Lab 1: Difficulty in Voting in the 2020 Election

Ria Mahajan, Shane Kramer, Viswanathan Thiagarajan

Contents

1. Who experienced more difficulty voting in the 2020 election - Democrats or Republicans?

1.1 Importance and Context

Did Democratic voters or Republican voters experienced more difficulty voting in the 2020 election?

The major U.S. political parties (Democratic and Republican) bitterly fought the 2020 elections. We have never seen such a level of acrimony between the parties in the prior elections. Extreme sides of both parties were fully active with their rhetoric during pre and post-election periods, with each side alleging major issues during the elections. Republicans alleged voting irregularities and fraud in the 2020 elections, while the Democrats alleged voter suppression and hurdles to vote in the 2020 election.

COVID-19 pandemic changed the norms and the way people vote, with more and more states opting to expand the mail-in voting and some states closing some voting locations. These changes required the states and counties to serve a higher number of requests for absentee ballots which could have caused some difficulty to voters in obtaining them. In addition, voters could have faced issues with registration, reaching polling stations, concerns about the identification card, confusion about the ballot, long wait times, work schedule, weather-related, other problems, or none at all. It will be helpful to understand if Democrats or Republicans faced more difficulty voting in the 2020 elections so that issues can be highlighted and addressed by the respective administrations in the future.

1.2 Description of Data

We will address the "who faced more difficulty" question using the data from the 2020 American National Election Studies (ANES). This study surveyed a cross-section of U.S. eligible voters in two waves (pre-election and post-election). ANES collects data from its surveys on voting, political participation, and public opinion to serve the research interests of social scientists, students, policymakers, and others.

This dataset is a time series study, and there are a total of 8280 samples in it from eligible voters who were the respondents. To be eligible to participate in the survey, a respondent had to reside at the sampled address where mail is delivered and be a U.S. citizen age 18 or older at the time of recruitment. These were considered eligible voters. As you can see from the below plot, the sample has Seven different values for voter party affiliation. We are only interested in respondents from the two major political parties in the U.S. (Democratic and Republican) to answer our question.

Voter Affiliation
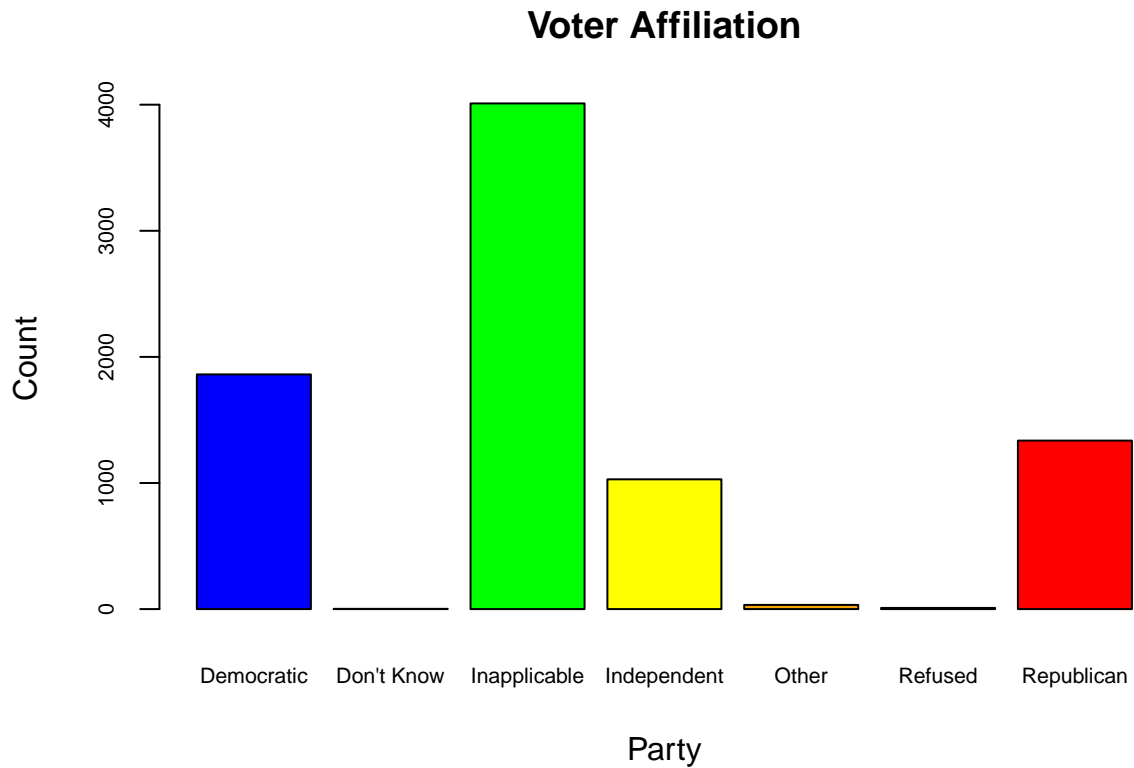
A Democrat is a person registered with the Democratic party, and this is a person who believes in the political or social equality of all people. A Republican is a person registered with the Republican party, and this is a person who favors or supports a republican form of government. The dataset was further subset to include only Democrat and Republican voters using the "Party of Registration" column, which reduced the sample size to 3197. The second column of interest to our analysis is "How difficult was it for the respondent to vote."

| Difficulty	Value
Refused	-9
No post-election data	-7
No post-election interview	-6
Interview breakoff	-5
Inapplicable	-1
Not difficult at all	1
A little difficult	2
Moderately difficult	3
Very difficult	4
Extremely difficult	5

Figure 1: Difficulty Scale.

3

The difficulty scale column has Ten different unique values, as shown in Figure 1. Since we are interested only in the valid values for the difficulty, we subset the sample by considering only the values between 1 and 5. This filter resulted in 2709 samples.

Our final dataset for analysis has 2709 samples as seen in figure 2. There are 1587 Democratic and 1122 Republican voters. We plot a side-by-side bar chart of the count of Democratic and Republican voters on the y-axis and the difficulty on an ordinal scale on the x-axis. The plot shows a similar distribution of Democratic and Republican voters with a higher number of Democratic voters responding that they faced difficulty in voting when compared to the Republican voters in the sample.

Below is the sample size audit table for the analysis.

Sample Size	Notes
8280	Initial sample size
3197	Subset only with Democratic and Republican voters
2709	Filter the values of 1 to 5 in difficulty scale (final dataset)
1587	Total Democratic voters in the final dataset
1122	Total Republican voters in the final dataset

Figure 2: Sample Size Audit.
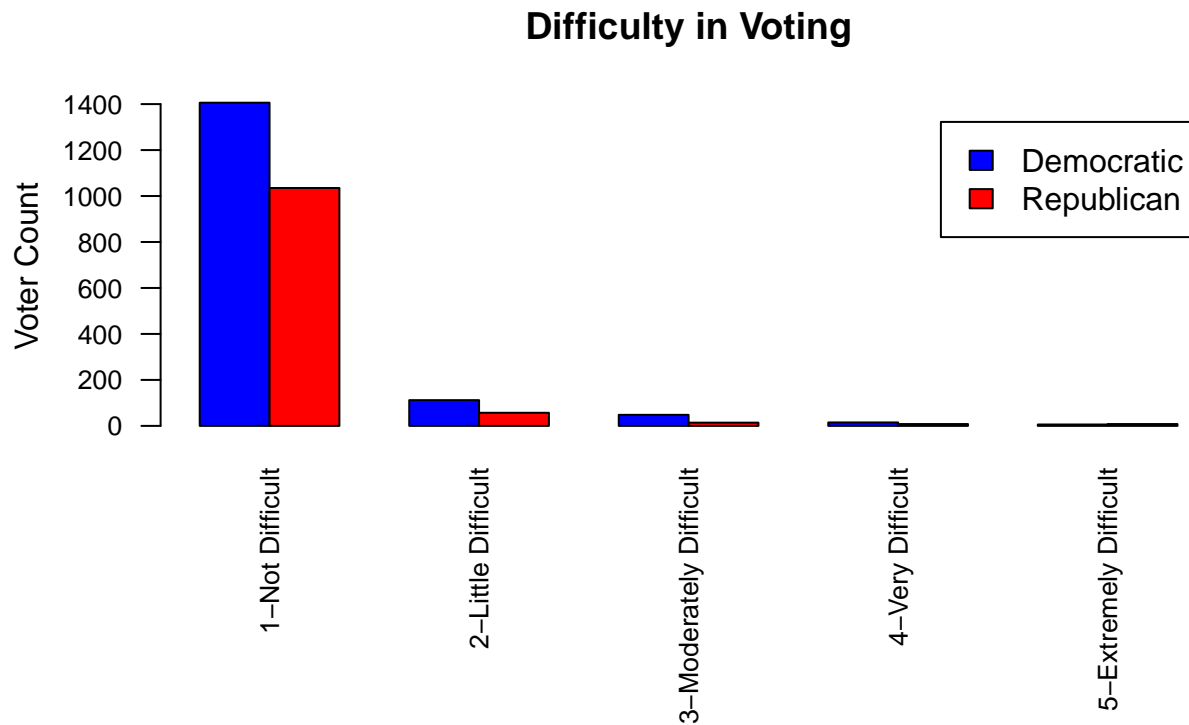
1.3 Hypothesis

Null Hypothesis H0: The probability that an observation from a Democratic sample (Dem) is larger than an observation from a Republican sample (Rep) is the same as the probability that an observation from a Republican sample is larger than an observation from a Democratic sample. In this test, the observation is the ordinal value of the difficulty in voting in each sample.

H0: P(Difficulty in Voting of Dem>Difficulty in Voting of Rep) = P(Difficulty in Voting of Rep>Difficulty in Voting of Dem).

Alternative Hypothesis Ha: The probability that an observation from a Democratic sample (Dem) is larger than an observation from a Republican sample (Rep) is not the same as the probability that an observation from a Republican sample is larger than an observation from a Democratic sample.

Ha: P(Difficulty in Voting of Dem>Difficulty in Voting of Rep) != P(Difficulty in Voting of Rep>Difficulty in Voting of Dem).

If the p-value is less than 0.05 we will reject the null hypothesis.

1.4 Most appropriate test

The grouping variable for the test of comparisons is party affiliation (Democratic or Republican). The test is an unpaired test as we compare the difficulty level between the groups of Democratic and Republican voters. Since the response variable indicating the difficulty level of the voters is ordinal data, a non-parametric test is appropriate. We plan to use the non-parametric Wilcoxon Rank Sum test in the analysis.

We assume that these are random samples from the population. Another assumption is that the samples are independent within the group, and there is mutual independence between groups. Since ANES claims that the survey respondents were selected based on a random selection of addresses from over 231 million records, the assumption of random sample and independence will hold. We also assume that the distribution of the two groups is similar in shape. We can say that this assumption is true from the above plot which shows the difficulty scale and counts of the two groups.

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  Dem and Rep
## W = 922827, p-value = 0.001738
## alternative hypothesis: true location shift is not equal to 0

##
##  Wilcoxon rank sum test with continuity correction
##
## data:  Rep and Dem
## W = 857787, p-value = 0.001738
## alternative hypothesis: true location shift is not equal to 0
```

1.5 Test, results and interpretation

The results above indicate a p-value of 0.0017; hence we reject the null hypothesis. This test indicates that there is a statistical significance that the probability that an observation from a Democratic sample (Dem) is larger than an observation from a Republican sample (Rep) is not the same as the probability that an observation from a Republican sample is larger than an observation from a Democratic sample.

Now we need to understand if democrats faced more difficulty than republicans or if Republicans faced more difficulty. To determine this, we use the W value from the test in both directions (Comparison of Democratic scores to Republican scores vs. Republican Scores to Democratic scores). W1 = 922827 when scores from Democratic samples are compared to scores from Republican samples. W2 = 857787 when scores from

Republican samples are compared to scores from Democratic samples. The Democratic sample size n1 is 1587. The Republican sample size n2 is 1122. The total number of pairs of observations is n1*n2.

The probability of a Democrat with a higher score is

$$W1/n1 * n2 = 0.52$$

The probability of a Republican with a higher score is

$$W2/n1 * n2 = 0.48$$

Hence, we can find evidence from the sample data that there is a statistical significance of Democratic voters facing more difficulty than Republican voters. The difference in probability from the above analysis is 0.04.

We can also validate the proportion of the samples who faced difficulty within the group. The proportion of Democrats who faced difficulty (scale value > 1) within the group is 181/1587 i.e. 0.114. The proportion of Republicans who faced difficulty (scale value > 1) within the group is 87/1122 i.e. 0.077. Even in this case, the difference in proportions is approximately 0.037.

From the current analysis, we cannot say that the difficulty did discourage voters or if the respondents failed to cast a vote due to the difficulty. We have less or no information from the sample to determine if the difficulties impacted casting votes and hence the election results. As the difference is so close, we cannot tell if there is enough practical significance in the results that it might have influenced the election results or caused other repercussions. Further analysis with additional data is required to answer these questions.