**Week 1 Lab Handout- STATA Review**

**PA 5033 – Multivariate Techniques**

March 12th, 2021

Wenchen Wang

**PART A: INTRODUCTION ~ 10min**

**PART B: STATA REVIEW~10min**

**PART C: SIMPLE REGRESSION REVIEW ~ 20min**

**PART D: MAKING REGRESSION TABLES ~10min**

**PART A: INTRODUCTION**

Welcome to Multivariate Techniques Lab! Each Friday we will be covering topics that will aid you in your independent quantitative project. We will lecture on new commands, expand upon items discussed in class, and give you all time to practice new skills.

Labs are meant to be an interactive experience, so please ask questions if you feel confused or lost. It is expected that you attend all labs. If you must miss due to illness, a family emergency, or other conflict, email me ahead of time so that I can make sure to get you up to speed on new material covered. All labs will also be recorded and saved on the cloud for you to access.

If you have not already, make sure you know how to access AppsToGo on your personal computers. We will be using AppsToGo during lab to access STATA, and it allows you to use STATA outside the University.

This link walks you through the three steps necessary to access AppstoGo: <https://it.umn.edu/self-help-guide/appstogo-use-umn-apps-your-personal>. Refer to the section “Access AppsToGo”.

**PART B: STATA REVIEW**

**Directories and log files**

***DO files*** show your code and your comments

***LOG files*** show your code and your comments, and also the output generated from your code. \*\*\*You will turn in a *text* *format* log file for every assignment in this course!\*\*\*

*Start with a do file:*

*clear/cls* Clear your work environment

*set more off*

*Creating a log file:*

In your .do file, write the code for opening a log file after your header:

|  |  |
| --- | --- |
| *cap log close* | This ensures that any previously open log files are closed |
| *cd [folder address]* | This sets the directory for the log file (where it will be saved). Use either your O: drive or your H: drive. You might want to make a folder specifically for PA 5033. |
| *log using lab1\_log, text replace* | This starts the log. The most important part of this command is “*text replace*”  *, text* : ensures the log is saved as a .log file rather than a .smcl file.  *, replace* : allows you to re-run your .do file and generate an updated log file as you make edits to your code  You may also want to use *, text append* if you are adding on to a previously existing log! |
| *use mus03.dta* | Using just the name only works if the file is in the same folder referenced above - otherwise you need to have the full directory reference |

Example Code:  **clear**

**set more off**

**cap log close**

**cd "** **O:\wang6054\PA5033\_Spring2021\Lab1 "**

**use "mus03data.dta"**

**log using Lab1\_log, text replace**

**……**

**log close**

**Commenting**

Commenting in your do-file can be useful to remind yourself why you did something in Stata or to point out important information. There are a couple of ways to comment in your do-file.

\* To create a comment on one line, use an asterisk at the beginning of the line. Comments will show up in green in your do-file, and Stata will recognize them as text rather than as commands

/\* If your comment will be more than one line long, use a forward slash (/) and then an asterisk.

You can even create a new line and the comment will continue until you close the comment with another asterisk and slash. \*/

/\*\*\*\*\*\*\*\*Using more stars can make your comment stand out. Just end the comment with a forward slash like you did above\*\*\*\*\*\*\*\*\*\* \*/

**Summarizing the data**

Now we are going to do some review of basic Stata commands to summarize your data.

If we want to see what is in our data (number of observations, number and descriptions of variables):

*describe*

*d*

For a summary of all of our variables:

*summarize*

*sum*

Let’s take a look at one of the key variables: suppins (1 = does have supplemental insurance)

*sum suppins*

For descriptive of suppins:

*sum suppins, d*

Now, let’s get an understanding of both having a supplemental insurance and the gender of the respondent by creating a crosstab

*tab suppins*

*tab suppins female*

*tab suppins if female == 1*

*tab suppins female, row col cell*

Let’s take a look at another key variable: totexp (total expenditures on medical care)

*sum totexp, d*

*hist totexp*

If we want to get the total number of observations under certain conditions of totexp

*count if totexp == 0*

*count if totexp > 500*

Let’s break down the summary for total expenditures into two categories: those who have supplemental insurance, and those who do not

*tab suppins, sum(totexp)*

The help command can tell you what a command does or how to obtain a particular type of output

*help*

*help scatterplot*

*help twoway scatter*

*two scatter totexp income*

**PART C: SIMPLE REGRESSION REVIEW**

Recall from Regression class, if we want to find statistical relationship between two variables, we will run regression to help us find that relationship. In this current dataset, let’s say we want to find the relationship between income and total medical expenditure, to be more specific, how would income impact total medical expenditure, holding family size, education, gender, age, and marital status constant, we can run a simple linear regression:

*reg totexp income famsze educyr female age marry*

How can we interpret the results? What about the test statistics?

**PART D: MAKING REGRESSION TABLES**

**Estout**

After we install the function of estout, first we run the targeted regression, and we save the regression result as m1:

*ssc install estout*

*reg educyr famsze age*

*est store m1, title(Model 1)*

Then we output the regression results into an output table inside Stata:

*estout m1, cells(b(star fmt(3)) se(par fmt(2))) ///*

*replace starlevels(\* 0.1 \*\* 0.05 \*\*\* 0.01) ///*

*legend label varlabels(\_cons Constant) ///*

*stats(r2 N)*

If we want to export more than one regression result, we can save other regression results and output the saved results altogether:

*reg totexp age famsze*

*est store m2, title(Model 2)*

*reg retire age*

*est store m3, title(Model 3)*

*estout m1 m2 m3, cells(b(star fmt(3)) se(par fmt(2))) ///*

*replace starlevels(\* 0.1 \*\* 0.05 \*\*\* 0.01) ///*

*legend label varlabels(\_cons Constant) ///*

*stats(r2 N)*

**Outreg2**

Similarly, with outreg2, we first install this function, and run our targeted regression, and output our regression result into a table outside of Stata in a saved file of our chosen format:

sscinstall outreg2

*reg educyr famsze age*

*outreg2 using reg1.doc, replace ctitle(Model 1)*

We can also append more results together:

*reg totexp age famsze*

*outreg2 using reg1.doc, append ctitle(Model 2)*

If we want to get separate table than reg1, we need to rename our file:

*reg totexp age famsze*

*outreg2 using reg2.doc, replace ctitle(Model 2)*