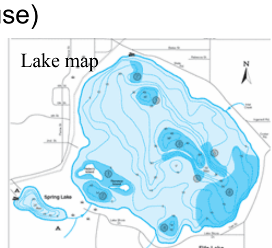
Area objects and spatial autocorrelation

Outline

- Introduction
- Geometric properties of areas
- Spatial autocorrelation: joins count approach
- Spatial autocorrelation: Moran’s I
- Spatial autocorrelation: Geary’s C
- Spatial autocorrelation: weight matrices
- Local indicators of spatial association (LISA)

Types of area object

- **Natural areas**: self-defining, their boundaries are defined by the phenomenon itself (e.g. lake, land use)
  - Fuzzy boundaries

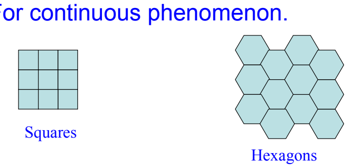


Types of area object

- **Imposed areas**: imposed by human beings, *e.g.* countries, states, counties etc.
  - Boundaries are defined independently of any phenomenon, and attribute values are enumerated by surveys or censuses
  - Potential Problems
    - may bear little relationship to underlying patterns
    - Arbitrary and modifiable (MAUP)
    - Danger of ecological fallacies (aggregated format)

Types of area object

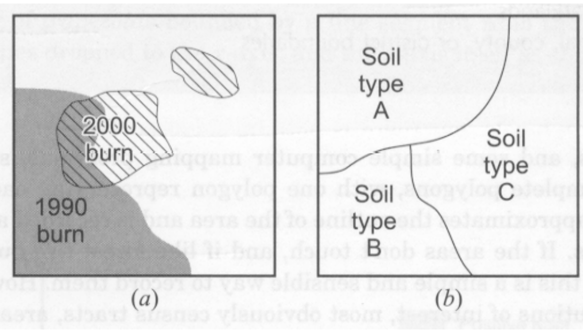
- **Raster**: space is divided into small regular grid cells.
  - Area objects are identical and together cover the region of interest.
    - Each cell can be considered an area object.
  - For continuous phenomenon.



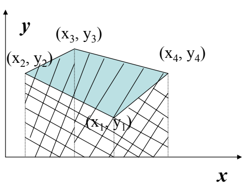
Types of area object

- **Planar enforced**: area objects mesh together neatly and exhaust the study region, so that there are no holes, and every location is inside just a single area;
  - e.g. soil type
- **Not planar enforced (non-planar)**: the areas do not fill or exhaust the space, the entities are isolated from one another, or perhaps overlapped
  - e.g. forest patches

Planar vs. non-planar



Geometric Properties of Areas - Area

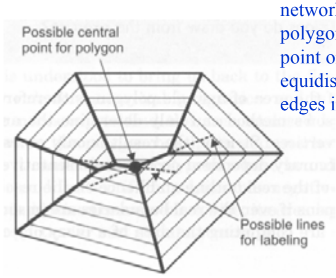


$$Area = \frac{1}{2}\sum_{i=1}^{n}(x_{i+1} - x_i)(y_{i+1} + y_i)$$

Assume $x_1 = x_{n+1}$

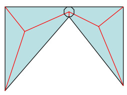
Geometric Properties of Areas - Skeleton

The skeleton of a polygon is a network of lines inside a polygon constructed so that each point on the network is equidistant from the nearest two edges in the polygon boundary.

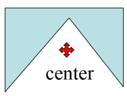


Geometric Properties of Areas - Skeleton

- **Skeleton →** centroid of an area?



Center derived by skeleton analysis



Arithmetic center

$$\begin{cases} \hat{x} = \sum_{i=1}^{n} x_i \\ \hat{y} = \sum_{i=1}^{n} y_i \end{cases}$$

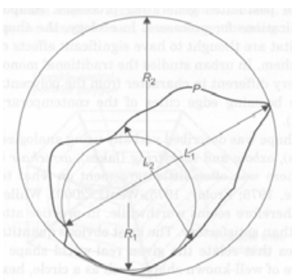
Exercise 12

- **Generate centroids of polygons**

Geometric Properties of Areas - Shape

- A set of relationships of relative position between points on their perimeters
  - In ecology, the shapes of patches of a specified habitat are thought to have significant effects on what happens and around them.
  - In urban studies, urban shapes change from traditional polycentric to multiple polycentric sprawl

Geometric Properties of Areas - Shape



Parameter: *P*

Area: *a*

Longest axis: L_1

Second axis: L_2

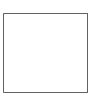
The radius of the largest internal circle: R_1

The radius of the smallest enclosing circle: R_2

Geometric Properties of Areas - Shape

Compactness ratio $= \sqrt{a / a_2} = 2\sqrt{\pi a} / p$

a is the area of the polygon
a_2 is the area of the circle having the same perimeter (P) as the object
p is the perimeter of the polygon



What is the compactness ratio for a circle?
What is the compactness ratio for a square?

Geometric Properties of Areas - Shape

- Other measurements
  - Elongation ratio: *L1/L2*
  - Form ratio: a / L_1^2

Review

- Area type:
  - Natural vs. arbitrary
    - Raster grids
  - Plannar vs. non-plannar
- Area properties:
  - Area
  - Skelton
  - Centroid
  - Shape

Reminder on Spatial Autocorrelation

- Value as a description of the geography
- Waldo Tobler's 1st Law of Geography
  - '*Everything is related to everything else but nearby things are more related than distant things*'
- Importance of spatial autocorrelation:
  - Impacts on standard statistics

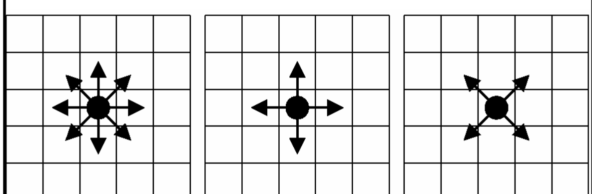
Spatial Autocorrelation - Joins count approach

- Developed by Cliff and Ord (1973) in their book: Spatial Autocorrelation
- The joins count statistic is applied to a map of areal units where each unit is classified as either black (B) or white (W): binary
- The joins count is determined by counting the number of occurrences in the map of each of the possible joins (e.g. BB, WW, BW) between neighboring areal units.

Spatial Autocorrelation
- Joins count approach

- **Neighbor definition**
  – Rook's case: four neighbors (North-South-West-East)
  – Queen's case: eight neighbors (including diagonal neighbors)

Queen vs Rook (occasionally bishop)

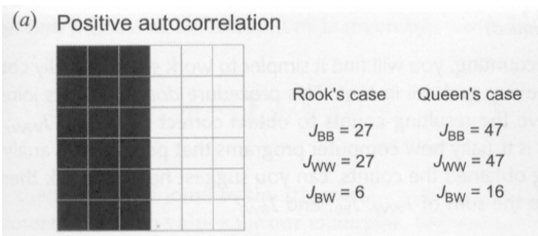


Spatial Autocorrelation
- Joins count approach

- **Possible joins:**
  – *JBB*: the number of joins of BB
  – *JWW*: the number of joins of WW
  – *JBW*: the number of joins of BW or WB

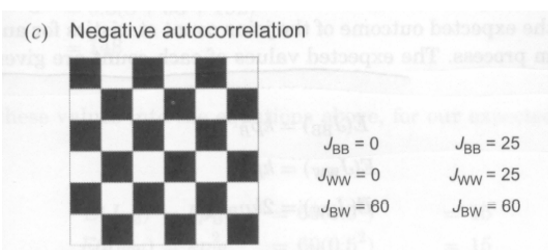
Spatial Autocorrelation
- Joins count approach

- Patterns (positive)?
  - Small JBW and large JBB & JWW



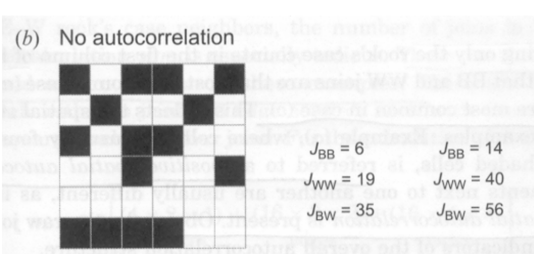
Spatial Autocorrelation
- Joins count approach

- Patterns (negative)?
  - Large JBW and small JBB & JWW



Spatial Autocorrelation
- Joins count approach

- Patterns (zero)?
  - Medium JBW and medium JBB & JWW



Spatial Autocorrelation
- Joins count approach

- Statistical tests for spatial correlation
- Under CSR:
  - Mean:

$$E(J_{BB}) = kp_B^2$$
$$E(J_{WW}) = kp_W^2$$
$$E(J_{BW}) = 2kp_B p_w$$

Where k is the total number of joins on the map
p_B is the probability of an area being coded B
p_W is the probability of an area being coded W

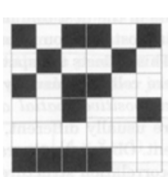
Spatial Autocorrelation
- Joins count approach

- Under CSR:
  - Standard Deviation

$$E(s_{BB}) = \sqrt{kp_B^2 + 2mp_B^3 - (k+2m)p_B^4}$$
$$E(s_{WW}) = \sqrt{kp_W^2 + 3mp_W^3 - (k+2m)p_W^4}$$
$$E(s_{BW}) = \sqrt{2(k+m)p_B p_W - 4(k+2m)p_B^2 p_W^2}$$

Where k is the total number of joins on the map
p_B is the probability of an area being coded B
p_W is the probability of an area being coded W

Spatial Autocorrelation
- Joins count approach



$$m = \frac{1}{2}\sum_{i=1}^{n} k_i(k_i - 1)$$

Rook case

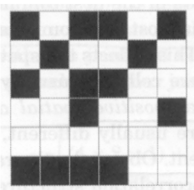
k_i is the number of joins to the ith area

$m = 0.5\ [(4\times2\times1) + (16\times3\times2) + (16\times4\times3)]$
corners edges center

= 148

Spatial Autocorrelation
- Joins count approach

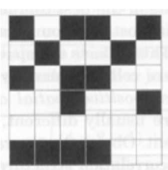
- Illustration



$$E(J_{BB}) = kp_B^2 = 60(0.5^2) = 15$$
$$E(J_{WW}) = kp_W^2 = 60(0.5^2) = 15$$
$$E(J_{BW}) = 2kp_Bp_W = 2(60)(0.5)(0.5) = 30$$

Spatial Autocorrelation
- Joins count approach

- Illustration



$$E(s_{BB}) = E(s_{WW}) = \sqrt{kp_B^2 + 2mp_B^3 - (k+2m)p_B^4}$$
$$= \sqrt{60(0.5)^2 + 2(148)(0.5)^3 - [60 + 2(148)](0.5)^4}$$
$$= \sqrt{29.75}$$
$$= 5.454$$

$$E(s_{Bw}) = \sqrt{2(k+m)p_Bp_W - 4(k+2m)p_B^2p_W^2}$$
$$= \sqrt{2(60+148)(0.5)(0.5) - 4[60 + 2(148)](0.5)^2(0.5)^2}$$
$$= \sqrt{15}$$
$$= 3.873$$

Spatial Autocorrelation
- Joins count approach

- Convert to *z*-scores

$$Z_{BB} = \frac{J_{BB} - E(J_{BB})}{E(s_{BB})}$$

$$Z_{BW} = \frac{J_{BW} - E(J_{BW})}{E(s_{BW})}$$

$$Z_{WW} = \frac{J_{WW} - E(J_{WW})}{E(s_{WW})}$$

A large *negative* *Z*-score on J_{BW} indicates *positive* autocorrelation since it indicates that there are fewer BW joins than expected.

A large *positive* *Z*-score on J_{BW} is indicative of *negative* autocorrelation.

Spatial Autocorrelation
- Joins count approach

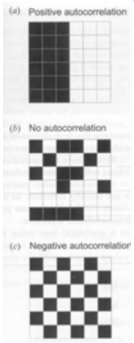


Table 7.2 *Z*-Scores for the Three Patterns in Figure 7.5 Using the Rook's Case

	Example		
Join type	*(a)*	*(b)*	*(c)*
BB	2.200	−1.650	−2.750
WW	2.200	0.733	−2.750
BW	−6.197	1.291	7.746

Exercise 13

- **Joins Count Statistics**

Spatial Autocorrelation
- Joins count approach

- Limitations:
  - Only applicable to binary data
    - not numeric data
  - Although the approach provides an indication of the strength of autocorrelation present in terms of *z*-scores, it is not readily interpreted, particularly if the results of different tests appear contradictory
  - The equations for the expected values of counts are fairly formidable.

Spatial Autocorrelation
- Moran's I

$$I = \frac{n}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}}$$

$$w_{ij} = \begin{cases} 1 & \text{If zone } i \text{ an zone } j \text{ are adjacent} \\ 0 & \text{otherwise} \end{cases}$$

Spatial Autocorrelation
- Moran's I

- Spatial autocorrelation measure: if near things similar (or dissimilar) to each other.
  - Nearness measure: w_{ij}
  - Similarity measure: co-variance $(y_i - \bar{y})(y_j - \bar{y})$
- w_{ij} switches on-off the covariance based on certain definition of nearness:

$$w_{ij}(y_i - \bar{y})(y_j - \bar{y})$$

Spatial Autocorrelation
- Weighting Matrix

a	b
2	0
2	0
c	d

$$A = \begin{array}{c|cccc} & a & b & c & d \\ a & 0 & 1 & 1 & 0 \\ b & 1 & 0 & 0 & 1 \\ c & 1 & 0 & 0 & 1 \\ d & 0 & 1 & 1 & 0 \end{array}$$

a	b
2	0
0	2
c	d

$$A = \begin{array}{c|cccc} & a & b & c & d \\ a & 0 & 1 & 1 & 0 \\ b & 1 & 0 & 0 & 1 \\ c & 1 & 0 & 0 & 1 \\ d & 0 & 1 & 1 & 0 \end{array}$$

Spatial Autocorrelation
- Moran's I

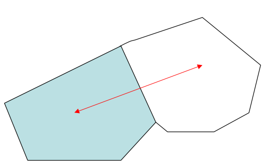
- Normalization by discounting
  - # of joins by neighbors: $\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}$
  - Variance of value *y*: $\dfrac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n}$

Spatial Autocorrelation
- Moran's I

- For Moran's *I*, a positive value indicates a positive autocorrelation, and a negative value indicates a negative autocorrelation.
- Moran's *I* is not strictly in the range of -1 to +1.

Spatial Autocorrelation
- Other Weighting Matrices

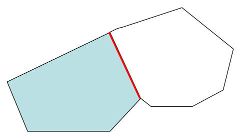
- **Using distance**



$$w_{ij} = \begin{cases} d_{ij}^{z} & \text{Where } d_{ij} < D \text{ and } z < 0 \\ 0 & \text{Where } d_{ij} > D \end{cases}$$

Spatial Autocorrelation
- Other Weighting Matrices

- Using the length of shared boundary

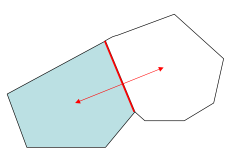


$$w_{ij} = \frac{l_{ij}}{l_i}$$

Where: l_i is the length of the boundary of zone i
l_{ij} is the length of boundary shared by area i and j

Spatial Autocorrelation
- Other Weighting Matrices

- Using both distance and the length of shared boundary



$$w_{ij} = \frac{d_{ij}^{z} l_{ij}}{l_i}$$

Where: l_i is the length of the boundary of zone i
l_{ij} is the length of boundary shared by area i and j

Exercise 14

- Moran's I

Spatial Autocorrelation
- Geary's C

- Proposed by Geary's contiguity ratio C

$$C = \frac{n-1}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i - y_j)^2}{2\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}}$$

$$w_{ij} = \begin{cases} 1 & \text{If zone } i \text{ an zone } j \text{ are adjacent} \\ 0 & \text{otherwise} \end{cases}$$

Spatial Autocorrelation
- Geary's C

- Spatial autocorrelation measure: if near things similar (or dissimilar) to each other.
  - Nearness measure: w_{ij}
  - Similarity measure: squared distance

$$(y_i - y_j)^2$$

Spatial Autocorrelation
- Geary's C

- The value generally varies between 0 - 2.
- The theoretical value of C is 1 under CSR. values < 1 indicate positive spatial autocorrelation while values > 1 indicate negative autocorrelation

Spatial Autocorrelation
- Local Indicators

- Global statistics tell us whether or not an overall configuration is autocorrelated, but not where the unusual interactions are.
- Local indicators of spatial association (**LISA**) were proposed in Getis and Ord (1992) and Anselin (1995).
- These are disaggregate measures of autocorrelation that describe the extent to which particular areal units are similar to, or different from, their neighbors.
- LISA: **mapable**

Spatial Autocorrelation
- Local Indicators

- **Local *Gi***
  - Used to detect possible non-stationarity in data, when clusters of similar values are found in specific subregions of the study area.

$$G_i = \frac{\sum_{j \neq i} w_{ij} y_i}{\sum_{i=1}^{n} y_i}$$

$$w_{ij} = \begin{cases} 1 & \text{If zone } i \text{ an zone } j \text{ are adjacent} \\ 0 & \text{otherwise} \end{cases}$$

Spatial Autocorrelation
- Local Indicators

- **Local Moran's *I***

$$I_i = z_i \sum_{j \neq i} w_{ij} z_j$$

- Local Moran's I decomposes Moran's I into contributions for each location, I_i. The sum of I_i is proportional to Moran's I
- Two interpretations:
  - ✓indicator of localized clusters
  - ✓diagnostic for outliers in global spatial patterns.

Where $z_i = (y_i - \bar{y}) / s$

W matrix can be row-standardized (i.e. scaled so that each row sums to 1)

Exercise 15

- **Local Moran's I**

Spatial Autocorrelation - Local Indicators

- ***Local Geary's C***

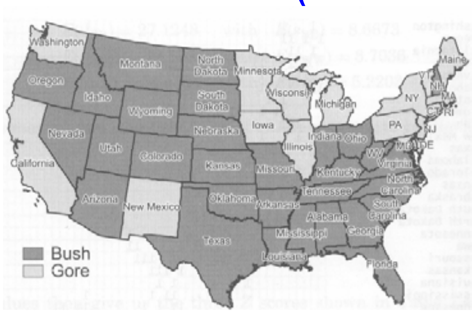
$$C_i = \sum w_{ij}(y_i - y_j)^2$$

- Local Geary's C decomposes Geary's C into contributions for each location, C_i. The sum of I_i is proportional to Geary's C
- Two interpretations:
  - ✓indicator of localized clusters
  - ✓diagnostic for outliers in global spatial patterns.

Review

- Global:
  - A simple test: Joins Count
  - Moran's I
  - Geary's C
- Local:
  - Local Moran's I
  - Local Geary's C
  - Local Getis's Gi

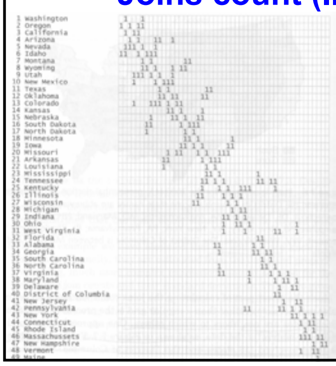
Spatial Autocorrelation
- Joins count (Illustration)



B: Bush
W: Gore

State-level results for the 2000 U.S. presidential election

Spatial Autocorrelation
- Joins count (Illustration)



Adjacency (joins) matrix: if two states share a common boundary, they are adjacent.

Spatial Autocorrelation
- Joins count (Illustration)

Bush
48,021,500
$p_B = 0.49885$

Gore
48,242,921
$p_w = 0.50115$

$E(J_{BB}) = 27.1248$ with $E(s_{BB}) = 8.6673$
$E(J_{WW}) = 27.3755$ with $E(s_{WW}) = 8.7036$
$E(J_{BW}) = 54.4997$ with $E(s_{BW}) = 5.2203$

$J_{BB} = 60$
$J_{WW} = 21$
$J_{BW} = 28$

Join type	*z-score*
BB	3.7930
WW	−0.7325
BW	−5.0763