

Point Pattern Analysis

Frances F. Burden

This document provides a brief overview of Point Pattern Analysis.

Definition

Point Pattern Analysis involves the ability to describe patterns of locations of point events and test whether there is a significant occurrence of clustering of points in a particular area.

Overview

Historically, Point Pattern Analysis was first noted in the works of botanists and ecologists in the 1930s (Chakravorty, 1995). However, in the intervening years, many different fields have also started to use point pattern analysis, such as archeology, epidemiology, astronomy, and criminology. In general, Point Pattern Analysis can be used to describe any type of incident data. For instance, we may want to conduct “Hot Spot” analysis in order to better understand locations of crimes, or else we may want to study breakouts of certain diseases to better see whether there is a pattern. In both of these cases, Point Pattern Analysis can be of great help to institutions and policymakers in their decisions on how to best allocate their scarce resources to different areas.

5 Criteria

In order to conduct Point Pattern Analysis, your data must meet five important criteria:

1. The pattern must be mapped on a plane, meaning that you will need both latitude and longitude coordinates.
2. A study area must be selected and determined prior to the analysis.
3. The Point Data should not be a selected sample, but rather the entire set of data you seek to analyze.
4. There should be a one-to-one correspondence between objects in the study area and events in the pattern.
5. The Points must be true incidents with real spatial coordinates. For example, using the centroids of a census tract would not be an especially useful process.

Techniques to Analyze Point Pattern Data

When we are examining incident data, we often need to first get the coordinates of each incident and determine the study area that we wish to use. For instance, if we were examining one hundred robberies within a square mile, we would not want to use a study area of 5 square miles. Although this may sound obvious, we also want to examine our data and make sure that we are not estimating beyond areas, for which we have no data. In general, when we are examining areas to see whether incidents are clustered we are using a null hypothesis that there is no clustering present and that incidents are evenly spread throughout the study area. Sometimes, we may specify that incidents are evenly clustered, controlling for certain variables, such as population density.

In general there are Three Types of Techniques: 1.) Quadrant Count Methods, 2.) Kernel Density Estimation (sometimes called K-Means), and 3.) Nearest Neighbor Distance

- 1.) Quadrant Count Methods: This method involves simply recording and counting the number of events that occur in each quadrant.
 - In general, it is important to remember that large quadrants produce a very coarse description of the pattern, but as quadrant size is reduced, many areas may become too small and some may contain no events at all.
 - Two examples of this type of point pattern analysis are Mode and Fuzzy Mode.
- 2.) Kernel Density Estimation: This method counts the incidents in an area (a kernel), centered at the location where the estimate is made. This analysis is a partitioning technique, meaning that incidents are partitioned into a number of different clusters. Oftentimes the user is able to specify the number of clusters. In some forms of this analysis, all incidents, even the outliers, are assigned to one and only one group. However, other techniques allow for a form of “clumping” analysis, where there are groups that have overlapping membership.
 - This method is very good for analyzing the point patterns to discover the Hot Spots.
 - Also, this method provides us with a useful link to geographical data because it is able to transform our data into a density surface.
 - Our choice of r , the kernel bandwidth strongly affects our density surface.
 - Also, we can weight these patterns with other data – such as density of populations and unemployment rates.
 - In Dual Kernel Estimates, you are able to weight the estimates against another set of incidents. For instance you might want to analyze the number of assaults against establishments that are allowed to serve liquor.
- 3.) Nearest-Neighbor Distance: This method measures the distance from one point to the nearest neighbor point. In general there are three different functions that users are able to employ in Nearest Neighbor Analyses:
 1. **G Function**: This is the simplest measure and is similar to the mean, however instead of summarizing with a mean, the G function allows us to examine the cumulative frequency distribution of the nearest neighbor distances. The shape of this function can tell us a lot about the way the events are clustered in a point pattern. If events are clustered together, G increases rapidly at short distances, and if events are evenly spaced, G increases slowly up to the distance at which most events are spaced, and only then increases rapidly.
 2. **F Function**: Instead of accumulating the fraction of the nearest-neighbor distances between events, this measure selects point locations anywhere in the study region at random, and the minimum distance from them to any event in the pattern is determined.
 3. **K Function**: Imagine placing circles of a defined radius centered on the event in turn. Then, the number of events inside the circle’s radius is totaled, and the mean count for all of the incidents is totaled. This mean count is then divided by the overall study area. Because all of the incidents are used, the K function provides more information about patterns and clusters than either G or F.

Future Learning Opportunities

Every semester PRI offers classes on CrimeStat, a freeware program that can be downloaded over the Internet and is a good tool for analyzing point pattern data. For related information see the GIS Resource Documents on CrimeStat, ArcGIS Geostatistical Analyst, ArcGIS Spatial Analyst, and SpaceStat.