Econometrics 1: Microeconometric Methods of Impact Evaluation.
Exercise 1: Data manipulation, omitted variable bias

1. Download the Davis database from the following location:

   http://www.ats.ucla.edu/stat/stata/examples/ara/arastata_dl.htm

   Explore this database and study its codebook available at the same address

   (a) Regress the logarithm of the weight on the logarithm of the height. Why use logarithms? Give an interpretation of the results. Comment the results.

   (b) Explore further the database by doing some graphs for example. What do you discover?

   (c) Make the necessary corrections and go back to question a). Give a graphical interpretation of your results: is the human being a cube or a stick?

   (d) Try to define a variable measuring lies by individuals about their height or weight. Is it linked to gender? In what way?

2. Download at http://www.er.uqam.ca/nobel/r11720/Sciences_Po.htm ("matthieu chemin" in google, then teaching) under "Lecture 1" the databases merge1.dta and merge2.dta. These databases contain 1000 observations about individuals identified with the variable "id". Explore this database by looking at the variables. sex_merge1 is the sex of the individual in merge1.dta, sex_merge2 is the sex of the individual in merge2.dta. Verify visually that for the same individual (with the same "id"), sex_merge1=sex_merge2. merge1.dta contains religion, and merge2.dta contains socgroup. The individuals are the same, and we thus want to merge the databases to get access to all information in one file.

   (a) Open merge1.dta, type "merge using merge2.dta" (with the correct path on your computer).

   (b) Explore the result (with tab, inspect, graphs, codebook, or visually). Is there something weird in the final database?

   (c) Read carefully the explaainations for the merge command in stata.

   (d) According to these instructions, carefully merge the databases

Conclusions:

- You must carefully read the instructions in stata.

- You must never use the stata commands like black boxes, you must always understand the underlying functioning of the commands.

- From now on, I encourage you to explore yourself stata, now that you know how to open databases in stata. Use the help section in stata, or use the internet (type your question in google: somebody must have asked the ame question before you).

3. Omitted Variable Bias.

   Download the "Data NSLY" under "cours 1". Regress LGEARN (logarithm of earnings) on:

   (a) S and ASVABC

   (b) S only

   (c) ASVABC only

   Calculate the correlation between S and ASVABC. Compare the coefficients between S in regressions (a) and (b). Give a mathematical and intuitive explanation about the direction of the change. Same exercise for ASVABC.

4. Introduce the variable MALE, then the variables ETHBLACK, ETHHISP to test for sexual and racial discrimination. Criticize these regressions.

5. This exercise is about the controversy raised by Foote et Goetz (2005) in "Testing Economic Hypotheses with State-Level Data: A Comment on Donohue and Levitt (2001)" about Donohue, J. and Levitt, S. (2001), "The Impact of Legalized Abortion on Crime", Quarterly Journal of Economics, 116(2), 379-420.

   (a) Read the two papers. What is the controversy about the state-year dummies? What is the adavantage of including state-year dummies, compared to simply state dummies and year dummies.

   (b) Visit the webpage of John J. Donohue, III (http://islandia.law.yale.edu/donohue/pubsdata.htm) Open the do-file corresponding to the paper. Find the error mentionned by Foote and Goetz about the state-year dummies.

   (c) Open the corresponding database in stata.

   (d) Generate the state-year dummies (using a loop, hint: forvalues).

   (e) Include these state-year dummies in the regressions of columns 1,3,5,7.

   (f) Interpret the coefficients.