Why Is It There?

Getting Started with Geographic Information Systems

Chapter 6

Dursun Z. ªeker

Objective of this chapter;

The ultimate purpose of using GIS is to analyze data. The power of GIS is that we can do spatial analysis in addition to basic statistical analysis. We introduce these methods, and discuss how they can be used to seek out new information.

6 Why Is It There?

- 6.1 Describing Attributes
- 6.2 Statistical Analysis
- 6.3 Spatial Description
- 6.4 Spatial Analysis
- 6.5 Searching for Spatial Relationships
- 6.6 GIS and Spatial Analysis

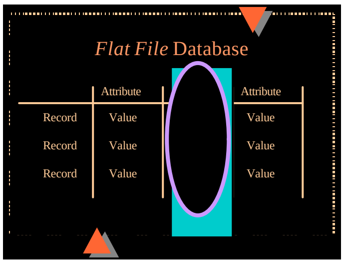
Dueker (1979)

- **"a geographic information system is a special case of information systems where the database consists of observations on spatially distributed features, activities or events, which are definable in space as points, lines, or areas. A geographic information system manipulates data about these points, lines, and areas to retrieve data for ad hoc queries and analyses".**

GIS is capable of data analysis

- Attribute Data
  - Describe with statistics
  - Analyze with hypothesis testing
- Spatial Data
  - Describe with maps
  - Analyze with spatial analysis

Describing one attribute



Attribute Description

- The extremes of an attribute are the highest and lowest values, and the range is the difference between them in the units of the attribute.
- A histogram is a two-dimensional plot of attribute values grouped by magnitude and the frequency of records in that group, shown as a variable-length bar.
- For a large number of records with random errors in their measurement, the histogram resembles a bell curve and is symmetrical about the mean.

ITU Photogrammetry Division

If the records are:

- Text
  - Semantics of text e.g. “Hampton”
  - word frequency e.g. “Creek”, “Kill”
  - address matching
- Example: Display all places called “State Street”

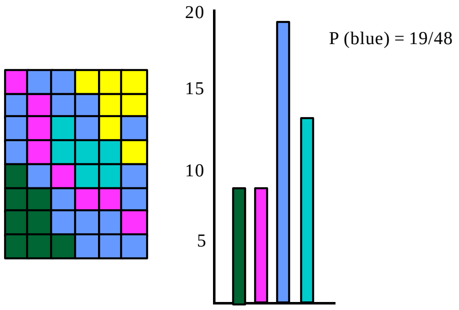
ITU Photogrammetry Division

If the records are:

- Classes
  - histogram by class
  - numbers in class
  - contiguity description, e.g. average neighbor (roads, commercial)

ITU Photogrammetry Division

Describing a classed raster grid



ITU Photogrammetry Division

If the records are:

- Numbers
  - statistical description
  - min, max, range
  - variance
  - standard deviation

ITU Photogrammetry Division

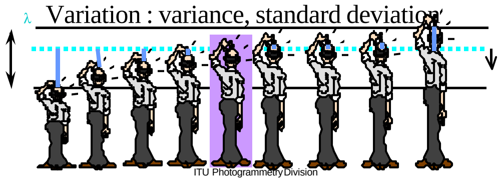
Measurement

- One: all I have! [6:00pm]
- Two: do they agree? [6:00pm;6:04pm]
- Three: level of agreement [6:00pm;6:04pm;7:23pm]
- Many: average all, average without extremes
- Precision: 6:00pm. “About six o’clock”

ITU Photogrammetry Division

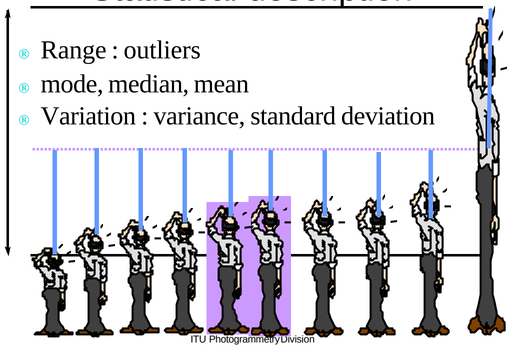
Statistical description

- Range : min, max, max-min
- Central tendency : mode, median (odd, even), mean
- Variation : variance, standard deviation



Statistical description

- Range : outliers
- mode, median, mean
- Variation : variance, standard deviation



Computing the Mean

Sum of attribute values across all records, divided by the number of records.

Add all attribute values down a column, / by # records

A representative value, and for measurements with normally distributed error, converges on the true reading.

A value lacking sufficient data for computation is called a missing value. Does not get included in sum or n.

Mean

- Statistical average
- Sum of the values for one attribute divided by the number of records

$$\bar{X} = \sum_{i=1}^{n} X_i / n$$

Variance

The total variance is the sum of each record with its mean subtracted and then multiplied by itself.

The standard deviation is the square root of the variance divided by the number of records less one.

For two values, there is only one variance.

Standard Deviation

- Average difference from the mean
- Sum of the mean subtracted from the value for each record, squared, divided by the number of records-1, square rooted.

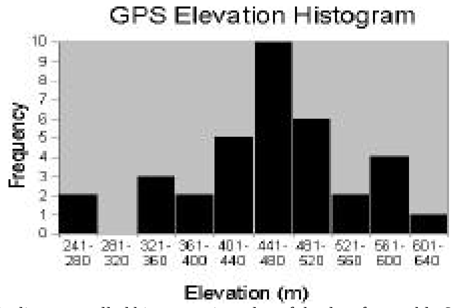
$$\text{st.dev.} = \sqrt{\frac{\Sigma (X_i - \bar{X})^2}{n-1}}$$

GPS Example Data: Elevation

Table 6.2: Sample GPS Readings

Data Extreme	Date	Time	D M S	D M S	Elev
Minimum	6/14/95	10:47am	42 30 54.8	75 41 13.8	247
Maximum	6/15/95	10:47pm	42 31 03.3	75 41 20.0	610
Range	1 Day	12 hours	00 8.5	00 6.2	363

Elevation (book example – p.172)



This diagram, called histogram is a plot of the data from table 6.1.

GPS Example Data: Elevation Standard deviation

Same units as the values of the records, in this case meters.

Average amount readings differ from the average

Can be above of below the mean

Elevation is the mean (459.2 meters)

plus or minus the expected error of 82.92 meters

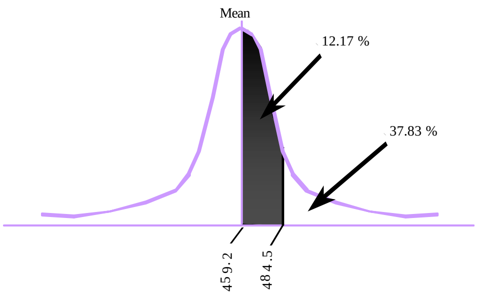
Elevation is most likely to lie between 376.28 meters and 542.12 meters.

These limits are called the error band or margin of error.

Samples and populations

- A sample is a set of measurements taken from a larger group or population.
- Sample means and variances can serve as estimates for their populations.
- Easier to measure with samples, then draw conclusions about entire population.

The Bell Curve



Testing Means

Mean elevation of 459.2 meters

standard deviation 82.92 meters

what is the chance of a GPS reading of 484.5 meters?

484.5 is 25.3 meters above the mean

0.31 standard deviations (Z-score)

0.1217 of the curve lies between the mean and this value

0.3783 beyond it

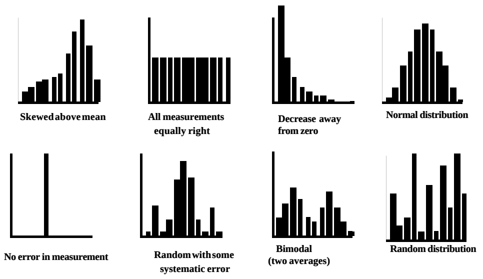
Hypothesis testing

- Set up NULL hypothesis (e.g. Values or Means are the same) as H_0
- Set up ALTERNATIVE hypothesis. H_1
- Test hypothesis. Try to reject NULL.
- If null hypothesis is rejected alternative is accepted with a calculable level of confidence.

Testing the Mean

- Mathematical version of the normal distribution can be used to compute probabilities associated with measurements with known means and standard deviations.
- A test of means can establish whether two samples from a population are different from each other, or whether the different measures they have are the result of random variation.

Alternative attribute histograms



Accuracy

- Determined by testing measurements against an independent source of higher fidelity and reliability.
- Must pay attention to units and significant digits.
- Can be expressed as a number using statistics (e.g. expected error).
- Accuracy measures imply accuracy users.

The difference is the map

- GIS data description answers the question: Where?
- GIS data analysis answers the question: Why is it there?
- GIS data description is different from statistics because the results can be placed onto a map for visual analysis.

Spatial Statistical Description

- For coordinates, the means and standard deviations correspond to the mean center and the standard distance
- A centroid (mean center) is any point chosen to represent a higher dimension geographic feature, of which the mean center is only one choice.
- The standard distance for a set of point spatial measurements is the expected spatial error.

Spatial Statistical Description

- For coordinates, data extremes define the two corners of a bounding rectangle.

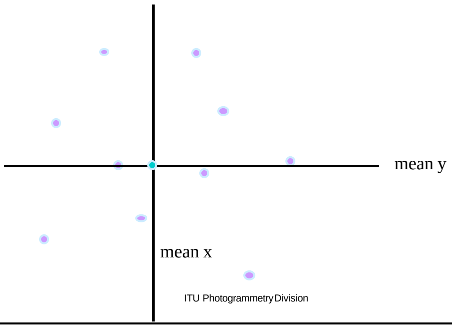
ITU Photogrammetry Division

Geographic extremes

- Southernmost point in the continental United States.
- Range: e.g. elevation difference; map extent
- Depends on projection, datum etc.

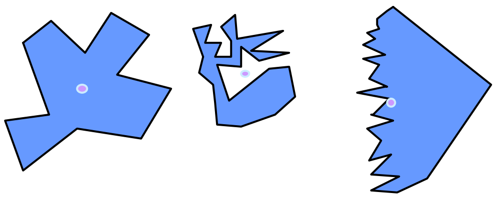
ITU Photogrammetry Division

Mean Center



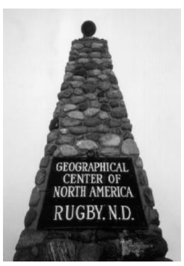
ITU Photogrammetry Division

Centroid: mean center of a feature



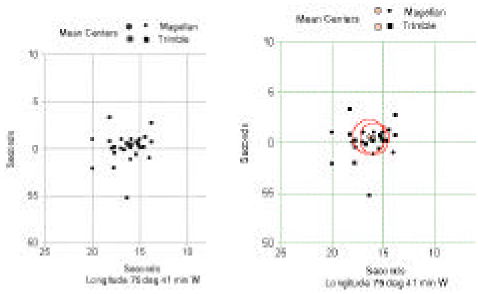
ITU Photogrammetry Division

Mean center?



ITU Photogrammetry Division

Comparing spatial means

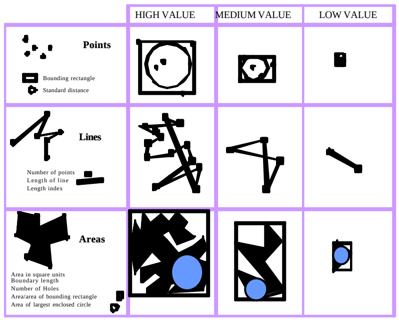


ITU Photogrammetry Division

GIS and Spatial Analysis

- Descriptions of geographic properties such as shape, pattern, and distribution are often verbal
- Quantitative measure can be devised, although few are computed by GIS.
- GIS statistical computations are most often done using retrieval options such as buffer and spread.
- Also by manipulating attributes with arithmetic commands (map algebra).

Statistics and features

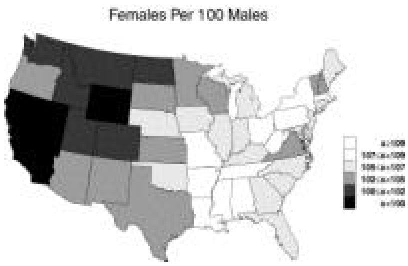


Numbers that describe points, lines, and areas.

An example

- Lower 48 United States
- 1996 Data from the U.S. Census on gender
- Gender Ratio = # females per 100 males
- Range is 96.4 - 114.4
- What does the spatial distribution look like?

Gender Ratio by State: 1996

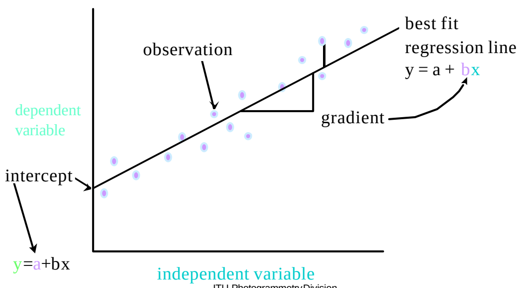


Searching for Spatial Pattern

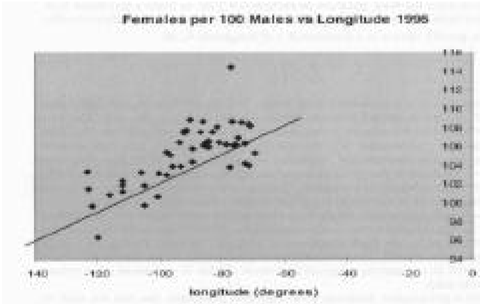
A linear relationship is a predictable straight-line link between the values of a dependent and an independent variable. (y = a + bx) It is a simple model of the relationship.

A linear relation can be tested for goodness of fit with least squares methods. The coefficient of determination r-squared is a measure of the degree of fit, and the amount of variance explained.

Simple linear relationship



Testing the relationship



gr = 117.46 + 0.138 long.

Patterns in Residual Mapping

Differences between observed values of the dependent variable and those predicted by a model are called residuals.

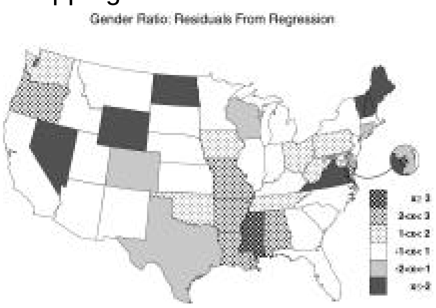
A GIS allows residuals to be mapped and examined for spatial patterns.

A model helps explanation and prediction after the GIS analysis.

A model should be simple, should explain what it represents, and should be examined in the limits before use.

We should always examine the limits of the model's applicability (e.g. Does the regression apply to Europe?)

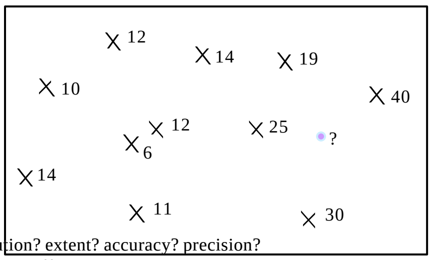
Mapping residuals from a model



Unexplained variance

- More variables?
- Different extent?
- More records?
- More spatial dimensions?
- More complexity?
- Another model?
- Another approach?

Example: Spatial Analysis



resolution? extent? accuracy? precision?
boundary effects? point spacing?

meters to water table

GIS and Spatial Analysis

Geographic inquiry examines the relationships between geographic features collectively to help describe and understand the real-world phenomena that the map represents.

Spatial analysis compares maps, investigates variation over space, and predicts future or unknown maps.

Many GIS systems have to be coaxed to generate a full set of spatial statistics.

Analytic Tools and GIS

Tools for searching out spatial relationships and for modeling are only lately being integrated into GIS.

Statistical and spatial analytical tools are also only now being integrated into GIS, and many people use separate software systems outside the GIS.

Real geographic phenomena are dynamic, but GISs have been mostly static. Time-slice and animation methods can help in visualizing and analyzing spatial trends.

GIS places real-world data into an organizational framework that allows numerical description and allows the analyst to model, analyze, and predict with both the map and the attribute data.

You can lie with...

Maps

Statistics

- Correlation is not causation!
- Hypothesis vs. Action

Coming next ...

Making Maps with GIS