Spatial Statistics: Understanding Concepts and Variables

Zhengxuan Deng

Spatial statistics is a specialized field that plays a crucial role in analyzing and interpreting data with geographical components. The consideration of spatial relationships is fundamental in various disciplines, including geography, ecology, epidemiology, and environmental science. In this comprehensive exploration, we delve into key concepts, mathematical expressions, and the meaning behind variables in spatial statistics.

1. Descriptive Statistics in Spatial Data:

In spatial statistics, descriptive statistics are adapted to account for geographical locations associated with each observation.

2. Inferential Statistics and Spatial Relationships:

In the context of spatial autocorrelation, Moran's I statistic (I) is employed to measure the degree of similarity or dissimilarity between nearby locations. The variables involved include:

$$I = \frac{n}{W \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(z_i - \bar{z})(z_j - \bar{z})}$$

where n is the number of spatial units, W is the sum of spatial weights, w_{ij} are spatial weights between locations i and j, z_i is the value of the variable at location i, and $\bar{z}$ is the mean of the variable.

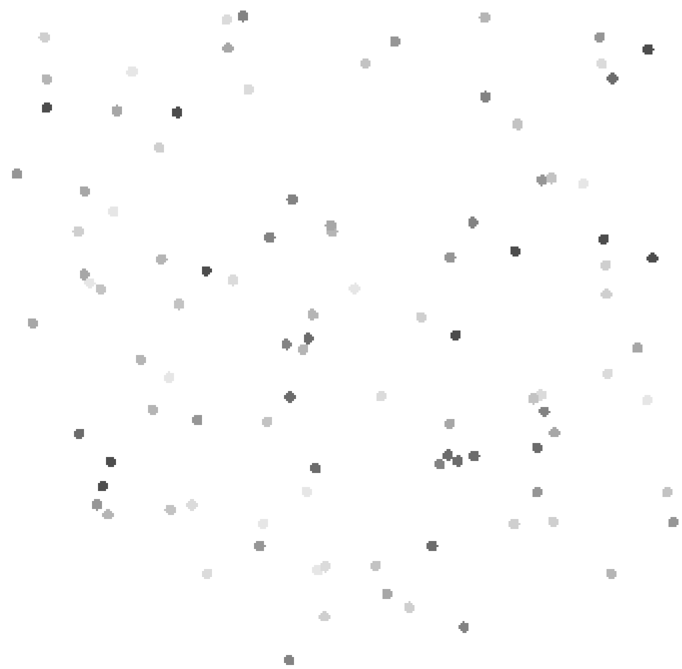
3. Geostatistics and Kriging:

Geostatistics focuses on spatial variability, and kriging is an interpolation method with the following variables:

$$Z(u) = \sum_{i=1}^{n} \lambda_i Z_i$$

where $Z(u)$ is the estimated value at an unobserved location u, λ_i are the kriging weights, and Z_i is the observed value at location i.

Spatial Data



4. Spatial Regression Models:

Spatial regression models incorporate spatial dependencies with the following variables:

$$y = \rho W y + X\beta + \varepsilon$$

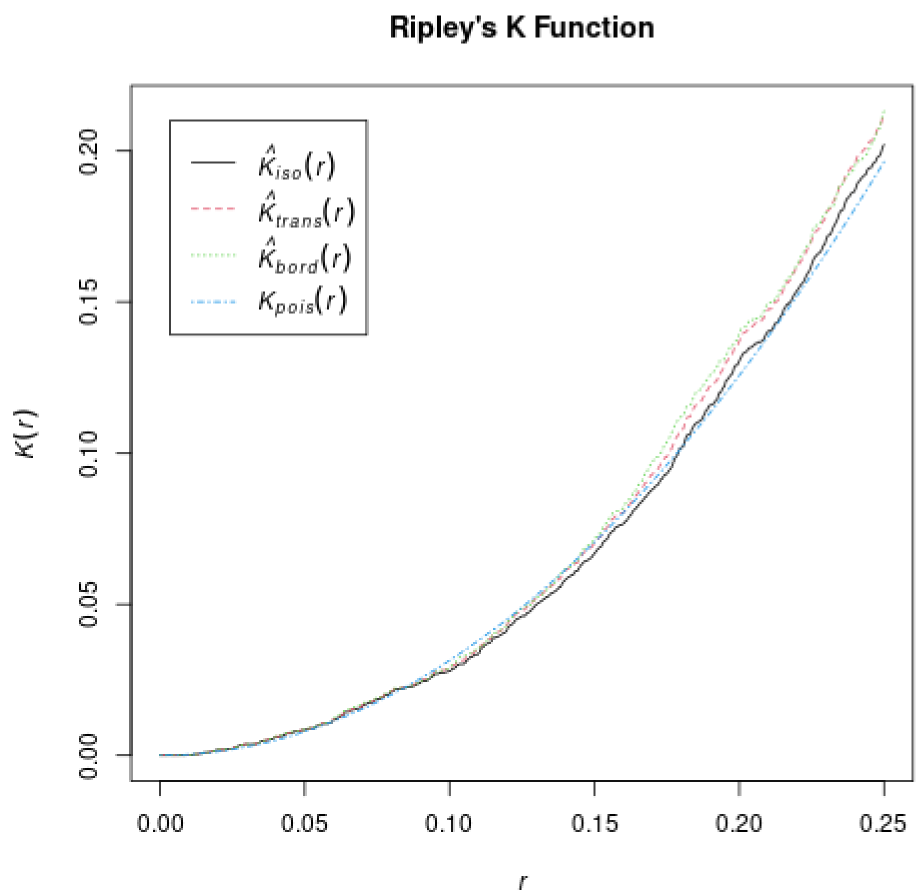
where y is the dependent variable, ρ is the spatial autocorrelation coefficient, W is the spatial weights matrix, X is the matrix of independent variables, β is the vector of coefficients, and ε is the error term.

5. Spatial Point Patterns and Ripley's K Function:

Ripley's K function assesses spatial patterns in point datasets. Relevant variables include:

$$K(t) = \frac{1}{\lambda} E \left[\sum_{i=1}^{n} \sum_{j=1}^{n} \mathbf{1}_{(0<t<\|\mathbf{x}_i - \mathbf{x}_j\| \leq t+\Delta t)} \right]$$

where $K(t)$ is Ripley's K function, λ is the intensity of the point process, $\mathbf{x}_i$ is the location of point i, and $\mathbf{1}_{(A)}$ is the indicator function for event A.



6. Cluster Detection with Getis-Ord Gi* Statistic:

The Getis-Ord Gi* statistic identifies clusters in spatial data. Key variables are:

$$Gi^*(i) = \frac{\sum_{j=1}^{n} w_{ij} Z_j - \bar{Z} \sum_{j=1}^{n} w_{ij}}{s \sqrt{\frac{\sum_{j=1}^{n} w_{ij}^2 (Z_j - \bar{Z})^2}{(n-1) \sum_{j=1}^{n} (Z_j - \bar{Z})^2}}}$$

where $Gi^*(i)$ is the Getis-Ord Gi* statistic, w_{ij} are spatial weights, Z_j is the variable value at location j, $\bar{Z}$ is the mean of the variable, and s is the standard deviation of the variable.

7. Landscape Metrics and Fragmentation Index:

Landscape metrics, such as the Landscape Fragmentation Index (LFI), provide insights into spatial patterns. Variables include:

$$LFI = \frac{\sum_{i=1}^{n} P_i \cdot \log(P_i)}{\log(n)}$$

where LFI is the Landscape Fragmentation Index, P_i is the proportion of landscape covered by patch i, and n is the total number of patches.

8. Moran's Eigenvector Maps in Spatial Analysis:

Moran's Eigenvector Maps (MEM) aid in spatial regression models. Relevant variables are:

$$\rho W y = \sum_{i=1}^{n} \lambda_i \mathbf{u}_i$$

where $\rho W y$ is the spatial lag term, λ_i is the eigenvalue, and $\mathbf{u}_i$ is the i-th eigenvector.

9. G Function for Spatial Point Patterns:

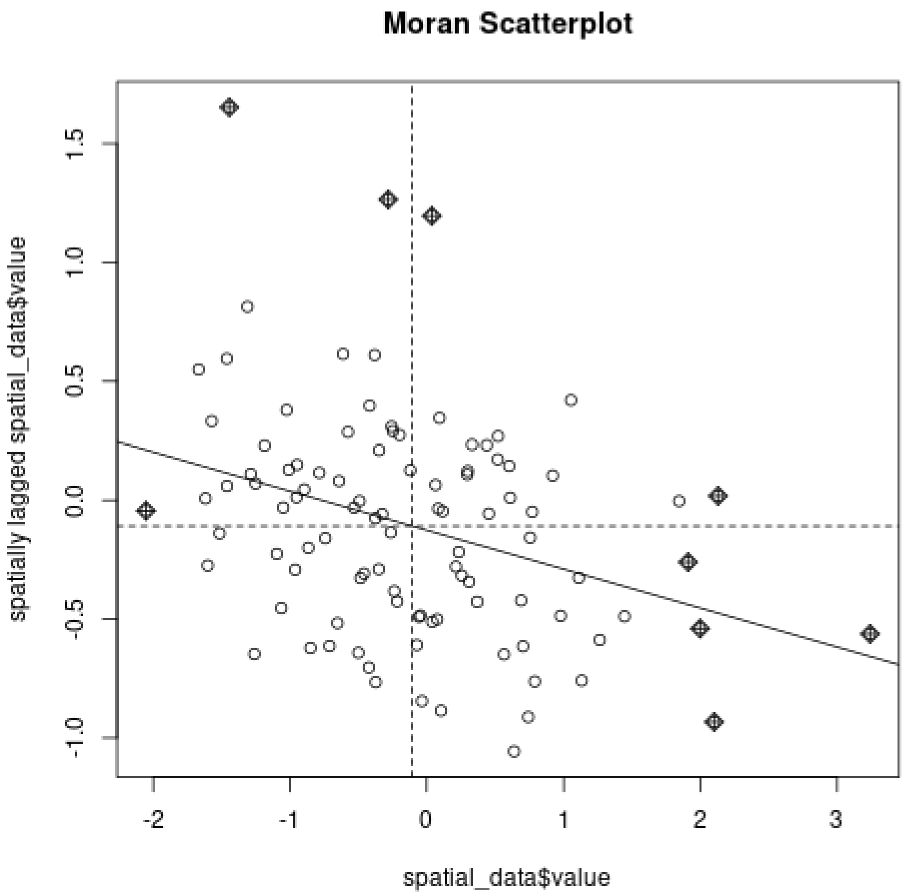
The G function assesses spatial point pattern dispersion. Variables include:

$$G(t) = \frac{1}{\lambda^2} E \left[\sum_{i=1}^{n} \sum_{j=1}^{n} \mathbf{1}_{(0<t\leq\|\mathbf{x}_i - \mathbf{x}_j\|)} \right]$$

where $G(t)$ is the G function, λ is the intensity of the point process, $\mathbf{1}_{(A)}$ is the indicator function for event A, and t represents the distance.

10. Spatial Data Visualization: Moran Scatterplot:

Moran Scatterplots visually represent spatial autocorrelation, aiding in the interpretation of spatial relationships.



In practical applications, these variables and concepts are implemented using statistical software such as R, Python, or specialized GIS software. Spatial statistics continues to evolve, providing powerful tools for gaining insights into complex spatial patterns and relationships across diverse fields.

11. Problem Set

Problem 1: Spatial Autocorrelation:

- Load a spatial dataset of your choice.

- Calculate Moran's I to assess spatial autocorrelation.
- Interpret the Moran Scatterplot to understand the nature of spatial relationships.

Problem 2: Geostatistics and Kriging:

- Generate a simulated dataset with spatial coordinates.
- Apply kriging to estimate values at unobserved locations.
- Visualize the original and kriged values on a map.

Problem 3: Spatial Regression:

- Obtain a dataset with a spatial component and relevant variables.
- Perform a spatial regression analysis considering spatial autocorrelation.
- Interpret the results, paying attention to coefficients and their significance.

Problem 4: Point Pattern Analysis:

- Create a synthetic point pattern dataset.
- Calculate and visualize Ripley's K function.
- Interpret the K function plot to understand spatial patterns.

Problem 5: Cluster Detection:

- Choose a spatial dataset with a variable of interest.
- Apply the Getis-Ord Gi* statistic to identify hotspots or clusters.
- Interpret the results and map the identified clusters.

Problem 6: Landscape Metrics:

- Obtain a spatial dataset representing a landscape with patches.
- Calculate the Landscape Fragmentation Index (LFI) to assess landscape fragmentation.
- Interpret the LFI values in the context of landscape structure.

Problem 7: Moran's Eigenvector Maps:

- Perform a spatial regression using Moran's Eigenvector Maps (MEM).
- Examine the results and analyze the impact of spatial eigenvectors on the model.

Problem 8: G Function for Point Patterns:

- Create a spatial point pattern dataset.
- Compute and visualize the G function to analyze point pattern dispersion.
- Interpret the G function plot.