Voting Difficulty for Repulican vs Democratic Voters

Ed Brown, Srila Maiti, Lisa Wu

6/28/2022

Contents

1 Importance and Context **1**

2 Data and Methodology **1**

3 Results **3**

4 Discussion **3**

1 Importance and Context

In the United States, voter turnout is a key measure of civic participation by the government and also is a key question to many political campaigns. Among the potential factors that affect voter turnout, there has been increasing attention on how voting access may impact voter turnout and even the election outcome, especially in the swing states. Access to voting has been a key debate between Democrats and Republicans, with the Republican party taking the stance that more safeguards should be established to protect the voting process by limiting access to polling places and the establishment of voter ID laws, and limied access to mail-in voting.

The Brookings Institute has commissioned us to understand which party has more difficulty voting. As part of their larger study on how difficulty voting by the two major political parties influences policy decisions within the party. We address the following research question in this study:

> *Do members of the Democratic or Republican parties have more difficulty voting?*

Answering this question will provide insight into one of the factors affecting party policy decisions.

2 Data and Methodology

Our analysis leverages data from the 2020 American National Election Studies (ANES). This is an observational data set. Respondents received invitations including 10 dollars cash to participate in the study and 40 dollars upon completion of the study. Non responding households were offered escalating incentives up to 100 dollars later in the field period. The original ANES data set contains 8280 respondents. These original samples are filtered down to our final working set according to Table 1.

Table 1: Samples Removed from ANES 2020 Timeseries Data

Removal Reason	Number of Samples Removed
Drop repondents not identified as Republican or Democrat	1003
Drop non registered voters.	919
Drop respondents who did not use the web survey.	360
Drop repondents who did not complete both pre and post sample interviews.	32
If respondents participated in the validation study, drop respondents with eligibility concerns.	798
If respondents participated in the validation study, drop invalid respondents.	174
If respondents participated in the validation study, drop invalid pre-post samples.	0
Remove samples listing “Not Registered” as difficulty reason for not voting.	16
Remove samples where difficulty variable V202119 was coded Inapplicable.	356
Drop samples where we could not compute a value for voter difficulty.	1

Samples were removed for multiple reasons to ensure a consistent working data. Our analysis is looking at any difference in voting difficulty between Republicans and Democrats. We are not interested in respondents who identify as Independent and removed those samples from our data. We are only interested in respondents who are registered to vote, anyone who was not registered was dropped from our sample. Two of our variables for assessing difficulty were available for use only if the respondents participated in the web version of the survey and we removed samples that were obtained through other formats. Our methodology utilizes variables from both the pre and post voting portions of the survey to maintain consistency, respondents who did not complete both portions of the survey were dropped from our sample. ANES includes a subset of the respondents in a validation study if any eligibility concerns were raised for the samples which participated

in the study they were omitted from our final data. In reviewing the post election results, despite filtering out non registered respondents in the prior step, we identified samples where "Not Registered" was listed as a reason for not voting. Because our final sample should only include registered voters, these samples were removed. The primary variable used to identify difficulty included data listed as "Inapplicable", we removed those samples since we are unable to assess difficulty for these samples. Finally, we omitted a sample where we were unable to compute a custom difficulty variable. These steps resulted in the removal of 3659 samples and a final sample size of 4621, including 2163 Republicans and 2458 Democrats.

We identified a respondent as voter based on these criteria. Any respondent from the 2020 ANES Time Series study who was registered to vote for the 2020 election, cast their vote early or on election day, and completed both the pre and post portions of the ANES survey. We have not applied the weighting variables to the data and this classification cannot be extrapolated beyond the respondents in the original study.

Classification of voters as Independent, Republican, or Democrat was conducted according to the guidelines laid out by the ANES in 2009 [1]. This procedure is documented on pp. 13-14 and is applicable to the 2020 time series data set. This methodology is based on three survey questions V201228 V201229 V201230, the SPSS code provided by the ANES was adapted to the 2020 Time Series Study and used to create a 7pt Party classification variable PID7. We later discovered the ANES includes this classification as a summary variable V201231x. We validated our PID7 variable against this summary variable provided by ANES and showed 100% match for respondents classified as Republican or Democrat on the 7pt scale. After the voters were classified on the 7pt scale, a three classification party variable was established by taking each party's "Strong", "Not Very Strong" and "Independent" classifications and assigning them the base value of Republican or Democrat with Independents classified as Independents.

For this study, we define voting difficulty based on the V202119 field. Difficulty in the voting process can arise from various factors like an error in the registration process, concern about ID card, difficulty obtaining an absentee ballot, difficulty reaching the polling place, etc. In the V202119 field, a 5-point rating scale is used to gauge voters' feeling of the voting process with 1 being not difficult and 5 being extremely difficult. The V202119 field is used as our outcome variable. This study does not reflect the causal effect why respondents select a certain rating which could be of an interest in the future.

The graphs below show the distribution of voting difficulty for Republican and Democratic voters in our final sample. The majority (88% of Democratics and 91% of Republicans) of voters in our sample rated 1 (no difficulty), but there are still 9-12% of the voters who experienced some level of voting difficulty.

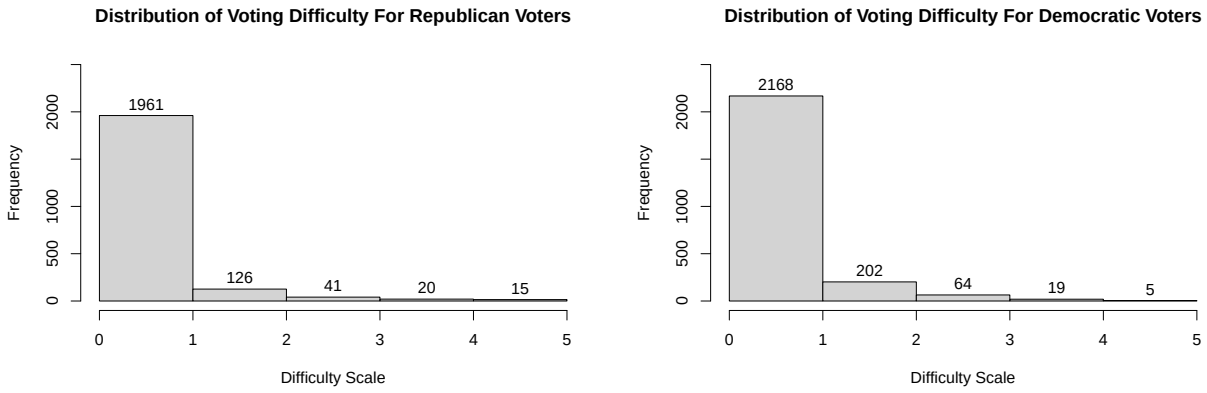

Figure 1: Voting Difficulty Values: 1- Not difficult at all; 2 - A little difficult; 3 - Moderately difficult; 4 - Very difficult; 5 - Extremely difficult

To answer our research question whether the Democratic or Republican voters have more or less voting difficulty, we evaluated what is the appropriate statistical test for our data set and research question. Our

[1] https://electionstudies.org/wp-content/uploads/2018/03/20091208PartyID.pdf

grouping variable (Party) is binary data (Republicans vs. Democrats) and our outcome variable (V202119 with a voting difficulty scale of 1-5) is a Likert scale. In this case, the most common test is the nonparametric Wilcoxon Rank Sum test for two independent samples. We also transformed several data fields (V202119, V202123, V202124) into a binary outcome data field "Had_Difficulty_Voting" (0 means no difficulty and 1 means having difficulty). So we could also perform the Welch two-sample t-test because both the grouping variable (Party) and the alternative outcome variable (Had_Difficulty_Voting) are binary data. We proceeded with the Wilcoxon Rank Sum test because the V202119 Likert scale data is provided directly.

The null hypothesis of our Wilcoxon Rank Sum test can be phased as follows:

> **Null Hypothesis:** *The probability of Republicans having more voting difficulty than Democrats is the same as the probability of Republicans having less voting difficulty than Democrats*

In order for the Wilcoxon Rank Sum test to produce reliable results, the following must be true: 1) the data must be drawn from an i.i.d. sample; 2) data must be at least ordinal, and 3) no substantial differences in group sample sizes

First, the ANES survey group consisted of respondents who had participated in the ANES 2016 Time Series Study. The remaining sample was based on a freshly drawn cross-section, and addresses were randomly assigned. Only one person from each selected address participated in the survey. So we can reasonably assume i.i.d in the data. Second, the outcome variable (V202119) with a 5-point scale and is ordinal data. Finally, the final sample size is 2163 for Republicans and 2458 for Democrats, no substantial difference.

3 Results

```
wilcox_result = wilcox.test(dplyr::filter(final_dataset_for_testing, PARTY == "R")$V202119,
                            dplyr::filter(final_dataset_for_testing, PARTY == "D")$V202119,
                            alternative = "two.sided", mu = 0, correct = TRUE,
                            exact = FALSE, conf.int = T, conf.level = 0.95)
```

The test shows p-value of 0.0098092 which is smaller than 0.05 and therefore we reject the null hypothesis that the probability of Republicans and Democrats having the same amount of difficulty voting is equal.

The 95% confidence interval was -1.78e-06 to 2.65e-05 and provided no insight into which party had a more difficult time voting. We can say there is evidence of difference of voting difficulty existing between the voters of the two parties. However, the test statistic does not indicate the direction which party might have more difficulty since our confidence interval straddles zero and is very small. We performed the Spearman rho test to evaluate the practical significance, the rho value is -0.0379945 which indicates very small effect. A rho value smaller than 0.1 to 0.3 (positive) or higher than -0.1 to -0.3 (negative) is consider small effect.

There are limitations in using the V202119 field to gauge the voting difficulty as it is observational data and reflects judgment. This field does not provide information on what caused the voting difficulty. Our final sample only included respondents who used the web survey. Additionally, our research focuses on the respondents who voted which may have missed people who did not vote or register to vote due to difficulty.

4 Discussion

This study found evidence that the probability that Republican and Democratic voters experience the same voting difficulty is not the same. However, the practice significance is very small and therefore this study by itself is not sufficient for the Brookings Institute to draw meaningful insight. An additional study will need to be conducted better quantify the difference in difficulty voting between two parties.

We believe that further study of non-registered people or registered non-voters may provide further insight for the Brookings Institute.