

Lab 1: Question 1

Jun Qian, Lucas Schroyer, Ryan Mitchell, Oliver Chang

Contents

Are Democratic voters older or younger than Republican voters in 2020?	2
Importance and Context	2
Description of Data	2
Most appropriate test	3
Test, results and interpretation	4
Test limitation	4

Are Democratic voters older or younger than Republican voters in 2020?

Importance and Context

Are Democratic voters older or younger than Republican voters in 2020?

The age profile is always one of the most important factors in forecasting and understanding American election outcomes. People in different generations are usually associated with different political ideologies. It has been widely believed that older American electorates are more likely to vote for Republicans. However, as we know, the American population is aging, is this stereotype still grounded on fact? We analyzed the 2020 pre-election data to confirm whether there is an age difference between Democratic voters and Republican voters.

Description of Data

To address the question, we downloaded the pre-election data from the 2020 American National Election Studies (ANES) site. We then extracted the following variables related to this question for further processing and exploration:

Voter: registered_to_vote_status, voted_early_in_gen_election, plan_to_vote_in_gen_election, Party: party_of_registration_str, party_of_registration, democrat_party_rating, republican_party_rating, Age: age

In this study, we defined “voters” as who registered to vote and either voted early in general election or plan to vote. Then, we looked into the registered party of the voters to determine the party affiliations of these voters.

However, we noticed that around 48% of the voters do not have applicable registered party information. Since it consists significant portion of the sample, we tried to infer their partisan leanings based on if they have a strong preference for one party. To determine such preference, we used the difference in party rating, a “feeling thermometer”. If one party is rated 50 higher than the other party (measured as in below chart for democratic voters), we assign a party to that voter. Figure 1 shows the distribution of the party preference.

This party assignment methodology assigned party information to 2105 data points out of 3929 respondents, bringing the total sample size of eligible voters with clear party tendency to 7624.

Democrat – Republican Party Ratings

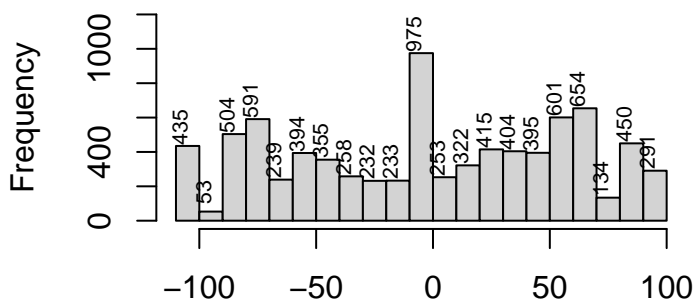


Figure 1: If a survey response's $\text{abs}(\text{Democratic Rating} - \text{Republican Rating}) > 50$ and they are listed as NA for their party affiliation, remap that respondent to either a Democrat/Republican from NA.

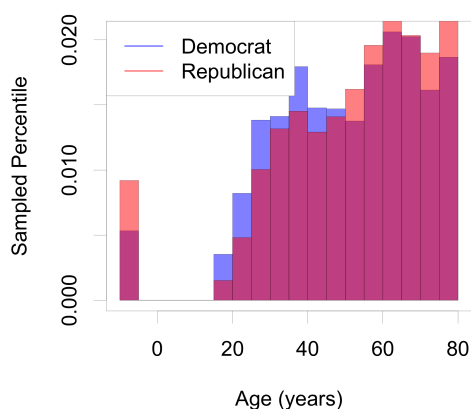
Then, we examined the age of the voters. The age is by year expect for all the people above 80. They are bucketed into the “80+” bin. To simplify the calculation, we assumed 80+ bin just as age of 80. More details of the impacts will be discussed in the test limitation section. In addition, people refused to answer this question is assigned as -9.

Figure 2 shows there are more Democratic voters(purple) in all the age buckets below 50 and more Republican(orange) voters in age buckets above 50 as well as “refuse to answer” bucket.

Since there is little information available to impute the age of the “refuse to answer” group. We excluded them from the following analysis and sample for statistics tests. Table 1 shows the mean and median of Democratic voters are smaller than Republican voters. We also took the 80+bin out and the trimmed mean without 80+ bin follow the same pattern.

Intuitively, we expect the democratic voters cohort are younger than the republic voter group. However, is the difference statistically and practically significant?

```
## pdf
## 2
```



Political Party	Median Voter Age	Mean Voter Age
Democrat	55	53
Republican	58	55
Democrat_filtered	53	51
Republican_filtered	56	54

Figure 2: Age of Survey Respondents by Political Party and Table 1: Age by Party for the Unfiltered Dataset and for with the 80+ Age Bin Removed

Most appropriate test

To determine whether there is a age difference between democratic voters and republic voters, we compared the mean of the age of two groups using the unpaired welch t two sample test. The reasons we selected this test are as follows:

- (1) The two sample group are unrelated from each other.
- (2) The data is interval except the 80+ bin. For the 80+ bin, as in the mean/median calculation, we just assumed it is 80 for testing purpose. The effect of actual mean deviation from 80 and different distributions for the two groups will be discussed in the test limitation section.
- (3) The data extracted from the population is I.I.D.. Clustering effect could skew the data during sampling process. The procedures how the survey was done are likely to have independent responses. The clustering effect should be minimal.
- (4) The sample size is very large. Without the “refuse to answer” respondents, the sizes of Democratic voters and Republic voters are 2701 and 2260 respectively. As a result, although the samples are not perfectly normally distributed, we decided the large sample size overweight the skewness.
- (5) The sample variance actually is unknown due to the 80+ bin. Under such circumstance, the Welch t test is generally regarded as more appropriate t test.

Therefore, we decided to use the Welch t two sample test. The Null Hypothesis for the test was the means of the age of two voters groups are equal. We used a two-tailed test in this study because we didn't want to assume one group is older or younger than the other group beforehand. After that, we also calculated the effect size to analyze the practical significance.

Test, results and interpretation

According to the results, we can not simply assume the age of Democratic voters is different from the Republican voters.

From the first glance, the t-test results indicated that the Null Hypothesis was rejected at 95% confidence level. The reported p-value was 0, close to zero. So statistically, we can conclude the mean age of two groups are different. With a t-value of -5.89, we anticipated the mean of Democratic voters is smaller than Republican voters. This also seems to be consistent with our observations earlier that the Democratic voters are younger.

However, the effect size of the test is $c(\text{Cohen's } d = -0.17)$. This meant the practical significance in the difference in the age is fairly small, almost negligible. The possible reason for the conflicting results might be our large sample size. As the sample size increases, the hypothesis test possesses increasing statistic power to detect a small difference in two sample groups. But actually the difference in practice is fairly minor compared to the age dispersion of the sample.

Test limitation

In our analysis, we simply assume the age "80+" bin as 80 years old since there is no way for us to know the exact distribution of data in this bin. However, this bin consists 5.04% of the Democratic voters' group and 5.81% of Republican voters' group. Different age patterns in this bin can vary the result significantly. For example, if the distributions of two groups are heavily skewed toward different directions, obviously assuming it as 80 can distort the test results. If we can get the birth year of the respondents, it will largely reduce the uncertainty.