



Computer Tools Used for Data Analysis

By Mark Newsom

January 2004

Center for Workers with Disabilities
American Public Human Services Association
810 First Street, NE Suite 500
Washington, DC 20002
Phone: 202.682.0100

Introduction

Collecting data via surveys, administrative records or other methods is just the first step in any statistical evaluation of a program or project. Once you have the data you have to do something with it. The data need to be analyzed to produce statistics that will help you demonstrate efficacy and effectiveness. There are a number of different computer tools for effective analysis and the following brief reviews some of the more common programs available to you.

SPSS®

The major players in data analytic software are SPSS and SAS. These two computer programs have existed for many years, and have tens of thousands of prominent users in government, corporate America and academe. SPSS uses a base spreadsheet format and has easy to use Windows based drop down menus for commands. The advanced user can do programming with the syntax command function if they wish. SPSS can read data files from common formats such as Excel, Access, SAS and dBase. The base package allows you to perform basic descriptive (e.g. frequencies, sums, mean, etc.) and univariate (e.g. Pearson correlation, one-way ANOVA, t-test, etc.) statistics. Add-on packages are available for advanced statistics such as multivariate regression modeling. Special prices are available for both state and federal government employees. Training is available via SPSS directly, and most colleges and universities have courses or training programs available. For more information see www.spss.com.

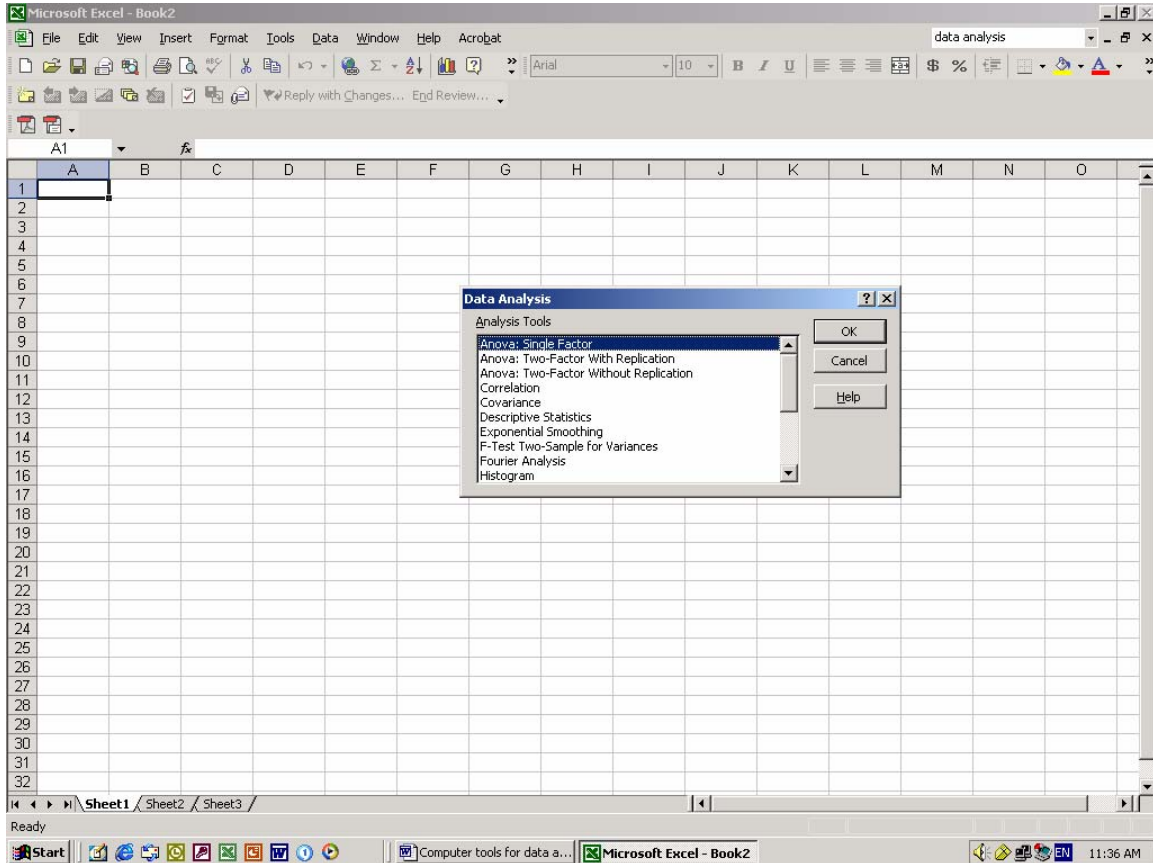
SAS®

SAS has been the software of choice for most hardcore data analysts in the area of health policy and health services research. This is because of its traditionally unique (SPSS now has the ability) ability to deal with large data sets from which population estimates are derived (i.e. the Census, Current Population Survey, National Health Interview Survey, etc.). Most SAS users perform analyses by writing up program statements, which require rather advanced computer and statistics skills. However, SAS has added a Windows based menu drive command function to make the program easier to use. SAS has a highly developed training program for use of its software products. Many universities also provided courses on SAS. For more information see www.sas.com/

Using Excel for Data Analysis

Although Excel is typically thought of as only a spreadsheet program, it can also be used to conduct most of the basic statistics. On the **Tools** menu, simply click **Data Analysis** to get a view of the possible statistics. It should look like Figure 1. If **Data Analysis** is not available, load the Analysis ToolPak. You can do this by going to the **Tools** menu and then clicking **Add-Ins**. In the **Add-Ins** available list, select the **Analysis ToolPak** box, and then click OK. The data analysis functions in Excel are a little bit clumsy, but if you are already comfortable with Excel you may find it easier to use than learning a whole new system like SPSS or SAS. The other advantage is that most people have Excel already so you can save the cost of purchasing new software and can easily share data files with professional colleagues. For more information about Excel's functions see: <http://office.microsoft.com/home/office.aspx?assetid=FX01085800>

Figure 1. Microsoft Excel Data Analysis Function



EpiInfo™

EpiInfo is a series of programs for Microsoft Windows for use in managing databases and conducting statistics analyses. With EpiInfo you can easily develop a questionnaire or form, customize the data entry process, and enter and analyze data. EpiInfo is written in Visual Basic and uses the Microsoft Access file format as a gateway to industry database standards. Although EpiInfo data is stored in Microsoft Access files for maximum compatibility with other systems, many other file types (e.g. Excel, dBase, etc.) can be analyzed, imported, or exported.

EpiInfo can do most basic descriptive statistics and some advanced statistics. It can analyze data from complex sampling surveys such as the National Health Interview Survey or the Behavioral Risk Factor Surveillance System. EpiInfo is also a geographic information systems (GIS) tool. A component called EpiMap displays geographic maps with data from EpiInfo. EpiInfo is in the public domain and can be downloaded **free** from the Internet. For more information or to download see: www.cdc.gov/epiinfo/

What about Microsoft Access?

Microsoft Access is primarily a data warehousing or storage tool. In other words, you use Access to create databases. Then the actual data are analyzed with other tools.