Stata Lab

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 the dataset**

1. Download CEOSAL2.DTA from the class' Moodle page.

- **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 "EE489") on your computer.

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

1. file -> open -> then, direct the program to the file "CEOSAL2.DTA".
2. On the command window, you will see a command to open this file. If you would like to open the file from your do-file in the future, you can use this command.

- **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 execute mathematical operations**

1. type: generate log_salary = log(salary)
2. type: gen log_profit = log(profit)
3. type: gen profit_2 = 3+5*profit
4. type: regress salary profits profit_2
5. type: gen profit_sq = profit^2
6. type: regress salary profits profit_sq

- **To perform a multiple regression analysis**

1. type: regress salary profits sales
2. type: regress salary profits sales ceoten

- **To exit your Stata**

1. Save your do-file
2. file -> exit -> don't save (never ever modify your master dataset!)

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

3 Tutorial 2 – Dummy, Heteroskedasticity, Specification Issues

3.1 Does "beauty" help increase wage?

1. Download the datafile "beauty.xlsx" from your EE325 Moodle page.
2. Open the STATA software program. Click on the "Data Editor" icon.
3. Copy the entire dataset from the excel file and paste it onto the STATA's Data Editor page.

   Choose "Treat first row as variable names".
4. Save the new STATA dataset.

   Choose File -> Save As -> (then name the new dataset "beauty_stata")
5. Open a new do-file and save it.

   Choose "New Do-file Editor" icon

   On the Do–file's top panel, choose File -> Save As -> (then name the new do-file "second_stata_lab")

   - **To explore and understand the data**
6. type: sum
7. type: codebook
8. type: tab look
9. type: tab look female

   - **We want to test whether "good look" has a positive impact on wage**
10. type: gen belavg = 0
11. type: replace belavg = 1 if look < 3
12. type: gen log_wage = log(wage)
13. regress log_wage belavg abvavg

    - **Seems like we may have the omitted variable bias. Let's take into account other variables.**
14. type: regress log_wage abvavg belavg educ
15. type: regress log_wage abvavg belavg educ exper expersq
16. type: regress log_wage abvavg belavg educ exper expersq bigcity
17. type: regress log_wage abvavg belavg educ exper expersq bigcity black
18. type: regress log_wage abvavg belavg educ exper expersq bigcity black union

19. type: regress log_wage abvavg belavg educ exper expersq bigcity black union female

    - **Check if we have the heteroskedasticity problem. (Let's use the White Test (special case))**

20. type: regress log_wage abvavg belavg educ exper expersq bigcity black union female

21. type: estat hettest

22. Do we reject H_0 : homoskedasticity at 5% level of confidence?

    - **Should we believe that the value of β are the same for female and male?** (Chow Test)
    - Chow statistic is a type of F-statistic $F = \frac{SSR_p - (SSR_1 + SSR_2)}{SSR_1 + SSR_2} \cdot \frac{[n-2(k+1)]}{k+1}$

23. To get SSR_p: regress log_wage abvavg belavg educ exper expersq bigcity black union

24. To get SSR_1: regress log_wage abvavg belavg educ exper expersq bigcity black union if female == 0

25. To get SSR_2: regress log_wage abvavg belavg educ exper expersq bigcity black union if female == 1

26. What is the value of the F-statistic? Can we reject H_0 (can use the same model)?