Professor Geoffrey Wallace
TA: Darrell Glaser
Fall 2002

STATA 6.0/7.0 Quick Reference for Routines Used in PA 818

This list only provides a brief introduction to STATA. All STATA commands in this summary are printed in **bold** typeface.

Before getting started, here are a few basic help commands that often will get you the information about a specific routine. In the command window you can type one of two requests: **search** and/or **help**. You may also use the windows help menu.

- For example, if you want to know whether a command exists for regression, type **search regression** in the Command window. A list of routines and commands available in STATA will pop-up in the Results window. Clicking on the text for one of these commands (usually printed in blue) will give you the details. Notice the command **regress** on this list. Click on this blue text to get the summary.
- If you have a more specific request, in other words, if you already know about the command **regress**, but you've forgotten the syntax or options available for the command, use the **help** command. For example, in the Command window, type **help regress**.

The Interface

When you open STATA you will see four windows. There is a Review Window, a Results Window, a Variables Window, and a Command Window.

Command Window – This is the window where commands are typed. This is the only window that you will actually type in.

Results Window – This is the window where the results of your commands are displayed. You can either save your results in a log file (more on this later) or cut and paste from the results window using the Edit Menu.

Review Window – The review menu list all of the commands that you have typed during your STATA session (since you opened STATA). If you want to reuse a command all that you need to do is double click on the command in the Review Window.

Variables Window – This window shows all of the variables that are in the data set you are working with. These variables will be variables that were in the data set when it was opened and variables that you have created since the data was opened.

Loading Data

To load a dataset that's already in STATA format, type **use** *<path\file>*. Alternatively, you can also open a dataset by clicking on the open folder icon on the STATA toolbar and following the given procedures. In order to open a file you will have to know where the file is located so when you download data for homework make sure you know where it is saved on the computer.

Saving Data

After making some changes to a data file you might want to save the changes. To save changes to an existing file simply click on the "disk icon" that is right below the File and Edit menus. If you want to save the changes in a new file then click on the File Menu and select "Save As."

Generating New Variables

To generate new variables use the **generate** command (or **gen**).

Example: To create a variable that I'll name "lncost", the natural log of a variable called "cost":

generate lcows=ln(cows)

or alternatively

gen lncost=ln(cost)

Editing Variables

To edit the value of a variable, use the **replace** command.

Example #1: To change the value of a particular variable named "lnq" to zero for all observations:

replace lnq=0

Example #2: To change the value of a particular variable named "lnq" to zero, but only for observations that have a value of "qlevel" less than 2:

replace lnq=0 if qlevel<2

OLS Regression.

Example: To estimate a linear equation by ordinary least squares type

regress lncost lnq lnpk lnpl

where **lncost** is the dependent variable, and **lnq, lnpk** and **lnpl** are independent variables (regressors).

Example: If you want to run a regress on only the observations where "qlevel"=1 then I would type

reg lncost lnq lnpk lnpl if qlevel==1

Example: If I wanted to run a regression by "qlevel" then I would type

sort qlevel
reg lncost lnq lnpk lnpl

This set of commands runs as many regressions as there are values of "qlevel".

Post Estimate Commands

There are a number of very useful commands that are available after you run a regressions (or any other estimation command). These commands allow you to do things like test hypotheses, obtain predicted values, obtain residuals, and access estimated coefficients.

Predicted values. To get the predicted values from a regression, follow-up the **regress** command with the predict command. This has two words, "predict" and whatever you want to name the variable that is the predicted value.

Example: Suppose I want to name the predicted value "lncosthat", then I would type the following set of commands:

regress lncost lnq lnpk lnpl
predict lncosthat

Residuals. To get regression residuals, you follow a similar format as you would for predicted values, simply adding an additional word at the end of the predict statement.

Example: Suppose I want to name the residuals "ehat", then I type:

regress lncost lnq lnpk lnpl
predict ehat, resid

<u>Accessing Coefficient Estimates</u> – In STATA the coefficient estimates can easily be accessed to and used to create new variables or references in hypothesis test

> **Example:** This is a pretty silly thing to do, but I could obtain predicted values by using the coefficient estimates to and the independent variables. To obtain the predicted values from the above regression I will generate a new variable called "lncosthat" and will use the fact that the _b[*independent variable name*] allows me to access the coefficient on the variable named inside the brackets.

> **regress lncost lnq lnpk lnpl**
> **gen lncosthat=_b[lnq]*lnq+_b[lnpk]*lnpk+_b[lnpl]*lnpl**

<u>Hypothesis Testing.</u> To test a set of linear restrictions in the context of a regression model I can use the **test** command following the regress command. The output print after the test command indicates the hypothesis being tested, the F-statistic associated with the test and its distribution, and the p-value.

> **Example:** Suppose that I want to test the null hypothesis $\beta_{\ln pk} = \beta_{\ln pl}$ against the alternative $\beta_{\ln pk} \neq \beta_{\ln pl}$. We would use a generalized F-statistic to test this hypothesis. In STATA we can test the hypothesis by typing

> **regress lncost lnq lnpk lnpl**
> **test lnpk=lnpl**

> alternatively we could test this hypothesis by referencing the coefficients

> **regress lncost lnq lnpk lnpl**
> **test _b[lnpk]=_b[lnpl]**

> **Example:** If I wanted to test the hypothesis that the coefficient on lnpk=0, I would type

> **regress lncost lnq lnpk lnpl**
> **test lnpk**

> **Example:** I can use the **test** command to test more than one hypothesis. For example if I am interested in the hypothesis that the coefficients on lnpk and lnpl are jointly equal to zero, I would type

> **regress lncost lnq lnpk lnpl**
> **test lnpk lnpl**

Example: I could use the test command and the accum option to test the hypothesis that the coefficients on lnpk and lnpl are jointly equal to 1

> **regress lncost lnq lnpk lnpl**
> **test lnpk=1**
> **test lnpl=1, accum**

Creating Plots

It is pretty easy to plot stuff in STATA. Unfortunately, it is more difficult to do plots that look very pretty.

> **Example:** Suppose that I wanted to plot the residuals for my regression against the variable lnq. To do this I would first create the residuals using the **predict** command and then use the **graph** command to plot the residuals
>
> > **regress lncost lnq lnpk lnpl**
> > **predict ehat, resid**
> > **graph ehat lnq**
>
> One the graph command is executed a graph will pop up on your screen. You can copy and post this graph into a word processing program by choosing "Copy Graph" from the Edit Menu.

Creating a Log File

You might be interested in saving you STATA output (the stuff that appears in the Results Window). To do this you will want to open a log file at the beginning of your session. If you are working off a floppy disk the log file "mylog.doc" can be opened with the following commands

> **log using a:\mylog.doc**

Once a log file is open everything that you type and all of the output that is produced will be saved to the file a:\mylog.doc. To stop saving your commands and output to a:\mylog.doc simply type

> **log close**

Of course there is not reason that you have to work on a disk. If you want the log file to be name hw10log and saved to c:\myfiles type

> **log using c:\myfiles\hw10log**

and **log close** when you are done.