: rook1.gal
Select ID: FIPS
Contiguity Weight: Rook
Order of contiguity: 1
```

The ID (key) variable must be specified to ensure a match between the data and the weights.

- Tools > Weights > Create

```
Input File: cpoverty48.shp
Output File: queen1.gal
Select ID: FIPS
Contiguity Weight: Queen
Order of contiguity: 1
```

The weight file can be opened and explored in WordPad.

Discussion and Other Tools

1. Select FIPS #35045 in the table (San Juan County, New Mexico). Look at the corresponding highlighted county on the map. How many neighbors do you count? How many neighbors are in each of the [rook and queen] weights files? How does this relate to the choice of weight matrix?

2. A connectivity histogram shows how many counties have 0,1,2,3... etc. neighbors. This reveals interesting features about connectivity such as how many areas are “islands” (have no neighbors) or whether the distribution of connectivity is bimodal (many counties with few neighbors and many counties with many neighbors). For the `rook1.gal` weights matrix, are there any islands? What are they?

- Tools > Weights > Properties

3. Higher-order contiguity weights matrices can also be easily constructed.

- Tools > Weights > Create

Input File: `cpoverty48.shp`
 Output File: `rook2.gal` OR `queen2.gal`
 Select ID: FIPS
 Contiguity Weight: Rook OR Queen
 Order of contiguity: 2

Check “Lower Order” box to include the 1st order neighbors.

Caution: Local Moran maps will not work for weight matrices that have islands. GeoDa will still report estimates, but they will not be correct.

Distance and K-nearest Neighbor Weights

- Tools > Weights > Create

Input File: `cpoverty48.shp`
 Output File: `mindist.gwt`
 Select ID: FIPS
 Distance metric: Euclidean

Check “Threshold Distance” to set minimum distance.

- Tools > Weights > Create

Input File: `cpoverty48.shp`
 Output File: `k6.gwt`
 Select ID: FIPS

Check “k-Nearest Neighbors”
 Number of neighbors: 6

The default minimum threshold is chosen to guarantee that every unit has at least one neighbor. Setting a lower value will result in islands

Caution: Nearest neighbor weights are asymmetric and cannot be used for spatial regression. GeoDa will still perform the analysis, but results will be inaccurate.

Discussion and Other Tools

1. The minimum threshold distance required to prevent islands may be too large for most points in the data set. K-Nearest Neighbors may be a better option in this case. This criterion ensures that every unit has exactly the same number (K) of neighbors.
2. Spatial weights can be used to smooth rates (as an alternative to Empirical Bayes smoothing).

Spatially Lagged Variables

A weights file must be opened first.

- Tools > Weights > Open → Navigate to queen1.gal
- *right click in Table* > Field Calculation → Select Lag Operations [Result: W_LOGPOV, Weight Matrix: DEFAULT, Variable: LOGPOV]

Discussion and Other Tools

1. Spatial autocorrelations can be explored by looking at a scatter plot of W_LOGPOV (Y variable) against LOGPOV (X variable). The slope of the regression line is the Global Moran's I statistic for LOGPOV using a given weights matrix.
2. The significance of the Moran I statistic cannot be assessed without constructing a null distribution.
3. Spatially lagged variables can alternatively be created "on the fly" using built-in autocorrelation functions.

Global Moran's I Statistic

- Space > Univariate Moran → Select LOGPOV and weights matrix queen1.gal
- *right click* > Save Results
- *right click* > Randomization > 999 permutations

Brushing and Exclude Selected tools can be used on the Moran Scatter plot.

Local Indicators of Spatial Autocorrelation (LISA) Maps

- Space > Univariate LISA → Select LOGPOV and weights matrix k6.gwt
- Select Cluster Map

Discussion and Other Tools

1. Both variables in the Moran scatter plot are standardized. The plot has four quadrants: high-high (upper right) and low-low (lower left) indicating positive spatial autocorrelation; and high-low (lower right) and low-high (upper left) indicating negative spatial autocorrelation.
2. The slope of the regression line is Moran's I. It simply tells us whether the variable of interest in an areal unit is associated with variable values in its neighbors (as defined by the choice of weight matrix).
3. The permutation test is used to assess the significance of the Moran's I statistic against the null hypothesis that there is no spatial autocorrelation. Small p-values indicate there is evidence of global spatial autocorrelation.
4. The Global Moran's statistic provides information about the presence and strength of any overall spatial association, but not about local cluster and spatial regimes.

- The LISA cluster map color codes significant local autocorrelations by type of association. The clusters displayed in the map likely extend to additional neighbors of the cluster.

Example 2: Lip Cancer in Scotland

Variables

CANCER: Count of cancer occurrences (event variable)

POP: Population (base variable)

Example 2

scotlip.shp
CODENO

56 districts in
Scotland



Exercises

- Construct a variable of raw cancer rates called R_RAWRATE.
- Create a 5-neighbor weights matrix called `scot5k.gwt`.
- Plot a box map of raw rates and Empirical Bayes smoothed rates.
 - Map > Box Map > Hinge = 1.5
- Plot a box map of rates smoothed using the 5-neighbor weights matrix.
 - Tools > Weights > Open → Navigate to `scot5k.gwt`
 - Map > Smooth > Spatial Rate
[Event: CANCER, Base: POP, Map Theme: Box Map (Hinges=1.5)]
- Calculate the Global Moran's I statistic for the raw rates R_RAWRATE and test its significance.
- Calculate the Empirical Bayes Adjusted Moran's statistic and assess its significance.
- Draw the Empirical Bayes Adjusted LISA map.
 - Space > Moran's I with EB Rate
 - Space > LISA with EB Rate

Variance instability in rates may lead to spurious inference for Moran's I. This can be corrected by implementing an Empirical Bayes standardization.

The EB Adjusted Moran's statistic is *not* the same as computing Moran's I for EB-smoothed rates.