Homework 2
Due Thursday, February 4, 1:30 PM
Full Credit will not be Given Unless You Show Your Work

Part I: from page 1-83 in the class notes, we have that

$$\hat{\beta}_1 = \beta_1 + \sum_{i=1}^{n} w_i u_i \ \text{ where } \ w_i = \frac{d_i}{SST_x} \ \text{ and } d_i = x_i - \bar{x}$$

a. Show that $\hat{\beta}_1$ and $\bar{u}$ are uncorrelated

b. Show that $\hat{\beta}_0 = \beta_0 - \bar{u} - \left(\hat{\beta}_1 - \beta_1\right)\bar{x}$

c. Show that $\ \text{Var}\left(\hat{\beta}_0\right) = \sigma^2\left(\dfrac{1}{n} + \dfrac{\bar{x}^2}{SST_X}\right) = \dfrac{\sigma^2 \sum_{i=1}^{n} x_i^2}{n \cdot SST_X}$

Part II – On Trunk, you will find a Stata data set **College Distance** that contains data from a random sample of high school seniors interviewed in 1980 and re-interviewed in 1986. A file that describes the variable in **College Distance** is also on Trunk.

In this exercise, you will use these data to investigate the relationship between the number of completed years of education for young adults and the distance from each student's high school to the nearest four-year college (proximity to college lowers the cost of education, so that students who live closer to a four-year college should, on average, complete more years of higher education).

1. Run a regression of years of completed education (*yrsed*) on distance to the nearest college (*dist*) where distance is in tens of miles. Display the regression results as in equation 4.14 in Wooldridge (page 132; of course, the variables from this regression are not in logs) and interpret the estimated coefficient for *dist*.
2. Is it statistically significant at the 5% significance level (why or why not)?
3. What is the associated p-value?

4. Is the estimated coefficient economically significant (why or why not)?
5. What is the 99% confidence interval estimate for the slope coefficient?
6. Run the same regression using data only for women and display the regression results as in equation 4.14 in Wooldridge (page 132). Interpret the estimated coefficient for *dist*.
7. Run the same regression using data only for men and display the regression results as in equation 4.14 in Wooldridge (page 132). Interpret the estimated coefficient for *dist*.
8. Is the effect of distance on completed years of education (statistically) different for men and women? Show your work.

Part III – Read through Section IV (up to page 1586) of "Where is the Land of Opportunity? The Geography of intergenerational Mobility in the United States," by Raj Chetty, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez.

Answer the following questions:

1. What does the concept "Land of Opportunity" refer to?

2. What does the evidence indicate about the United States as a "Land of Opportunity"?

3. On page 1554, the authors state "Although our descriptive analysis does not identify the causal mechanisms that determine upward mobility, the publicly available statistics on intergenerational mobility developed here can facilitate research on such mechanisms." Why is it the case that their analysis does not identify the causal mechanisms that determine upward mobility?

An important finding in the paper is that there is considerable variation across the U.S.in upward mobility.

4. One of the five factors that is strongly correlated with this geographic variation in upward mobility is inequality as measured by the Gini coefficient. What is the Gini coefficient and how does in reflect inequality in this case? Why might inequality impact upward mobility?

5. In order to be able to compare parental and child income over time, the authors put these variables in real terms. Do a similar exercise with the average hourly wage of all employees for the time period 2009:3 to 2015:10 (monthly). Obtain data from the Bureau of Labor Statistics – see below.

    i.   Generate real average hourly earnings where the base month is January 2015 (i.e. put hourly earnings in January 2015 dollars).

    ii.   Generate summary statistics for the cpi, average hourly earnings, and real average hourly earnings using the "tabstat" command. Make the table look presentable and give that table a title.

    iii.   What is the percent change in the average hourly earnings between the first period (March 2009) and the last period (October 2015)?

    iv.   What is the percent change in real average hourly earnings between the first period (March 2009) and the last period (October 2015)?

6. Questions about the rank-rank slope measure of relative mobility.

    i.   Explain the rank-rank slope measure of relative mobility.

    ii.   Is a large or small value associated with more mobility? Why?

    iii.   What is the estimate of the rank-rank slope using national data?

    iv.   How is this obtained?

    v.   How does this value for the U.S. compare to other countries?
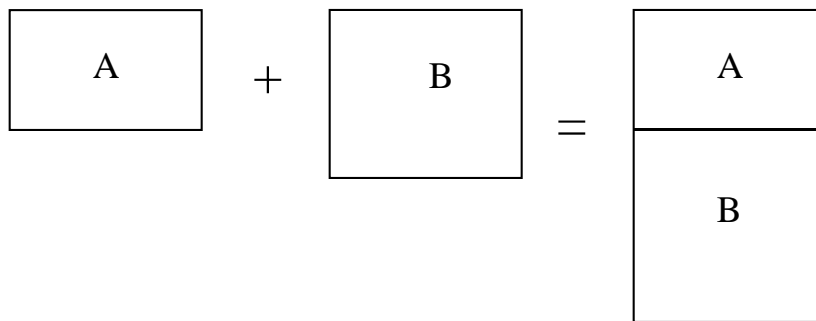
    vi.   Why do the authors prefer this over the IGE?

Stata command "merge": There are two ways to combine data:
1. Merge: Different variables for the same observations
2. Append: Different observations with the same variables

1. Merge

A + B = A | B

2. Append

A + B = A / B
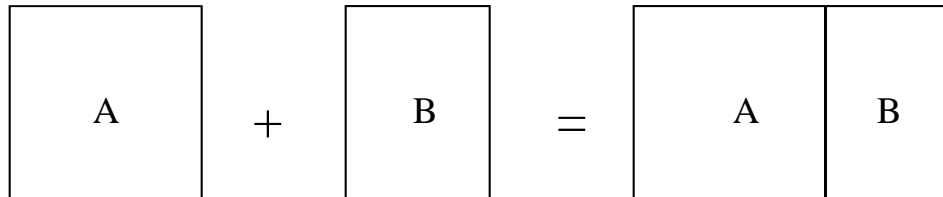
To merge two data sets, A and B, with the same observations but different variables requires that both A and B include the same variable that uniquely identifies each observation. Call this variable ID. Assume that there are two stata datasets A.dta and B.dta

```
use c:\A.dta
sort ID
save c:\A.dta, replace
clear
use c:\B.dta
sort ID
merge 1:1 ID using c:\A.dta
tab _merge
```

The "_merge" variable is created when the merge command is run. It takes on three values

$_merge$ = 1 if observation is from dataset A and NOT B
= 2 if observation is from dataset B and NOT A
= 3 if observation is from BOTH data sets A and B

Obtaining monthly data on the average hourly wage and the cpi from the BLS

1. Go to the BLS homepage at www.bls.gov
2. Click on the "Data Tools" on the top banner
3. Click on "Top Picks"
4. Under the heading "Employment, " click the box for "Total Private Average Hourly Earnings of All Employees - Seasonally Adjusted - CES0500000003"
5. Then click "Retrieve Data"
6. Click "More Formatting Options" (upper right)
7. Highlight "Column Format" and click "Retrieve Data"
8. Copy and paste the three columns of data **through 2015 M10** into the data editor in Stata and then close the data editor
9. Save the data as a Stata dataset named "'homework2_hourly_wages"
   - This is monthly data from March 2009 to October 2015
10. Clear the register
11. Back on the BLS website, click the back arrow three times and under the heading "Price Indices, " click the box for "CPI for All Urban Consumers (CPI-U) 1982-84=100 (Unadjusted) - CUUR0000SA0" (make sure to unclick the box for "Total Private Average Hourly Earnings of All Employees - Seasonally Adjusted - CES0500000003")
12. Then click "Retrieve Data"
13. Click "More Formatting Options"
14. Highlight "Column Format" and click "Retrieve Data"
15. Copy and paste the three columns of data for March 2009 through October 2015 into the data editor in Stata and then close the data editor
16. Save the data as a Stata dataset named "'homework2_cpi"
17. Open the do file homework2.do (download from trunk)
18. For the line that starts "cd" add in the location where you saved the two Stata datasets 'homework2_hourly_wages and 'homework2_cpi.
19. Follow the directions in III.5 to add Stata code to this do file and then run the do file.