

# Evaluation of Voting Difficulty Between Republicans and Democrats

Datasci 203: Lab 1, Part 2 Assignment

Sophie Chance, Amy Zhang, Maureen Fromuth

## Contents

|  |   |
|--|---|
| Introduction . . . . .                             | 1 |
| Conceptualization and Operationalization . . . . . | 1 |
| Visual Design . . . . .                            | 2 |
| Data Wrangling . . . . .                           | 2 |
| Hypothesis Testing . . . . .                       | 2 |
| Test Results and Interpretation . . . . .          | 3 |

## Introduction

It has been a growing concern for believers in equitable voting rights that there are unfair voting challenges being placed on one party over the other. If proven true, this may unfairly influence election results. Some examples of these voting challenges include changing district boundaries, adding specific voter regulations, and inaccessible voting environments that would make it more difficult for voters to physically vote<sup>1</sup>. It is important for us to understand if there is a difference in voting difficulty between the Democratic and Republican parties to inform our focus and funding priorities going forward. As the leading Democratic NGO, we are testing our hypothesis that Democratic voters experience more difficulty voting than Republican voters in the United States. As such, for this study, we are evaluating if there is a difference in difficulty voting between Democratic or Republican voters. We will then present this result to our party’s largest donors to request more funding to alleviate barriers for Democratic voters.

## Conceptualization and Operationalization

The data used in this study is the 2022 American National Election Studies (ANES) pilot study data, an online opt-in survey that asks various questions related to voter experience, choice, and election topic opinions<sup>2</sup>. The survey is broken into various sections, with the most relevant for this study being ‘Turnout and Choice’ and ‘Party Identification’. Questions in these sections acted as the basis for the key definitions of this study and were tied to the variables listed in the table below.

| Term              | Column with Feature                    | Definition Values  |
|-------------------|--|--|
| Voter             | turnout22; pipevote22a;<br>turnout22ns | rand_pipe=1,3 & turnout22=1,2,3; rand_pipe=1,3 & turnout22=5 & turnout22ns=1; rand_pipe=2, pipevote22a=2 |
| Republican        | pid1d; pid1r; pidlean                  | pid1d=‘Republican’; pid1r=‘Republican’; pidlean=1  |
| Democrat          | pid1d; pid1r; pidlean                  | pid1d=‘Democrat’; pid1r=‘Democrat’; pidlean=2  |
| Difficulty Voting | votehard                               | votehard = [1,5]   |

**Voter** Ideally, we would define a voter as a participant that completed any portion of the voting process from registering to completing a vote. A voter for this study, however, is identified as an individual who voted in the 2022 election. This is due to the structure of the survey, which only provided the question of hardness in voting to those that voted. The questions in the survey that inform the designation of a voter are the following: *“In the election held on November 8, did you definitely vote in person on election day, vote in person before Nov 8, vote by mail, did you definitely not vote, or are you not complete sure whether you voted in that election?”*, *“We talk to many people who tell us they did not vote. And we talk to a few people who tell us they did vote, who really did not. We can tell they did not vote by checking with official government records. What about you? If we check the official government voter records, will they show that you voted in the election held on November 8, or that you did not vote in that election?”*, and for those that were unsure if they voted, *“If you had to guess, would you say that you probably did vote in the election held on November 8, or probably did not vote in that election?”*

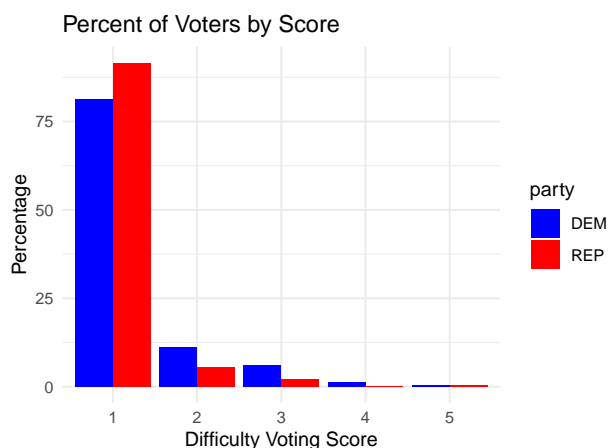
**Republican** Identifying with a political party has its challenges given the current political landscape. Ideally, the data would objectively evaluate a participant based on many factors including comprehensive voting history, alignment of values and ideologies. This study defined a Republican as a participant that identified as or leaned Republican. This self-identification is a strong indicator of political party belonging. For party indicators, the variable was tied to the answers to the following survey questions: *“Generally speaking, do you usually think of yourself as a Democrat, a Republican, an independent, or what?”*, *“Generally speaking, do you usually think of yourself as a Republican, a Democrat, an independent, or what?”*, *“Do you think of yourself as closer to the Republican Party or to the Democratic Party?”*

**Democrat** Similar to Republicans, this study defined a Democrat as a participant that identified as or leaned Democrat. To designate Democrats, this study evaluated the same questions as above.

**Difficulty Voting** Evaluating how difficult it is to vote objectively is challenging as it is dependent on the voters personal experiences. This study defines the difficulty voting as how hard a participant felt it was to vote on an ordinal scale of 1-5, with 1 being not difficult at all and 5 being extremely difficult. The survey question asked was, “*How difficult was it for you to vote?*”. Ideally, there would be more context given about why someone did not vote. This question was not presented to them, making it difficult to determine if their reason for not voting was due to difficulty or some other reason.

## Visual Design

The voter data showed that the mean, median, and 3rd quartile of the data have a value of 1, indicating that most voters did not find it difficult to vote (on average, and at least 75% of our sample). Yet, as seen to the right, the breakout of difficulty by party appears to be skewed to the right for both parties.



The "Percent of Voters by Score" metric represents the proportion of voters at each difficulty level split by party. This makes it clear that Republican’s have a significant portion of their population in the "Not difficult at all" category. For the Democratic voters, there is a noticeably lower percentage of voters in the "Not difficult at all" category.

## Data Wrangling

Four major steps were completed during data wrangling for this study: *remove invalid data*, *clean data for party*, and *clean data for voting*. The chart below provides an accounting of the data loss from this cleaning.

| Starting Entries | Not Weighted | No Party | Didn't Vote | Final Entry Count |
|------------------|--------------|----------|-------------|-------------------|
| 1585             | 85           | 267      | 257         | 976               |

**Remove invalid data** According to the documentation, there were there are 1500 that are used to represent the population but 85 that are not and consequently did not receive weights. To remove the 85 we removed the entries that do not have weights in the *weight* or *weight\_spss* columns. We also removed unnecessary variables for simplicity and readability, which did not result in any data loss.

**Clean data for party** To simplify the tracking of party, this study created a new variable, *party*. If the member was a Republican, they would be designated as ‘REP’ in the *party* column, and if a Democrat they would be designated as ‘DEM’ in the column. Participants that did not identify as either option retained the original ‘na’ value and were ultimately removed.

**Clean data for voting** Similar to above, we created a new variable *did\_vote* to quickly designate which participants voted ‘1’ and which did not ‘0’. Using the variables above we designated those that voted as a ‘1’. For those that were unsure, we entered ‘1’ for those that believed they did vote but retained a ‘0’ for those that were fairly confident they did not. For those that did not vote, we removed those entries as there was no way to identify clearly why they chose not to vote.

## Hypothesis Testing

For this study, we elected to use a non-parametric Wilcoxon ranked-sum test as the most ideal statistical test<sup>3</sup>. First, while the participants are grouped based on *party*, the survey does not lend itself to conducting

a paired test, as each participant is only measured once. Second, the response variable that is used for comparison, *votehard*, uses an ordinal Likert scale and thus is not interval or ratio data. Given there is no metric scale, the hypothesis of comparisons is the correct version of the Wilcoxon ranked-sum test to use.

There are two key assumptions for this type of test: that the data is at least in an ordinal scale and that the data is independent and identically distributed (IID)<sup>3</sup>. As identified, the *votehard* variable is ordinal and uses a Likert scale, satisfying the first assumption. The second assumption has two parts. The first part of IID, independence, loosely holds as the survey is conducted through a reputable online portal. However, there may be some concerns around grouping of participants who are aware of and participates in ANES surveys, which would lead potential dependence via geographical clustering, etc. The second part of IID, identical, holds as each Democratic sample is be drawn from the same distribution, and each Republican sample is drawn from the same distribution.

The null and alternative hypotheses of our Wilcoxon ranked-sum test can be phrased as follows:

**Null Hypothesis:**  $P(\text{Democrats difficulty voting} > \text{Republican difficulty voting}) = P(\text{Democrats difficulty voting} < \text{Republican difficulty voting})$

**Alternative Hypothesis:**  $P(\text{Democrats difficulty voting} > \text{Republican difficulty voting}) \neq P(\text{Democrats difficulty voting} < \text{Republican difficulty voting})$

Using the following code in R, we evaluated the null hypothesis using the Wilcoxon ranked-sum test<sup>4</sup>:

```
wilcox <- wilcox.test(votehard ~ party, data=voter_data)
```

The results of this test include a p-value of  $3.6163904 \times 10^{-6}$  and a W-value of 130622.

## Test Results and Interpretation

The p-value of  $3.6163904 \times 10^{-6}$  is below the conventional 0.05 threshold, and thus we can reject the null hypothesis. This means that there is a statistically significant difference between the distributions of the Democrat’s and Republican’s difficulty to vote scores.

To further assess the practical implications of this difference, Spearman’s rank correlation, a rank-based correlation coefficient, is used to evaluate difference in the ranks given the two variable, *party* values for each survey entry<sup>5</sup>. The results of the Spearman formula for this test is a value of -0.1483581. As it is negative, the association is negative, meaning that as you get closer to a value of ‘1’ in the *party* than the number in *votehard* gets lower. More directly, Republicans tend to have a lower *votehard* value indicating an easier time in voting vs. a Democrat. This correlation aligns with the rejection of the null hypothesis and the percentage of ranked sum suggesting that Democrats do have a greater challenge in voting.

In conclusion, the ANES survey data suggests a distinct difference in the difficulty experienced by Democratic and Republican voters during the 2022 election, with Democrats experiencing more difficulty when voting. In the context of voting rights, even minor disparities can have instrumental consequences. As the leading Democratic NGO, this information is essential when shaping our outreach programs, advocacy efforts, and in making a case for increased funding. Although this difference may seem insignificant when looking at this small sample, this difference could result in a high number of Democrats facing difficulties voting and potentially change the results of an election when extrapolated to the entire country of millions of voters. This study could further be improved by incorporating multiple years of data, incentivizing the participants, including locality, and using the weighting system. The United States relies on equitable voting processes as the foundation of our government. Addressing these barriers to voting is essential to the success of the system and should be tackled head-on ensuring a fair voting experience for everyone.

## Sources

1. American Civil Liberties Union, *Block the Vote: How Politicians are Trying to Block Voters from the Ballot Box*, <https://www.aclu.org/news/civil-liberties/block-the-vote-voter-suppression-in-2020>
2. American National Election Study, *2022 Pilot Study*, <https://electionstudies.org/data-center/2022-pilot-study/>
3. University of California, Berkeley DATASCI2003, *Cheat Sheet: Wilcoxon Rank Rank Sum Test*, [https://github.com/mids-w203/cheat\\_sheets/blob/master/wilcoxon\\_rank\\_sum\\_cheat\\_sheet.md](https://github.com/mids-w203/cheat_sheets/blob/master/wilcoxon_rank_sum_cheat_sheet.md)
4. University of Virginia, *The Wilcoxon Rank Sum Test*, <https://library.virginia.edu/data/articles/the-wilcoxon-rank-sum-test>
5. GeeksforGeeks.org, *Spearman Correlation Testing in R Programming*, <https://www.geeksforgeeks.org/spearman-correlation-testing-in-r-programming/amp/>