DATA EXERCISE #1 – *PRE-LAB*

Please do the following *before the start of class* on Tuesday, April 13

Your first step is to download the following to files from Canvas. They can both be found in the files section of canvas in the folder called "data."

- **de1-nls.dta**– a STATA dataset with information from the National Longitudinal Survey of Young Men. The dataset includes information from a survey of young men (aged 24-34) in 1976. The survey also collected information on family background and IQ scores, two of the key variables we discussed as likely important "omitted" variables from a regression attempting to estimate the causal effect of education on wages.
- **d1-template.do** – a template STATA do-file that you will use for the exercise.

1. Move both files into a new folder on your computer that you will use for this exercise (e.g., my folder is called "DE1").
2. Open the dataset by double clicking on it and use that to identify the path to the folder where you are working.
3. Copy the folder path (the part between "/"s on macs in the use command, or the "C:\ …\" or whatever on PCs).
4. Open the template do-file. Add the folder path inside the quotes of the **cd ""** command.
5. Add a command to the program to take the mean of the *momhs* dummy variable that is initially constructed by the template program. Think of an answer before class to the question: "what does the mean of this variable tell you?"
6. Now try running a regression of *lnwage* on *educ* and the *momhs* dummy variable. Interpret the slope coefficient on *educ.* Any guess on how to interpret the slope coefficient on the *momhs* dummy variable?