Hypothesis Tests about 2012 election

James Nguyen

Summary

In this experiment, I used the data from the ANES 2012 election dataset to try to answer some questions about voters's demographics, their political preference, the difference between repulican voters and democratic voters. . . I use the confidence level of 95% in all the tests.

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

In this test, I design coding scheme to quantify level of liberalism and conservativeness as Extremely liberal = -3, Liberal: -2, Slightly liberal =-1, Slightly conservative=1, Conservative = 2, Extremely conservative= 3, Others = 0 So the higher the score is, the more conservative a person is. I define the variable as X in R to represent the conservativeness before the election as bellow

```
X <-S$libcpre_self

X<-ifelse((X== "1. Extremely liberal"), -3,
        ifelse((X=="2. Liberal"),-2,
            ifelse((X=="3. Slightly liberal"),-1,
                ifelse((X=="5. Slightly conservative"),1,
                    ifelse((X=="6. Conservative"),2,
                        ifelse((X=="7. Extremely conservative"),3,
                            0))))))
```
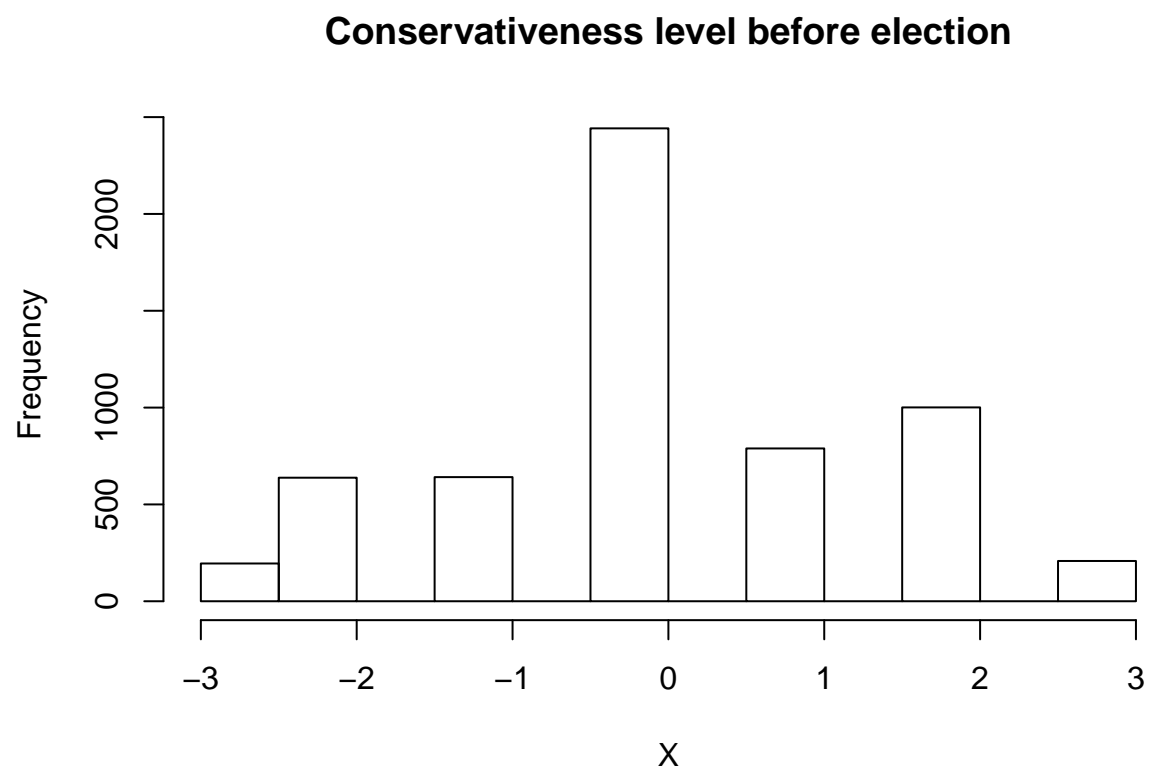
Similiarly, Y is defined for conservativeness after the election.

```
Y <-S$libcpo_self

Y<-ifelse((Y== "1. Extremely liberal"), -3,
        ifelse((Y=="2. Liberal"),-2,
            ifelse((Y=="3. Slightly liberal"),-1,
                ifelse((Y=="5. Slightly conservative"),1,
                    ifelse((Y=="6. Conservative"),2,
                        ifelse((Y=="7. Extremely conservative"),3,
                            0))))))
```
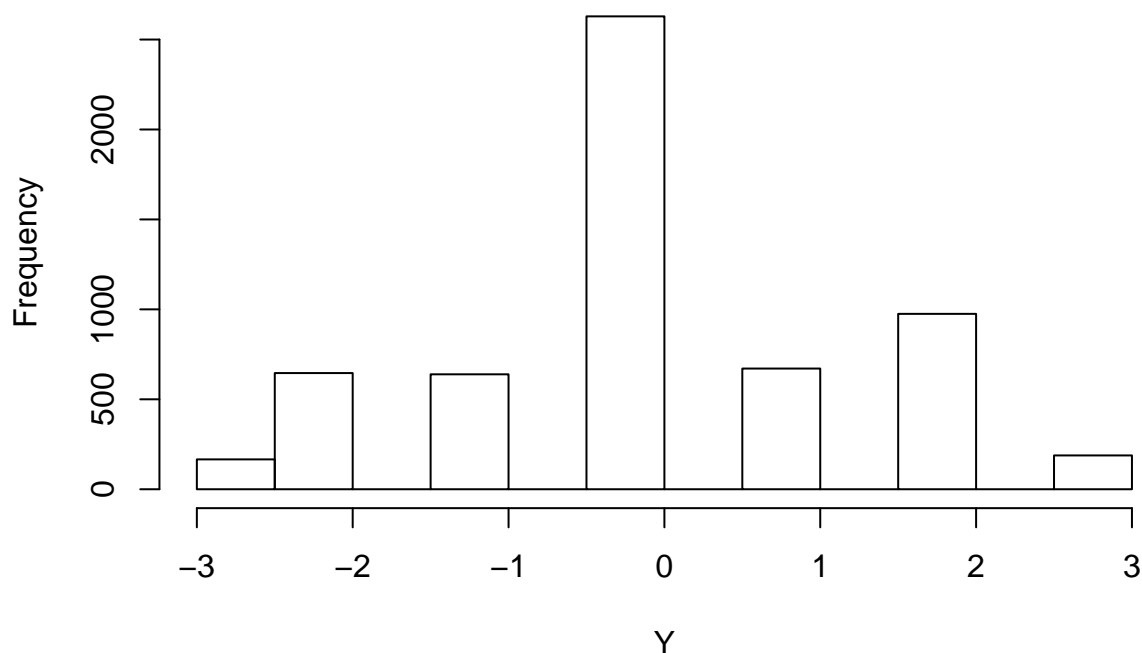
Let check the data distribution

Conservativeness level before election

The data is quite balanced between 2 sides.Similiarly I examine level of conservativeness after the election and also find that it's also quite balanced

Conservativeness level after election

Now we define the variable Z that indicate if a person became more liberal or not. If he/she was we give a score 1, if otherwise 0 is assigned.

```r
Z = ifelse(X-Y>0,1,0)
summary(Z)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0000  0.1879  0.0000  1.0000
```

```r
length(Z)
```

```
## [1] 5914
```

we see that in average 18.79% of voters changed toward being more liberal. Now we test to see if this average from the sample really means a shift of voters in general. This variable alone though not a normal distribibution but because the sample size is large so according to CLT, the mean should follow normal distribution. We now design a test to see if the mean really greater than 0.

```r
t.test(Z,mu=0, alternative = 'g')
```

```
##
##  One Sample t-test
##
## data:  Z
## t = 36.983, df = 5913, p-value < 2.2e-16
## alternative hypothesis: true mean is greater than 0
## 95 percent confidence interval:
##  0.1795028      Inf
## sample estimates:
```

```
## mean of x
## 0.1878593
```

With P value is really small ($< 2.2e-16$), we can reject H0 and conclude the test has statistical significance at confidence level 95% that Z mean is greater than 0. Practically speaking, 18.8% is far from zero and a good number to indicat that voters were more likely to change toward more liberal.

2. Were Republican voter older or younger, on the average, than Democratic voters in 2012?

Let first create two subsets of data, one represents democrats and one represent republicants

```
S_dem <-subset(S,(pid_x=="1. Strong Democrat"|pid_x=="2. Not very strong Democract"))

S_rep <-subset(S,(pid_x=="7. Strong Republican"|pid_x=="6. Not very strong Republican"))
```

Now we examine data for democrats in terms of their ages.

```
hist(S_dem$dem_age_r_x, main = "Democrats' age distribution")
```


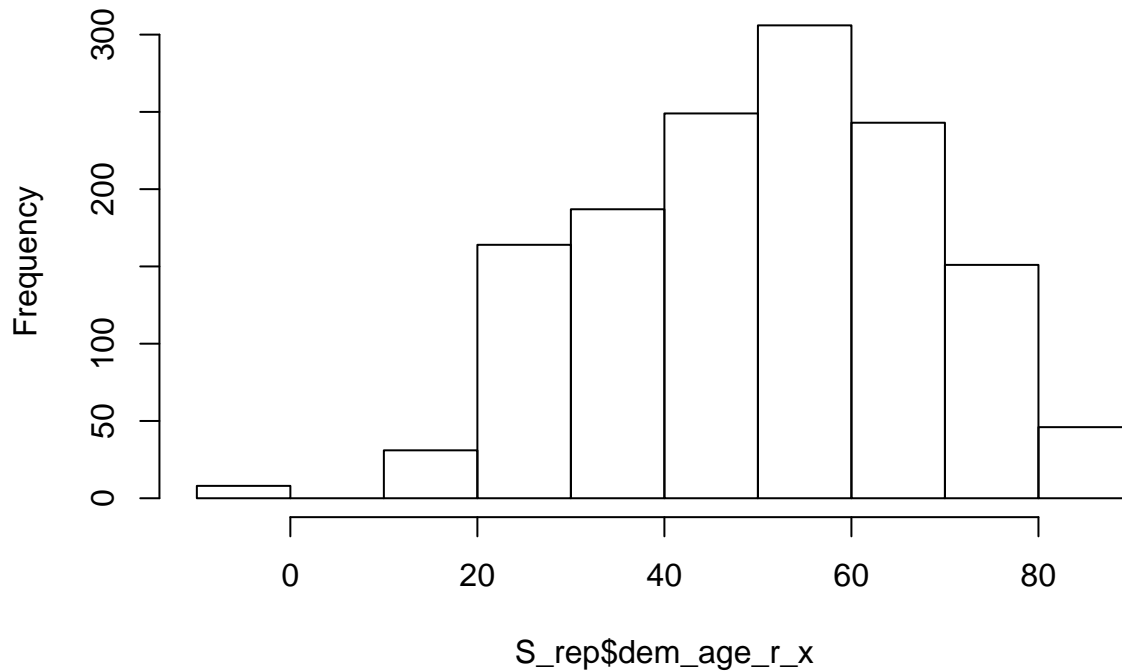**Democrats' age distribution**
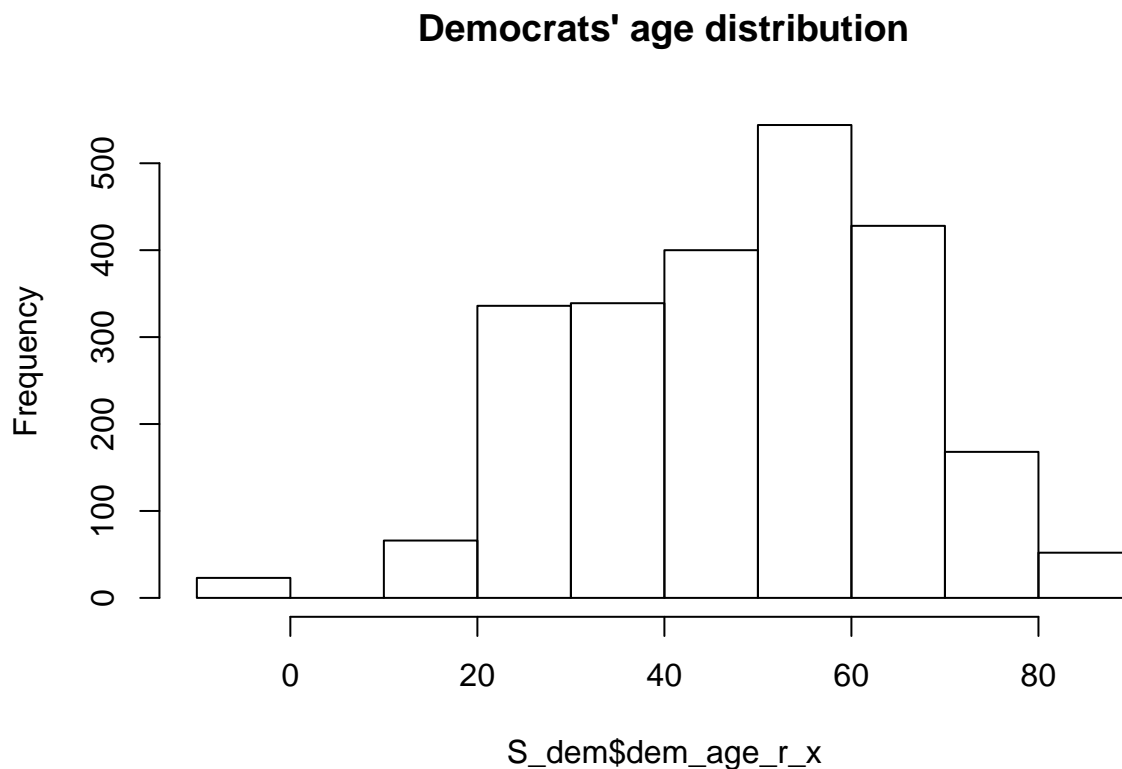
```
summary(S_dem$dem_age_r_x)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   -2.00   35.00   51.00   48.84   62.00   90.00
```

```
hist(S_rep$dem_age_r_x, main = "Republicans' age distribution")
```

Republicans' age distribution


```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   -2.00   38.00   52.00   51.06   63.00   90.00
```

They are both approximately normal. They have some very small portion with negative values which seems to be errors. However, as these are very small we can ignore and go ahead. From the summary, the democrat samples seems to have lower mean and median age, because of this, we will design a test with H0 to be mean of democrats to be older than that of republicans.

```
t.test(S_dem$dem_age_r_x, S_rep$dem_age_r_x, alternative = "l")
```

```
##
##  Welch Two Sample t-test
##
## data:  S_dem$dem_age_r_x and S_rep$dem_age_r_x
## t = -3.795, df = 2882.4, p-value = 7.535e-05
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -1.256097
## sample estimates:
## mean of x mean of y
##  48.84380  51.06137
```

The P value comes at very small compared to 0.05, and 48.2 age is visibly younger than 50.9 practically, we can conclude that democrat voters seems to be younger in average than republican voters.

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

From the above analysis of republican age distribution and the mean of 50.95639,we design a one tail test for the H0 which says republican voters were older than 51.

```
t.test(S_rep$dem_age_r_x, mu=51, alternative = "l")
```

```
##
##  One Sample t-test
##
## data:  S_rep$dem_age_r_x
## t = 0.13197, df = 1384, p-value = 0.5525
## alternative hypothesis: true mean is less than 51
## 95 percent confidence interval:
##      -Inf 51.82683
## sample estimates:
## mean of x
##  51.06137
```

The test bearly meet the statistical significance as p = 0.4552 which is close to 0.5 to reject the null hypothesis. So statistically, we can say republican voters are younger than 51. However, 51 and 50.956039 difference is tiny and does not have practical significance to come the same conclusion. In fact, two tailed test bellow fail to reject that the mean is equal 51. With 95% confidence interval, the range of mean age is from 50.19630 51.71648 which includes 51.

```
t.test(S_rep$dem_age_r_x, mu=51, alternative = "t")
```

```
##
##  One Sample t-test
##
## data:  S_rep$dem_age_r_x
## t = 0.13197, df = 1384, p-value = 0.895
## alternative hypothesis: true mean is not equal to 51
## 95 percent confidence interval:
##  50.14909 51.97366
## sample estimates:
## mean of x
##  51.06137
```

The reason for being statistically significant in the first test is the large sample size.

4. Were Republican voters more likely to shift their political preferences right or left, compared to Democratic voters during the 2012 election?

Let first build two datasets one for democrats, one for republicans and they have values that indicate if a voter shifted their preference.

```
S_dem_pref <- as.character(S_dem$libcpre_self) ==as.character(S_dem$libcpo_self)

S_rep_pref <- as.character(S_rep$libcpre_self) ==as.character(S_rep$libcpo_self)
```

Let examine data for democrats

```
summary(as.numeric(S_dem_pref))
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0000  0.4813  1.0000  1.0000
```

And data for republicans

```
summary(as.numeric(S_rep_pref))
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  1.0000  0.6289  1.0000  1.0000
```

As the data sample is large, we can assume that the mean (% of voters who shifted their preference) follow normal distribution. The democrat set has 47.89% voters shifted their preference, the republican set has 60.1% did the same. Because of this, let design a one tailed test with H0 to be democrat voters shifted more.

```
t.test(as.numeric(S_dem_pref), as.numeric(S_rep_pref), alternative = "l")
```

```
##
##  Welch Two Sample t-test
##
## data:  as.numeric(S_dem_pref) and as.numeric(S_rep_pref)
## t = -8.9038, df = 2979.1, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -0.120289
## sample estimates:
## mean of x mean of y
## 0.4813243 0.6288809
```

The test easily meets statistical significance requirement to reject H0 as p-value < 2.2e-16 <0.05 Practically, 47% vs. 60% is a substantial difference so we can safely say that on average republican voters were more likely to shift their political preference during the 2012 election more than deomcrat voters.

5. Were republican voters more likely to have colleage or higher education than democrat voters on average?

Let first build two datasets one for democrats, one for republicans and they have values that indicate if a voter shifted their preference.

```
S_dem_highedu <- S_dem$profile_educ =="4. Bachelor's degree or higher"

S_rep_highedu <- S_rep$profile_educ =="4. Bachelor's degree or higher"
```

Let examine data for democrats

```
summary(as.numeric(S_dem_highedu))
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0000  0.1969  0.0000  1.0000
```

And data for republicans

```
summary(as.numeric(S_rep_highedu))
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0000  0.2816  1.0000  1.0000
```

As the data sample is large, we can assume that the mean (% of voters who have colleage or higher education) follows normal distribution. The democrat set has 19.69% voters with colleage or higher education , the republican set has 28.16% of the same. Because of this, let design a one tailed test with H0 to be democrat voters have more colleage or higher education more.

```
t.test(as.numeric(S_dem_highedu), as.numeric(S_rep_highedu), alternative = "l")
```

```
##
##  Welch Two Sample t-test
##
## data:  as.numeric(S_dem_highedu) and as.numeric(S_rep_highedu)
## t = -5.7953, df = 2622.6, p-value = 3.817e-09
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##        -Inf -0.06061178
## sample estimates:
## mean of x mean of y
## 0.1969440 0.2815884
```

The test easily meets statistical significance requirement to reject H0 as p-value = 3.817e-09 <0.05 Practically, 19.6% vs. 28.1% is a good difference so we can safely say that on average republican voters were more likely to have colleage or higher education than democrat voters.