14.75, Extra Recitation, Introduction to STATA

1 The different windows

Everyone opens STATA

- output window - output and error messages
- command window - to directly enter and try commands; otherwise, use a dofile to write a program
- variables window - with labels
- past commands window

2 How to use and save data and how to program

2.1 Dofile

- ALWAYS USE A DOFILE
- enables to save the program and rerun it in one click (difference with excel or SPSS)

Everyone opens a dofile

- in which folder your databases are located: *cd "C:\TA\14.75\Stata_recitation"*
- comment on what you do, starting lines with *
- how to run part or whole of the dofile: Execute

2.2 Use and save data

- use data: *use stataintro_1475, clear.* "clear" enables to switch to a new database when another one was already opened
- save data: *save stataintro_1475_v2, replace.* "replace" enables to save the database even if it already existed

3 How to get help

- in Stata, if you know the command: *help regress*
- in Stata, if you don't know the command: *search prediction*
- online: http://stata.com/support/faqs/; type "introduction to stata" in google

4 Describe the data

- Click on Data Editor (Browse)
- *describe*
- there are different types of variables: numeric (different types, depending on the precision) or string
- *su education* (short for *summarize education*): number of observations, mean, standard deviation, etc.
- *ta education*: frequency of each value of the variable

5 Manipulate the data: 3 examples, "replace", "keep" and "gen"

- Change the level of education of the first person (data entry mistake): *replace education = 10 if name == "Alain Auguste"*
- We are only interested in people with less than 20 years of education: *keep if education <= 20*
- Generate ϵ_i, a random normal variable with mean 0 and standard deviation 300: *gen e = rnormal(0,300)*
- Generate the outcome variable wage w_i where we call education x_i, $\alpha = 1100$ $\beta = 70$,

$$w_i = \alpha + \beta education_i + \epsilon_i.$$

gen alpha = 1100

*gen wage = alpha + 70*education + e*

6 Analyze the data (1): linear regression, testing

- Correlation between education and e? *corr education e*
- Plot wage against education: *graph twoway scatter w education*
- Estimate $\hat{\beta}$: impact of education on wage? Linear model: *reg w education.* Note that the constant is included by default in the regression
- Test that $\beta = 0$: *test education*

7 Analyze the data (2): instrumental variables

- Instrumental variable: suppose we are suspicous that we are missing a variable and have found an instrument for education

  so name

  gen instrument = (_n <= 94)

  ivreg wage (education = instrument)

Remember to save your dofile and your data!

14.75 Political Economy and Economic Development

Fall 2012