Part Two

Steven Sung, Adam Weintraut, Lawrence Jiang, Tanmay Mahapatra

2/27/2022

Contents

1 Introduction

2020 was the year that COVID-19 forced many people and businesses to adapt to the new social norms of working from home, wearing masks, and isolating themselves to protect others. It was also the year when elections began. ANES aimed to gauge information about voters using a sequential mixed-mode design. In response to the challenges within the pandemic, data was gathered through self-administered online surveys, live video interviews, and telephone interviews.

John Richard Petrocik stated in his article *Measuring party support: Leaners are not independents* that in recent years, many middle class Americans self-identified as independents actually "acknowledge a party preference". "A bit of probing indicates that independence is more a matter of self-presentation than an accurate statement about our approach to elections'', giving rise to the term leaners.

By using the definition of leaners and preferential parties, this study aims to determine whether democratic or republican voters experienced more difficulty voting in the 2020 elections. Our research question serves as a follow-up to Petrocik's study, utilizing leaners and partisans within the two-party system to cross-examine political party's behaviors based on today's standards. Additionally, if we found that there was a difference in voter difficulty based on party, then one presidential candidate may have a higher chance of winning based on votes.

We will conduct a hypothesis test of Democrats and Republicans based on data from *ANES 2020 Time Series Study Full Release*. The study provides information on a participant's party affiliation and difficulty rating of the voting method which we will use to answer our question. Furthermore, we will analyze and interpret our findings through visual diagrams of our datasets in R.

2 Data

Our analysis leverages data from this dataset which includes 8280 individuals. In our research, we define voters as an entity/row and difficulty voting as the response to the key voting difficulty question in the ANES study. We also associate leaners with question "Do you think of yourself as closer to the Republican Party or to the Democratic Party?" because it aligns with the principles of a leaner. We only included pre-election questions since it would align better with V20230.

The terms republicans and democrats are defined as the following:

1. If a voter is pre-registered as a Republican, then the voter is mapped to be a Republican
2. If a voter is pre-registered as a Democrat, then the voter is mapped to be a Democrat.
3. If a voter is pre-registered as an Independent but leans towards Republican or Democrat, then the voter is mapped to Republican or Democrat respectively.
4. If a voter is pre-registered as one party and registered as the opposing party in post registration, then that voter is excluded.

In an ideal world, the best possible method of measuring the concept of party identification is going to each voter and asking for which party they affiliate with, but it would also be extremely timely because going to every house in the US to ask questions will take a lot of time. We would have to assume that they would give us an answer with integrity.

Therefore, the best available method to measure party affiliation is to use surveys to define how strongly they support a specific party. We address this by using the columns for party registration and no party identification within the dataset.

We have not made any changes to the integrity of the dataset such as filling in missing data or changing individual cells.

3 Methodology

Within our data wrangling method, we limited voting difficulty from 1 to 5 and removed all other values to reduce complications of addressing unanswered or invalid answers.

Our data wrangling code can be seen in our Rmd file.

An assumption we made about the data is that the sample is I.I.D. According to the ANES study, addresses were randomly assigned out of the cross-section of the 231 million US citizens of legal age, thus we can assume the identical property of I.I.D. We can also say that voters are independent because they can be either democrat or republican but not both, and one person's vote will not influence another person's vote. We also assumed that the data is ordinal because each column is taken from a survey of set numbers like a Likert scale which is also ordinal. The distance between each numeric option of the survey does not represent the same metric distance. It is also challenging to quantify voting difficulty equally among voters. Some limitations of the data is that we only considered a person's affiliation with a party under certain columns, hence we are not utilizing all of the data to properly determine one's party.

Therefore, the hypothesis test we conducted is the Wilcoxon Rank Sum test. We also ran two different versions of the text, a two-sided test and a one-sided test. Our null hypothesis for both tests states that Democrats and Republicans experienced the same amount of difficulty in voting for the 2020 election. The alternative hypothesis for our two-sided test is that there is a difference of voting difficulty between Democrats and Republicans. The alternative hypothesis for our one-sided test is that Republicans have less difficulty voting than Democrats. We thought our tests were most suitable because it allows us to not only understand whether or not the two groups are different but also the magnitude of difference.

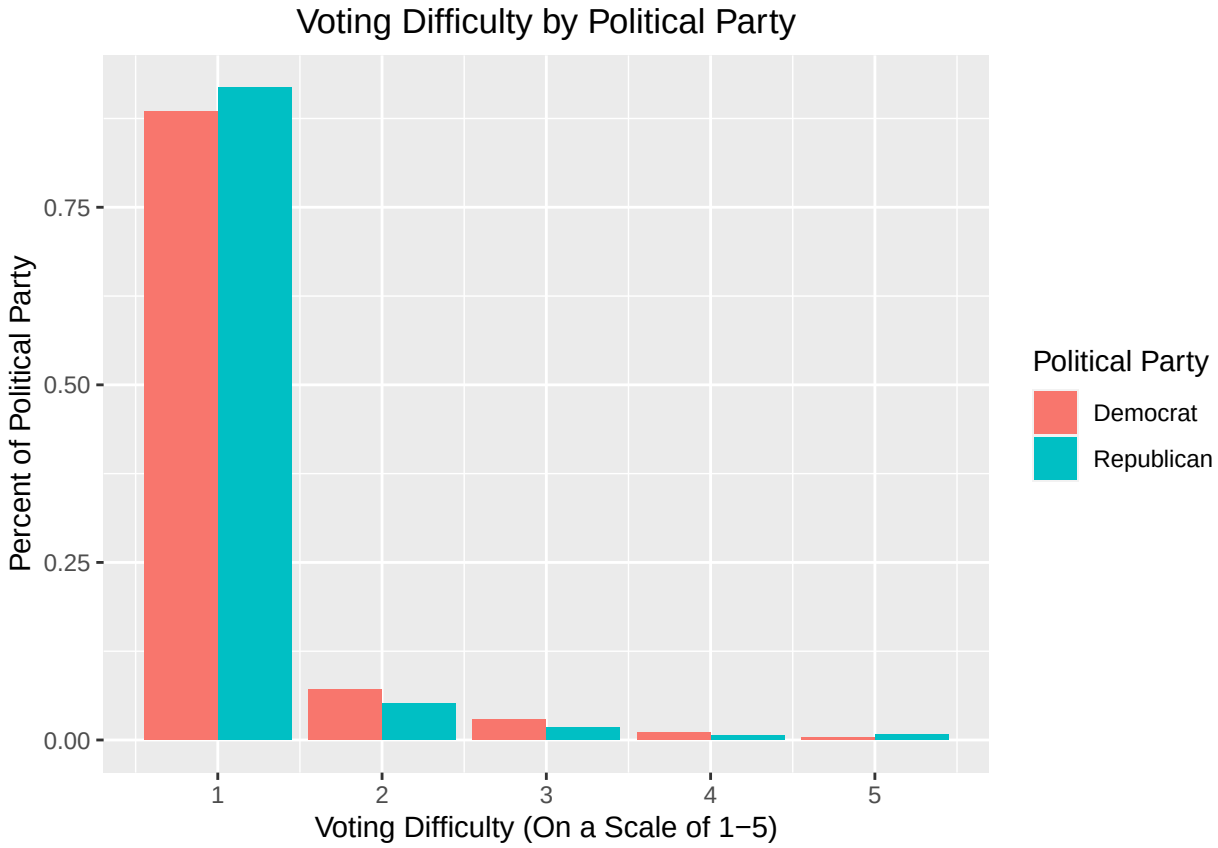
```
# HYPOTHESIS TESTING

# H0: reps and dems had the same voting difficulty
# Ha: reps and dems had different voting difficulty
two_sided_test = wilcox.test(reps$voting_difficulty, dems$voting_difficulty, alternative = "two.sided")
two_sided_test
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  reps$voting_difficulty and dems$voting_difficulty
## W = 1165900, p-value = 0.002665
## alternative hypothesis: true location shift is not equal to 0
```

```
# H0: reps and dems had the same voting difficulty
# Ha = reps had less difficulty voting than dems
one_sided_test = wilcox.test(reps$voting_difficulty, dems$voting_difficulty, alternative = "less")
one_sided_test
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  reps$voting_difficulty and dems$voting_difficulty
## W = 1165900, p-value = 0.001333
## alternative hypothesis: true location shift is less than 0
```

While the Democrats defeated the incumbent Republican President in the 2020 presidential elections, the data suggests that the Democrat respondents faced more difficulty in casting their ballot as compared to the other party affiliates. In our two-sided test, we obtained a p-value of 0.0026655, so we reject the null hypothesis and state that there is a difference in voting difficulty between parties. To be more specific, we ran a one-sided test and saw the p-value of 0.0013327, thus failing to reject the null hypothesis and stating that the Republicans had less difficulty voting than Democrats.



In our graph that compares democrats with republicans, we can see that the diagram agrees with our hypothesis test at first glance. Even though both parties mainly voted 1, the particular column illustrates that more republicans have an easier time voting compared to their democratic counterparts. Besides the majority of voters saying that voting was fairly easy, there were others who thought otherwise. However, the sample size of the votes greater than 1 pale in comparison, thus the effect size was less significant.

We initially thought that there would be correlation between difficulty voting since challenges of submitting a ballot can potentially cost a candidate the presidential seat. In 2020, we have seen many theories on the uprisings and downfalls of those running for presidency, some of which may have been attributed to leaners and demographics. Our study has shown that although the Republicans had less difficulty voting, the Democrats won in the presidential election.