CSISS WORKSHOP

Introduction to Spatial Pattern Analysis in a GIS Environment

Measures of Spatial Pattern: Global and Local Statistics

Arthur Getis
Pattern Statistics

• GLOBAL
  \( I, c, K, G, Knox, Mantel, Tango, \)
  Grimson, Cuzick and Edwards, Kernels, Scan

• LOCAL
  \( I_i, c_i, G_i, G_i^*, GWR, O_i \)
Global Statistics

• Nearest Neighbor
• K-Function
• Global Autocorrelation Statistics
  Moran’s I
  Geary’s c
  Semivariance
Matrix Representation: $WY$

- **$W$**
  - The Spatial Weights Matrix
  - The Spatial Association of All Sites to All Other Sites
  - $d$, $d^2$, $1/0$, $1/d$

- **$Y$**
  - The Attribute Association Matrix
  - The Association of the Attributes at Each Site to the Attributes at All Other Sites
  - $+,-,/,x$
The Spatial Weights Matrix

$W$ is the formal expression of the spatial association between objects

(it is the pair-wise geometry of objects being studied).
Typical W

- Spatially contiguous neighbors (rook, queen: one/zero)
- Inverse distances raised to a power: \(1/d, 1/d^2, 1/d^5\)
- Geostatistics functions (spherical, gaussian, exponential)
- Lengths of shared borders (perimeters)
- All centroids within distance \(d\)
- \(n^{th}\) nearest neighbor distance
- Links (number of)
The Attribute Matrix

$Y$

The variable under study. One variable at a time. Interval scale (other scales under special conditions).

For example, residuals from regression; a socio-economic variable (number of crimes, household income, number of artifacts, etc.)
Attribute Relationships

Y

• **Types of Relationships**
  Additive association (clustering): \((Y_i + Y_j)\)
  Multiplicative association (product): \((Y_i Y_j)\)
  Covariation (correlation): \((Y_i - \bar{Y})(Y_j - \bar{Y})\)
  Differences (homogeneity/heterogeneity): \((Y_i - Y_j)\)
  Inverse (relativity): \((Y_i/Y_j)\)

• **All Relationships Subject to Mathematical Manipulation**
  (power, logs, abs, etc.)
**WY: Covariance**

- Set $W$ to preferred spatial weights matrix
- (rooks, queens, distance decline, etc.)
- Set $Y$ to $(x_i - \mu) (x_j - \mu)$
- Set scale to $n/W \sum(x_i - \mu)^2$
- $I = n \sum \sum W_{ij} (x_i - \mu)(x_j - \mu) / W \sum(x_i - \mu)^2$
  
  where $W$ is sum of all $W_{ij}$ and $i \neq j$

*This is Moran’s $I$.*
**WY: Additive**

- Set $W$ to 1/0 spatial weights matrix
- 1 within $d$; 0 outside of $d$
- Set $Y$ to $(x_i + x_j)$
- Set scale to $\Sigma W_{ij}(d) / \Sigma (x_i)$
- $G(d) = \Sigma W_{ij}(d) (x_i + x_j) / \Sigma (x_i)$ and $i \not\sim j$

*This is Getis and Ord’s $G.*
WY: Difference

- Set $W$ to preferred spatial weights matrix
- Set $Y$ to $(x_i - x_j)^2$
- Set scale to $(n-1)/2W \sum (x_i - \mu)^2$
- $c = (n - 1) \sum \sum W_{ij} (x_i - y_i)^2 / 2W \sum (x_i - \mu)^2$
  where $W$ is sum of all $W_{ij}$ and $i \neq j$

This is Geary’s $c$. 
WY: Difference

- Set $W$ to 1/0 weights matrix; 1 within $ah$ and 0 otherwise; $a$ is an integer; $h$ is a constant distance
- Set $Y$ to $(x_i - x_j)^2$
- Set scale to $1/2$
- $\chi(ah) = 1/2 \sum \sum W_{ij} (x_i - x_j)^2$

This is the semi-variogram.
Local Statistics

• Global Statistics reworked for focussing on $i$

• LISA statistics (Local Indicators of Spatial Association)
  
  Moran’s $I_i$, Geary’s $c_i$

• Clustering Statistics
  
  Getis and Ord’s $G_i$ and $G_i^*$
The Getis-Ord Approach

\[ G_i^*(d) = \frac{\sum_j w_{ij}^*(d) x_j}{\sum_j x_j} \]

- Normally distributed
- Tests for statistical significance
The $G_i^*$ Statistic

- The $G_i^*$ statistic is local, that is, it is focused on sites and is normally distributed. It is designed to yield a measure of pattern in standard normal variates.
- Indicates the extent to which a location (site) is surrounded to a distance $d$ by a cluster of high or low values.
- The input is a file containing coordinates for each house and, for example, the Y variable. The user specifies maximum search distance and number of increments.
- The output file contains a listing of the $G_i^*(d)$ value for sample point at a specified distance $(d)$. 
The Critical Distance

- The $G_i^*$ values are computed around each observation as distance increases.

- When the absolute values fail to rise, the cluster diameter is reached. This is the critical distance $d_c$.

- Spatial association weakens beyond $d_c$. 
Example Ranges

(SA = Santa Ana)