

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 will analyze the age difference between Democratic voters and Republican voters in 2020 election.

Description of Data

To address the question stated earlier, 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:

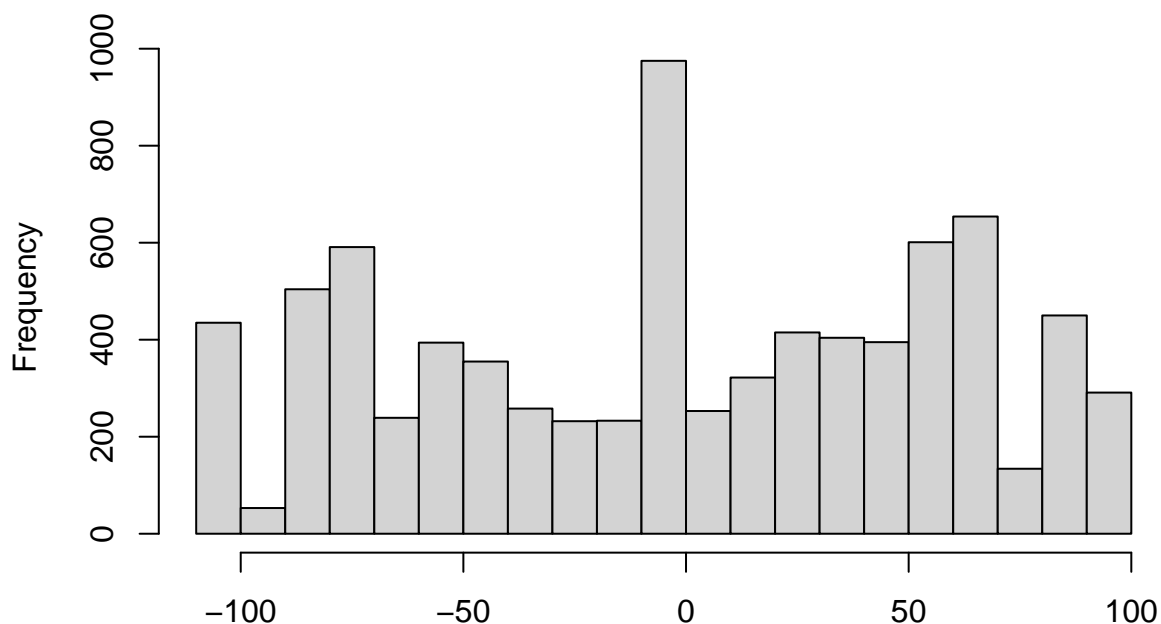
registered_to_vote_status, voted_early_in_gen_election, plan_to_vote_in_gen_election, party_of_registration_str, party_of_registration, democrat_party_rating, republican_party_rating, age

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

However, we notice 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 leaning if they have a strong preference for one party. To determine the preference, we used the difference of party rating. If one party is rated 50 higher than the other party, we assign a party to the voter. This party assignment methodology assigned 3929 data points with a specific party.

Eventually, we have 7624 sample data, 2775 for Democratic voters group and 2369 for Republican voters group. The large sample size will help in choosing the statistics test we will perform.

Dem Rating – Rep Rating

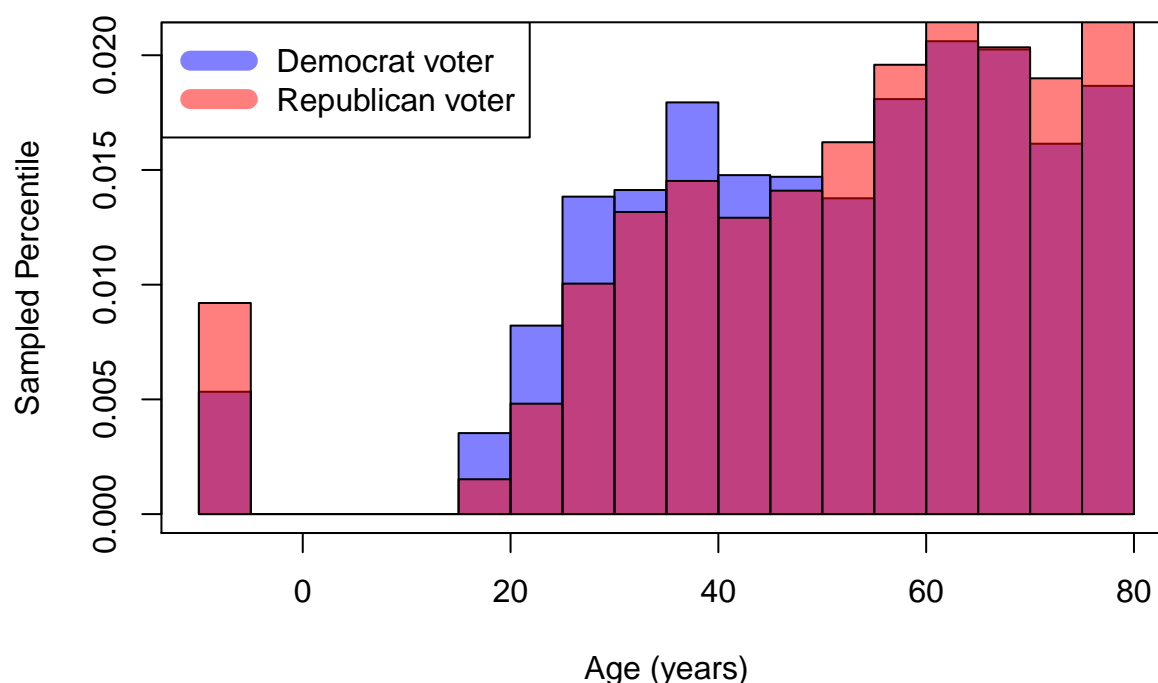


The age of the voters is in years expect all the people above 80 are bucketed into the “80+” bin. From the plot below, there are more Democratic voters in all the age buckets below 50 and more Republican voters in age buckets above 50 as well as “refuse to answer” bucket. Also the mean/median of Democratic voters with/without “80+” bin are smaller than Republic. 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.

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

(to add a table of mean/median here)

Age Distribution of Survey Respondents (Dem v.s. Rep)



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 groups 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. In addition, the weights that accounted for the population size of each sampling location were not applied in this study for simplicity reasons.
- (4) As described above, the sample size is very large. Although the samples are not perfectly normally distributed, we decided the large sample size overweight the skewness and t test is appropriate.
- (5) The sample variance actually is unknown due to the 80+ bin.

Thus, 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.

Test, results and interpretation

The test results indicated that at 95% confidence level, the Null Hypothesis was rejected. The reported p-value was 0 and t-value -5.89. Therefore, statistically, we anticipated the Democratic voters are younger than Republican voters. This seems to be consistent with our observation earlier.

However, the effect size of the test is merely 0.17, which meant the difference in the age is negligible. . . . (to add)

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 x of Democratic voters and x of republic. 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. To solve this issue, we can further look into the distribution of 2016 election data for all the voters above 76. This might be a good methodology to simulate the 2020 scenario.