

# A Tutorial with a Sample Data Set

The following is a short tutorial on entering data into *CrimeStat IV* and running routines. We will use one of the sample data sets that were provided (General Sample Data.zip). Unzipping this file reveals two individual files called *Incident.dbf* and *BaltPop.dbf*.

The incident file is a collection of incident locations that have been randomly simulated while the other file includes the 1990 population of census block groups, both from the Baltimore region.<sup>1</sup> Both files have locations coded in spherical (longitude-latitude) coordinates. The X/Y coordinates for the incident file is the location where the incident (crime) occurred. The X/Y coordinates for the block groups is the centroid location.

1. Start the *CrimeStat* program by either double-clicking on the *CrimeStat* icon on the desktop (if installed) or else opening Windows Explorer and locating the directory where *CrimeStat* is stored and double-clicking on the file called *crimestat.exe*.
2. Once the program splash page closes, the user will be looking at the **Data Setup** page with the Primary File page open.
3. Click on 'Select Files' followed by 'Browse'. Locate the file called *Incident.dbf* and click on 'Open' followed by 'OK'.
4. The file name will now be listed for the X, Y, Z (intensity), Weight, and Time fields. This variable, however, only has three fields - ID, Lon, Lat, indicating an record number, the longitude and latitude of the incident location.
5. Identify the appropriate fields under the Column heading by clicking on the cell and scrolling down to the appropriate name. For the X variable, the relevant name is Lon and for the Y variable, the relevant name is Lat (i.e., that is the names used for coordinates in this file. However, the variables will not always be simply named). For this example, there are no intensity, weight or time variables.
6. Under Type of Coordinate System, be sure that 'Longitude/latitude (spherical)' is checked since this data set use spherical coordinates.

---

<sup>1</sup> Note: the incident locations have had random coordinates assigned so this file should not be used for research.

7. Because the coordinate system is spherical, the data units are automatically decimal degrees. If they were projected, one would have to choose the particular units - feet, meters, miles, kilometers, or nautical miles. This finishes the setup for the primary file.
8. Next, Click on the Secondary File tab.
9. Again, click on select files, locate and open the *BaltPop.dbf* file. This is a file of census block groups. You are going to treat each block group as a 'pseudo-point', that is, as a single point which represents the block group. That point is the centroid of the block group. The population will be treated as residing exactly at that point.<sup>2</sup>
10. Once loaded, this file has six variables: Blockgroup, lon, lat, area, density, and Totpop.
11. Define the particular variables. For this file, the X variable is Lon and the Y variable is Lat. Also, define a Z (intensity) variable with Totpop. Note, that you could also assign this name to the Weight variable. Whether the population variable is assigned to the Intensity or Weight variable does not matter to the calculation. However, do not assign this name to both the intensity and the weight (i.e., only use one). This finishes the setup for the secondary variable.
12. Click on the Reference File tab. For these data, you will define a rectangle that covers the study area by identifying the X and Y coordinates for the lower-left corner of the rectangle and the upper-right corner of the rectangles. The following coordinates will work (Table Tutorial.1):

**Table Tutorial.1:  
Coordinates for Corners of Sample Data Set**

	X	Y
Lower-left corner	-76.91	39.19
Upper-right corner	-76.32	39.72

---

<sup>2</sup> The population does not live at the centroid, of course, unless the block group is a single building. But by treating the block group as a pseudo-point, we can analyze the population (or any other characteristic of the block group).

13. You will also need to tell the program how many columns you want it to calculate. The default value of 100 is fine. If you want it finer, type in a larger number. If you want it cruder, type in a smaller number. This finishes the Reference File setup.
14. Click on the Measurement Parameters tab. There are three parameters that have to be defined.
  - A. For many routines, an area estimate is needed. For this sample set, 684 square miles works.
  - B. For the linear nearest neighbor statistic only, the program needs the total length of the street network. In this data, the total street length of the Tiger Files for Baltimore City and Baltimore County is 4868.9 miles.
  - C. Finally, the type of distance measurement has to be defined, direct or indirect. For this example, use direct measurement.
15. The data setup is now finished. If you want to re-use this data setup, click on the Options page and 'Save parameters'. Define a file name and be sure to give it a 'param' extension (e.g., SampleData.param). The next time you want to run this data set, all you'll need to do is click on the **Options** page, click on 'Load parameters', and click on the name of the parameters file that you saved.
16. You are now ready to run some statistics. For this example, you will run only four statistics.
17. First, click on the **Spatial Description** page and then click on the Spatial Distribution tab.
  - A. Check the Mean center and standard distance (Mcsd) box. Then, click on the 'Save result to' button and identify which GIS program you are writing to (ArcGIS® 'shp'; several Ascii formats; MapInfo® 'MIF) and give it a name (e.g., SampleData).
  - B. Also, check the Standard deviational ellipse (Sde) box and, similarly, choose a file output with a name. You can use the same name (e.g., SampleData). *CrimeStat* will assign a **unique** prefix to each graphical object.

18. Second, click on the **Hot Spot Analysis** tab followed by the 'Hot Spot' Analysis I sub-heading. Then, check the Nearest Neighbor Hierarchical Clustering (Nnh) box. For this example, keep the default search radius, minimum points per cluster, and number of standard deviations for the ellipses. Also, click on 'Save ellipses to', select a GIS file output, and give it a name. Again, you can use the same name as with the other statistics.
19. Third, click on the **Spatial Modeling I** page and then the 'Interpolation I' tab. Check the dual kernel density interpolation box. This routine will interpolate the incident distribution (primary file) relative to the population distribution (secondary file). For this example, keep the default kernel parameters (these are explained in more detail in Chapter 10). Because the secondary variable is weighted by population (defined as the 'Intensity' variable) be sure to check the 'Use intensity' variable box towards the bottom of the page. This ensures that the dual kernel routine will interpolate the population variable that you assigned when you set up the secondary file.
20. You are now ready to run the statistics. Click on the 'Compute' button. The routine will run until all four routines that you selected are finished; the time will depend on the speed of your computer.
21. Each of the outputs is displayed on a separate results tab. You can print any of these results by clicking on 'Save to text file' (one at a time).
22. You can also display the graphical objects created by the routine in your GIS. Click on 'Close' to close the results window. Then, bring up your GIS and find the objects created by this run. There will be a number of graphical objects associated with the mean center routine (having prefixes of Mc, Xyd, Sdd, Gm, and Hm; see Chapter 4 for details). There will be two graphical objects associated with the nearest neighbor clustering routine (with prefixes of Nnh1 and Nnh2). Finally, there will be a grid object created by the dual kernel routine with a Dk prefix. You can load these objects into a GIS and display them along with the data file. For the dual kernel grid, you will need to graph the variable called "Z" to see the pattern.

23. For example, Figure Tutorial.1 shows an *ArcGIS*<sup>®</sup> map of 1996 vehicle thefts in Baltimore City and Baltimore County along with the standard deviational ellipse of the vehicle thefts, calculated with *CrimeStat*. *CrimeStat* outputs the ellipse as a shape file, which is then brought directly into *ArcGIS*<sup>®</sup>. A similar output could have been done for *MapInfo*<sup>®</sup>. Most of the statistics in *CrimeStat* have similar visual representations that can be displayed in a GIS program.
  
24. When you are finished with *CrimeStat*, click on 'Quit' to exit the program.

This finishes the short tutorial. *CrimeStat* is very easy to set up and to run. The chapters of the *CrimeStat* manual discuss the various statistics in the program and provide many examples. In addition, applications developed by other researchers that use *CrimeStat* routines are presented at the end of each chapter.

Tutorial Figure 1

# Baltimore Vehicle Thefts: 1996 Location of Incidents and Standard Deviational Ellipse

