Stata Lab 1 – Introduction

1 What is STATA?

- A statistical software package used mostly in economics, sociology, political science and epidemiology.
- Stata can be used to manage database, run regressions, generate graphics, do simulations, etc.
- The user should have their own dataset. The Stata data file is usually saved in the .dta format.
- Data of any other formats (like excel) can be imported and/or converted into .dta format.

1.1 STATA supports

- Stata's own website: http://www.stata.com/support/faqs/
- UCLA Academic Technology Services: http://www.ats.ucla.edu/stat/stata/
- Stata program's help function: For example, suppose you would like to know more about the "regress" command, then... Open the stata program > in the "command" box > type "help regress" without the " "> press enter.
- Stata's official manual (can be found in the library)
- Other Stata's user's manual: My favorite one is "An Introduction to Modern Econometrics Using Stata" by Christopher F. Baum.
- or... simply type your question(s) into a search engine.

1.2 Data files and Do-files

- The Stata's data file keeps all the data points. For example, each Y_i and the corresponding X_i, $\forall\ i = 1, 2, 3..., n$.
- The do-file records all the commands that you use to analyze the data.
- Once the data is cleaned, it is best not to keep on re-saving the original data file. If several steps have to be done before analyzing the data (like running a regression), do it on the do-file.

2 Tutorial 1: Exploring the Data and Running a Simple Regression

- **Download Wooldridge datasets**

1. Download Wooldridge's data – go to thomsonedu's website:
   "http://www.thomsonedu.com/aise/economics/wooldridge_2e_datasets/".
2. save "statafiles.zip" on your "H:\" drive (this is your personal folder). Unzip "statafiles.zip".
3. save "excelfiles.zip" in your "H:\" drive (this is your personal folder). Unzip "excelfiles.zip".

- **To open the STATA program**

1. Double click on the STATA icon.
2. Click on the "Do-file" icon on the top panel of the Stata program. "Save As" your Do-file (and name it "EE425") in your H:\ drive.

- **Open the data file using Do-file**

1. Indicate your working folder type: cd "H:\"
2. Create a "log-file" in order to record all your Stata activities (this is optional) type: log using "EE425_tutorial_1.txt"
3. When you would like to stop recording your work on Stata, type: log close
4. To open your log file again, type: log using "EE425_tutorial_1.txt", append
5. Open the data file type: use "CEOSAL2.DTA", clear

- **To explore and understand the data**

1. type: browse
2. type: describe
3. type: summarize
4. type: sum
5. type: codebook
6. type: describe salary
7. type: tabulate college
8. type: tab college
9. to use the "if" command to find conditional mean (average) type: sum if grad == 1
10. type: sum salary if age <= 40

11. type: correlate salary sales profits
12. type: correlate salary sales profits, covariance
13. type: plot salary profits
14. type: twoway scatter salary profits

- **To run a simple (OLS) regression (one explanatory variable)**

1. type: regress salary profits

- **To create the fitted value** $(\hat{Y}_i)$ **and the residual** $(\hat{u}_i)$

1. type: predict y_hat, xb
2. type: predict u_hat, residual

- **To see how well we do at finding a Sample Linear Function**

1. type: twoway scatter salary profits || line y_hat profits
2. type: twoway scatter u_hat profits
3. To check if the OLS estimation makes X_i uncorrelated with $\hat{u}_i$ (by the OLS calculation, they should not correlate), type: correlate sales u_hat

- **To exit your Stata**

1. type: log close
2. Save your do-file
3. file -> exit -> don't save

- **To find out what all the above commands mean** – (type in the command box) help summarize, help predict, help twoway, etc etc.

2.1 Examples from Wooldridge(2009)

C2.2 The data set in CEOSAL2.dta contains information on chief executive officers for U.S. corporations. The variable *salary* is annual compensation, in thousands of dollar, and *ceoten* is prior number of years as company CEO.

1. Find the average salary and the average tenure in the sample
2. How many CEOs are in their first year as CEO (that is, $ceoten = 0$)? What is the longest tenure as a CEO?
3. Estimate the simple regression model

$$\log(salary) = \beta_0 + \beta_1 ceoten + u.$$

and report your results in the usual form. What is the (approximate) predicted percentage increase in salary given one more year as a CEO?

C2.1 Use the data in WAGE2.dta to estimate a simple regression explaining monthly salary (*wage*) in terms of IQ score (IQ).

1. Find the average salary and average IQ in the sample. What is the sample standard deviation of IQ? (IQ score are standardized so that the average in the population is 100 with a standard deviation equal to 15)
2. Estimate a simple regression model where a one percentage point increase in IQ changes *wage* by a constant dollar amount. Use this model to find the predicted increase in *wage* for an increase in IQ of 15 percentage points. Does IQ explain most of the variation in *wage*?
3. Now, estimate a model where each one percentage point increase in IQ has the same percentage effect on *wage*. If IQ increases by 15 percentage points, what is the approximate percentage increase in predicted *wage*?