CrimeStat III

Part I: Program Overview
Chapter 1
Introduction to CrimeStat

CrimeStat® is a spatial statistics package that can analyze crime incident location data. Its purpose is to provide a variety of tools for the spatial analysis of crime incidents or other point locations. It is a stand-alone Windows® XP Professional® program that can interface with most desktop geographic information systems (GIS). It is designed to operate with large crime incident data sets collected by metropolitan police departments. However, it can be used for other types of applications involving point locations, such as the location of arrests, motor vehicle crashes, emergency medical service pickups, or facilities (e.g., police stations).

Uses of Spatial Statistics in Crime Analysis

Most GIS packages, such as MapInfo®, ArcView®, ArcGIS®, ARC/INFO®, Atlas®GIS™, and Maptitude®, have very sophisticated data base operations. They do not, however, have statistical methods other than means and standard deviations of variables. For most purposes, GIS can provide great utility for crime analysis, allowing the plotting of different incident locations and the ability to select subsets of the data (e.g., incidents by precinct, incidents by time of day). Most crime analysts visually inspect incident maps and, based on their experience, draw conclusions about shifts over time, ‘hot spots’ and other patterns suggested by the data.

There are times, however, when a more quantitative approach is needed. For example, an analyst wishing to examine patterns of streets robberies over time will need indices which document how the robberies may have shifted. For a neighborhood showing an apparent sudden increase in auto thefts, there needs to be a quantitative standard to define the ‘typical’ level of auto thefts. In assigning police cars to patrol particular major arteries, the center of minimum travel needs to be identified in order to maximize response time to calls for service. For research, as well, quantification is important. In examining correlates of burglaries, for example, a researcher needs to determine the exposure level, namely how many residences or commercial buildings exist in a community in order to establish a level of burglary risk. Or a precinct may want to target areas for which there is a high concentration of incidents occurring within a short time (‘hot spots’). While some of these analyses can be conducted with GIS queries, quantification can allow a more precise identification and the ability to compare different types of incidents. In short, there are many uses for quantitative analysis for which a statistical program becomes important.

The CrimeStat III Spatial Statistics Program

CrimeStat is a tool designed to provide statistical summaries and models of crime incident data. The tool kit provides crime analysts and researchers with a wide range of spatial statistical procedures that can be linked to a GIS. The procedures vary from the simple to some very sophisticated ‘cutting edge’ routines. The reasoning is that different audiences vary in their needs and requirements. The program should be of benefit to different organizations. For many crime analysts, simple descriptions of the spatial
distribution will be sufficient with the aim being practical intervention over a short time period. For these persons, many of the techniques provided in *CrimeStat* will be unnecessary.

For other analysts, statistical tools can supplement a much larger GIS effort, such as the Regional Crime Analysis System (RCAGIS) that was developed by the U.S. Department of Justice in cooperation with a number of police departments in the Baltimore-Washington metropolitan area (US DOJ, 2000). For other researchers, even more demanding techniques may be needed to detect the underlying spatial structure as a means for formulating a temporal-spatial theory. A pattern in and of itself has little meaning unless it is linked to some framework. The ability to quantify relationships with a large amount of data can address problems that previously were avoided and can be a first step in developing an explanatory framework or interventionist strategy. *CrimeStat* attempts to address both types of needs by providing statistics in a ‘toolbox’ framework. We recognize that today’s exotic statistical techniques may become tomorrow’s practical diagnostics and want the program to be useful for many years.

**Input and Output**

*CrimeStat* is a full-featured *Windows* XP Professional program using a graphical interface with database and expanded statistical functions. It can read files in various formats - *dBase* (III, IV, or V), which is a common file format in desktop GIS programs, *ArcView* Shape (shp) files, *MapInfo* data (dat) files, and files conforming to the ODBC standard, such as Excel, Lotus 1-2-3, Microsoft Access, and Paradox (Borland.Com, 1998; ESRI, 1998a; Microsoft, 1999). In addition, many other GIS packages, such as *Maptitude* can read ‘dbf’, ‘shp’, ‘bna’ or ‘mif’ files.

Output includes both displayed tables, which can be printed as text or copied to a word processing program, and graphical output. *CrimeStat* can write graphical objects to the *ArcView*, *ArcGis*, *MapInfo*, and *Atlas* GIS programs and can write interpolation files to these programs, to programs that read Ascii grid files (e.g., *Vertical Mapper*), and to the *Surfer* for *Windows* and *ArcView Spatial Analyst* programs (Golden Software, 1994; ESRI, 1998a; 1998b; 1998c; 1997; MapInfo, 1998).

**Statistical Routines**

*CrimeStat III* includes routines for:

- **Type of distance measurement**
  - Direct distance
  - Indirect distance
  - Network distance

- **Spatial distribution**
  - Mean center
  - Standard distance deviation
Standard deviational ellipse
Median center
Center of minimum distance
Directional mean and variance
Convex Hull
Moran’s I spatial autocorrelation index
Geary’s C spatial autocorrelation index
Moran Correlogram

**Distance analysis**
Nearest neighbor analysis
Ripley’s K statistic
Assign primary points to secondary points
Within primary file distance matrix
Between primary file and secondary file distance matrix
Between primary file and grid distance matrix
Between secondary file and grid distance matrix

**Hot spot analysis**
Mode
Fuzzy mode
Nearest neighbor hierarchical clustering
Risk-adjusted nearest neighbor hierarchical clustering
Spatial and temporal analysis of crime routine (STAC)
K-mean clustering
Anselin’s local Moran test

**Interpolation**
Single variable variable kernel density interpolation
Dual variable variable kernel density interpolation

**Space-time analysis**
Knox index
Mantel index
Correlated walk model

**Journey-to-Crime analysis**
Calibrate Journey-to-crime function
Journey-to-crime estimation
Draw crime trips

**Crime Travel Demand: Trip Generation**
Skewness diagnostics
Calibrate model
Make prediction
Balance predicted origins & destinations
Crime Travel Demand: Trip Distribution
Calculate observed origin-destination trips
Calibrate impedance function
Calibrate origin-destination model
Apply predicted origin-destination model
Compare observed and predicted origin-destination trip lengths

Crime Travel Demand: Mode Split
Calculate mode split

Crime Travel Demand: Network Assignment
Check for one-way streets
Create a transit network from primary file
Network assignment

Many of these routines allow variations yielding an even larger number of statistics to be calculated. Also, CrimeStat has Dynamic Data Exchange (DDE) capabilities so that it can be accessed from within another program.

CrimeStat is a program that specializes in the analysis of point locations. Over the years, many statistical tools have been developed for analyzing point locations. Many of these have either not been implemented as computer programs or were collected together as part of a specialized statistical system. They have been typically unavailable to crime analysts and the major statistical packages (e.g., SAS®, SPSS™, Systat®) do not include these routines. Consequently, we have collected those that are most appropriate for crime analysis and detection and organized them into a single package with a common graphical interface. They represent a wide variety of tools that can be used for crime analysis. CrimeStat can also analyze zonal data by treating them as 'pseudo' points. For example, the centroid of a census tract can be treated as a point and a value associated with the tract (e.g., its population) can be treated as an Intensity value (see chapter 3).

Program Requirements

Required Hardware and Operating System

CrimeStat III was developed for the Windows®XP Professional® operating system, though it will also work with the Windows®2000®, or Windows®NT® operating system; it is not hardware dependent so that any processor that can run Windows XP Professional/ 2000/ NT will suffice. Some of the routines can also run on the Windows®95® (Microsoft, 1995) or Windows® 98® (Microsoft, 1998c) operating systems. However, the program was not designed around nor fully tested for those operating systems. It is highly recommended that the program be run on a more current version of Windows.

While it can run on a relatively slow computer (e.g., 250 MHz clock speed) with limited RAM (e.g., 64 MB), it will run much better on a 1.6 GHz computer (or faster) with more than 256 MB of RAM. The faster the processor used, the quicker the program will
run. The more RAM the computer has, the quicker the program will run. The program is very intensive with respect to calculations. Some of the statistics produce very large matrices (e.g., the trip distribution routines in the Crime Travel Demand module). Depending on the size of the data files that will be processed, there may be hundreds of millions of calculations on any one run. It is critical, therefore, that the computer be fast and have sufficient amounts of RAM. The program was designed on an Windows® XP Professional® system with 1 GB of RAM running a single processor 1.6 GHz computer.

For most of the simple statistics, a reasonably fast computer will be adequate. However, several of the trip distribution routines will push the limits of most computer systems. The current 32 bit Windows operating system has a maximum limit of 4 Gb of RAM (actual and virtual). With a trip distribution matrix, there are M x N cells where M is the number of rows (origins) and N is the number of columns (destinations). With 8 bits being assigned to a number, practically a square matrix of about 10,000 x 10,000 would be close to the theoretical maximum allowed. Aside from taking a very long time to be calculated (days, if not weeks), the storage space required to save such a matrix will be very large. In short, the size of the files that can be processed will depend on the particular routines being run.

CrimeStat is a multi-threaded application written to take advantage of multiple processors if the hardware and operating system support multiple processors. The program is designed to be multi-threading which means that it will take advantage of multiple processors using Windows® XP Professional®, Windows® 2000®, or Windows® NT®. These operating systems support two processors. Windows 2003 Server® supports up to four processors. Thus, if there are two processors and Windows® XP Professional® or Windows® 2000® is the operating system, CrimeStat will calculate routines in about half the time. If there are four processors and Windows® 2003 Server® is the operating system, CrimeStat will calculate routines in about a quarter of the time. The multiples are not exact since processing time must be allocated for input of data and output of tables.

For small data sets, this feature is not important as most runs will be very quick. However, for large data sets (e.g., 3000 cases or larger), the speed of calculations become important. For example, on a 1.6 GHZ single-processor Pentium M® computer with 1 GB of RAM running Windows® XP Professional®, it takes about 4 minutes to complete a nearest neighbor analysis on 14,853 cases involving the calculating of distance from every point to every other point multiple times (for different neighbors). On a similar 1 GHZ dual-processor Pentium® computer with 1 GB of RAM running Windows® XP Professional®, it takes about 2 minutes to complete the same task. Slower systems will produce correspondingly slower times. The larger the file that is being processed, the more critical becomes the calculating efficiency of the computer.

If a police department is expecting to run large data sets, it would benefit them to purchase fast multiple-processor computers with lots of RAM and fast hard disks to speed calculating times. The evolution of new processors is moving in this direction anyway so that a multi-processor computer will become the norm in the next couple of years.

1.5
**Required Software**

*CrimeStat* needs a Windows environment to operate. The program was designed for a *Windows*® *XP Professional*/2000*/NT®* operating system so it is better optimized for that system. In particular, *Windows*® *XP Professional*/2000*/NT®* has two features that allows *CrimeStat* to run more efficiently. First, it is a multi-threading operating system and can utilize multiple processors, as mentioned above. Neither *Windows*® *XP Home*®, *Windows*® *95®* nor *Windows*® *98®* can utilize multiple processors. Second, it addresses memory in a more efficient way, as a large flat block. *Windows*® *95®* cannot handle cache memory above 64 MB. *Windows*® *98®* can handle RAM above 64 MB, but still has poorer memory management than *NT®*. Consequently, for the same machine, *CrimeStat* will run more efficiently (i.e., more quickly) in *XP Professional*/2000*/NT®* than in older or more limited operating systems.

*CrimeStat* is a stand-alone program. Hence, it does not require any other program other than a Windows operating system. However, to be maximally useful, there should be an accompanying GIS program. While point data can be obtained from a non-GIS system (e.g., census files include lat/lon coordinates for the centroid of census units), the use of the GIS to assign the coordinates is almost necessary. Further, many of the outputs of *CrimeStat* are for GIS programs. Thus, to view an ellipse or to view a three dimensional interpolation produced by *CrimeStat* will require an appropriate GIS package.

**Installing the Program**

*CrimeStat* comes compressed in a zipped file called *CrimeStat.zip*. To install the program, it is necessary to have a compression program that recognizes the ‘zip’ format:

1. Create a directory using *Windows Explorer* and copy the file to that directory.

2. Double click on the file name in *Explorer*. When the name *CrimeStat.zip* is visible in the dialog box name field, double click the name with the left mouse button. *CrimeStat* will be installed in that directory.

3. The program help menu can also access the manual. For this feature to work, however, it is important the chapters of the manual be kept in the same directory as the program.

**Adding an Item to the Start Menu**

To add *CrimeStat* to the start menu:

1. Click on the *Start* button in Windows followed by *Settings* then *Taskbar*. Click on *Start Menu Programs* followed by *Add*.

2. In the dialog box, click on *Browse*, point to the directory where *CrimeStat*
resides, and click on its name followed by Open. When the name CrimeStat is in the dialog box name field, click on the Next button.

3. Double-click on the folder to which CrimeStat is to be assigned.

4. Finally, type a name for CrimeStat (e.g., CrimeStat) followed by Finish.

**Adding an Icon to the Desktop**

To add CrimeStat to the desktop:

1. Double-click on My Computer.

2. Double-click on the drive in which CrimeStat resides followed by the directory that it is in (it may be several levels down).

3. Click once on the name CrimeStat with the left button and then hold down the right mouse button.

4. While holding the right mouse button, scroll to Create Shortcut.

5. The name Shortcut to CrimeStat will be placed at the end of the list of files.

6. Highlight the name by clicking on it once. Hold the left mouse button down and drag this name on to the desktop.

7. You can rename it CrimeStat by clicking on its icon with the right mouse button followed by Rename.

8. Alternatively, you can use Windows Explorer to create a shortcut and then drag the shortcut to the desktop.

**Installing the Sample Data Sets**

There are four sample data sets that can be used to run the program, also in 'zip' format. Since the data are simulated, they should not be used for real applications. They are provided to allow a user to become familiar with the program quickly. However, ultimately, the value of the program must be tested on real data, rather than simulated data.

1. **General Sample Data.zip.** The data are simulated incident points from Baltimore City and Baltimore County in Maryland.¹

   A. **Incident.dbf** - A simulated data set of 1061 incidents (e.g., robberies) in Baltimore County and the City of Baltimore
2. **Jtc Sample Data.zip.** There are three files of simulated data for use with the Journey-to-crime routine (chapter 8):

A. **JtcTest1.dbf** - A simulated data set of 2000 robberies in Baltimore County that can be used for calibrating a travel demand function. Each record has a crime location and a residence location of the offender.

B. **JtcTest2.dbf** - A simulated data set of 2500 burglaries in Baltimore County that can be used for calibrating a travel demand function. Each record has a crime location and a residence location of the offender.

C. **Serial1.dbf** - A simulated data set of the location of seven incidents committed by a single serial offender. To become familiar with the journey to crime routine, they can be treated as either robberies or burglaries.

3. **CWA Sample Data.zip.** These are three files of simulated data for use with the Correlated Walk Analysis routine (chapter 9):

A. **PredictableOffender1.dbf** - A simulated data set for an algorithmic offender who committed 13 incidents.

B. **PredictableOffender2.dbf** - A simulated data set for an algorithmic offender who committed 12 incidents.

C. **RealOffender1.dbf** - A data set for a real offender who committed 12 incidents - 10 larceny thefts, 1 robbery and 1 burglary.

D. **RealOffender2.dbf** - A data set for a real offender who committed 15 incidents - 10 larceny thefts, 2 assaults, 2 burglaries and 1 robbery.

4. **Crime Travel Demand Sample Data.zip.** There are 13 files of data, parameter files, and a spreadsheet file for modeling travel behavior in Baltimore County, Md. They are examples used in the crime travel demand module (chapter 11-17):

A. **Crime Travel Demand read me.pdf** - A file that explains the three data sets and their fields and describes the eight parameter files.

B. **BCOrigins.dbf** - A data set on 532 origin zones in both Baltimore County and the City of Baltimore from the late 1990s. There are data
C.  *BCDestinations.dbf* - a data set of 325 destination zones in Baltimore County only. There are data on crimes occurring in each zone and demographic, economic and land use variables associated with those zones.

D.  *ObservedODTrips.dbf* - the actual trip distribution indicating the number of trips from each origin zone to each destination zone.

E.  *Trip generation origin model.param* - Runs trip generation model using the Poisson regression for the origin zones affecting Baltimore County.

F.  *Trip generation destination model.param* – Runs trip generation model for Baltimore County destinations.

G.  *Make predicted origins.param* – Applies modeled coefficients for the origin model to the same data set from which it was modeled. Then the routine adds in external trips.

H.  *Balance Origins and Destinations.param* – Balances the number of trips by origin and by destination. In the example, the number of predicted destinations are held constant.

I.  *Calibrate Origin-Destination Model Coefficients.param* – Using the predicted origins and predicted destinations from the trip generation stage, estimates coefficients for distributing trips from origin zones to destination zones.

J.  *Apply Origin-Destination Model.param* – Inputs the predicted origins and predicted destinations from the trip generation stage as well as the modeled coefficients from H above. Outputs predicted trips for each origin-destination zone combination. For the graphic display, outputs top 200 trips.

K.  *CompareObserved and Predicted Trip Lengths.param* – Inputs observed (actual) and predicted trip distribution and compares them by trip lengths. Calculates coincidence ratio and then compares the top 200 origin-destination links.

L.  *Mode Split Model.param* – Inputs predicted origins, predicted destinations, and predicted trips along with estimates of the mode split function (see Excel spreadsheet below). Splits trips by origin-destination pair into specific travel modes. The output is both a table
of origin-destination trips by mode as well as five ArcView shape files representing zone-to-zone trips by mode.

M. Mode split impedance defaults.xls - An Excel spreadsheet for estimating the coefficients of the mode split stage. This should be used in establishing the parameters for the mode split routine.

To install any of these sample data files, it is necessary to have a compression program that recognizes the ‘zip’ format:

1. Create a data directory using Windows Explorer and copy the files to that directory.
2. In Windows Explorer, double-click on its name and then follow the instructions.

Step-by-Step Instructions

This manual will go through the program step-by-step to address how it can be used by a crime mapping/analysis unit within a police department. Chapter 2 provides a quick guide for all the data definition and program routines and chapter 3 provides detailed instructions on setting up data to run with CrimeStat. The statistical routines are described in parts II, III, and IV. Part II presents a number of statistics for spatial description, part III presents a number of statistics for spatial modeling, while part IV presents a crime travel demand module. The different statistics are presented and detailed examples of each technique are shown.

Options

There is an option tab that allows the saving and loading of program parameters and the setting of colors for each of main headings: Data setup, Spatial description, and Spatial modeling. One can also output simulated data during the simulation runs; this will be explained in the appropriate section.

Short Applications

The manual also includes a number of applications conducted by other researchers and analysts. These are presented as one page sidebars in the various chapters. Most of these are from criminal justice. But, applications from other fields have also been included. The aim is to show the diversity of applications that researchers and analysts have used with the various routines in CrimeStat.
On-line Help

In addition, there is on-line help for the program. There is a Help button that can be pushed to access all the help items. In addition, the program has context-sensitive help. On any page or routine, typing *F1* will pop up an appropriate help item. The on-line help can also access the program manual. For this to be available, be sure to store the chapters of the manual in the *same directory* as the program.
Endnotes for Chapter 1

1. The data were simulated by a random number generator following the distribution of several types of crime incidents. Because the data were selected by a random generator, the points do not necessarily fall on streets or even stay within the boundaries of Baltimore City and Baltimore County; some even fall into the Chesapeake Bay! Their purpose is to provide a simple data set so users can become familiar with the program.