GEOG/Math574G Introduction to Geostatistics

Introduction

Daoqin Tong

School of Geography & Development
408 Harvill Building
Email: daoqin@email.arizona.edu

Spatial Data

- Properties
  - Location (coordinates)
  - Attribute (variables)
- Data matrix

FID	Shape*	AREA	TRACT	POPULATION
0	Polygon	15.157422	009740	3931
1	Polygon	18.441607	009590	4354
2	Polygon	4.801734	009510	2093
3	Polygon	4.83201	009450	2720
4	Polygon	41.93699	0098	6590
5	Polygon	20.616666	008160	7755
6	Polygon	1.889164	008130	8571
7	Polygon	6.341103	008140	2735
8	Polygon	1.745984	008350	3593
9	Polygon	0.941483	008311	4057
10	Polygon	0.816985	008120	4026

Spatial Data

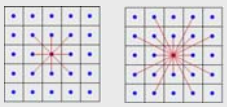
- Bad news: many of the standard techniques and methods documented in standard statistics textbooks have significant problems when we try to apply them to the analysis of the spatial data.
- Good news: Geospatial referencing provides us with a number of new ways of looking at data and the relations among them. (e.g. distance, adjacency, interaction, and neighbor)

Spatial Is *Special*

- Properties
  - Distance
  - Adjacency

  - Interaction

  - Neighborhood



Pitfalls of spatial data

- Spatial data always violate the fundamental requirement of conventional statistical analysis
  - Spatial autocorrelation
  - Modifiable areal unit problem
  - Ecology fallacy
  - Scale
  - Nonuniformity of space
  - Edge effect

Spatial

- "After choosing the area we usually have no guidance beyond the widely verifiable fact that patches in *close proximity* are commonly *more alike*, as judged by the yield of crops, than those which are *far apart*"

-----R.A. Fisher (1935) on analyzing agricultural field trials

- "Everything is related to everything else, but near things are more related to each other"

------W. Tobler, First law of geography

MAUP

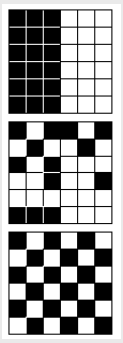
- Scale problem
  - Results vary when areal units are progressively aggregated into fewer and larger units
- Aggregation problem
  - Results vary due to the use of alternative aggregation schemes at equal or similar scales (resolutions)

Spatial autocorrelation

- *"Everything is related to everything else, but near things are more related than distant things" (Waldo* Tobler, 1970)
- If the "first law" was not generally true, a geography major would be irrelevant
  - Think of cities, storms, landforms
  - The opposite would just be ***random chaos***

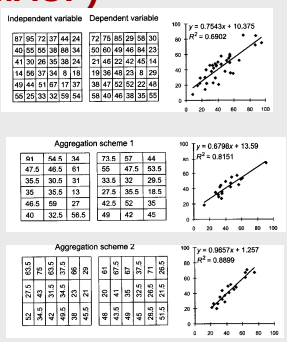
Spatial autocorrelation (cont.)

- This leads to biased parameter estimates
- This also introduces redundancy in samples
  - n=50 is not really what your statistics textbook tells you
- Can be
  - Positive – nearby locations similar
  - Negative – nearby locations different.
  - Zero – no correlation in space



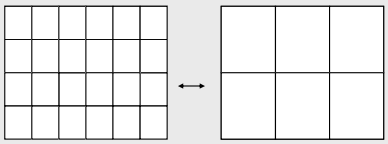
Modifiable areal unit problem (MAUP)

- Areal units often arbitrary
- Changing the unit borders or regrouping units (sometimes) lead to very different results
- More than academic or theoretical interest



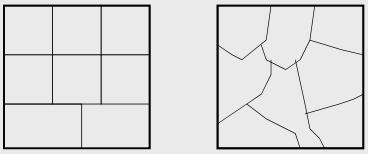
MAUP (cont.)

- Scale effect
  - Involves the aggregation of smaller units into larger ones.
  - Generally speaking, the larger the spatial units, the stronger the relationship among variables.



MAUP (cont.)

- Zoning effect
  - Units are arbitrary defined and different organization of the units may create different analytical results.

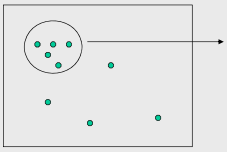


MAUP (cont.)

- Possible solutions to MAUP
  - A basic entity approach
    - ⇒ Identify entities that are meaningful and not modifiable
  - An optimal zoning approach
    - ⇒ A system that maximizes interzonal variation and minimizes intrazonal variation
  - A sensitivity analysis approach

Nonuniformity

- Space is not uniform

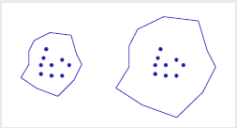


Area with high crime rates?

Crime locations

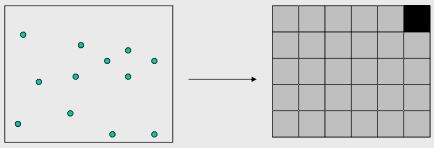
Edge effect

- Study areas are often bounded due to practical constraints or research interests
- However, spatial phenomena under study may not be bounded
- Both the size and shape of boundary can affect the results of geographic analysis



Edge effect (cont.)

- Another problem associated with artificial boundary is that sites in the center of the study area can have nearby observations in all directions, whereas sites at the edges only neighbors toward the center of the study area



Spatial interpolation

Statistics for Spatial Data

- Data are observed at (sometimes imprecise) spatial locations
- Data and/locations are modeled as random
- Spatial locations are used in statistical analysis, often to model statistical dependence

Spatial Statistical Model

$$\{Z(s) : s \in D \subset R^d\}$$

$Z(s)$: random variable/vector/set located at s

D : random set in R^d

Spatio-temporal statistical model

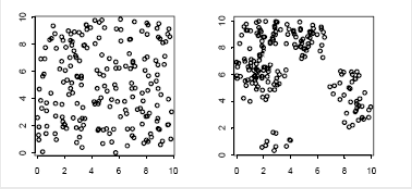
$$\{Z(s;t) : s \in D(t), t \in T\}$$

$D()$temporal process of sets in R^d

T random set in time - dimension

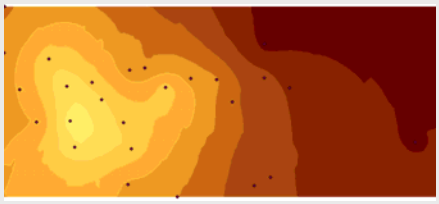
Point Pattern

- Point patterns
  - Locations are random variables of interest
  - Attributes (marked) or No Attributes (unmarked)



Spatially Continuous Data

- Attributes are random variables of interest
- Attributes exist at infinite many points but only observed at a finite number of points



Area Data (Lattice Data)

- Attributes are random variables of interest
- Attributes exist and are observed on finite locations
- Area can be regular or irregular