ECON 104 **Truth in Numbers**

Homework 2

In the first homework you evaluated the effect of a get out the vote effort on voting. Hopefully you are convinced that the analysis you conducted as part of the first homework produced unbiased estimates of the effect of the get out the vote effort. This is because in the data you analyzed in the first homework the random assignment to treatment resulted in the comparison group providing a good estimate of the probability that people in the treatment group would have voted without the encouragement.

For this homework you will determine if you can get an unbiased estimate of the treatment effect using a treatment that people have to opt into. The treatment for this homework is actually receiving and listening to the entire call encouraging you to vote. The point of this homework is to see if you can get the "right" (experimental answer we believe is correct from the first homework) in a non experimental setting.

Use the same dataset as you used for HW1. The variable *contact* is equal to 1 for people who actually received and listened to the entire call and 0 otherwise. We can think of these people as opting into the treatment in the sense that they were available to receive the call and did not just hang up.

1. Create a Table 1 that lets you compare the means of the sample characteristics in the treatment and comparison group. Please create a clean table that includes columns with the means of each group, the difference between the two groups and the p-value of the difference. The table should be comprehensible on its own.

2. Does Table 1 suggest that the control group will provide a good counter factual for the treatment group's voting potential outcomes (can we just compare the fraction voting in each group)? Why or why not?

3. How does Table 1 differ from the table you created in answer to balance of covariates question on first homework? Why do you think they differ?

4. Lets see if regression will get us the right answer by adjusting for the differences between the two groups you documented in Table 1. Create a carefully labeled Table 2 where each column corresponds to a regression. The first column contains the parameters (and their standard errors) of the regression $vote02 = B_0 + B_1 contact + u$. In each column after the first add one more covariate to the regression.

5. What effect does adding covariates have on your estimate of the treatment effect? What does this tell you about the relationship between the covariates and the outcome?

6. How do the results in Table 2 differ from the regression results from the first homework? Carefully examine both the standard errors and the point estimates for the treatment effect. Why do these two tables differ?

7. Did adding covariates to the regression as you did in answer to question 4 reduce the bias? If yes why do you think it did? Which variable reduced the bias the most? Why do you think this is the case?

8. Did regression eliminate all the bias? Why or why not?

9. Please attach all your code