Spatial Statistics using GIS

An Introduction
What are spatial statistics?

- They are similar to traditional statistics, but they integrate spatial relationships into the calculations.
Spatial statistics will allow you to answer the following questions about your data:

- How are the features distributed?
- What is the pattern created by the features?
- Where are the clusters?
- How do patterns and clusters of different variables compare on one another?
- What are the relationships between sets of features or values?
Types of data often analyzed

- Location of crimes, animals, retail, industry, etc.
- Land cover
- Land use
- Census/social science data
Conceptual Models

- Inverse distance (spatial autocorrelation) – all features influence all other features, but the closer something is, the more influence it has

- Distance band – features outside a specified distance do not influence the features within the area

- Zone of indifference – combines inverse distance and distance band
Conceptual Models

- K Nearest Neighbors – a specified number of neighboring features are included in calculations
- Polygon Contiguity – polygons that share an edge or node influence each other
- Spatial weights – specified by user (ex. Travel times or distances)
IDENTIFYING GEOGRAPHIC DISTRIBUTION
**Finding the center of features**

| Mean Center | • Average x and y-coordinates for all features  
|            | • Useful for comparing distributions of different features or over time |
| Central feature | • **Feature** having the shortest total distance to all other features  
|               | • Useful for finding the most accessible feature |
Finding the center of features

The mean center, median center, and central feature for a group of high schools in a region.
Measuring Compactness

- Standard distance – the extent to which the distance between the mean center and the features vary from the average distance

- The greater the standard distance, the more the distances vary from average → features are more widely dispersed around the center
In a spatial normal distribution (most features are concentrated in the center):

- a one standard deviation circle will cover about 68% of the features.
- a two standard deviation circle will cover about 95%.
- a three standard deviation circle will cover about 99%.

Measuring Compactness

SD of American Indian population = 35,685 ft.
SD of African American population = 21,621 ft.
Dispersion of crimes

Dispersion of auto theft (blue), assaults (red), and other thefts (brown)

How does the dispersion of each crime compare to the others?
Orientation

**Linear directional mean** identifies general mean direction of a set of lines.

Migration paths for elk (left) and moose. Different movements, but same direction.
Standard deviational ellipse is useful for comparing distributions of features and comparing one type of feature at different times.

Standard deviational ellipses for financial businesses (left) and manufacturing businesses.
IDENTIFYING PATTERNS
Patterns

Useful to:

- Better understand geographic phenomena (ex. Habitats)
- Monitor conditions (ex. Level of clustering)
- Compare different sets of features (ex. Patterns of different types of crimes)
- Track change

Level of clustering of cut areas within a forest
You can measure the pattern formed by the location of features or patterns of attribute values associated with features (ex. median home value, percent female, etc.).
Average Nearest Neighbor

- Measures how similar the actual mean distance between locations is to the expected mean distance for a random distribution.

- Measures clustering vs. dispersion of feature locations.

- Can be used to compare distributions to one another.
Ripley’s K-function

GIS counts the number of neighboring features within a given distance to each feature based on location. The test compares the observed K value at each distance to expected K value for a random distribution.

Assaults are clustered until about 13,000 ft and then they are dispersed beyond 15,000 ft.
IDENTIFYING CLUSTERS
Global vs. Local Statistics

- Global statistics – identify and measure the pattern of the entire study area
  - Do not indicate where specific patterns occur
- Local Statistics – identify variation across the study area, focusing on individual features and their relationships to nearby features (i.e. specific areas of clustering)
Spatial Autocorrelation (Moran’s I)

• Measures whether the pattern of feature values is clustered, dispersed, or random.

• Global Statistic

• Calculates I values to test for statistically significant clustering

0<\text{I}=\text{Values Clumped}

\text{I}<0=\text{Values Dispersed}

\text{I}=0=\text{Random Distribution}
Anselin Local Moran’s I

- **Local statistic**

- Measures the strength of patterns for each specific feature.

Census tracts for percentage of 65 years old and above  

<table>
<thead>
<tr>
<th>I values</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.6 - 6.8</td>
</tr>
<tr>
<td>6.9 - 13.0</td>
</tr>
<tr>
<td>13.1 - 19.3</td>
</tr>
<tr>
<td>19.4 - 25.5</td>
</tr>
<tr>
<td>25.6 - 31.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Z-scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; -1.645</td>
</tr>
<tr>
<td>-1.645 - 1.645</td>
</tr>
<tr>
<td>-0.810 - 1.445</td>
</tr>
<tr>
<td>0.810 - 1.445</td>
</tr>
<tr>
<td>1.446 - 3.220</td>
</tr>
</tbody>
</table>
Anselin Local Moran’s I

- **Positive I value:**
  - Feature is surrounded by features with similar values, either high or low.
  - Feature is part of a **cluster**.
  - Statistically significant clusters can consist of high values (HH) or low values (LL).

- **Negative I value:**
  - Feature is surrounded by features with dissimilar values.
  - Feature is an **outlier**.
  - Statistically significant outliers can be a feature with a high value surrounded by features with low values (HL) or a feature with a low value surrounded by features with high values (LH).
Getis-Ord General G

• **Global** statistic that indicates whether similar *values* (either high or low) are *clustered*.

• Works best when either high or low values are clustered (but not both).

• Value of G score indicates statistically significant relationships.

High G score: Statistically significant clustering of high values.

Low G value: Slight clustering of low values.
Hot Spot Analysis (Getis-Ord Gi*)

- **Local** version of the G statistic that indicates hot (cluster of high values) or cold spots (clusters of low values).

- To be statistically significant, the hot spot or cold spot will be surrounded by features with similar values, but have significantly higher/lower values than its neighbors.

- $G=$ high value = hot spots
- $G=$ low value = cold spots
IDENTIFYING RELATIONSHIPS
Regression Analysis

- With other statistical tools you ask **WHERE** something is happening?

- With Regression Analyses, you ask **WHY** something is happening.
  - Why are there places in the United States where people persistently die young? What might be causing this?
  - Can we model the characteristics of places that experience a lot of crime, 911 calls, or fire events to help reduce these incidents?
Regression analysis allows you to...

- Model, examine and explore spatial relationships
- Predict

Coefficients for percent rural and low-weight births

T-scores show where this relationship is significant
Linear Regression

- Used to analyze linear relationships among variables.
- Linear relationships are positive or negative
- Regression analyses attempt to demonstrate the degree to which one or more variables potentially promote positive or negative change in another variable.
Linear Regression Techniques

- Ordinary Least Squares (OLS) is the best known technique and a good starting point for all spatial regression analyses.
  - **Global** model = provides 1 equation to represent the entire dataset

- Geographically Weighted Regression (GWR)
  - **Local** Model = fits a regression equation to every feature in the dataset
  - Regional variation incorporated into the regression model
Software

- ArcGIS
- GeoDa – open source
  - Standalone
  - Shapefiles for inputs
  - User friendly, graphical interface
  - Available with registration at: http://geodacenter.asu.edu/