Unit 8 Case Study

Problem Statement:

The League of Women Voters asked PhD and Masters' students enrolled in a statistical consulting course in the Department of Statistical Science at SMU to help them perform an experiment to determine whether sending either a postcard or a voter's guide (flyer) to potential voters increased the percent who actually voted in the 2014 election. The LWV was particularly interested in the effect on voting behavior among young and Hispanic voters, since this group tended to be less likely to vote in past elections.

The project spanned three semesters. In the spring of 2014, students estimated the sample size needed for the study and developed the experimental design and sampling methodology. In the fall of 2014, the data were gathered and cleaned, and in the spring of 2015, the data were analyzed.

The sample consisted of 24,000 randomly selected individuals from a population of 531,735. The population was low-propensity voters, determined ahead of the 2014 election. A low propensity voter is someone who has only voted in 0 or 1 of the last three elections. The sample was randomly partitioned into three treatment groups.

- 8000 received a postcard reminding them to vote
- 8000 received a flyer with a reminder and voting instructions
- 8000 did not receive any mailing. These are the control group.

After the 2014 election, information regarding voter participation was collected for all voters.

When the data were analyzed, it was found that 10% of the low propensity voters who had received the flyer voted, approximately 12% of those who received the postcard voted, and approximately 23% of the control group voted.

The LWV was not happy with these results, as they had expected that some sort of reminder, whether a flyer or a postcard, would increase the chances of voting. The LWV asked the students to go back and examine the data. And guess what? There was a problem.

Assignment:

You are to figure out the problem. There is a good (albeit subtle) reason that the expectations do not match the results, and it can be found in the data. The reason has something to do with the topic for Unit 8, which is statistical sampling.

The LWV data is supplied to you in a CSV file and a SAS file. All that are required to figure out the problem are simple descriptive statistics and graphics that you have already encountered in the program. You may use either R or SAS to examine the data. You may also work in groups, as long as there are no more than three people in a group.

All you are to worry about is finding the problem itself. Do not worry about the remedy for the problem. It is somewhat complicated, and it will be discussed in the live session for Unit 10.

Your deliverable is a Word or Rmd document containing a well-written explanation for the problem, along with any supporting graphics and tables. Your document should be no more than two pages.

We will discuss the case study in Unit 10 live session.