

# Lab 3: Hypothesis Tests about the Mean.

w203: Statistics for Data Science

*Tako Hisada*

*11/08/2017*

## The Data

The American National Election Studies (ANES) conducts surveys of voters in the United States before and after every presidential election. You are given a small subset of the 2012 ANES survey, contained in the file `ANES_2012_sel.csv`.

There are a number of special concerns that arise whenever statisticians work with survey data. In particular, the complete ANES survey data assigns a survey weight to each observation, which corrects for differences in how likely individuals are to be selected, and how likely they are to respond. For the purposes of this assignment, however, we have removed the survey weights and we ask you to assume that the observations you have are a random sample from the voting population.

For a glimpse into some of the intricacies that go into survey design, take a look at the introduction to the ANES User's Guide and Codebook.

```
S = read.csv("ANES_2012_sel.csv")
```

Following is an example of a question asked on the ANES survey:

Where would you place YOURSELF on this scale, or haven't you thought much about this?

Possible answers included:

- 1. Extremely liberal
- 2. Liberal
- 3. Slightly liberal
- 4. Moderate; middle of the road
- 5. Slightly conservative
- 6. Conservative
- 7. Extremely conservative
- -2. Haven't thought much about this
- -8. Don't know
- -9. Refused

The variable `libcpre_self` records answers before the election, while `libcpo_self` records answers after the election.

WARNING: If you coerce these variables directly into numeric vectors (or if such coercion happens automatically), the levels may not translate into the right numbers. Consider the following code.

```
table(S$libcpre_self, as.numeric(S$libcpre_self))
```

```
##
##              1      2      3      4      5      6      7
## -2. Haven't thought much about this 556      0      0      0      0      0      0
## -8. Don't know                     0     26      0      0      0      0      0
```

##	-9. Refused	0	0	32	0	0	0	0
##	1. Extremely liberal	0	0	0	195	0	0	0
##	2. Liberal	0	0	0	0	638	0	0
##	3. Slightly liberal	0	0	0	0	0	641	0
##	4. Moderate; middle of the road	0	0	0	0	0	0	1828
##	5. Slightly conservative	0	0	0	0	0	0	0
##	6. Conservative	0	0	0	0	0	0	0
##	7. Extremely conservative	0	0	0	0	0	0	0
##								
##		8	9	10				
##	-2. Haven't thought much about this	0	0	0				
##	-8. Don't know	0	0	0				
##	-9. Refused	0	0	0				
##	1. Extremely liberal	0	0	0				
##	2. Liberal	0	0	0				
##	3. Slightly liberal	0	0	0				
##	4. Moderate; middle of the road	0	0	0				
##	5. Slightly conservative	789	0	0				
##	6. Conservative	0	1001	0				
##	7. Extremely conservative	0	0	208				

You can find explanations for other variables in the ANES codebook.

## Assignment

You will use the ANES dataset to address the following questions:

1. Did voters become more liberal or more conservative during the 2012 election?
2. Were Republican voters (examine variable `pid_x`) older or younger (variable `dem_age_r_x`), on the average, than Democratic voters in 2012?
3. Were Republican voters older than 51, on the average in 2012?
4. Were Republican voters more likely to shift their political preferences right or left (more conservative or more liberal), compared to Democratic voters during the 2012 election?
5. Select a fifth question that you are interested in investigating.

Prepare a report addressing these questions. A successful submission should include:

1. A brief introduction.
2. A suitable hypothesis test for each question above.
3. For each test, include:
  - A brief exploratory analysis targetted to check the assumptions needed for your test.
  - A justification for why the test is the most appropriate choice
  - An explanation of test results, including *BOTH* statistical significance and practical significance.
4. A brief conclusion with a few high-level takeaways.

Please limit your submission to 10 pages. Be sure to submit both your pdf report as well as your source file.