

Chapter 1

Basic Commands

Aim: Upon completing the chapter, the learner should be able to understand the general form of the basic commands of Stata for reading and saving databases.

1.1 Introduction

Stata is a computer program designed to perform various statistical procedures. Among the basic statistical procedures that can be performed are the following: calculation of summary measures, construction of graphs, and frequency distribution using contingency tables. Furthermore, using Stata, you can perform parameter estimation in generalized linear models and survival analysis models using uncorrelated and correlated data. The program also has the ability to perform arithmetic operations on matrices. Its ability to export and import databases in the Excel format gives Stata great versatility. This program is regularly used in biostatistics courses in public health schools in different countries. It is also often cited as one of the main programs used for statistical analysis in scientific publications related to public health research.

This chapter will provide an introduction to the Stata program, version 14.0. We assume that readers of this book have a basic knowledge of both biostatistics and epidemiology.

1.2 Entering Stata

After selecting the Stata icon on your computer, the program responds with five windows (Figure 1.1), which have the following utilities:

1. *Command*: In this window the user can write or enter “commands” or instructions to perform various operations with an active database. Not all commands can be executed in this area; there is also a taskbar with executable commands.
2. *Results*: This window shows the results obtained after the execution of the commands introduced or requested via the taskbar.
3. *Variables*: In this window the variables of an active database are displayed. If this window is blank, that is an indication that there is no active database.
4. *Review*: This window lists all the commands used during the current open session of the program and allows them to be repeated without rewriting them in the command area.
5. *Properties*: This window displays the properties of the user’s variables and dataset.

1.3 Taskbar

The taskbar provides common access to all windows-based program commands, such as *File*, *Edit*, *Data*, *Graphics*, and *Statistics*; these options can be found at the upper part of the main window. The most frequently used icon is the *Data Editor* icon, with which

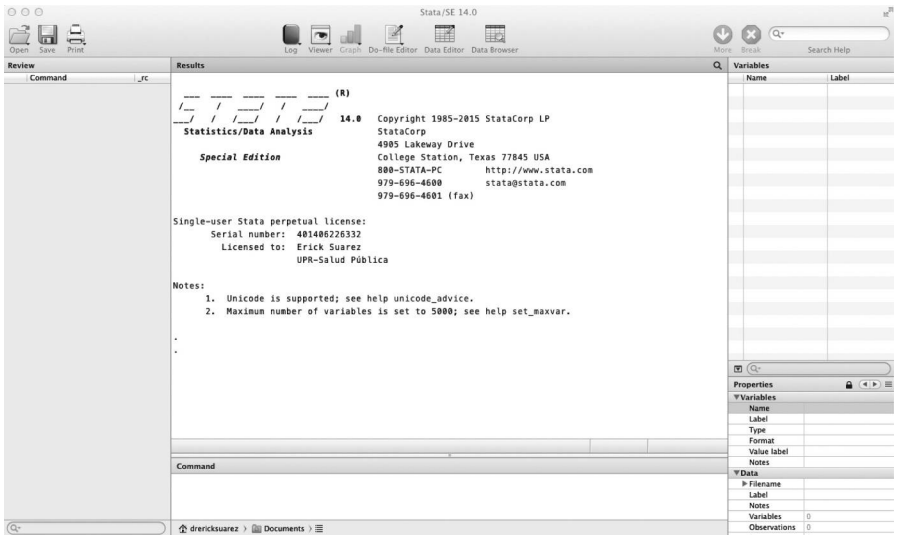


Figure 1.1 Main Stata 14 window.

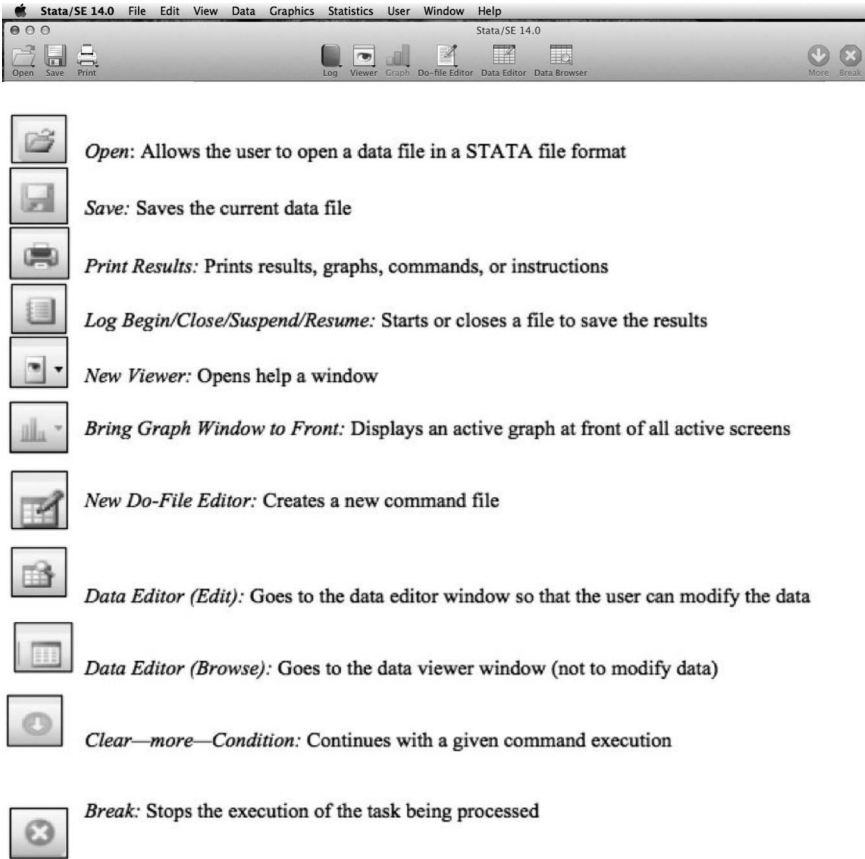


Figure 1.2 Taskbar and icons.

it is possible to enter values and identify the variables in a given project. The *Graphics* button provides access to the window used to generate different types of graphs. The *Statistics* option allows the user to perform statistical mathematical operations through the execution of the commands. Below the taskbar are icons that allow the user to open, save, and print, along with icons that facilitate the observation of graphics (Figure 1.2).

1.4 Help

One of the most useful attributes of Stata is its support system, which allows the user to find the commands and their ways of execution, according to that user's specific needs. The help menu can be accessed by clicking on the "New Viewer" icon on the toolbar or by typing either *help* or the letter *h* in the command area and following that with a keyword that represents the topic about which the user requires more information (see Figure 1.3).



Advice on finding help

Let's say you are trying to find out how to do something. With over 2,000 help files and 12,000 pages of PDF documentation, we have probably explained how to do whatever you want. The documentation is filled with worked examples that you can run on supplied datasets. Whatever your question, try the following first:

1. Select **Help** from the Stata menu and click on **Search...**
2. Type some keywords about the topic that interests you, say, "logistic regression".
3. Look through the resulting list of available resources, including help files, FAQs, Stata Journal articles, and other resources.
4. Select the resource whose description looks most helpful. Usually, this description will be a help file and will include "(help <xyz>)", or, as in our example, perhaps "(help logistic)". Click on the blue

Figure 1.3 Help window.

For example, if we want to learn how to perform an analysis of variance (ANOVA), we can use one of the following commands:

```
help anova
```

or

```
h anova
```

Upon entering those commands, a specific window for ANOVA will appear (see [Figure 1.4](#)).

1.5 Stata Working Directories

When working with Stata, files and results can be saved to a specific directory, which is defined during the installation instructions. For example, to view the working directory for a project, enter the command *pwd* (path of the current *working directory*), and the following results will be displayed:

```
. pwd
/Users/Documents/students
```

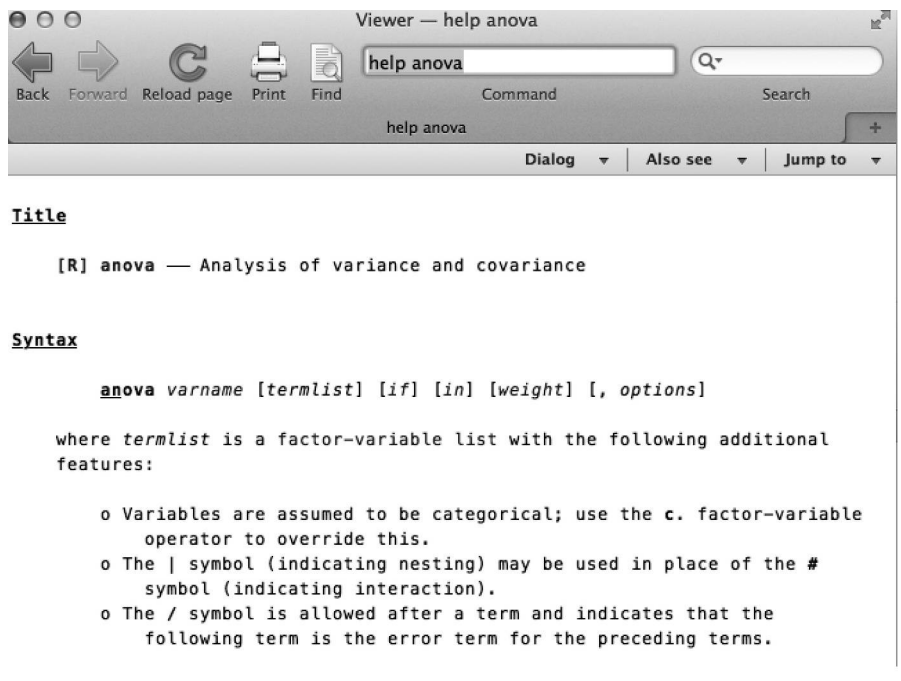


Figure 1.4 Help ANOVA.

It is important to keep the working files in a directory that is different from the default directory that Stata assigns, because during the regular program updates files located in the default directory may be removed.

To create a particular file, the *mkdir* and *cd* commands must be used to navigate to that directory again. The sequence of commands to create a directory is as follows:

cd C:\	Navigate to the main directory of your hard drive or to the location where you wish to create your home directory
mkdir new_folder	Create a new working directory
cd new_folder	Navigate to the new working directory

To use Stata in the new working directory, you need to restart the program and immediately move to the desired directory. For example, assuming that the name of the working directory is “students” and assuming, as well, that this

directory is located in your computer's *Documents* folder, the following will take you to that folder:

```
cd "/Users/Documents/students"
```

1.6 Reading a Data File

After creating the working directory in which, outside the Stata program, we have previously copied a data file (i.e., the file named "Cancer.dta"), we proceed to open the file. This can be done in two different ways: using the command area or using the icon on the toolbar. For the former, we would write the following command sequence:

cd "C:\new_folder" dir use cancer	Command to navigate to the new working directory Command to browse the contents of the folder The use command indicates the name of the file that will be used
---	---

For the latter, on the other hand, it is necessary to click , the *Open* icon, and browse the folder that contains the working file. The *describe* command can be used to view the information contained in the data file, which might include the number of observations, variables, and file size, among others, as shown below (assuming that the active database being used contains the anthropometric measurements of 10 subjects):

describe

Output

```
. describe
```

Contains data

```
obs:      10
vars:      5
size:     200
```

variable name	storage type	display format	value label	variable label
var1	float	%9.0g		
var2	float	%9.0g		
var3	float	%9.0g		
var4	float	%9.0g		
var5	float	%9.0g		

1.7 *insheet* Procedure

Another way to read a database in Stata is to import existing databases created in other formats. Delimited text files using the .txt (can be opened by most text editors), .raw (a raw image file), and .csv (an MS Excel file) extensions can be imported into Stata. The most commonly used is .csv, which, as indicated, is created using MS Excel. In Excel, you must save the data file using the .csv file extension instead of the .xls extension. When you have the data saved with .csv, you can then proceed to use the *insheet* command in Stata:

```
insheet using "c:/data.csv", replace
```

The *replace* option that has been placed after the comma (above) is used to clear the program if another database was being used. Stata does not open a database if there is another one that is already open. The *clear* command can also be used in Stata to remove a database, therefore clearing the way to use a new one.

1.8 Types of Files

Below is a list of the different types of archives that can be created in Stata; the left-hand column contains the file extensions that correspond to each archive.

.dta	data files
.do	command files
.ado	programs
.hlp	help files
.gph	graphs
.dct	dictionary files

1.9 Data Editor

In the Data Editor window, you can input data for the creation and identification of the study variables. One advantage of Stata's data editor is its ability to import databases built in Excel. This is done by the simple operation of selecting the entire database in Excel and copying and pasting it into the Stata data editor.

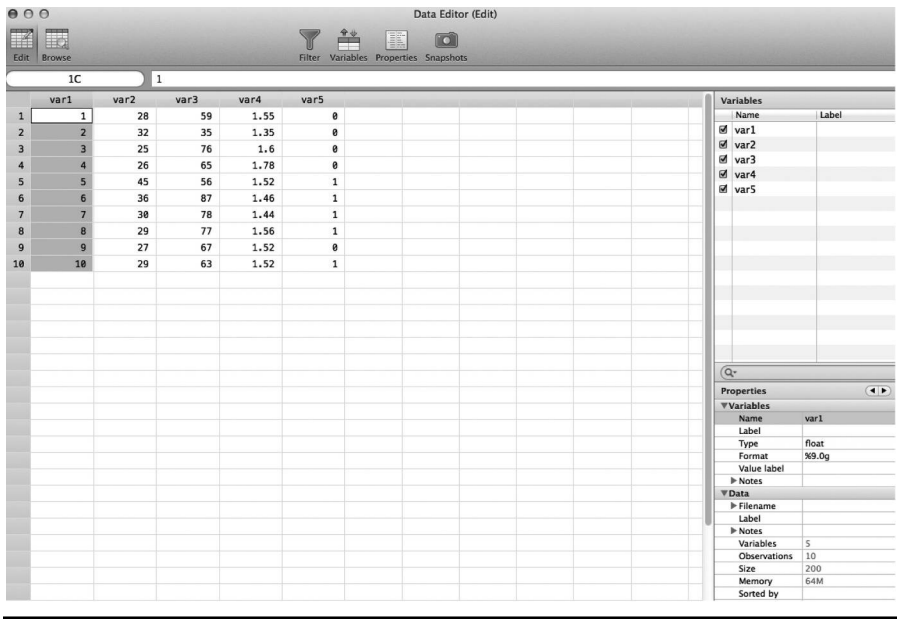


Figure 1.5 Data Editor window.

To access the Data Editor window (Figure 1.5), click the “Edit” icon, , on the taskbar located in the main window.

At the beginning of the data entry process, the program automatically assigns a name to the column that defines each variable (var1, var2, ..., vark). This name can be changed in the Variables Manager window after clicking the *Data Editor* icon, using the box “Name” (Figure 1.6). To return to the main window of Stata, you close or minimize the Data Editor window.

Constructing a user-friendly database requires that each variable be named in such a way as to be easy to identify. This can be done using the “Label” box in the properties window. When building a database, it is possible for the values assigned to the variables to be represented by codes. The coding of the variables can be done using the “Value Label” option. With this option you can assign numerical values to alpha-numeric variables, thereby allowing better management of the database. This coding can be done in the Variables Manager window. The steps to do this are as follows:

1. Click “Manage” in the *Variables Manager* window, and a new window appears (Figure 1.7). Then click “Create Label” to assign each code a label.
2. After creating the value labels, return to the *Variables Manager* window, in which you will be able to assign labels to each variable in the “Label” box (if they were not assigned previously in the *Properties* window) (Figure 1.8).

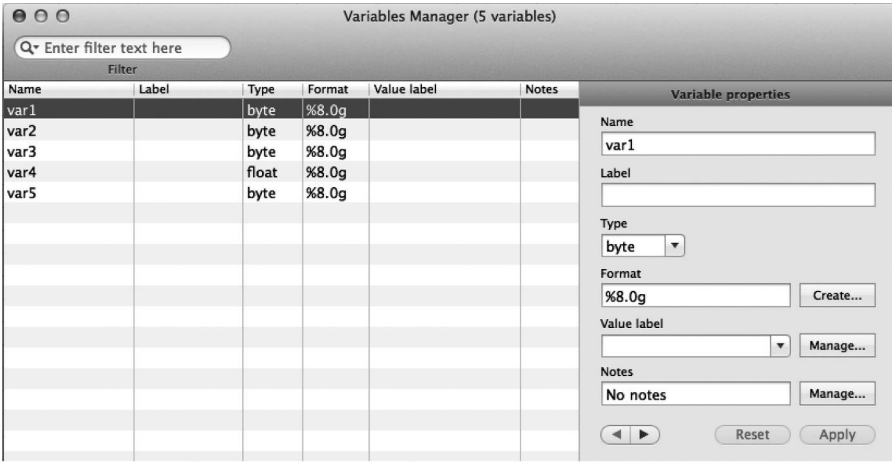


Figure 1.6 Variable name change.

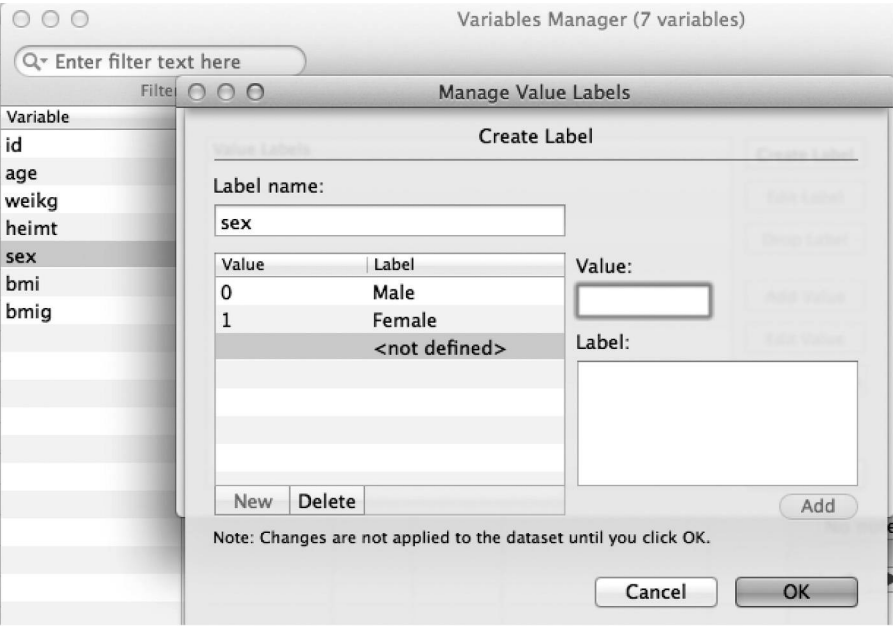


Figure 1.7 Assigning value label.

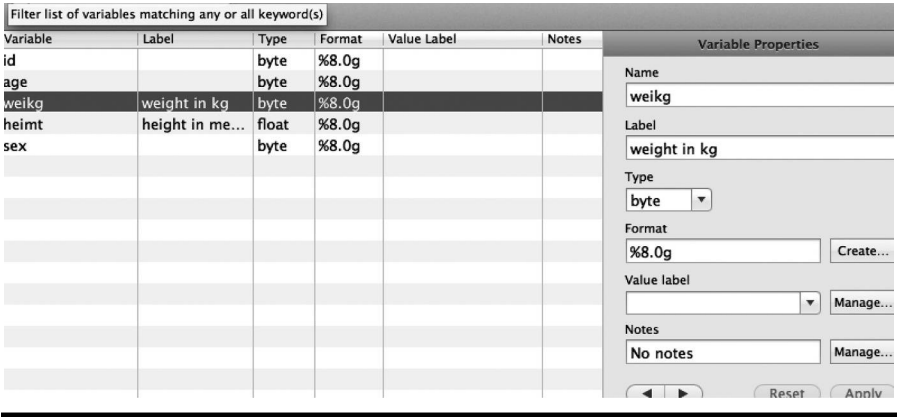


Figure 1.8 Assigning labels to each variable.

To continue working in Stata after having created a database, the user needs to ensure that the data have been saved. To that end, the user will need to assign a name to the file to continue working on the database. Clicking on “File” (on the toolbar) followed by “Save As” (on the subsequent dropdown menu) begins this process. After that, select the working folder or directory and assign a name to the database. The default file extension is .dta.