Lab 3: Hypothesis Tests about the Mean.

w203: Statistics for Data Science

Introduction

The American National Election Studies (ANES) conducts surveys of voters in the United States before and after every presidential election. Using the data taken from the 2012 elections, I will try to answer the following questions of interest via various hypothesis tests:

1. Did voters become more liberal or more conservative during the 2012 election?
2. Were Republican voters older or younger, 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. Right before the 2012 election, were women voters more or less liberal than men voters?

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("/Users/shanhe/Desktop/w203/Lab/Lab_3/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

Analysis

I used the ANES dataset to address the following questions:

1. Did voters become more liberal or more conservative during the 2012 election?

```r
# EDA on libcpre_self & libcpo_self
unique(S$libcpre_self)
```

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

```r
unique(S$libcpo_self)
```

```
##  [1] -6. Not asked, unit nonresponse (no post-election interview)
##  [2] 2. Liberal
##  [3] -8. Don't know
##  [4] 4. Moderate; middle of the road
##  [5] -2. Haven't thought much {do not probe}
##  [6] 6. Conservative
##  [7] 5. Slightly conservative
##  [8] 3. Slightly liberal
##  [9] 7. Extremely conservative
## [10] 1. Extremely liberal
## [11] -7. Deleted due to partial (post-election) interview
## [12] -9. Refused
## 12 Levels: -2. Haven't thought much {do not probe} ...
```

Notice that 1) we have survey responses like "-2. Haven't thought much" and 2) we have different levels of liberalness and # conservativeness. Hence, I 1) add numeric variables that represent the level of conservativeness/liberalness ans 2) exclude answers with no applicability in terms of levels of liberalness and conservativeness

```r
#temporarily assign 0 to the NA answers, but will be excluded in the analysis
S_n <- mutate(
        S,
        libcpre_self_n = as.numeric(ifelse(substr(S$libcpre_self,0,1) == "-", 0, substr(S$libcpre_self,0
        libcpo_self_n = as.numeric(ifelse(substr(S$libcpo_self,0,1) == "-", 0, substr(S$libcpo_self,0,1
)

summary(S_n[, c("libcpre_self_n", "libcpo_self_n")])
```

```
##  libcpre_self_n  libcpo_self_n
##  Min.   :0.000   Min.   :0.000
##  1st Qu.:3.000   1st Qu.:2.000
##  Median :4.000   Median :4.000
```

```
##  Mean   :3.739   Mean   :3.537
##  3rd Qu.:5.000   3rd Qu.:5.000
##  Max.   :7.000   Max.   :7.000
```

```r
# subsetting dataset, exclude voters with non-applicable answers, either pre or post the election
S_n_1 <- subset(S_n, libcpre_self_n != 0 & libcpo_self_n != 0, select= c(libcpre_self_n,libcpo_self_n))
```

In this case, we have a very natural pairing between our observations. Each voter has a data point before and after the election. Since the variable is ordinal, we will want to use Wilcoxon Signed-Rank test, depending on whether the sample meets its assumption

We first examine the means directly

```r
c(mean(S_n_1$libcpre_self_n), mean(S_n_1$libcpo_self_n))
```

```
## [1] 4.178857 4.159731
```

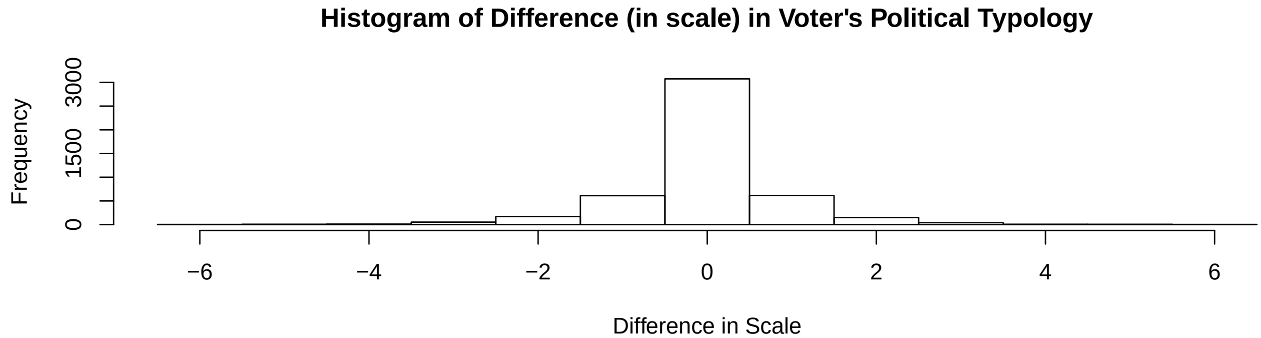
Notice that the two means are very similar, which shows low practical significance

We then investigate the difference in values

```r
D_1 = S_n_1$libcpo_self_n - S_n_1$libcpre_self_n
summary(D_1)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -6.00000  0.00000  0.00000 -0.01913  0.00000  6.00000
```

```r
hist(D_1, breaks = -6:7 - 0.5,
     main = "Histogram of Difference (in scale) in Voter's Political Typology",
     xlab = "Difference in Scale"
     )
```



The sampling distribution of the differences seem to have a symmetric distribution. With an approximately symmetric distribution and a large sample size on ordinal scale, we can use a Wilcoxon Signed-Rank test.

We then run a two tailed Wilcoxon Signed-Rank test with: Null Hpypothsis = voters didn't become either more liberal or conservative during the 2012 election Alternative Hypothesis = voters did become either more liberal or conservative during the 2012 election

```r
# Wilcoxon Signed-Rank test
wilcox.test(S_n_1$libcpo_self_n, S_n_1$libcpre_self_n, paired = TRUE)
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  S_n_1$libcpo_self_n and S_n_1$libcpre_self_n
## V = 682330, p-value = 0.1662
## alternative hypothesis: true location shift is not equal to 0
```

Notice that we have a p-value of 0.1662, showing weak statistical signifincance. Hence, we can't reject the null hypothesis, stating that the voters didn't become either more liberal or conservative during the 2012 election, with a 95% confidence level.

We then look at the pratical significance of our hypothesis test:

```
#Investigate Practical Significance
cohen.d(S_n_1$libcpo_self_n, S_n_1$libcpre_self_n, paired = TRUE)
```

```
##
## Cohen's d
##
## d estimate: -0.02006681 (negligible)
## 95 percent confidence interval:
##         inf         sup
## -0.06025672  0.02012310
```

Notice that the estimated effect size is -0.02, which is negligible

2. Were Republican voters (examine variable pid_x) older or younger (variable dem_age_r_x), on the average, than Democratic voters in 2012?

We first do an EDA on our variables of interest

```
# EDA on pid_x & dem_age_r_x
unique(S_n$pid_x)
```

```
## [1] 1. Strong Democrat            3. Independent-Democrat
## [3] 4. Independent                6. Not very strong Republican
## [5] 5. Independent-Republican     2. Not very strong Democract
## [7] 7. Strong Republican          -2. Missing
## 8 Levels: -2. Missing 1. Strong Democrat ... 7. Strong Republican
```

```
summary(S_n$dem_age_r_x)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   -2.00   35.00   51.00   48.92   62.00   90.00
```

Notice that in pid_x we have "-2. Missing" and different levels of Demoracts and Republicans I will 1) exclude records with "-2. Missing" or "4. Independent" as pid_x since they can't be categrized as either Republican voters or Democratic voters and 2) create a categorical value based on pid_x to categorize Republican voters or Democratic voters

```
S_n_2 <- mutate(subset(S_n,substr(pid_x,0,1) != '-' & substr(pid_x,0,1) != '4',
                       select = c(pid_x,dem_age_r_x)),
                voter_cat = factor(ifelse(substr(pid_x,0,1) < 4, 'Democrat', 'Republican'))
              )

# Next, examine dem_age_r_x variable
summary(S_n_2$dem_age_r_x)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   -2.00   35.00   51.00   49.32   62.00   90.00
```

Notice that we see negative age, which doens't make sense and should be excluded

```
S_n_2 <- subset(S_n_2,dem_age_r_x > 0)
```

We then look at the distribution of dem_age_r_x for both Republican voters and Democrat voters

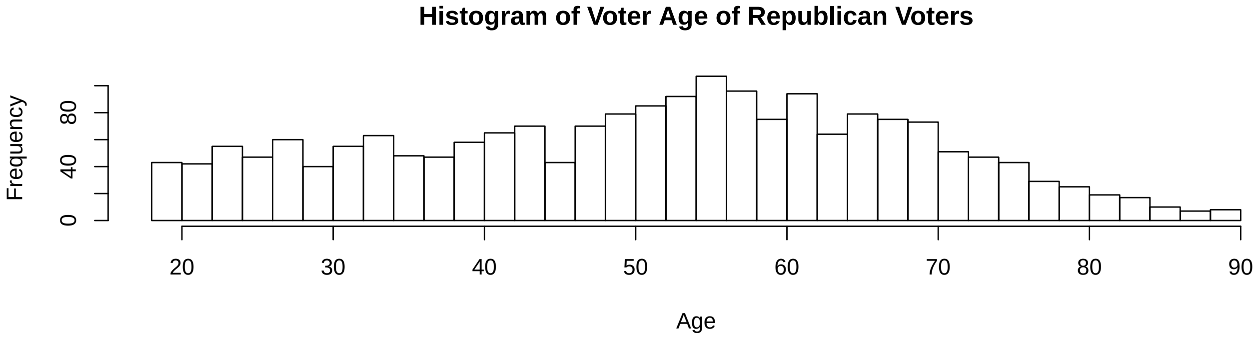
```r
# For Republican Voters
length(subset(S_n_2, voter_cat == "Republican")$dem_age_r_x)
```

```
## [1] 1981
```

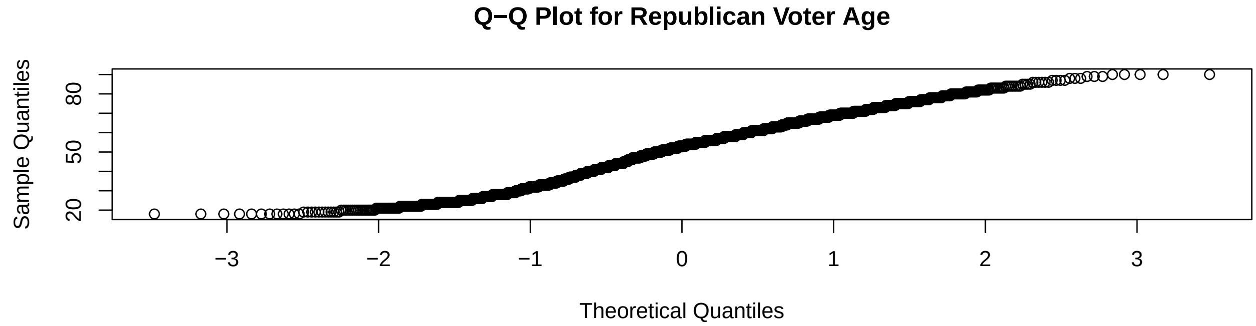
```r
summary(subset(S_n_2, voter_cat == "Republican")$dem_age_r_x)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   18.00   38.00   53.00   51.33   64.00   90.00
```

```r
hist(subset(S_n_2, voter_cat == "Republican")$dem_age_r_x, breaks = 50,
     main = "Histogram of Voter Age of Republican Voters",
     xlab = "Age")
```




```r
qqnorm(subset(S_n_2, voter_cat == "Republican")$dem_age_r_x, main = "Q-Q Plot for Republican Voter Age")
```



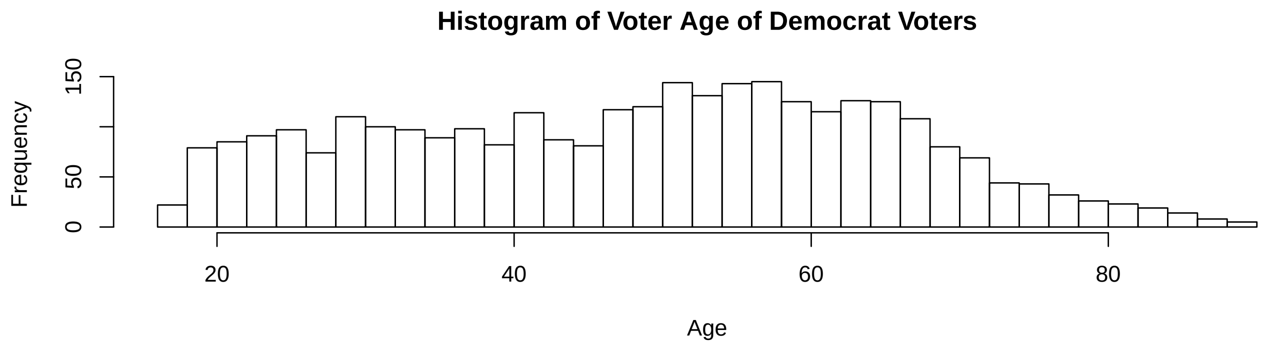

```r
# For Democrat Voters
length(subset(S_n_2, voter_cat == "Democrat")$dem_age_r_x)
```

```
## [1] 3068
```

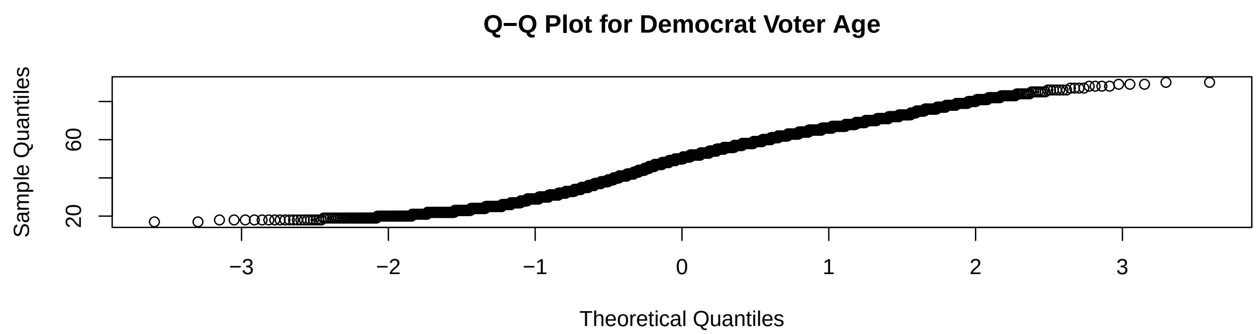
```r
summary(subset(S_n_2, voter_cat == "Democrat")$dem_age_r_x)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   17.00   35.00   50.00   48.84   62.00   90.00
```

```r
hist(subset(S_n_2, voter_cat == "Democrat")$dem_age_r_x, breaks = 50,
     main = "Histogram of Voter Age of Democrat Voters",
     xlab = "Age")
```




```r
qqnorm(subset(S_n_2, voter_cat == "Democrat")$dem_age_r_x, main = "Q-Q Plot for Democrat Voter Age")
```



Notice that both distributions are skewed. Here we want to compare the age of the Democrat voters to that of the Republican voters. Although we don't have normal distributions, we do have large sample sizes allowing us to perform a two-sample t-test for the continuous vairble, age, with:

Null Hypothesis = the average age of Democrat voters and Republican voters were the same in 2012 Alternative Hypothsis = the average age of Democrat voters and Republican voters were different in 2012

```r
t.test(dem_age_r_x ~ voter_cat, data = S_n_2)
```

```
##
##  Welch Two Sample t-test
##
## data:  dem_age_r_x by voter_cat
## t = -5.1653, df = 4206.5, p-value = 2.512e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.439637 -1.546939
## sample estimates:
##   mean in group Democrat mean in group Republican
##                 48.83735                 51.33064
```

Notice that we have a p-value as 2.512e-07, showing strong statistical significance We hence are confident to reject the null hypothese that the average age of Democrat voters and Republican voters are the same.

Also, by observing that we have a smaller mean age for Democrat Voters and a low p-value in the two-tailed test. We know that we can also reject the null hypothese against an alternative hypothese as "The mean age of Democrat voters are younger than that of the republican voters"

We can do the one-tailed test just to confirm, here we have:

Null Hypothesis = the average age of Democrat voters and Republican voters were the same in 2012 Alternative Hypothsis = the average age of Democrat voters was younger than that of Republican voters in 2012

```r
t.test(dem_age_r_x ~ voter_cat, data = S_n_2, alternative = "less")
```

```
##
##  Welch Two Sample t-test
##
## data:  dem_age_r_x by voter_cat
## t = -5.1653, df = 4206.5, p-value = 1.256e-07
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -1.69914
## sample estimates:
##   mean in group Democrat mean in group Republican
##                 48.83735                 51.33064
```

This confirms our analysis

We then look at the practical significance for our hypothesis test:

```r
# Investigate Practical Significance
cohen.d(dem_age_r_x ~ voter_cat, data = S_n_2)
```

```
##
## Cohen's d
##
## d estimate: -0.149074 (negligible)
## 95 percent confidence interval:
##         inf         sup
## -0.20565352 -0.09249442
```

Although our sample statistics show strong statistical significance, the effect size is quite small. With a Cohen's d value of -0.149, it shows small practicial significance.

3. Were Republican voters older than 51, on the average in 2012?

We first create a subset of the data used in question 2 to only contain the age for Republican voters

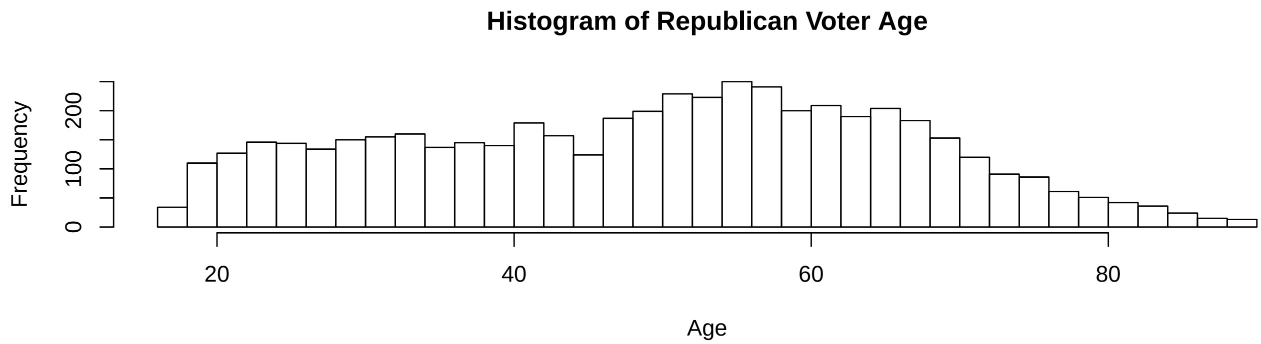
```r
S_n_3 <- subset(S_n_2, voter_cat == "Republican", select = c(voter_cat, dem_age_r_x))
```

We then perform a EDA on the variables of interest

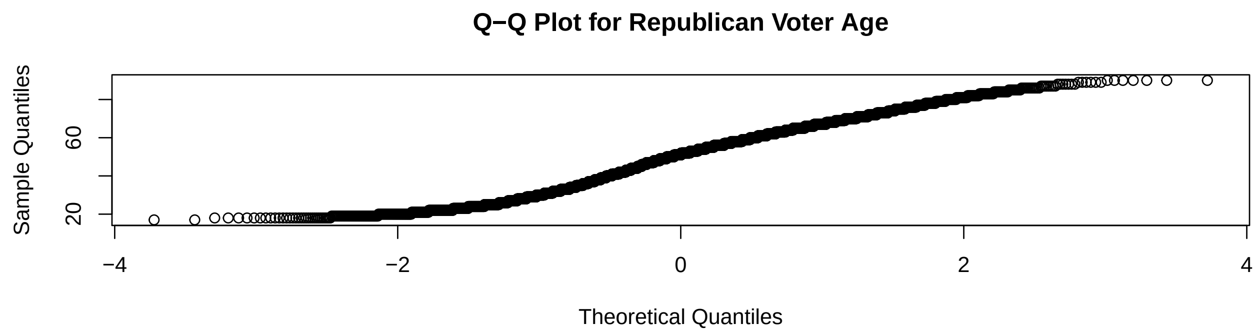
```r
# Examine dem_age_r_x
summary(S_n_3$dem_age_r_x)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   18.00   38.00   53.00   51.33   64.00   90.00
```

```r
hist(S_n_2$dem_age_r_x, breaks = 50,
     main = "Histogram of Republican Voter Age",
     xlab = "Age")
```




```r
qqnorm(S_n_2$dem_age_r_x, main = "Q-Q Plot for Republican Voter Age")
```



Notice that the distribution is skewed with a peak at around 55. Here we want to compare the age of the Republican voters to a constant. Although we don't have a normal distribution, we do have a large sample size allowing us to perform a parametric t-test

However, although we were asked that whether the average age of the Republican voters was higher than 50, we should still perform a two-tailed test because 1) there is no solid explanation that why we would treat a large observed difference in the unexpected direction no differently from a difference in the expected direction that was not strong enough to justify rejection of the null hypothesis and 2) if the sample mean is indeed larger than 50, a two-tailed test decreases the porbability of a Type I error.

Null Hypothesis = In 2012, the average age of the Republican voters was 51 Alternative Hypothesis = In 2012, the average age of the Republican voters was not 51

```r
t.test(S_n_3$dem_age_r_x - 50)
```

```
##
##  One Sample t-test
##
## data:  S_n_3$dem_age_r_x - 50
## t = 3.5279, df = 1980, p-value = 0.0004284
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.5909323 2.0703498
## sample estimates:
## mean of x
##  1.330641
```

Notice that with a two-tailed test, our t-statistic shows strong statistical significance rejecting the null hypothesis stating that the average age for Republican voters is equal to 50 with up to 99.95% confidence . Moreover, for an one-tailed test where the alternative hypothesis is that average age for Republicanvoters is greater then 50, we would have a p-value that's a half of the p-value in the two-tailed test. Hence, we can confidently (up to 99.97%) reject the null hypothesis in that scenario

We can to a one-tailed test to confirm comments above:

Null Hypothesis = In 2012, the average age of the Republican voters was 51 Alternative Hypothesis = In 2012, the average age of the Republican voters was greater than 51

```
t.test(S_n_3$dem_age_r_x - 50, alternative = "greater")
```

```
##
##  One Sample t-test
##
## data:  S_n_3$dem_age_r_x - 50
## t = 3.5279, df = 1980, p-value = 0.0002142
## alternative hypothesis: true mean is greater than 0
## 95 percent confidence interval:
##  0.709947      Inf
## sample estimates:
## mean of x
##  1.330641
```

This confirms our analysis

We then look at the practical significance of our hypothesis test:

```
# Investigate Practical Significance
(mean(S_n_3$dem_age_r_x) - 50) / sd(S_n_3$dem_age_r_x)
```

```
## [1] 0.0792632
```

We have an effect size of 0.08, which means small practical significance.

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?

We first create a subset of the data to exclude non applicable survey answers and create a variable to identify Demcrat or Republican voters, similar to what's been done previously

```
# Create subset
S_n_4 <- mutate(subset(S_n, libcpre_self_n != 0 & libcpo_self_n != 0 & substr(pid_x,0,1) != '-' & substr
                       select = c(pid_x,libcpre_self_n,libcpo_self_n)),
                voter_cat = factor(ifelse(substr(pid_x,0,1) < 4, 'Democrat', 'Republican'))
              )
```

We then examine the differences between the libcpre_self_n and libcpo_self_n, representing voter's swing in political typology before and after election, for both Republican and Democrat voters

```
D_4_R <- subset(S_n_4, voter_cat == 'Republican')$libcpo_self_n - subset(S_n_4, voter_cat == 'Republican
D_4_D <- subset(S_n_4, voter_cat == 'Democrat')$libcpo_self_n - subset(S_n_4, voter_cat == 'Democrat')$l

# For Republican Voters
length(D_4_R)
```

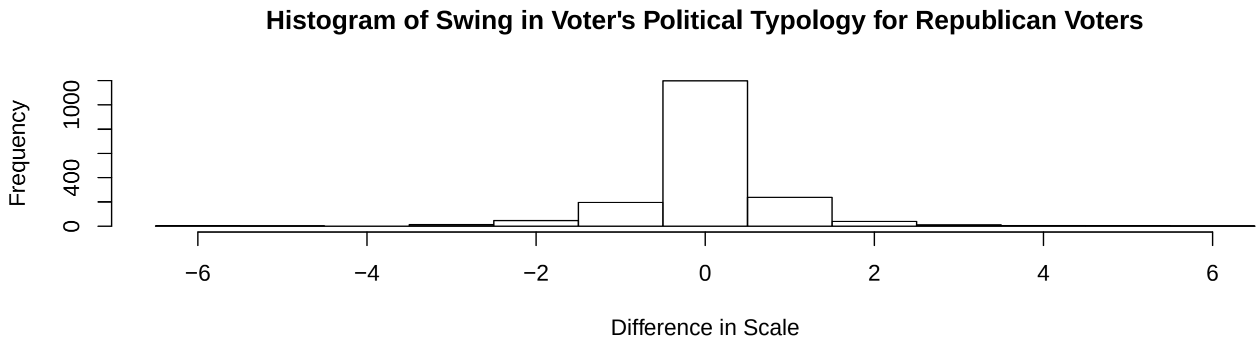
```
## [1] 1748
```

```
summary(D_4_R)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -6.00000  0.00000  0.00000  0.01888  0.00000  6.00000
```

```
hist(D_4_R, breaks = -6:7 - .5,
     main = "Histogram of Swing in Voter's Political Typology for Republican Voters",
```

```
     xlab = "Difference in Scale"
     )
```



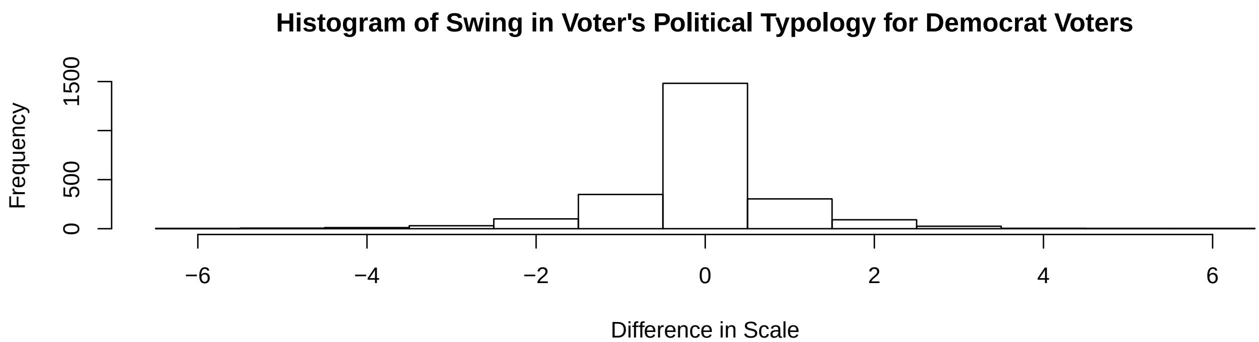

```
# For Democrat Voters
length(D_4_D)
```

```
## [1] 2410
```

```
summary(D_4_D)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -6.00000  0.00000  0.00000 -0.04772  0.00000  6.00000
```

```
hist(D_4_D, breaks = -6:7 - .5,
     main = "Histogram of Swing in Voter's Political Typology for Democrat Voters",
     xlab = "Difference in Scale"
     )
```



It seems like the differences between the libcpre_self_n and libcpo_self_n for both Republican and Democrat voters are symmetrically distributed with similar spread, allowing us to perform the Wilcoxon Rank Sum test.

Since our data is in ordinal scale, we will do the Wilcoxon Rank Sum test with a null hypothesis stating that the differences between the libcpre_self_n and libcpo_self_n for both Republican and Democrat voters are the same

We can first look at the means directly

```
c(mean(D_4_R), mean(D_4_D))
```

```
## [1]  0.01887872 -0.04771784
```

We then perform two-tailed Wilcoxon Rank Sum test

Null Hypothese = Republican voters weren't more likely to shift their political preferences right or left compared to Democratic voters during the 2012 election

Alternative Hypothese = Republican voters were more likely to shift their political preferences right or left compared to Democratic voters during the 2012 election

```r
wilcox.test(D_4_R, D_4_D)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  D_4_R and D_4_D
## W = 2189300, p-value = 0.01096
## alternative hypothesis: true location shift is not equal to 0
```

Notice that with the null hypothese being that the mean differences between the libcpre_self_n and libcpo_self_n for both Republican and Democrat voters are the same and an alternative hypothese saying that the mean differences between the libcpre_self_n and libcpo_self_n for both Republican and Democrat voters are not the same, we have a p-value of 0.01096. It shows strong statistical significance of our test statistic, rejecting the null hypothese at ~99% confidence level

In fact, since we have a p-value of 0.01096 for our two-tailed test, we know that we will get a p-value of 0.00548 if we were to do a one-tailed test with the same null hypothese but a different alternative hypothesis stating that the mean difference between the libcpre_self_n and libcpo_self_n for Republican is higher than that for Democrat voters

We can do a one-tailed test just to confirm:

Null Hypothese = Republican voters weren't more likely to shift their political preferences right or left compared to Democratic voters during the 2012 election

Alternative Hypothese = Republican voters were more likely to shift their political preferences right compared to Democratic voters during the 2012 election

```r
wilcox.test(D_4_R, D_4_D, alternative = "greater")
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  D_4_R and D_4_D
## W = 2189300, p-value = 0.005479
## alternative hypothesis: true location shift is greater than 0
```

Test results confirms our comments above

We then investigate the practical significance of our hypothesis test:

```r
# Investigate Practical Significance
cohen.d(D_4_R, D_4_D)
```

```
##
## Cohen's d
##
## d estimate: 0.0700727 (negligible)
## 95 percent confidence interval:
##         inf         sup
## 0.008460392 0.131685002
```

We see an effect size of 0.07, which shows small practical significance.

5. Right before the 2012 election, were women voters more or less liberal than men voters?

We first do an EDA on variables of interest:

```r
# Investigate data quality
table(S_n$profile_gender, S_n$libcpre_self_n)
```

```
##
##                       0   1   2   3   4   5   6   7
##   -1. Inapplicable 590  71 206 191 477 227 235  57
##   1. Male           11  66 199 223 632 317 434  77
##   2. Female         13  58 233 227 719 245 332  74
```

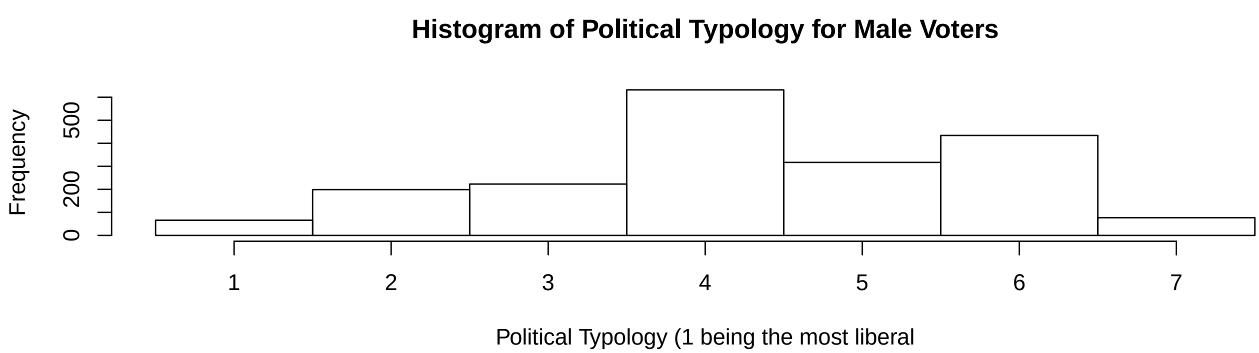
```r
# Create subset that includes records with applicable gender and libcpre_self_n response
S_n_5 <- mutate(subset(S_n, libcpre_self_n != 0 & substr(profile_gender,0,1) != "-",
                       select = c(libcpre_self_n, profile_gender)),
                gender = factor(ifelse(profile_gender == '1. Male', "Male", "Female"))
              )

# Examine distributions of libcpre_self_n for male and female voters

# For Male Voters
length(subset(S_n_5, gender == "Male")$libcpre_self_n)
```

```
## [1] 1948
```

```r
summary(subset(S_n_5, gender == "Male")$libcpre_self_n)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   3.000   4.000   4.306   6.000   7.000
```

```r
hist(subset(S_n_5, gender == "Male")$libcpre_self_n, breaks = 1:8 - 0.5,
     main = "Histogram of Political Typology for Male Voters",
     xlab = "Political Typology (1 being the most liberal")
```



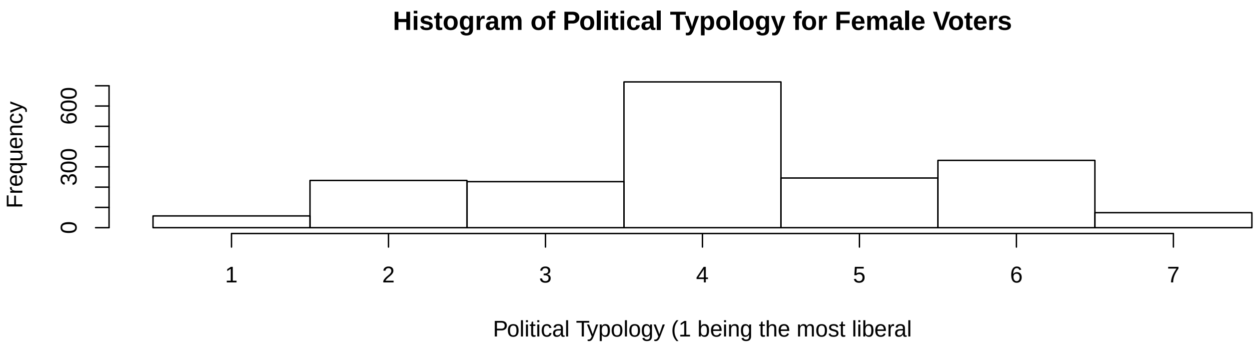

```r
# For Female Voters
length(subset(S_n_5, gender == "Female")$libcpre_self_n)
```

```
## [1] 1888
```

```r
summary(subset(S_n_5, gender == "Female")$libcpre_self_n)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.00    3.00    4.00    4.14    5.00    7.00
```

```r
hist(subset(S_n_5, gender == "Female")$libcpre_self_n, breaks = 1:8 - 0.5,
     main = "Histogram of Political Typology for Female Voters",
     xlab = "Political Typology (1 being the most liberal")
```



Although they don't seem to be normally distributed, the large sample sizes involes Central Limit Theorem and approximate the sampling distributions for the sample means to be normally distributed. With that, we can run a Wilcoxon Rank Sum test for the ordinal value libcpre_self_n

We first look at the means directly:

```
c(mean(subset(S_n_5, gender == "Female")$libcpre_self_n), mean(subset(S_n_5, gender == "Male")$libcpre_s
```

```
## [1] 4.139831 4.306468
```

We then perform a two-tailed Wilcoxon Rank Sum test:

Null Hypothesis: Right before the 2012 election, women voters were not more or less liberal than men voters
Alternative Hypothesis: Right before the 2012 election, women voters were more or less liberal than men voters

```
wilcox.test(libcpre_self_n ~ gender, data = S_n_5)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  libcpre_self_n by gender
## W = 1707700, p-value = 8.057e-05
## alternative hypothesis: true location shift is not equal to 0
```

Notice that with the null hypothese being that before the 2012 election, women and men were equally liberal and an alternative hypothesis stating that before the 2012 election, women voters and men voters weren't equally liberal, we have a p-value of 8.057e-05, which shows strong statistical significance of our test statistic, rejecting the null hypothese at $> 99\%$ confidence level

In fact, since we have a p-value of 8.057e-05 for our two-tailed test, we know that we will get a p-value of 4.028e-05 if we were to do a one-tailed test with the same null hypothese but a different alternative hypothesis stating that right before the 2012 election, women voters were more liberal than men voters

We can do an ont-tailed test just to confirm:

Null Hypothesis: Right before the 2012 election, women voters were not more or less liberal than men voters
Alternative Hypothesis: Right before the 2012 election, women voters were more liberal than men voters

```
wilcox.test(libcpre_self_n ~ gender, data = S_n_5, alternative = "less")
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  libcpre_self_n by gender
## W = 1707700, p-value = 4.028e-05
## alternative hypothesis: true location shift is less than 0
```

Test results confirmed our analysis

We then look at the practical significance of our hypothesis test

```
# Investigate Practical Significance
cohen.d(libcpre_self_n ~ gender, data = S_n_5)
```

```
##
## Cohen's d
##
## d estimate: 0.1149731 (negligible)
## 95 percent confidence interval:
##        inf        sup
## 0.05160251 0.17834361
```

We see an effect size of 0.11, which shows small practical significance. “‘

Conclusions

After performing hypothesis testings, whether parametric or non parametric, for the questions of interests. I 1) rejected null hypotheses that provides poor explanations of the data in favor of an alternative hypothesis that shows good evidence or 2) didn’t reject the null hypotheses that provided good explanation of the data.

Although we saw strong statistical significances for a few of the hypothesis tests, all of them had small practical significances. This is due to a large sample size that we have for our dataset. But please note that we are using a simplied version of the ANES survey data with the survey weight removed. The results of the analyses above are done based purely on the simplied sample and may not reflect actual situations for 2012 election.