Question 2 - Did Democratic voters or Republican voters report experiencing more difficulty voting in the 2020 election?

Matt Pribadi, MaKenzie Muller, Mon Young

Contents

1 Importance and Context

In the United States, one of the basic rights of a citizen is the right to vote. From representatives of congress, to the president, voters have the power to choose their government. The American National Election Studies (ANES) conducts surveys of voters in the United States, with their main survey occurring right before and after each presidential election, which occurs every four years.A premise of Democracy is the a fair and equal opportunity for all citizens to vote. The statement in question is **Did Democratic voters or Republican voters report experiencing more difficulty voting in the 2020 election**? This question is significant because skewed difficulty in one population of voters versus another could result in a biased election result, which was heavily contested topic in the 2020 presidential election.

The 2020 ANES survey used a contactless, mixed-mode design in response to the COVID-19 pandemic. This included online-surveys, live video interviews, and telephone interviews. The 2020 sample group consisted of respondents that participated in the 2016 ANES Time Series Study as well a freshly drawn sample depending on the survey mode. The target population was a subset of the 231 milion U.S. citizens aged 18 or older living within the United States. Two surveys were taken: pre-election and post-election; with 90% of respondents who filled out the pre-election survey filling out the post-election survey.

2 Description of Data

Utilizing the 2020 ANES study with combined Pre-Election and Post-Election data, we started off with a sample set of 8280 respondents. Using information provided from the User Guide and CodeBook document, we narrowed down our dataset of 1381 variables to the those of interest for exploring the question above:

- **V202064: POST: PARTY OF REGISTRATION**
- **V201018 - PRE: PARTY OF REGISTRATION**
- **V202066 - POST: DID R VOTE IN NOVEMBER 2020 ELECTION**
- **V202119 - POST: HOW DIFFICULT WAS IT FOR R TO VOTE**

One part of the question is to determine whether or not Democrats or Republicans had more difficulty voting. Thus it was required to utilize **V201018** and **V202066**. These two datasets provided the party affiliation of the respondent with **1 - Democratic Party**, **2-Republican Party**, and anything less than 0 being invalid or incomplete data. After cleaning up the data and aggregating the data to only have respondents with the Democratic and Republican parties, our sample data set dropped to 3271 from the original 8280 total respondents.

The other main data point in question was **V202119** to determine how difficult it was for a respondent to vote. We could have aggregated data from **V202120a - V202120k**, however the data was unclear in the sense that the responses were either "mentioned problem or not". This was not a good indication of having a problem voting or not. Thus we determined that **V202119** is a good aggregate of how difficult was for a respondent to vote. Again, the data was filtered for values greater than 0, to ensure only valid survey data was used. Difficulty was ranked from a Likert scale like choice from **1 (least difficulty voting)** to **5 (most difficulty voting)**. Filtering this data set dropped our sample size to 2757 from 3271 total respondents.

In understanding difficulty of the voting process, we need to ensure the respondent actually voted. Here we used **V202066** to determine if the respondent voted. Again, the data was filtered for values **greater than 0**, to ensure only valid survey data was used. This survey gave the respondents four choices:

1. I did not vote (in the election this November)
2. I thought about voting this time, but didn't
3. I usually vote, but didn't this time
4. I am sure I voted

From the wording of these response choices, choices 1,2, and 3 are indications that the respondent did not vote. Therefore we took data points only from those who had a response of 4 for this survey question. Filtering this data set dropped our sample size to 2598 from 2757 total respondents.

Figure 1 shows a distribution of the difficulty responses when separated by party affiliation. The majority of voters expressed little to no difficulty voting, but the distribution looked very similar between the two party affiliations.

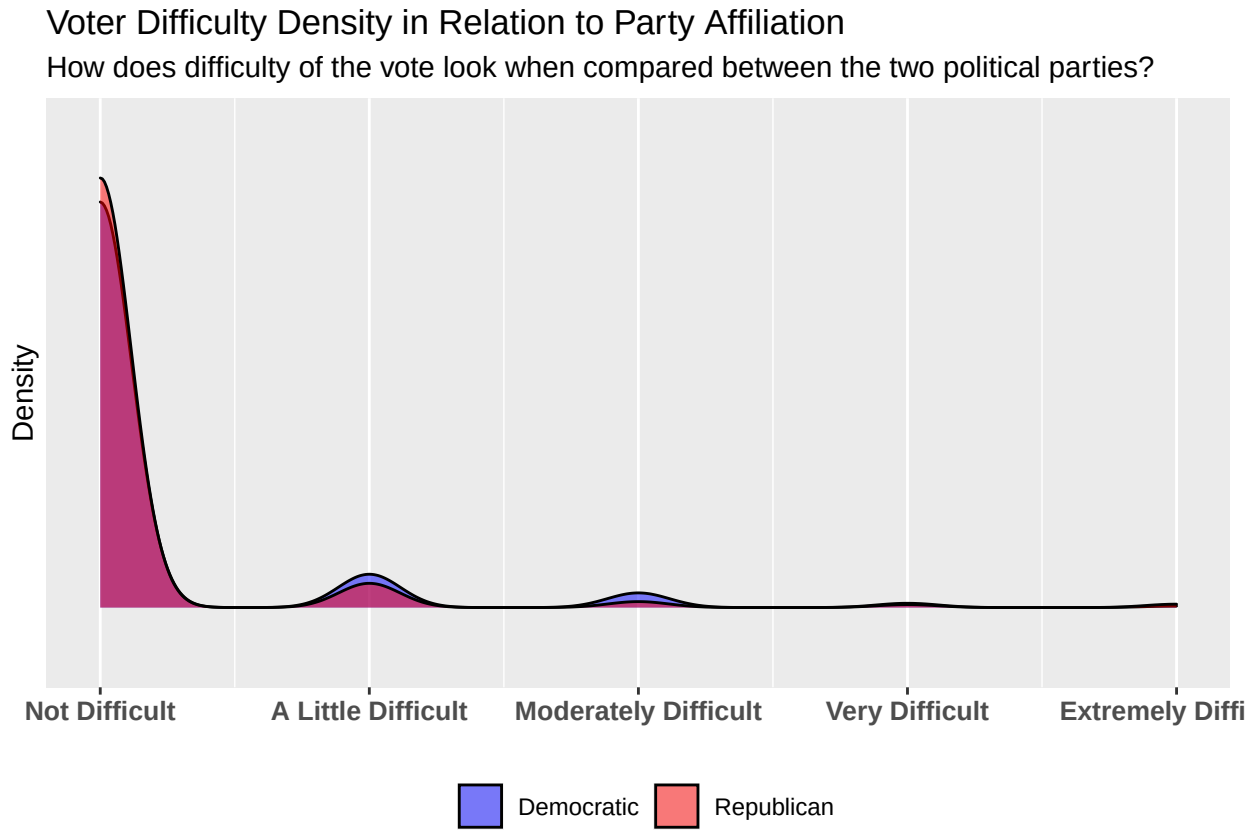

Figure 1: (#fig:graph Voter)Distribution of Difficulty Voting by Party Affiliation

3 Sample Testing

3.1 Most appropriate test

As briefly mentioned above, the data around difficulty is a Likert variable and is not metric. Therefore the base assumptions for all parameteric tests are violated. In addition, between the Democratic and Republican respondents, we can assume that these data sets are unpaired, though they could be taken from similar communities or households which would not make the data independent. The data sets are also not identical in sample size. Therefore based on these properties, we determined that the Wilcoxon Ranked Sum Test would be sufficient for this analysis. This test relies on the following assumptions to be true:

1. Ordinal Scale - this is met by utilizing the difficulty levels in terms of the Likert scale
2. i.i.d data - this assumes that all individual data points are drawn independent from the other data points within the same distribution.

3.2 Test and results

To answer this question, we begin with the two hypothesis:

- $H_o : P(Difficulty|Democrat) = P(Difficulty|Republican)$; There is no difference in the median of difficulty in voting between Democrats and Republicans
- $H_a : P(Difficulty|Democrat) \neq P(Difficulty|Republican)$; There is a difference in the median of difficulty in voting between Democrats and Republicans

Utilizing the Wilcoxon rank-sum test, we checked to see if a Democratic voter had more difficulty voting than a Republican voter. We ran three Wilcoxon rank-sum tests: a two-tailed and one-tailed tests for whether or not the the Democratic group had more or less difficulty than the Republican voter.

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  Difficulty by Voter
## W = 855022, p-value = 0.001207
## alternative hypothesis: true location shift is not equal to 0
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  Difficulty by Voter
## W = 855022, p-value = 0.0006035
## alternative hypothesis: true location shift is greater than 0
```

The result of the Wilcoxon tests shows a p-value of 0.0012 for the two-tailed test, 0.0006 for the one-tailed test (for Democrats having more difficulty voting than Republicans), and 0.9994 for the one-tailed test (for Democrats having less difficulty voting than Republicans). Based on this result and this small sample of filtered data, there is statistical significance to reject the null hypothesis that there is no difference in the medians of the two difficulties between the voters of each party. The data also shows to reject the null hypothesis that Democratic voters had less difficulty than Republican voters.

3.3 Practical Results

For a practical significance, we test out correlation methods to determine to summarize the strength of the bivariate relationship between a voter's party affiliation and difficulty voting. Below, we ran the Spearman Rho Correlation test due to the ordinal, non-normal nature of the *Difficulty* data.

```
cor.test(as.numeric(filter_difficulty$Voter), filter_difficulty$Difficulty,
         method = "spearman", exact = FALSE)
```

Running the correlation between difficulty voting and party affiliation showed a p-value of 0.0012, which is below the alpha of 0.05. With slight statistical significance, we can reject the null hypothesis that there is no correlation between the two variables. With the Rho value of -0.06, we can also conclude that there is a small strength of association between a voter's party affiliation and their difficulty voting.

Therefore, from a practical significance there is evidence that the difficulty in the voting process between voters of the Democratic Party and Republican Party are not equal. Overall, with **Democratic** voters, there is a slight correlation with with higher voting difficulty and this is reflected in the conclusion stated above that indicates that the median between the two groups is not equal. Limitations of this data include the fact that difficulty voting was based on only one survey question. To further pursue this conclusion, more data and analysis of what exact difficulty these voters were experiencing will be required.