


**GIS Analysis:
Spatial Statistics for Public Health**

Lauren M. Scott, PhD
Mark V. Janikas, PhD
ESRI



Section A

Spatial Statistics Toolbox

Spatial Data Analysis

- Traditional GIS analysis
 - Spatial and temporal queries
 - ▶ Do ER visits for respiratory illness increase seasonally?
 - Buffer analysis
 - ▶ Where should evacuation efforts focus after a toxic chemical spill?
 - Overlay analysis
 - ▶ Which facilities should be targeted for quarantine?
- Spatial statistics toolbox
 - Hot spot (cold spot) analysis
 - ▶ Where is the disease outbreak concentrated?
 - Spatial and temporal pattern analysis
 - ▶ Is the problem remaining geographically fixed or is it spreading?
 - Spatial regression analysis
 - ▶ Why are cancer rates so high in particular areas?

3

Objectives

- Introduce the spatial statistics toolbox through a variety of public health applications
- Present the new OLS and GWR regression analysis tools
- Outline the challenges of regression for spatial data
- Discuss regression analysis diagnostics
- Highlight resources for learning more about spatial statistics in ArcGIS
- DEMO
 - Mortality and average age of death

4

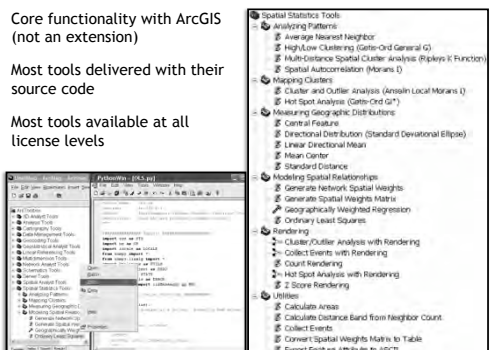
What Are Spatial Statistics in a GIS Environment?

- Software-based tools, methods, and techniques developed specifically for use with geographic data
- Spatial statistics:
 - Describe and model spatial distributions, spatial patterns, spatial processes, and spatial relationships
 - Incorporate space (area, length, proximity, orientation, and/or spatial relationships) directly into their mathematics
- In many ways spatial statistics extend what the eyes and mind do intuitively to assess spatial patterns, trends, and relationships

5

Spatial Statistics Toolbox

- Core functionality with ArcGIS (not an extension)
- Most tools delivered with their source code
- Most tools available at all license levels



6

Spatial Statistics Toolbox


- Measuring Geographic Distribution
- Analyzing Patterns
- Mapping Clusters
- Modeling Spatial Relationships

Measuring Geographic Distribution

- Central Feature
- Directional Distribution (Standard Deviational Ellipse)
- Linear Directional Mean
- Mean Center
- Standard Distance

Questions

- Which site is most accessible?
- Is there a directional trend to the spatial distribution of the disease?
- What is the primary wind direction for this region in the winter?
- Where is the population center?
- Which strain has the broadest distribution?



7

Spatial Statistics Toolbox

- Measuring Geographic Distribution
- Analyzing Patterns
- Mapping Clusters
- Modeling Spatial Relationships

Analyzing Patterns

- Average Nearest Neighbor
- High/Low Clustering (Getis-Ord General G)
- Multi-Distance Spatial Cluster Analysis (Ripley's K Function)
- Spatial Autocorrelation (Moran's I)

- Is there an unexpected spike in pharmaceutical purchases?
- Are incidents spatially clustered?
- Are new AIDS cases remaining geographically fixed?
- Does the spatial pattern of the disease mirror the spatial pattern of the population at risk?

8

Analyzing Global Patterns

Analyzing Patterns

- Average Nearest Neighbor
- High/Low Clustering (Getis-Ord General G)
- Multi-Distance Spatial Cluster Analysis (Ripley's K Function)
- Spatial Autocorrelation (Moran's I)

1994 New AIDS Cases

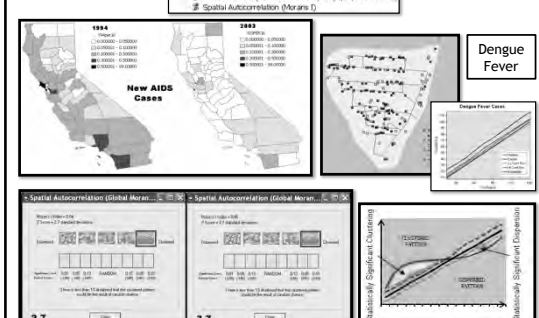
2002 Dengue Fever

Spatial Autocorrelation (Global Moran's I)

2.7

3.7

Statistically Significant Clustering



9

Spatial Statistics Toolbox

- Measuring Geographic Distribution
- Analyzing Patterns
- Mapping Clusters
- Modeling Spatial Relationships

Mapping Clusters

- Cluster and Outlier Analysis (Local Moran's I)
- Hot Spot Analysis (Getis-Ord G*)

High poverty

High poverty surrounded by low poverty

Low poverty

Low poverty surrounded by high poverty

- Are fire/police units well located in relation to 911 call hot spots?
- Where are their sharp boundaries between affluence and poverty?
- Where do we see unexpectedly high rates of diabetes?
- Where do we find unexpected resistance to illness?

10

Spatial Statistics Toolbox

- Measuring Geographic Distribution
- Analyzing Patterns
- Mapping Clusters
- Modeling Spatial Relationships

Modeling Spatial Relationships

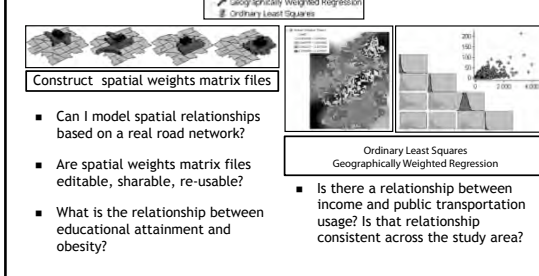
- Generate Network Spatial Weights
- Generate Spatial Weights Matrix
- Geographically Weighted Regression
- Ordinary Least Squares

Construct spatial weights matrix files

- Can I model spatial relationships based on a real road network?
- Are spatial weights matrix files editable, sharable, re-usable?
- What is the relationship between educational attainment and obesity?
- Is there a relationship between income and public transportation usage? Is that relationship consistent across the study area?

Ordinary Least Squares

Geographically Weighted Regression



11

Regression Analysis

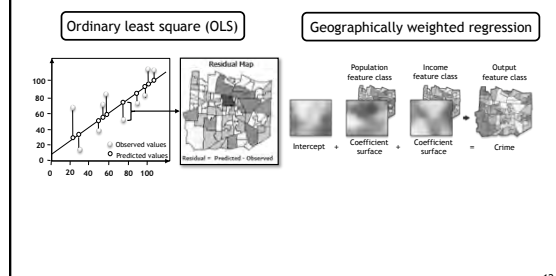
- Regression analysis allows you to model, examine, and explore spatial relationships, in order to better understand the factors behind observed spatial patterns or to predict outcomes

Ordinary least square (OLS)

Geographically weighted regression

Population feature class + Income feature class = Output feature class

Intercept + Coefficient surface + Coefficient surface = Crime



12

Regression Analysis Terms and Concepts

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon$$

Residential Burglary = β_0 + β_1 (Income) + β_2 (Vandalism) + β_3 (Households) + Residual Error

- **Dependent variable (Y):** what you are trying to model or predict (residential burglary, for example)
- **Explanatory variables (X):** variables you believe cause or explain the dependent variable (income, vandalism, households)
- **Coefficients (B):** values, computed by the regression tool, reflecting explanatory to dependent variable relationships
- **Residuals (ϵ):** the portion of the dependent variable that isn't explained by the model; the model under and over predictions

13