

Point Pattern Analysis

Prepared by Brian McManus

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. 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. As such, 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.

Criteria and Techniques

In order to conduct Point Pattern Analysis, data must meet these important criteria: 1.) The pattern must be mapped on a plane, and so must include 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 especially useful.

Quadrant Count Methods: This method involves simply recording and counting the number of events that occur in each quadrant. 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. Examples of this type of point pattern analysis are Mode and Fuzzy Mode.

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 Hot Spots and it provides with a useful link to geographical data because it is able to transform data into a density surface. These patterns can be weighted with other data, such as density of populations and unemployment rates. Dual Kernel Estimates can be weighted against another set of incidents.

Nearest-Neighbor Distance: This method measures the distance from one point to the nearest neighbor point. There are three different functions that users are able to employ in Nearest Neighbor Analyses. The **G Function** is the simplest measure and is similar to the mean, however instead of summarizing with a mean, the G function can be used to examine the cumulative frequency distribution of the nearest neighbor distances. The shape of this function can indicate 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. The **F Function** selects point locations anywhere in the study region at random, rather than accumulating the fraction of the nearest-neighbor distances between events, and the minimum distance from them to any event in the pattern is determined. The **K Function** provides more information about patterns and clusters than either G or F because it considers all incidents. Its use involves 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.

To access the original extended version of this document by Frances F. Burden, see http://www.pop.psu.edu/gia-core/pdfs/gis_rd_03-41.pdf.