

Setting Spatial Weights

Sources: Anselin, 'Spatial data analysis with GIS: An introduction to application in the social sciences'; Terraseer products, spatial weights; 'Statistical/modeling tools for design and analysis of conservation monitoring data'

Why?

- Relationships among neighbors—called neighbor relationships, contiguity relationships, adjacency sets, or spatial weights— underlie spatial statistics.
- Fundamental question: how to define 'neighbors'? Based on your knowledge of the topic at hand, *what is the anticipated structure of spatial dependence?*
 - "For each data point, a relevant 'neighborhood' is defined as those locations surrounding it that are considered to interact with it" (Anselin, 'Spatial data analysis').
 - In a strict sense, two spatial units are either neighbors or not: e.g., polygons that share a border or data points that are within some critical distance of each other (Anselin).
 - Less strictly, "the meaning of 'contiguity' is generalized to any measure of potential interaction between the two units" (Anselin).
 - Basic problems of spatial scale ('modifiable areal unit problem').
- Create a spatial weights matrix with n as the number of objects or locations, with a row and column representing each object or location (e.g., district). If a cell's value is non-zero, then the location or object labeled in the column header is a neighbor of the row location ('ego').
 - Weighting methods differ between polygons and points, though a weighting scheme can combine different methods.
 - Binary weights: 1=neighbor, 0=not neighbor
 - Continuous-value weights: such as length of a shared border between polygons or inverse distance between points.
 - Orders of contiguity.
 - A weighting scheme can be symmetrical or asymmetrical, and standardized (by neighbor count or shared border length) or not.
 - Standardizing the weights gives each object or location the same influence.
 - See Anselin, Table 2, for a spatial weighting example.
 - See Anselin about 'spatially lagged variables' as "An easy way to compare the value at a location to that of its 'neighbors'."
 - A *spatial lag variable* is a weighted average of the values in neighboring locations.
- *Caution:* "Digitizing errors may lead to the creation of spurious contiguities, and other GIS operations, such as the conversion between a raster and a vector format (and vice versa) affect the constructed weights matrix" (Anselin).
- What is the greatest distance over which pairs of locations significantly affect each other? Consider the following:
 - Estimate this distance by finding the distance at which spatial autocorrelations turns from statistically significant to insignificant: choose the last distance at which there is significance, but compare it to other, slightly more distant points.
 - If, at the selected distance, histogram detects a couple or so unconnected pairs of points, drop them from consideration. If there are more unconnected pairs, increase the distance to diminish the number of unconnected pairs.

In GeoDa:

- Set spatial weights: *Tools/Weights/Create* or click the *Create Weights* button on the toolbar.
- Before creating a distance-based weights matrix, create the centroids for the spatial unit's shapefile.
 - Add the centroids by right-clicking the map and selecting *Add Centroids*.
 - GeoDa automatically creates spatial weights for *Moran's Scatterplot*.
- What is the greatest distance over which pairs of locations significantly affect each other? Consider the following:
 - Estimate this distance by finding the distance at which spatial autocorrelations turns from statistically significant to insignificant: choose the last distance at which there is significance, and compare it to other, slightly more distant points.
 - If, at the selected distance, GeoDa's histogram detects a couple or so unconnected pairs of points, drop them from consideration. If there are more unconnected pairs, increase the distance to diminish the number of unconnected pairs.
 - Set GeoDa's histogram to more than 7 bins (using *Options*). Try using the highest number of connections that the histogram displays.