GEOG 413/613

LECTURE 10

1

Spatial Autocorrelation

- Nearest Neighbor Analysis, Quadrat Analysis and Cluster Analysis are limited as they only consider either the location or the attribute of the observations
- Spatial autocorrelation detects spatial patterns of a data distribution by considering both the ***locations*** and the ***attributes*** of the observations

2

2

Spatial Autocorrelation

- Spatial Autocorrelation – correlation of a variable with itself through geographic space.
  - The first law of geography: "everything is related to everything else, but near things are more related than distant things" – Waldo Tobler
  - If there is a systematic pattern in the spatial distribution of a variable, then there is spatial autocorrelation
    - *positive spatial autocorrelation* – neighbors are more alike
    - *negative autocorrelation* – neighbors are not alike/different
    - *no spatial autocorrelation* – Random distribution

3

3

Spatial Autocorrelation

- Most statistical measures are based on the assumption that the values of observations in each sample are independent of one another (this is why we carry out random sampling)
- This independence may be violated by spatial autocorrelation. For example, positive spatial autocorrelation may violate the independence if the samples were taken from nearby areas
- Therefore it may lead to incorrect conclusions about relationships between variables
- Goals of spatial autocorrelation are
  - To measure the strength of spatial autocorrelation for a given distribution
  - To test the assumption of independence or randomness

4

4

Moran's I Statistic

- Moran's I Statistic is a spatial autocorrelation technique that calculates the relationship between locations of observations (w_{ij}) the similarity between the attributes (c_{ij}) at those locations

- Given a set of features and an associated attribute, it evaluates whether the spatial data distribution is clustered, dispersed or random.

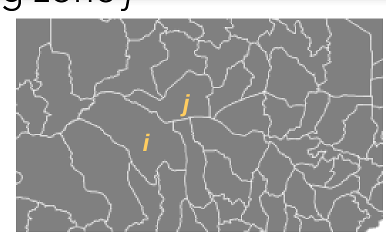
- It is applied to zones or points with continuous variables associated with them.

5

5

Moran's I Statistic

If zone i is neighboring zone j



Similarity of Attribute c_{ij} $$c_{ij} = (x_i - \overline{x})(x_j - \overline{x})$$

Proximity of Location w_{ij} $$w_{ij} = 1/d_{ij}$$

Sample Variance s^2 $$s^2 = \sum (x_i - \overline{x})^2$$

6

6

- Combines the measurement for attribute similarity and location proximity

$$I = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} c_{ij}}{s^2 \sum_{i=1}^{n}\sum_{i=1}^{n} w_{ij}}$$

7

7

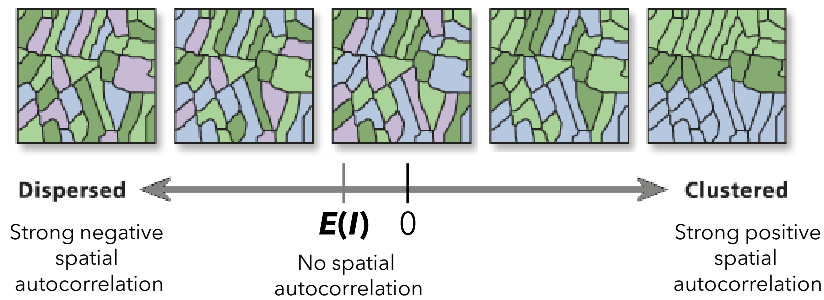
- The numeric scale of Moran's I is anchored at the expected value of

$$E(I) = \frac{-1}{n-1}$$

- A Moran's I value equal to *E(I)* demonstrates that the pattern has no spatial autocorrelation and hence can be considered random

8

8



9

9

Testing for Significance

$$Z = \frac{I - E_I}{S^2}$$

Assuming a two-tailed test at the 0.05 significance level, the observed degree of spatial autocorrelation is significant if it is beyond the critical value of
Z = -1.96 or +1.96

10

10

Local Indicators of Spatial Association

- Moran's I index we see that it can be disaggregated to provide a series of local indices
- LISA calculates for each and index value and a Z score
  - A high negative Z score indicates that the feature is adjacent to features of dissimilar values

11

11

Local Indicators of Spatial Association

- Moran's I index we see that it can be disaggregated to provide a series of local indices
- LISA calculates for each and index value and a Z score
  - A high negative Z score indicates that the feature is adjacent to features of dissimilar values
- Moran's I, Local or Global can only detect the presence of clustering of similar values
  - It cannot tell whether the clustering is made of high values or low values

12

12

G-Statistic for Measuring High/Low Clustering

- The G-Statistic can separate clusters of high values from clusters of low values.

$$G(d) = \frac{\sum\sum w_{ij}(d) x_i x_j}{\sum\sum x_i x_j}, i \neq j \qquad E(G) = \frac{\sum\sum w_{ij}(d)}{n(n-1)}$$

$$\begin{gathered} \textit{where } x_i \textit{ is the value at location } i, \\ x_j \textit{ is the value at location } j \textit{ if } j \textit{ is within } d \textit{ of } i, \\ \textit{and } w_{ij}(d) \textit{ is the spatial weight} \\ E(G) \textit{ is the expected value of } G \end{gathered}$$

Minimum Sample Size

- In statistics, it is often necessary to determine if our sample size is large enough to consider our results valid
- If the number of observations is too small the sample will not adequately represent the population
- Therefore, it is important to calculate the *minimum sample size* if there is spatial autocorrelation.
  - Generally the sample size will be larger than if spatial autocorrelation is not present
  - Different types of statistics have their methods for way calculating the *minimum sample size*

References

- David O'Sullivan, David Unwin, 2010 *Geographic Information Analysis*. Hoboken, NJ, John Wiley

- David W. S. Wong, Jay Lee, 2005. *Statistical Analysis of Geographic Information with ArcView GIS and ArcGIS*. Hoboken, NJ, John Wiley