

17.871  
Spring 2007  
Problem Set # 1

Handed out: February 28, 2007  
Due back: March 8, 2007

1. Comment on the research designs of the following two studies. Discuss whether they are designed in a way that would allow the researcher to draw the conclusion that is drawn. Comment on the sampling and whether the sampling method may introduce biases into the analysis. State what are the dependent and independent variables in these designs, and what any confounding variables might be. If the research design was insufficient, write a short paragraph indicating why not, and what could, or should, have been done to improve the design.

1.a. MIT faculty members were interested in determining whether ending spring-term freshman Pass/No Record had been a success. They decided to answer this question by comparing the GPA of spring-term freshmen before and after the change in Pass/No Record grading had taken effect. The average freshman GPA in the spring of 2002 is 4.0; the average freshman GPA in the spring of 2003 is 4.4. The faculty concluded that the change was a success. (Note the obvious: these are made-up data.)

1.b. Researchers were interested in determining whether postcards sent to registered voters encouraging them to vote actually worked. The researchers took the list of registered voters in a town and randomly assigned them to one of two samples—T, a sample of voters who were sent the get-out-the-vote postcard, and C, a sample of voters who were not sent the get-out-the-vote postcard. After the election, the researchers went to the town clerk to see who voted. They discovered that 75% of the T group voted, whereas 55% of the C group voted. The researchers conclude that the causal effect of sending the postcards is to increase turnout by  $75\% - 55\% = 20\%$ .

2. Find two tables that interest you in the Statistical Abstract of the United States that meet the following criteria: (1) they have between 25 and 52 observations and (2) they have the same units of analysis (e.g., states, years, nations), (3) the subject matter of the two tables are conceivably linked. You can find the abstract here:

<http://www.census.gov/prod/www/statistical-abstract.html>

2.a. Call these two tables Table A and Table B. Create a STATA data set that contains one variable, plus the identifying variable (like state or year), from Table A. Save it. Create a STATA data set that consists of one variable, plus the identifying variable (like state or year), from Table B. Save it. Merge the two data sets. Save the merged dataset.

- 2.b. Turn in the following: i. The “do file” that shows how you create the data sets and merged them. ii. A printout of the data. iii. A short (one paragraph, 2 or 3 sentences) description of the tables you got your data from.
3. Using the CCEs.dta file in the Examples folder of the 17.871 locker, draw a publishable histogram of income. By publishable, I mean label the axes and title the graph appropriately.
4. Using the CCEs.dta file in the Examples folder of the share locker:
- 4.a. Create a publishable scatter plot of partisan identification by income at the state level. Before doing so, drop the District of Columbia. Hint: first use the collapse command to aggregate the data by state. Turn in the figure and the do file.
- 4.b. Create a publishable scatter plot of partisan identification by income for each region. Hint: first use the collapse command to aggregate the data by income and region. Turn in the figure and the do file.
- 4.c. Does the relationship between partisan identification and income differ between the state level and the within region level? Briefly suggest an explanation for any difference.