1   Course Resources

During this semester, we will cover a wide range of data management and basic analysis skils using Stata. This guide is intended to get you to start exploring Stata (hopefully *before* the first class…if you have access to computer with Stata) so that you are familiar with the basic Stata environment and so that you can spend time "updating" a couple of Stata components before class, rather than during class. I have included many useful (and free) Stata and statistics resources to help you navigate the pitfalls of exploring a new software. It also provides some brief comments about Stata versus other software packages as well as some prepatory activities using Stata.

If you have not yet installed Stata or are having trouble installing it, you can refer to the FAQs on Stata's Technical Support pages for help (http://www.stata.com/support/) (or I can help you install the program if you are still having issues). Stata has a lot of great resources that are constantly updated by Stata programmers and the user community.

1.1   Internet Resources for Stata

UCLA Stata Help Pages: This provides step-by-step instructions on how to run and interpret most Stata commands.

http://www.ats.ucla.edu/stat/stata/

Statalist: This is a community of Stata users and programmers who post questions and answers to variety of Stata issues. If you run into an issue with Stata, chances are that someone else has had a similar issue and posted about it.

http://www.stata.com/statalist/

Stata Journal: There is an entire journal devoted to Stata:

Stata Journal: http://www.stata.com/bookstore/sjdetails.html

Stata Meetings: http://www.stata.com/meeting/

Graph examples: This webpage is an invaluable resources on types of graphs that Stata can produce and if you click on the type of graph you want, it will direct you to the Stata code to produce it.

http://www.ats.ucla.edu/stat/stata/library/GraphExamples/default.htm

1.2   Resources Included in the Stata Software

Now we need to open up the Stata program. Stata contains probably the most comprehensive help system included in a statistics software. Click on the Stata icon and you should see windows that look similar to this:[1]

---

[1] A note about the Operating System and Stata Version: I use Mac Os X 10.6 and Stata 11 MP, so the screenshots in lectures and handouts will reflect this. However, since Stata 7, the Mac, Windows, and Linux versions of Stata look virtually the same and all the commands are exactly the same (except when you include the file path to your data).
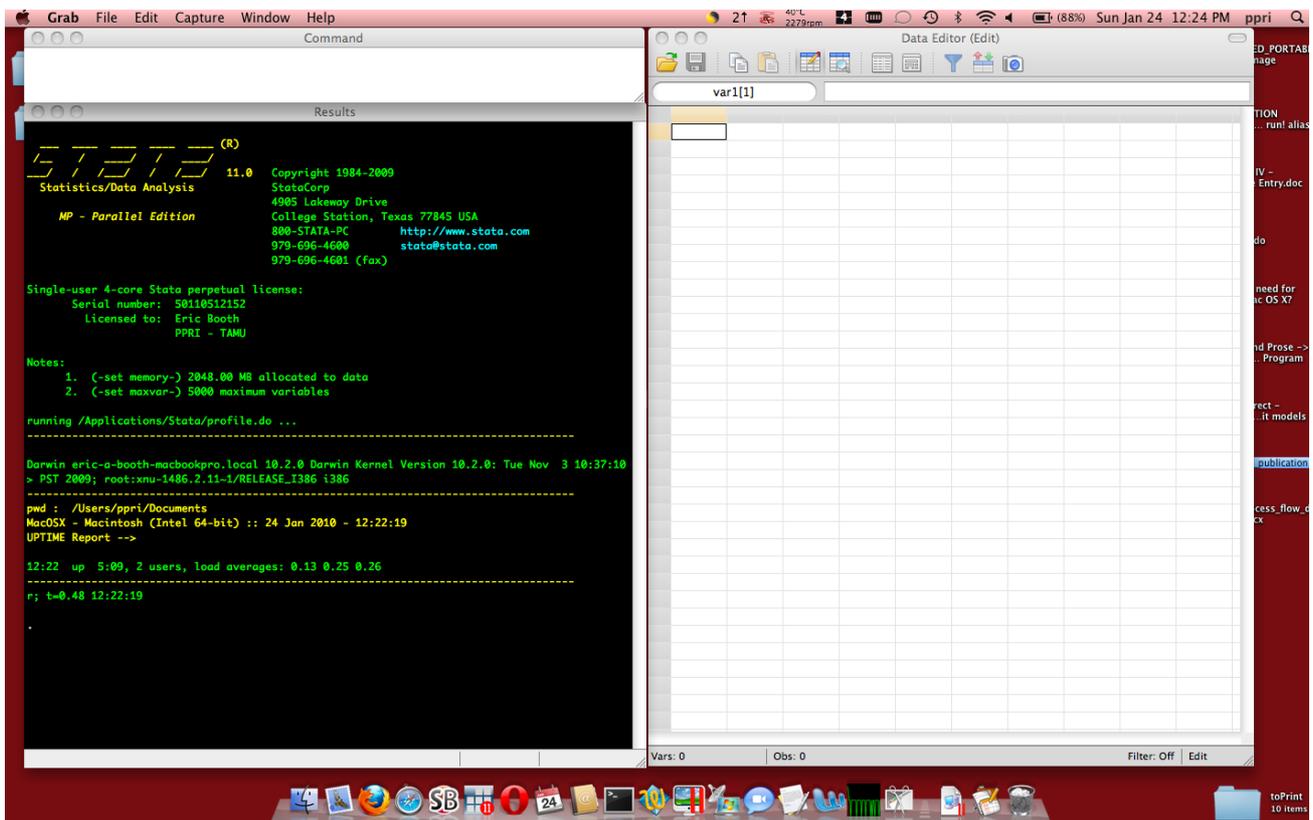
There are 4 main Stata windows. Clockwise: The **Command** window is where you can input commands to run on the dataset in memory (in "interactive mode"); the **Results** window displays the results of your analysis, and the Data Editor is where the data are displayed & edited (similar to Excel interface). (NOT SHOWN: There is also a **Variables** window shows the name and description of all of the variables for the dataset in memory; a **Viewer** window that shows help files, and finally the **Review** window allows you to see, copy, or rerun all the past commands you have run during this session. There are other windows that show the Results of a graphical command, and allow you input and run a set of commands all at once (in "batch mode")…We will cover these in class).

For now, you can use the **Command** window to access the help resources (there is also a "Help" menu in the menu bar at the top of the screen; however, the help commands combine all of the search results into one window and are therefore faster and easier to use). All the help documentation is available from the help menu and on the Stata installation disk in .pdf format.

There are 3 different versions of a help command and which one you use depends on the type of results you want to see.

Scenario 1: If you want to search for help on how to structure the syntax of specific command then simply type "help *command*" (no parantheses) where *command* is the command you are looking for. For instance, if you want to know how to write the syntax for running ***ANOVA***, type "help anova" and you will see the following window:
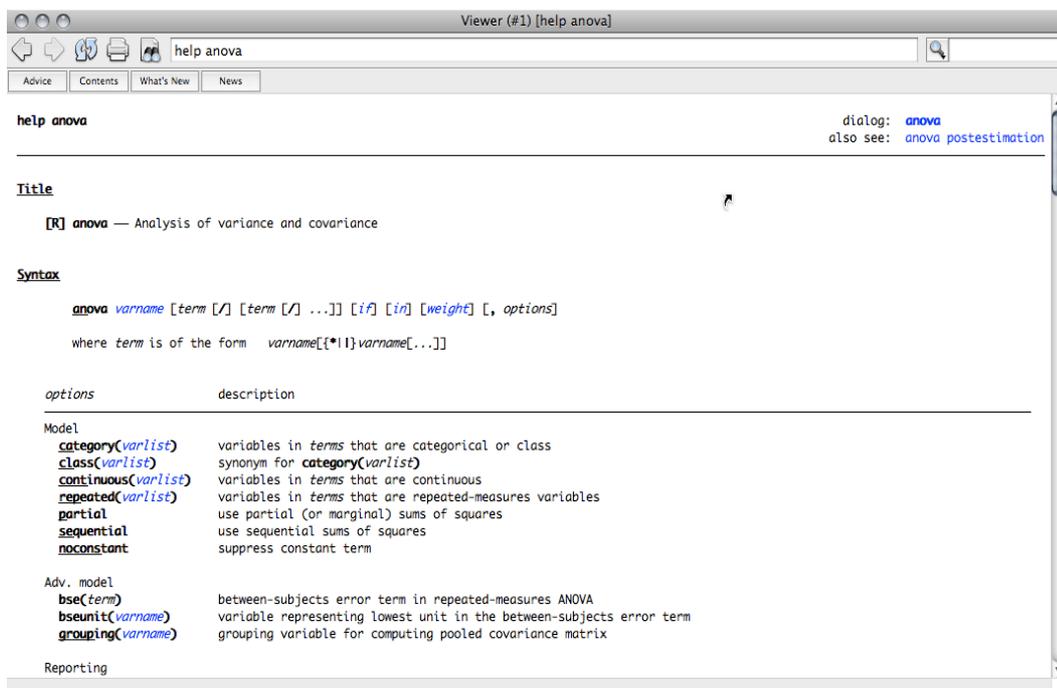
The help results window will always give you the following: 1) the command synax; 2) all options for running the command; 3) some examples; 4) and usually a couple of web resources for examples on how to run the command (generated by Stata users and programmers). There is help documentation for every Stata command and package.

Scenario 2: If you are not sure about the name of the command you want to run, you can search more broadly for a certain topic by using the "findit" command. Simply type "findit *command*". The results window this time will present a list of commands, organized by search topic, and then a list of web packages and web resources on the topic. **For example**, if you want to summarize the data in memory, you can search "findit summarize" or "findit summarize data", etc. It's search logic is Boolean and so the search functions similar to a web search engine.

Scenario 3: Finally, if you are looking for a subcommand, option, or descriptive word that would fall inside the description of a command, then you should use the "search *command*" option. Search looks at the words inside each help document for your search string. For example, if you are looking for commands that return a f-test statistic in the output, knowing that there is no f-test command, you could input "search f-test" or "search f-test statistic". If you are looking for commands that include the R^2 try typing "search r squared" or "search r-squared" or even "search r2"

2 Do File (or "batch file") Editors

Before the first class, I recommend you explore you options for downloading a good text editor. Text editors are used to run Stata in "batch mode" where you can run an entire set of commands all at once. Mac, Window, and Linux/Unix OS's already have basic text editors installed on the system (textedit, notepad, and abiword), but there are text editors out there with far better functionality. First, you want to use a text editor instead of a word

processor because word processors insert a lot of hidden formatting language and markers that Stata cannot interpret. Second, Stata has a built in text editor (called "Do File editor"), but it has some limitations in terms of character limits and user friendliness. Finally, freeware text editors are better at handling regular expressions and syntax (e.g. they often use visual cues for the commands you are using and highlight errors in the command before you run it, etc) and cross-platform commands (note: as of version 11, there is syntax highlighting in the Windows version of Stata).

The website below gives you a great overview of free text editors that are freely available. Before the first class, I suggest downloading one, or a few, and getting in a habit of saving all your data manipulation and analysis commands using that text editor. Visit this website for a comprehensive list of text editors for each OS platform: http://fmwww.bc.edu/repec/bocode/t/textEditors.html

(I use **Text Wrangler** as my text editor. It has never given me an error or problem, it saves files in virtually any file format, it has regular expression find/replace/search, and it has version control / document comparison tools. I would recommend it, but it certainly is not required that you download or use **Text Wrangler**)

3   'A Stata Mind' : A note on Stata vs. other software packages.

I use Stata everyday and *I think* it is the best tool for *most* of my data management or analysis needs. However, there are many other software packages out there, and sometimes Stata is not the best tool for the job. However, we will encounter several limitations of Stata where other software packages may be more be equipped to handle the issue. Many of you may have used SAS or SPSS in the past, so this section focuses on the comparing Stata to these packages.

If you currently are a SAS user, you may want to reference these web resources that help to explain how to do some things in Stata that you already know how to do in SAS:

http://www.cpc.unc.edu/services/computer/presentations/sas_to_stata   OR
www.cpc.unc.edu/services/computer/presentations/sas_to_stata/Stata_for_the_Struggling_SAS_Mind.pdf

If you have some datasets already in SAS format(or any other statistical program) you can convert it to Stata format by either using the "Stat Transfer" program included on most Stata CD's or using Stata's commands for converting data structure. A helpful tutorial for converting a SAS dataset to a Stata dataset can be found here:

http://www.ats.ucla.edu/stat/sas/faq/sas2stata.htm

Further more, you can run SAS commands from within a Stata command file, or do-file, by using a package called –saswrapper- to run any SAS code you've written from Stata.

3.1   Useability: Command Structure & Syntax

Whereas SPSS is heavily a drop-down menu driven software (a somewhat limited command line or programming interaction), SAS and Stata have a strong command line and programming interfaces. Stata's user GUI or drop-down menus have vastly improved in recent years, now rivalling the SPSS user friendly, menu-driven environment, and also, now producing publication quality graphics (Stata's graphical capabilities have probably now surpassed SPSS's). There is no Mac version SAS (there are linux/unix versions that can run in emulation in a Mac environment through the X11 framework, however the greatly reduces performance), some of the SPSS software runs on a Mac, and Stata runs on all major OS platforms.

Syntax: Stata commands are usually on a single line, with options after a comma at the end of the line, and are issued **interactively** using the Command window or in a file of **batch** commands. A SAS program is equivalent to a Stata do file (batch file). The SAS **macro** is essentially a Stata **program**. Data handling is different because Stata puts all of its data in memory at one time, and SAS has its data in memory one line at a time. This helps account for differences in processing time for large datasets or complex equations. A major advantage for Stata is that, in contrast to SAS, you can examine Stata's programs, procedures, and code (and make changes or add your own, if necessary). One downside of Stata is that you need enough physical memory on your machine to load all the data (unless you want to use virtual memory, which can get slow).

Consider the example of running a basic OLS regression in Stata: you just type **regress y x** to regress y on x and see the output and postestimation statisitcs. It's often more difficult to run basic commands in other software programs.

For the simple task of keeping observations in a dataset if the value of a variable (gender) is 1:

In **Stata** you type
"Keep if gender ==1"

in **SPSS** you type
"filter off.
use all.
select if(gender=1). execute "; and

in **SAS** you type
"Data new;
Set old;
If gender=1 then output;
Run;"

Another useful feature of Stata is the ability to output descriptive and analysis output or tables directly into your Word, Excel, PDF, Latex, or HTML document in a format/style of your choosing. No cutting or pasting or reformatting, just one line of code, and if you decide to change the data or variables and rerun the command it will reinsert the table with the new results instantly. There are many presets for table styles for certain style sheets (e.g. APA, APSA, etc) and common academic journals. SAS has the ability to export to certain programs; SPSS requires you to copy/paste output.

Activity: Before class, try updating your Stata software by typing "update all" into the Command window (make sure you have an internet connection). This should update Stata with all the current packages. If there has been an update to the Stata 'executable' file (or the Stata program) it will ask you to type "update swap" to install the free updated Stata program file. When it is complete it should say "updates s uccessfully installed."