

# Introduction to Stata

ARE C253/ PP253 Fall 2003

Copyright ©2003 James Manley All Rights Reserved.

version 3: 9/8/03

(now describing some differences between Stata 7 & Stata 8, including the **graph** command)

This handout is designed to give you an idea of how things work in Stata: it was NOT written to do any particular assignment. The sample commands you type through the course of this tutorial were not chosen because the particular operation they perform is useful: once we get started manipulating the data, the sample commands mostly just generate a bunch of clutter that you won't want to save, so don't bother saving the modified dataset unless you have some specific reason for doing so.

Before we get started, though, what *is* Stata? Stata is a software package designed to manipulate and analyze data. (Some similar programs are SAS, SPSS, and Shazam. Excel also has a bit of statistical analysis functionality, but its emphasis is different.)

## ORGANIZATION

Let's set up some workspace. I recommend that you create three directories in an easily accessible place. For myself, working with Windows, I opened My Computer and then opened my C: directory. There I created a directory called PP253. Within that I created three directories: data, do, and logs.

(You can do this wherever you want, or you can follow my example. It will be easier if you avoid directories like "My Documents" that have spaces in them, since Stata's not smart enough to deal with the more recent innovations in file naming. The three directories data, do, and logs are useful organizationally, but not strictly necessary.)

## UNIX

If you're using the Evans 616 lab, you will need to know a few commands to manipulate that environment. The next few commands are to be typed into the command windows that pop up right after you log in. These windows are called "X-term" windows, and you can make another one whenever you need it by right clicking (and holding) on the bluish background area, selecting programs, and then selecting X-term. To close an X-term window (or any other window) you should right click on the gray bar at the top of the window, again hold, and click Quit.

You can use the command **mkdir** to create directories. Example: **mkdir logs** creates a directory called logs. To change into another directory, use **cd**, so for example to go into the newly created **logs** directory type **cd logs**. When you're logged into these machines, the data is available at */class/data/a253/[yourlogin]/data* where *[yourlogin]* looks something like a253ai. This data will not be something you can modify, so if you want to save a modified dataset, give it a different name and save it in your own home directory. \*\*\*You only have 15 megabytes of storage space not including the original dataset, which is in shared space. Saving a modified dataset is almost guaranteed to take at least 5 megabytes, so be careful and don't try to save too many versions of the data! You really shouldn't need to save any, or one at the most, if your program does everything it's supposed to.

I don't think you should need many other Unix commands, but in case you do, the system has help files - at the prompt, type **help- -l** to get the list of help filenames, and then type **help- [file]** to read it. There is also a Unix tutorial at [elsa/eml/emltutorials.html](http://elsa/eml/emltutorials.html) if you feel the need. To run Stata in a Unix environment, type **xstata**.

To access the machines in Evans, you have two alternatives. One option is to go to there yourself and type directly into the machine of your choice. You can do this anytime: the lab is open 24/7. (Remember, the code to get in the door is 1-3, 5, 2.) The one catch is that the outside doors to the building aren't open all the time: unless you have a friend with a key to Evans, you'll only be able to get in until 7:30 pm on weekdays. There are some weekend hours, but I'm not sure of the schedule. Don't count on getting in on a Sunday.

Fortunately, the other way to get into the machines is always accessible. Any computer with internet access, Exceed, and SSH [Secure Shell] can remotely log in to these machines. (These programs are available on the Connected@Berkeley CD available through the Scholar's Workstation on campus. I think it's even legal for me to encourage you to steal the software from friends, since I think the University has a site license.) If you're having trouble getting **xstata** to run, go into "Edit Profiles" on SSH and make sure the box is checked marked "X11 Tunneling" for the profile you're using to log on.

To log on, use the machine name *emilyX.berkeley.edu* where X is a number between 1 and 15. (for example, use *emily6.berkeley.edu*) You should have a login name and password. What's going on here is that you're actually connecting to a machine in Evans remotely, even though someone else may be working on that machine: the computers can handle many people simultaneously. Because of this, when you're working directly on the Evans computers, NEVER POWER OFF your machine! If you do, you could destroy someone else's work. Anyway, back to Stata.

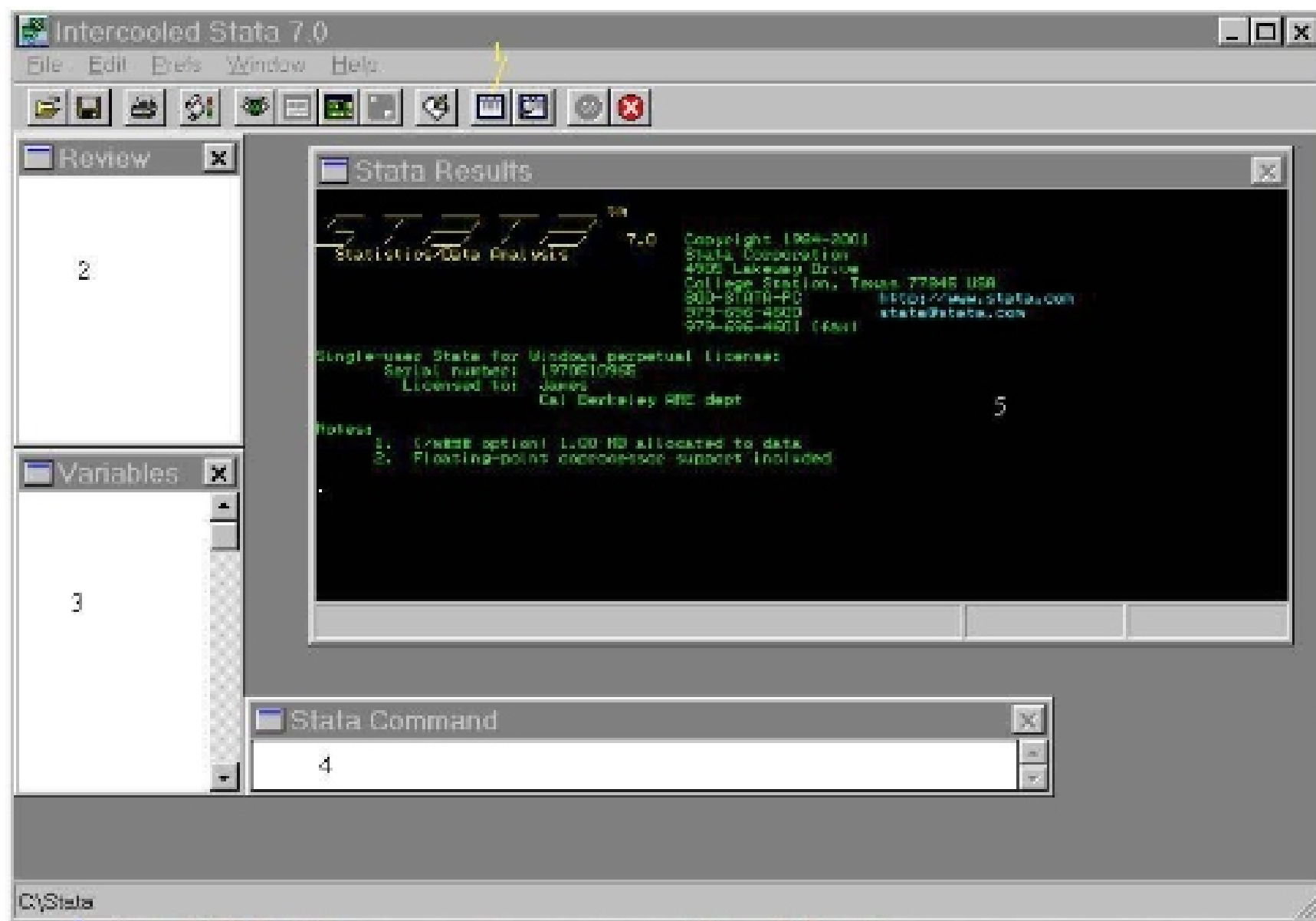
## **Part I: Checking out the data**

### **LOADING DATA**

You can't do much without any data, so let's load your dataset. In order to do this on your home computer, you need to have saved the data from the internet into a directory you have access to [unless you're on a Unix machine and already have direct access]. The data can be found on the class website (<http://are.berkeley.edu/courses/ARE253/2003/index.html>). The file is called **hhpov6**; put it in your new "data" directory.

If you're using (or logged into) the Evans 616 lab, the data's accessible to your account already. Click on the "open" folder, then make sure the directory says */class/a253/[yourlogin]*. Double-click on *data* and then open the file by double-clicking on it. Once you have saved the data in a convenient place, we can take Stata for a quick spin to see how things work. Open Stata by clicking on the icon in a Windows environment or by typing **xstata** in a Unix environment.

When you start the program, the screen will look something like this (without the numbers 1-5, of course) [Also, screen shots may differ from what you see in areas like the relative sizes of the windows, etc. Don't worry about that!]



A few familiar icons are in the upper left for opening or saving files. Here they're for opening or saving datasets rather than programs. The other icon I use a lot is labeled "1". It displays the dataset. (Advanced users may also make use of the last icon, which is a "break" key.)

The four windows are the heart of the interface. In the upper left hand corner labeled "2" is the Review window, which will list all the commands you type in or otherwise execute, even the mistakes. The Variables window, number "3", lists the variables in the dataset. It lists them in the order they appear in the dataset. Window "4" is the command window, which is where you can tell Stata what to do. Window "5" is what Stata uses to respond to you: it's the output window.

To open a data file on a Windows computer, we just click on the "opening" file folder in the upper left corner of the screen and move to the directory where you saved the data file. (Alternatively just type `use c:/pp253/data/hhpov6` in the Stata Command box, #4 in the above diagram). Here's what happened when I tried that: