Voting Difficulty: A Comparative Analysis of Democratic and Republican Voters

Datasci 203: Lab 1

Neeha Kotte, Sohail Khan, Michael Guldberg

Contents

1 Importance and Context

Participation in political processes and local, midterm and national elections is essential for representing the breadth and diversity of individuals' views in the United States, as this has a direct effect on government officials and policies. In a democratic political system such as the United States, high voter turnout is essential for both the health of the democracy and including the political views of all citizens in elections [1]. We would assume this results in all eligible voters in the United States rushing to polling centers to cast their ballots, but this is often not the case. While voter turnout in the elections of 2018, 2020, and 2022 marked some of the highest turnouts of their respective types in U.S. elections in decades, only 37% of the 70% of eligible U.S. adult citizens participated in all three elections.[2]. Although we have seen an increase in voter turnout, there is still significant room for improvement. This warrants investigation into the difficulty that American citizens experience when voting. Additionally, since election results are affected by the voter turnout for each party, it is valuable to understand how voting difficulty may be different for Democrats versus Republicans. Therefore, this analysis aims to address the following research question:

Do Democratic voters or Republican voters experience more difficulty voting?

The answer to this question could help assess fairness in voting accessibility for Democrats and Republicans and provide direction for further analysis regarding fair voting accessibility amongst political parties.

2 Data and Methodology

In order to test our hypothesis, we will utilize data from the American National Election Studies (ANES) 2022 Pilot Study. ANES' mission is to gather vital political information and behaviors through incentive-driven surveys. Our analysis is specifically focused on voting behaviors and voter characteristics, such as party affiliation, drawn from an online survey completed by 1,585 respondents [3] We will remove any individuals who are not classified as voters or do not have a party affiliation of either Democrat or Republican. This leaves 1,044 observations.

As we report in Table 1, 91.91% of Republican respondents reported voting to be "Not difficult at all" as opposed to 82.38% of Democrats with the same response. We can visualize this further in Figure 1 below.

Table 1: Self-Reported Voting Difficulty in 2022

	Republican	Democrat
1- Not difficult at all	91.91	82.38
2- A little difficult	5.39	10.50
3- Moderately difficult	2.07	5.69
4- Very difficult	0.21	1.07
5- Extremely difficult	0.41	0.36

To filter the dataset in this way and to ultimately conduct statistical testing on our null hypothesis, we must first operationalize three key concepts:

1. Who is a voter?
2. Who is a Democrat/Republican?

[1] MIT Election Data + Science Lab. "Voter Turnout." (2021).
[2] Pew Research Center. "Republican Gains in 2022 Midterms Driven Mostly By Turnout Advantage." (2023).
[3] American National Election Studies (ANES). "2022 Pilot Study." (2022).

3. What is difficulty voting?

To operationalize the concept of who we will consider a voter for this analysis, we chose to use the variables **turnout22** and **pipevote22a** from the ANES dataset. More specifically, we define a voter to be all observations where the respondents reported that they did vote in the 2022 midterm elections in either the **turnout22** or **pipevote22a** question. We chose this strategy for defining voters since only one of these questions is presented to each participant, and combining these variables allows us to assess 2022 voting status for every participant.

For assigning Democrat/Republican designations, we will use the **pid_x** (Summary Party ID) variable which assess party identification on a Likert scale from "Strong Democrat" (1) to "Strong Republican" (7). Additionally, we will consider those who lean towards a particular party to be functionally considered to be members of those parties, respectively.

Lastly, we will operationally define voting difficulty based on the **votehard** variable, which represents the self-reported voting difficulty that each voter experienced in the 2022 midterm election. This variable's values are assessed on a Likert scale ranging from "Not difficult at all" (1) to "Extremely difficult" (5). Additionally, this question is presented to the same subset of participants which we define as voters in this analysis (i.e. by combining the **turnout22** and **pipevote22a** variables), making it a good choice for our analysis. We also considered other options for operationalizing "difficulty voting", such as combining the **vharder0-12** variables or adding **triptime** and **waittime** to provide a "total" voting difficulty score. However, if we simply combined the **vharder0-12** variables, we would assume that they are all weighted the same, i.e. bad weather is the same as experiencing difficult registering. Similarly, simply "adding" **waittime** and **triptime** is unclear, because the combined ranges become conceptually less meaningful and too varied.

Since we have ordinal data, we will use the Wilcoxon Rank-Sum statistical test (Hypothesis of Comparisons version). This is a nonparametric test for unpaired data, which is relevant to our analysis since we are observing two different groups, Democratic voters and Republican voters.

The null hypothesis of our Wilcoxon Rank-Sum test (Hypothesis of Comparisons) can be phrased as follows:

> **Null Hypothesis:** *There is no difference in difficulty voting between Democrats and Republicans in the 2022 midterm election. (The probability that a Democratic voter has more difficulty voting than a Republican voter in the 2022 midterm election is the same as the probability that a Republican voter has more difficulty voting than a Democratic voter in the 2022 midterm election).*

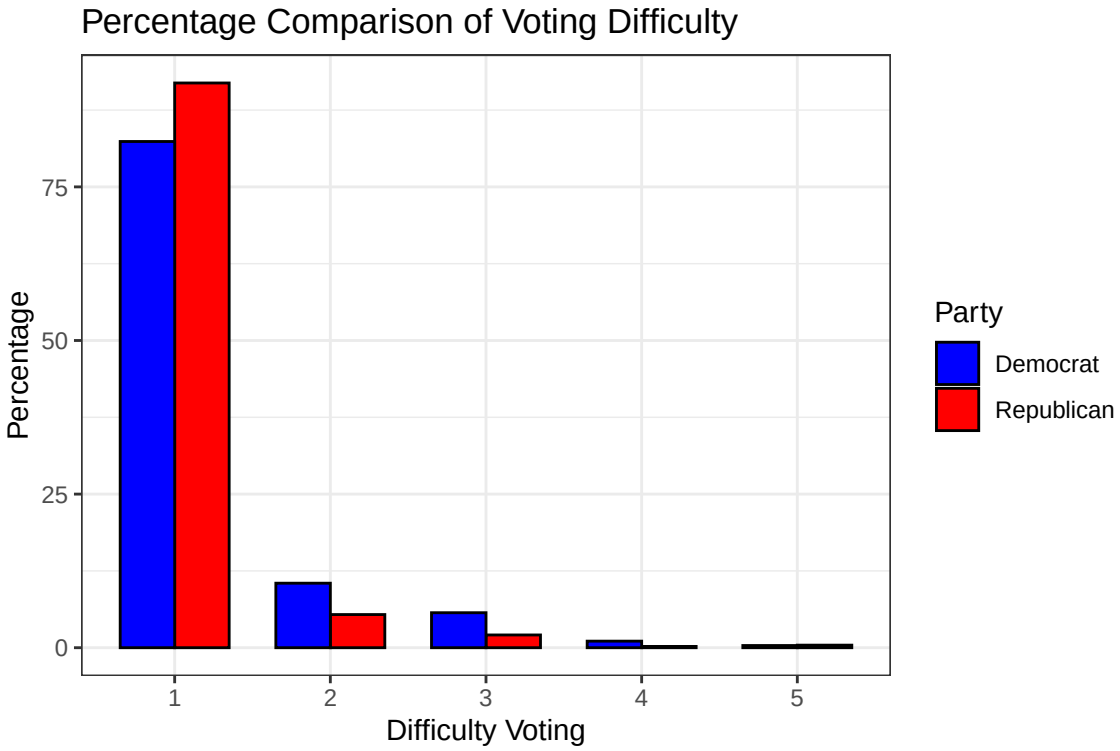
However, in order to use this statistical test, we must ensure our problem space meets the following assumptions: 1) at least an ordinal scale and 2) independent and identically distributed (i.i.d.) data.

1. In this case, our measurement variables are on an ordinal scale. As mentioned above, the **votehard** variable ranges from "Not difficult at all" (1) to "Extremely difficult" (5) which is on a Likert scale. There is no meaningful way to perform any mathematical operations on an ordinal variable. For example, we cannot necessarily conclude that the difference between "Not difficult at all" and "A little difficult" is the same as the difference between "Moderately difficult" and "Very difficult". Our variable for party identification (**pid_x**) is also ordinal, with binary values of either Democrat or Republican. Hence, any statistical tests which rely on metric data will not be applicable to our scenario.
2. The ANES data was gathered via a cross-sectional survey, which implies that the information about the sample subset is diverse (in fact, 1500 of the 1585 individuals' data points are weighted accordingly). There may be inherent dependencies if related individuals, neighboring individuals or same household individuals complete the survey, but ANES ensures random sampling. Although it is not necessarily nationally representative, given the number of samples gathered, we can assume for our purposes that this data is i.i.d.

3 Results

```
test <- wilcox.test(dem_vote_hard$votehard, repub_vote_hard$votehard)
```

The results of our Wilcoxon Rank-Sum test (Hypothesis of Comparisons) are significant, given the p-value 5.1281142×10^{-6}. Therefore, we can reject our null hypothesis, which implies that there was a difference in difficulty voting between Democrats and Republicans in the 2022 midterm election. The Wilcoxon Rank-Sum test results however, will not tell us which group has more difficulty voting. Therefore we performed additional analysis to understand the spread of voting difficulty (where we compare relative percentages of difficulty values within each group) between Democrats and Republicans. See figures below:



These visualizations suggest that Democrats may have had more difficulty voting - comparing relative percentages, we notice a higher percentage of Democrats who responded with above a "1- Not difficult at all". However, these conclusions are not causal and they do not necessarily reflect the behaviors of the broader U.S. population – our hypothesis testing only provides us information on associative relationships.

4 Discussion

Our study found evidence that there was a difference in difficulty voting between Democrats and Republicans in the 2022 midterm election. In particular, we found relatively more difficulty in voting amongst Democratic voters, although the effect size may be quite small. These results may be important for leaders in both political parties so they can assess – perhaps through more detailed and causal studies – which factors make voting more difficult for their target voter populations. Identifying and mitigating these factors may lead to increased participation in local, state and national elections, ensuring a nationally representative and diversely functioning democracy.