Economics 6160: Econometric Analysis
Assignment 5
Due date March 25th

In this assignment set, you will become acquainted the data used in Krueger (1999), a seminal study of the effects of class size on student achievement. These data are from the Tennessee Student-Teacher Achievement Ratio (STAR) experiment. Over a four-year period starting the fall of 1985, Project STAR randomly assigned elementary school students and teachers to classes of different sizes.

To do these exercises, you will use the data set stardata_Kindergarten.dta. This is a subset of a larger dataset that has been restricted to only include Kindergarteners. The codebook for this data is given in file readme-star.txt.

You may work on this assignment in pairs or independently. You should do this assignment set by writing a program (a *.do file in Stata). This will be turned in to your TA (via email) and should be well annotated. In the same email, you should also include a document that displays any relevant Stata output and responses to questions. This is a group homework, only one of the group members need to submit but you should clearly indicate your name and your members' name in the email.

As usual, keep in mind that your assignment will be returned to you for revision if it does not run on our machines unaltered or when a mistake in your analysis has been discovered.

Data notes:

- The provided data set contains 5,408 observations. The unit of observation is an individual student. Each student in the data set participated in Project STAR in Kindergarten.

- Missing values are coded to 9, 99, 999, etc. Before using any variables, you will therefore need to recode these values to missing (with a "."). (See the codebook for more details. Use the recode or replace command in Stata.) Note: if you create some variable "x2" based on some variable "x1" and the latter has missing values, you will want to check that x2 also has missing values where appropriate.

- Outcome variables:
  mathk, readk: Stanford Achievement Test (SAT) math and reading scores in kindergarten
  Note: These scores are in units that are not easily interpreted. As discussed in Krueger's article (pages 507-08, footnote 11), these scores should be converted to percentile scores before conducting the analysis. The "starter program" I provide you includes code that does so in creating the variable "mnscorek."

- Treatment variables:
  stark, ctypek: In a STAR classroom and classroom type (small, regular, or regular with aide) in K

- To see variable labels, use the describe command in Stata. (If desired, add names to values of categorical variables using label define and label values.)

0. Getting Started

Create a folder for this problem set. Put the data set in this folder along with a do file titled Ass1_LastnameFirstname.do. Your do file should begin with the code that I provided so that it will run on my computer and your TAs computer and so that you have the necessary variables.

1. Getting the data ready for analysis

a. As noted above, our analysis will focus on the variable "mnscorek" that has been constructed from the raw scores on the reading test and math test. As such, this variable should positively correlated with reading test scores and math test scores. Show that this is the case in two figures using the command "scatter." (You may want to use the "graph export" command to save the figures.)

b. Create the following variables: 0/1 indicator that takes 1 if a student receives free/reduced price lunch and zero if she does not; 0/1 indicator that takes a 1 if a student is white or Asian and 0 if she is not. Use the "summarize var1 var2 …" command to make sure that these variables make sense.

c. Create an indicator variable that takes a 1 if a student's teacher is white or Asian and 0 if she is not. Also create a variable (or modify the existing variable) for the teacher's number of years of experience. Again use the summarize command to verify that these data make sense.

d. Calculate means of these variables and class size by treatment status (small class, regular class, and regular class with aide). [Hint: Use the "tabstat var1 var2 …, by(ctypek)" command.)] Do these statistics support the assertion that the treatment was randomly assigned to students and teachers?

2. To be continued in future assignments…