Spatial Autocorrelation

Prepared by Rebecca E. Curtis

Definition: Spatial autocorrelation measures the relationship among values of a variable, according to the spatial arrangement of the values (Cliff and Ord, 1973). The relationship may be described as highly correlated if like values are spatially close to each other and independent or random if no pattern can be discerned from the arrangement of values. The absence of significant spatial autocorrelation can validate the use of standard statistical tests of hypotheses. On the other hand, the presence of significant spatial autocorrelation should encourage the researcher to incorporate spatial dependency in the analysis (Legendre, 1993). Spatial autocorrelation is well suited to raster data analysis because cells in a grid follow a well-defined spatial arrangement.

One common spatial autocorrelation measure is Moran's I, which can be computed by

$$\frac{\sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij}\,(x_i - x_m)\,(x_j - x_m) \Big/ \sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij}}{\sum_{i=1}^{n} (x_i - x_m)^2 / n}$$

where x_i is the value of cell i, x_j is the value of cell i's neighbor j, x_m is the mean cell value of the grid, w_{ij} is a coefficient, and n is the total number of cells in the grid. The coefficient w_{ij} has a value of 1 if j is one of the four cells directly adjacent to i and a value of 0 for other cells or cells with no data. Moran's I is positive when nearby areas have similar attribute values, negative when they have dissimilar values, and close to zero when attribute values are arranged randomly. The values Moran's I takes on tend to range between -1 and 1, but are not restricted to that range.

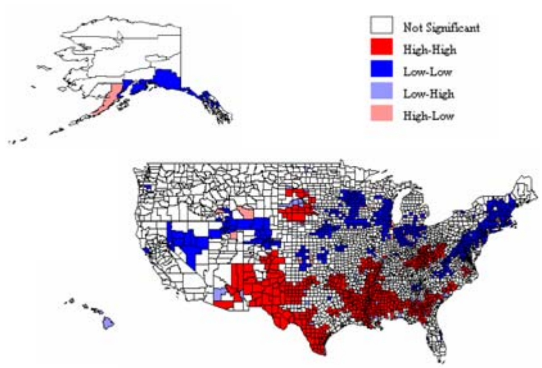
Another common spatial autocorrelation measure is Geary's c, which can be computed by

$$\frac{\sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij}\,(x_i - x_j)^2 \Big/ \sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij}}{\sum_{i=1}^{n} (x_i - x_m)^2 / (n-1)}$$

The notations in the equation are the same as those for Moran's I. Whereas Moran's I uses the covariance $(x_i - x_m)(x_j - x_m)$ in the computation, Geary's c uses the variance $(x_i - x_j)^2$. Geary's c has a value of 1 for random pattern, less than 1 for a positively correlated pattern, and greater than 1 for a negatively correlated pattern.

Application:

The example to the right, using a Moran's I analysis, confirms that child poverty is highly clustered regionally. Hotspot clusters of high rates of child poverty (surrounded by neighbors with high poverty) are the Mississippi Delta region, the Black Belt, Appalachia, southwest Texas and New Mexico, and Indian populations in southern South Dakota and northern Nebraska (Voss, et. al., 2004).



References/Sources:

Cliff, A.D. and J.K. Ord. 1973. *Spatial Autocorrelation*. New York: Methuen.

Haining, R. 1990. *Spatial Data Analysis in the Social and Environmental Sciences*. Cambridge University Press.

Legendre, P. 1993. Spatial Autocorrelation: Trouble or New Paradigm? *Ecology* 74: 1659-73.

Voss, P. et. al. 2004. County Child Poverty Rates in the U.S.: A Spatial Regression Approach. CDE Working Paper. University of Wisconsin-Madison.