Cornell University
Cornell Statistical Consulting Unit

StatNews #57

Location, Location, Location…or Spatial Data
April 2003
Revised 2012

Over the past few years, there has been a sharp increase in the use of spatial data by researchers at Cornell. A spatial data set contains information regarding the location of objects in space. Some researchers use spatial data to produce maps on which they can superimpose information regarding phenomena such as the spread of a disease. Others will use spatial data to incorporate spatial information in model building and testing. The use of location in the prediction of housing prices is an example of this.

The first step is to build the spatial information. There are many different ways by which the location can be referenced in a spatial data set. First, the location can be either point or areal referenced. A point location can be described by its latitude or longitude, for example, whereas an area could be referenced by the longitude and latitude of its center. Both the point and the areal locations can be sampled from a spatial continuum or from a discrete set. In addition, they can be taken regularly, for example, along a grid or defined by fixed blocks, or they can be sampled irregularly within a space.

Powerful Geographic Information Systems (GIS) software will help build a spatial data set by adding longitude and latitude variables to existing data. Another possibility is to use one of the numerous existing spatial databases and merge the location information present in these data sets with one's own data. The Cornell University Geospatial Information Repository (CUGIR - http://cugir.mannlib.cornell.edu/) at Mann library is a good place to start looking for spatial data. A new website (http://css.cals.cornell.edu/) will also give you information regarding geospatial research here at Cornell. The most commonly used GIS software on campus is ArcInfo, for which a Cornell site license is available. Other software, such as SAS, also offers GIS capabilities. A two-hour introductory GIS workshop is offered every semester at Mann Library[1], and semester-long courses on GIS are taught in both the Crop and Soil Sciences department and the City and Regional Planning department.

Once the spatial data set is completed, data can be visualized or explored via more or less detailed maps created with GIS software. Besides the GIS software, routine statistical methods offered in the more common statistical packages can often also be used to obtain a first exploratory analysis.

Besides visualizing and exploring spatial data, many researchers are interested in pursuing more advanced data analysis. Depending on the data, this might include: assessing spatial correlation, studying patterns of clustering, interpolating attribute data for predictive purposes, or modeling

[1] For current workshop information, check the website at http://mannlib.cornell.edu/research-help/gis for updated information. 2012, May

attribute data as depending on covariates while taking into consideration spatial correlation. Like time series data, spatial data are usually correlated such that points closer together are more similar than those further apart. Unlike one-dimensional time, spatial data can be correlated in two dimensions. Both S-Plus and SAS, which are commonly used at Cornell, can run most spatial data analyses. Two relevant courses are offered: Spatial Statistics (in Natural Resources) and Spatial Modeling and Analysis (in Crop and Soil Sciences). Please contact Francoise Vermeylen in the Office of Statistical Consulting for assistance with the statistical analysis of spatial data.

Useful references for spatial data are:

- Statistics for Spatial Data, Noel Cressie, New York, Wiley, 1993
- Interactive Spatial Data Analysis, Trevor Bailey and Anthony Gatrell, Harlow, Longman Scientific & Technical, 1995
- http://www.gis.com

Author: Francoise Vermeylen

Back to StatNews Table of Contents

(This newsletter was distributed by the Cornell Statistical Consulting Unit. Please forward it to any interested colleagues, students, and research staff. Anyone not receiving this newsletter who would like to be added to the mailing list for future newsletters should contact us at cscu@cornell.edu. Information about the Cornell Statistical Consulting Unit and copies of previous newsletters can be obtained at http://www.cscu.cornell.edu).