Stata .do Files

Why ".do files"?

Up until this point in your Stata use, most of you have probably used Stata as "dynamic", "call and response" data analysis. In other words, for every command you enter at the command window you get some output in the results window, or a graph window opens. You can archive all of this resulting output in a log file – however, if you were interested in duplicating your analysis efforts at a different time, you would need to start from ground zero, re-entering each of the commands in the command window, one by one.
If this need to duplicate a previously used set of commands ever arises, wouldn't it be nice to have all of the commands archived so that you just run them again quickly, without having to go re-enter things on a command by command basis?

Luckily, there is tool for archiving commands and re-running them: it is called a Stata .do file. Before we get into the details of the *.do file* itself, let's talk about some situations where you may want to have a ready to go archive of your commands:

1. Sometimes, an analysis you have prepared for one set of data would be useful and appropriate to apply to other data sets as well. For example, if you were testing a hypothesis regarding factors linked to school performance in Boston high schools, you may be interested in testing this same set of hypotheses to schools in 3 other cities as well.
2. Sometime, especially in the exploratory stage, it is desirable to run a series of basic tabulations and univariate summary statistics on all variables in a dataset. (For example, all items collected in a survey). Usually, the exploration stage begins prior to the cleaning and collection of the entire data set, and therefore it may be desirable to re-run the same analysis as the data set gets updated.

What is a .do file?

It is essentially just a text file, consisting of a list of Stata commands, and (optional) user comments to document the commands.

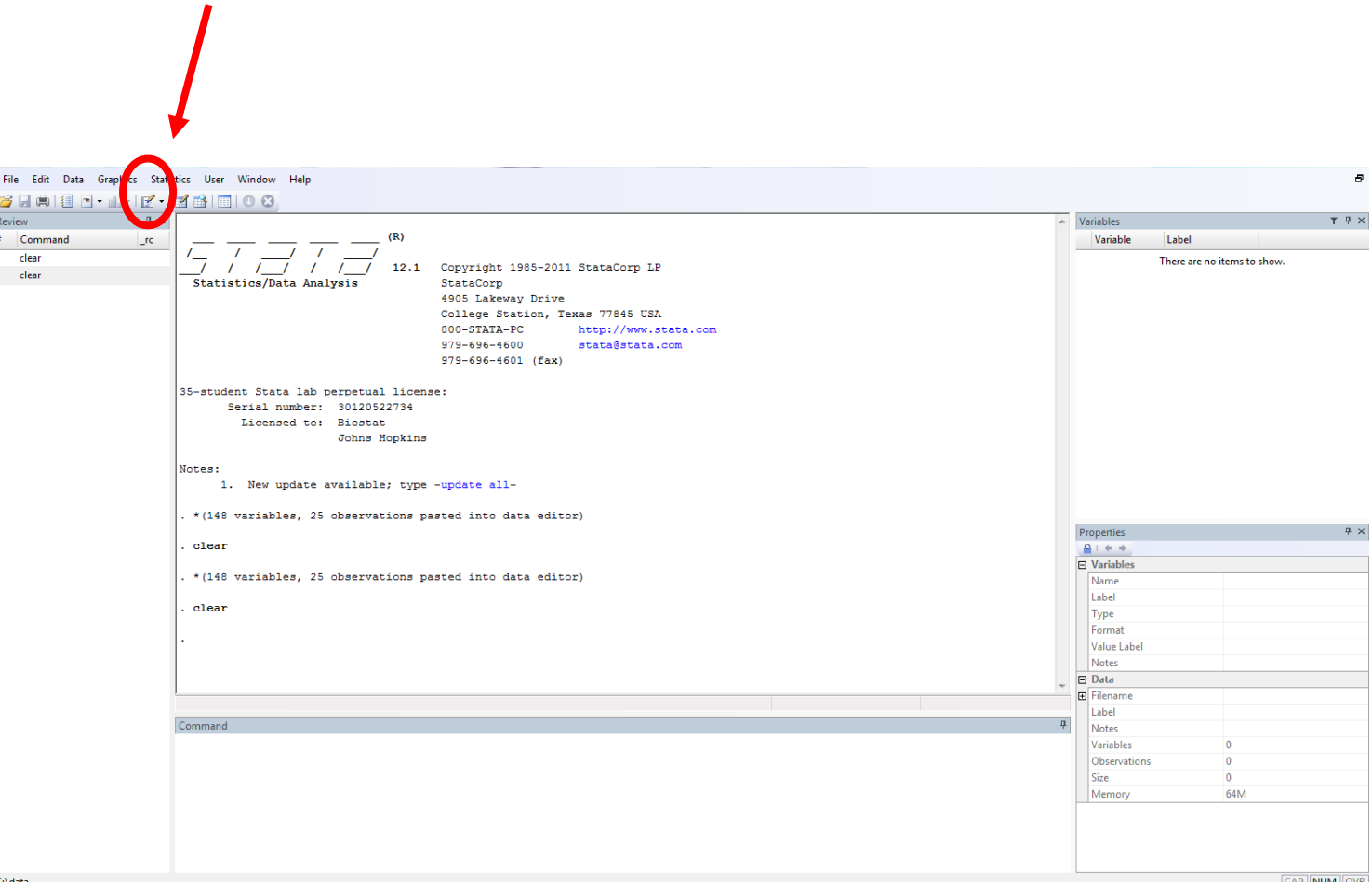
A *.do file* can be
1) created from scratch, with the analyst hand typing each command into a text file, or
2) constructed by cutting and pasting command statements from a *.log file* or
3) created by saving the commands from the review window.

Among these options, the last (number 3) is the most efficient!

First Approach: How do I get to a text editor to create a do-file from scratch?

Stata has a *.do file* editor built right in – to access this "blank canvas", you just need click the *.do file* editor icon on the command icon bar – the *.do file* editor icon looks like an ring notebook. You can type all of your commands into this editor, and then save the resulting .do file.

.do file editor icon



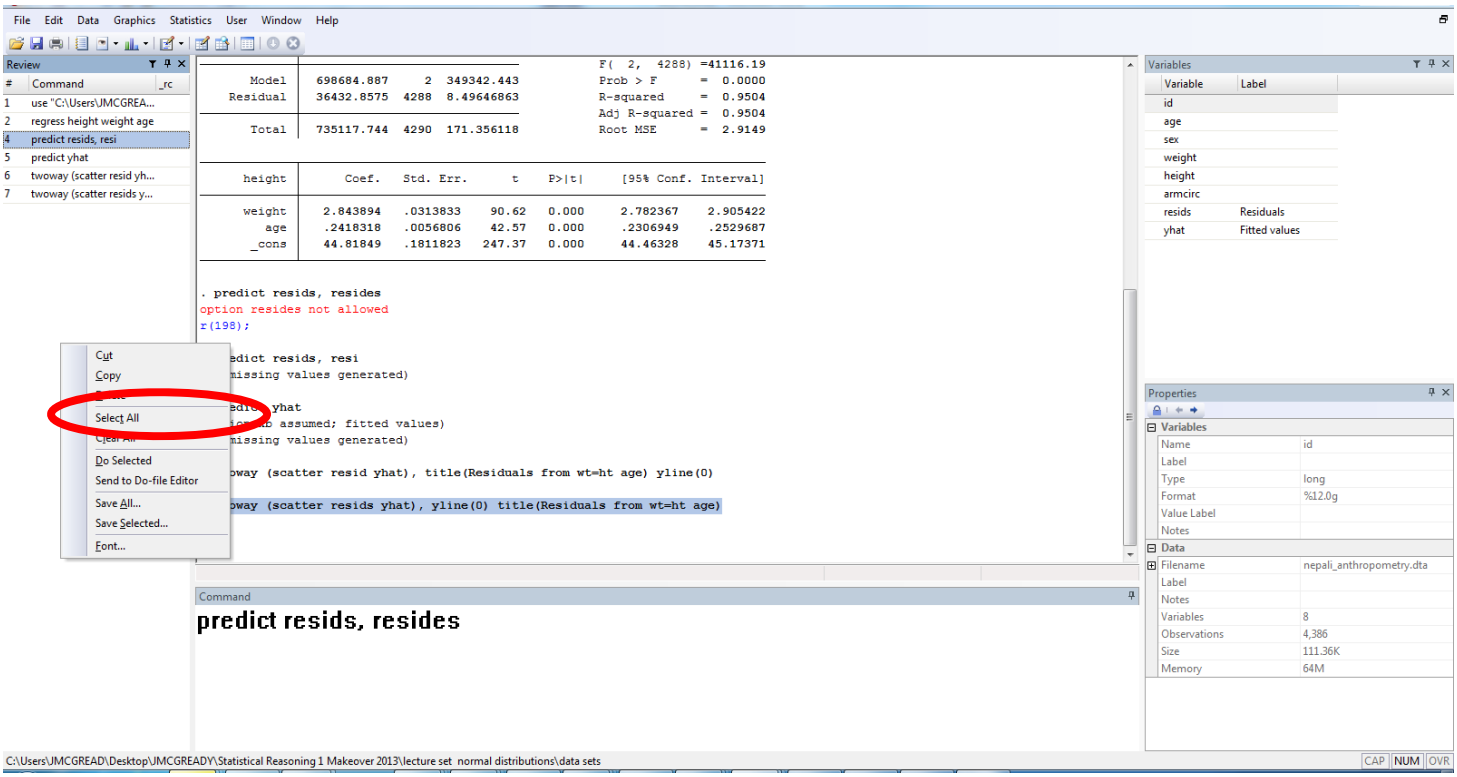
Second Approach: How do I create a do-file from a log file?

One can cut and paste all of the *.log file* into the *.do file* editor, and then delete everything but the relevant commands.

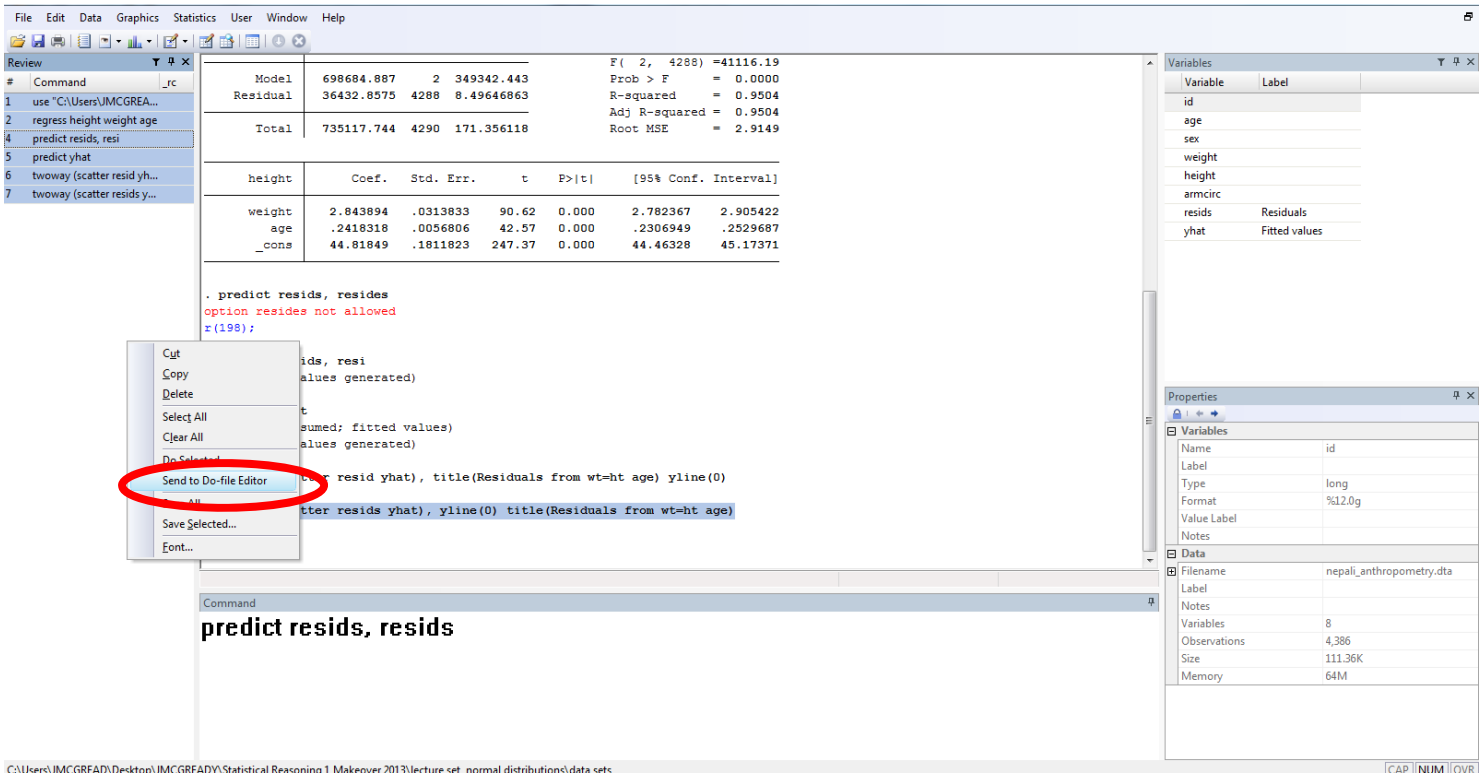
Third Approach: How do create a do-file from the commands in the Review Window?

First, one "right-clicks" within the Review Window:

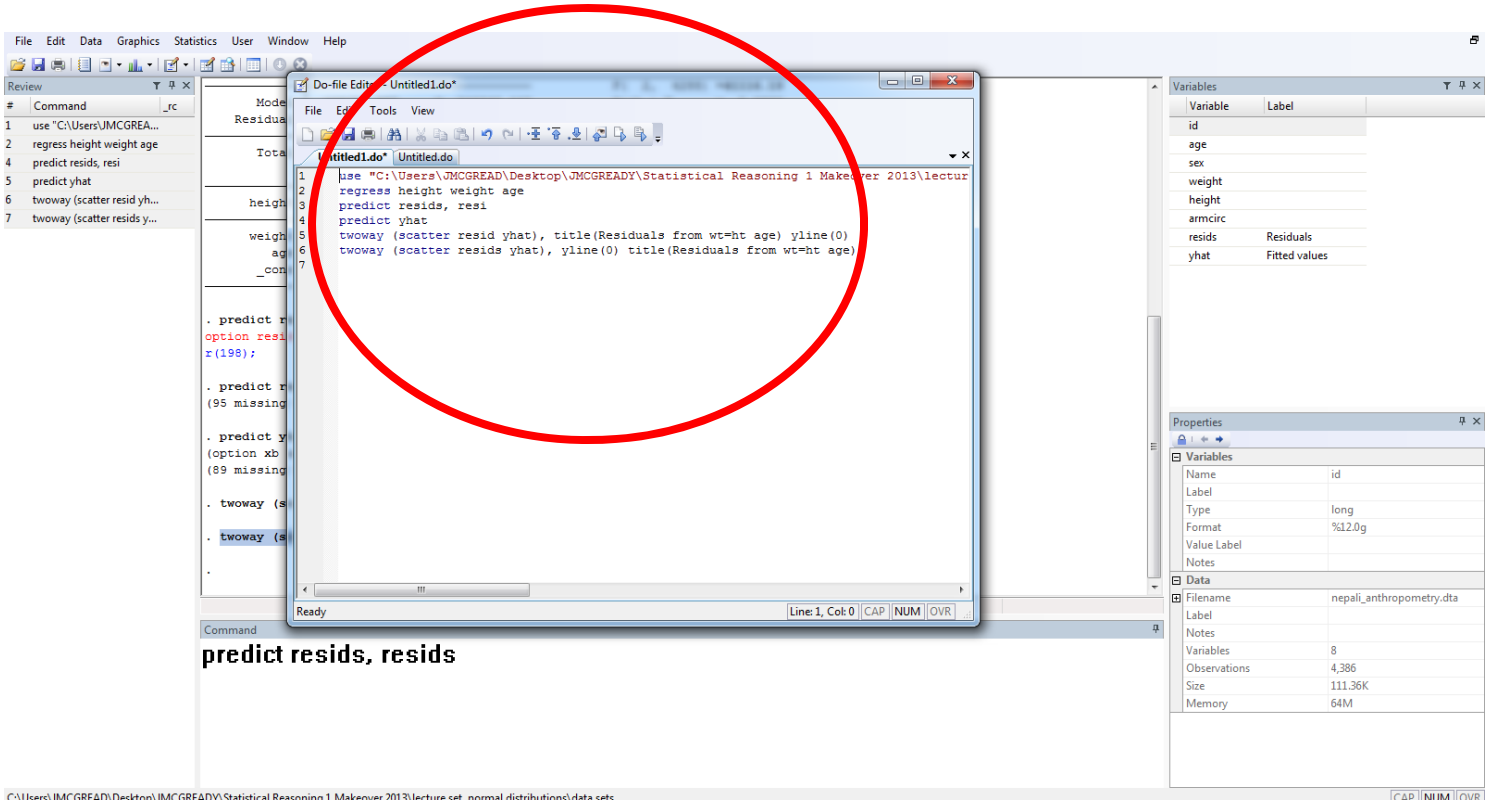
Second, one chooses "Select All" to highlight the commands.



Third, one then “right-clicks” on the highlighted commands and chooses “Send to Do-File Editor”



The commands are then copied into a do-file which can be edited and saved:

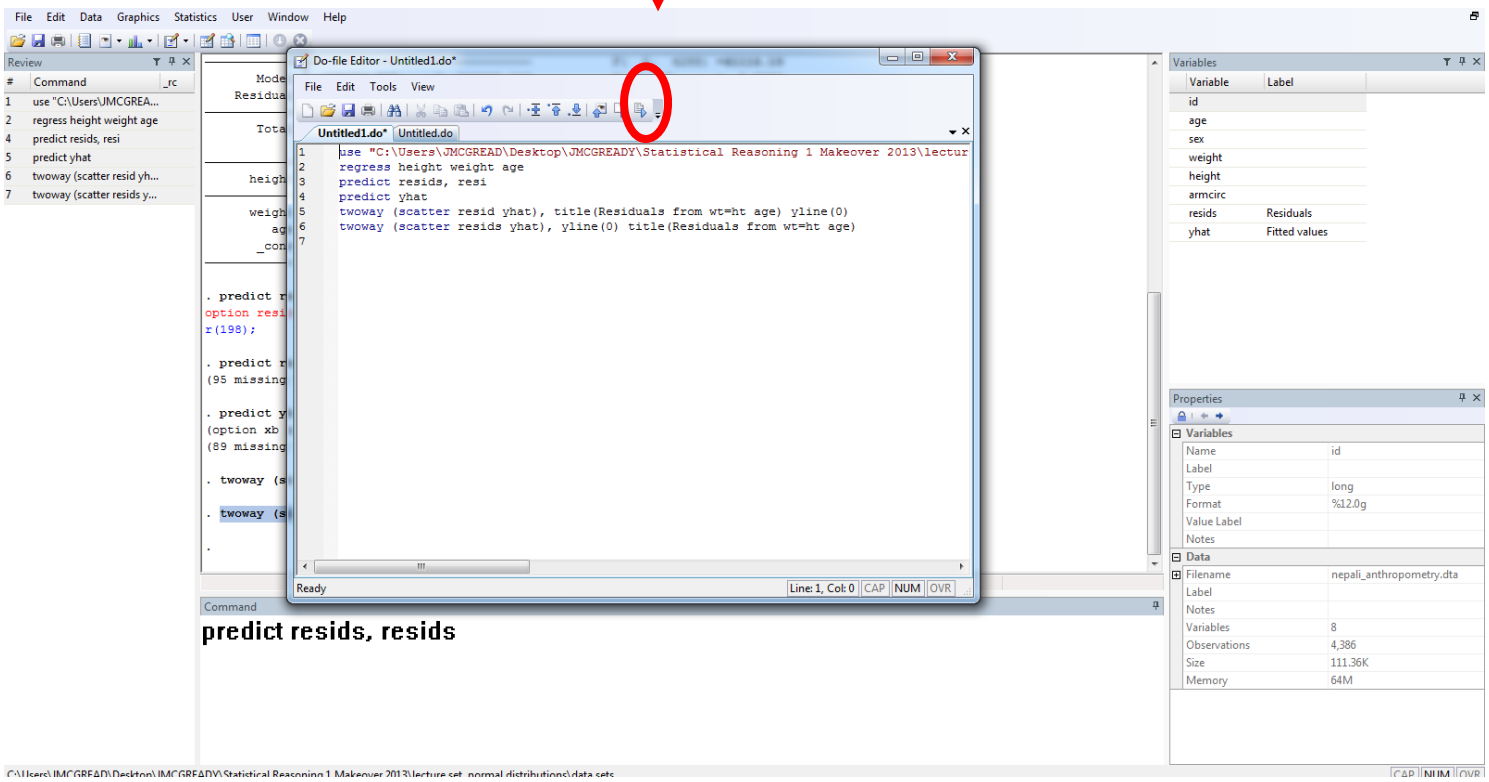


What does a *.do file* look like?

It is really just a file with a list of commands!

Great, now how do I get this .do file to run?

This icon is the "run current file" icon: click on this to run the *.do file*.



A less exiting, but equally effective way to run your do file is to type:

run *filename*

at the Stata command line. (see upcoming section about directory paths)

What's the Catch? The Directory Path Issue!

The only potential "sticky issue" when creating a *.do file* is concerned with directory path recognition issues. Many times when doing a particular analysis, your data files are stored in directories specific to your project, and it would be desirable to place any resulting output (graphs, log files) in the same directory path. Without being specifically told, Stata has no idea where to find your data, or to place your output files. Stata has its own "default" directory, a place where it looks to find things, and posts output unless being specifically told to do otherwise.

For example, suppose you have your data stored in a file called "hw2data" in folder called "bio623" on your disk drive A.

The full name of your file is:

A: \ bio622\hw2data.dta

In order to access this data in Stata, you may be tempted to type:

use "hw2data"

However, chances are good that Stata is not looking in your "A:\bio622" folder for this file. *(in fact, if you wish to know what Stata uses as it's "go to", default director, look in lower right hand corner of the Stata window to see the default working directory)* You will probably get an error message which says "file hw2data not found".

One way to remedy this problem is to modify your statement to:

use "A:\bio622\hw2data"

That is to tell Stata EXACTLY where to find the data file.

Another option is to change the default directory where Stata looks for files if a directory is not specified: This can be done using the "cd" command:

cd "A:\bio622\"

Now when you type "use hw2data", Stata will be looking for the file in the "A:\bio622" directory. Similarly, if you wanted to begin a log file called "mylog", and typed

log using mylog

Stata would open the file in the A:\bio622 directory.

Stata would continue using A:\bio622 as its default until you change it at the command line, or log out of Stata.

How do I make sure that Stata knows where to look for my files, and direct any output I want to save?

There are 3 options:

1) *Every time you reference a file, include the full name (directory path and all)*

```
use  A:\studentname\dofile.dta"
log using "C:\myfiles\term2\bio622\hw1.log"
```

   This option is really only appropriate if you are using files from multiple directories in one analysis – if there is one directory serves as your "base of operations" this approach would be very inefficient (and lead to unnecessary extra typing)

2) *Prior to running your .do file, change your default directory at the Stata command line.*

   Somewhat useful if you are doing your work from one directory – a pain, though, because you must consciously remember to do this every time you want to run your *.do file*.

3) *Include a change directory command at the beginning of your .do file. This will automatically change the default directory path EVERY TIME you run the .do file, on any computer.*

   The best, most efficient option if you are running everything out of one directory (getting your data etc…).

Of course, you can "mix and match" the above options: for example if you are doing most of your work from the directory "A:\bio622\hw1" but want to save your log file in "C:\mystuff\log1.log", you could change your default directory to "A:\bio622\hw1", and then explicitly type the full filename when opening the *.log file*.