Analysis of Democratic and Republican Voting Difficulty

Lab 1 Hypothesis Testing - Part 2: Group Analysis

Team Fishers: David Solow, Lin Nicholas Lin, Rong Huang, Summer McGrogan

Contents

1   Importance and Context

The United States is a nation built upon the principles of democracy, where the voice of every eligible citizen should be heard through the act of voting. However, recent research has shown that various states have adopted regulations that could create impediments to voting [1]. The implications of these rules have attracted concerns about the fairness of the electoral system and raised a critical question:

Do Democratic voters or Republican voters experience more difficulty voting?

Understanding and addressing the obstacles faced by voters of different political affiliations is crucial. Such disparities in voting experiences not only impact individual citizens but also carry wider implications for the overall well-being of our nation's democratic institutions. They bring into question the fundamental equity of our democracy, raising concerns about the potential erosion of public trust in the electoral system.

To provide a comprehensive, evidence-based response to this question, this report draws upon the dataset provided by the American National Election Studies (ANES), a highly regarded source of data on the American electorate. This report utilizes rigorous statistical techniques and methodologies to reveal patterns and trends that shed light on the potential disparities, if any, in the voting experiences of Democrats and Republicans. In doing so, we hope to inform policy discussions, promote transparency, and contribute to the ongoing dialogue on the state of American democracy.

2   Data and Methodology

The ANES study leveraged survey data collected by the YouGov panel. The YouGov panel survey consisted of respondents who signed up for this online questionnaire, in exchange for a monetary reward.

Throughout our report, "a voter" describes someone who voted on or before November 8th, 2020, responded to the ANES survey, and was selected by YouGov's sample matching procedure. A Republican is a voter that believes their ideas align more closely with the Republican party than any other political party. A Democrat is a voter that believes their ideas align more closely with the Democratic party than any other political party. "Difficulty voting" describes the challenges faced by a voter when attempting to cast a ballot.

The best possible method for measuring which voters experience more difficulty voting should rely on an objective, quantitative metric for measuring voting difficulty. Unfortunately, these metrics are not available in the data that are currently accessible. However, a lot can still be learned about voting difficulty between Democrats and Republicans by substituting the best available data into this metric. To do this, we will assume that difficulty voting can be examined by comparing the distributions of voter difficulty reported by voters in the "votehard" variable of the ANES dataset. While this variable represents a Likert Scale for which respondents may have different subjective understandings of the ratings, we will assume that with a large enough sample size a comparison of this variable between Democrats and Republicans will indicate whether or not voters of each party experience differing levels of difficulty voting.

Below is a summary of the relevant variables from the ANES dataset with the following changes made, resulting in a final sample dataset of 684 samples from an original dataset of 1,585 samples:

1. Samples that are not weighted to correct for demographic proportions that do not match the broader US population have been removed (85 samples removed)
2. Samples with missing data have been removed (2 samples removed)
3. Samples that report no leaning towards the Democratic or Republican party (pid_x = 4 & pid_x = .) have been removed (266 samples removed)
4. Samples for which the certainty that they had voted was less than 100%, as reported in the ANES dataset, have been removed (548 samples removed)

A summary of the total number of Democratic and Republican votes can be seen in Figure 1. For both political party groups, Democrat and Republican, a voting difficulty level of 1, indicating no difficulty voting, was the most common response.

---

[1]Brennan Center for Justice. "Voting Laws Roundup 2020." (2020).

After removing the appropriate samples, the following variables are selected for analysis:

1. **Party** - the party affiliation of the voter based on interpretation of the "pid_x" variable, which contains reported voter leaning. Values of 1, 2, or 3 are assigned "Democrat", while values of 5, 6, or 7 are assigned "Republican". Values of 4 are assumed to have not party affiliation and have been removed
2. **Voting Difficulty** - the reported difficulty that the voter experienced on a Likert Scale from 1-5
3. **Number of Voters** - count of the number of voters falling into each Party and Voting Difficulty value

The final sample for analysis is summarized below.

Based on this sample, Figure 1 illustrates the number of Democrats and Republicans for each level of voting difficulty (a lower number indicates less difficulty voting). Figure 1 shows that, for category 1 (no difficulty voting) and category 5 (extreme difficulty voting), Republicans scored the highest in these self-reported categories. We can also see that Democrats scored the highest for categories 2 (some difficulty voting), 3 (moderate difficulty voting), and 4 (high difficulty voting). Figure 2 illustrates the probability of a voter experiencing difficulty voting, aggregated by political party and level of voting difficulty. Figure 2 indicates that Democrats may experience higher difficulty voting based on the available data.

Figure 1 shows that Voting Difficulty is positively skewed for both parties, with most voters reporting no difficulty voting (Voting Difficulty = 1).
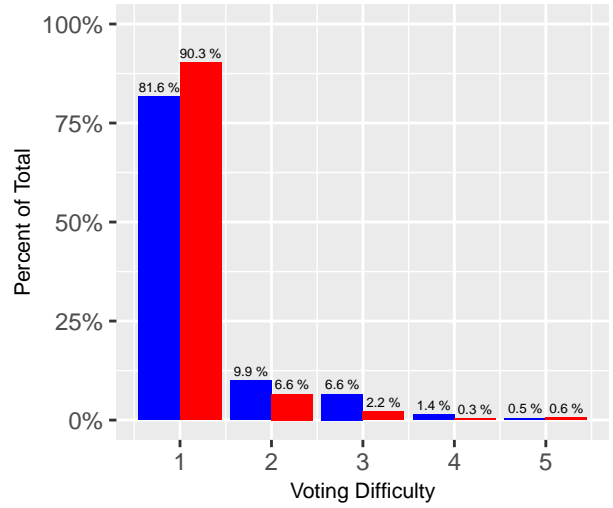
In order to normalize for the difference in sample size of Democrats vs. Republicans, Figure 2 is added below to illustrate the proportion of voters experiencing each level of difficulty voting by political party.
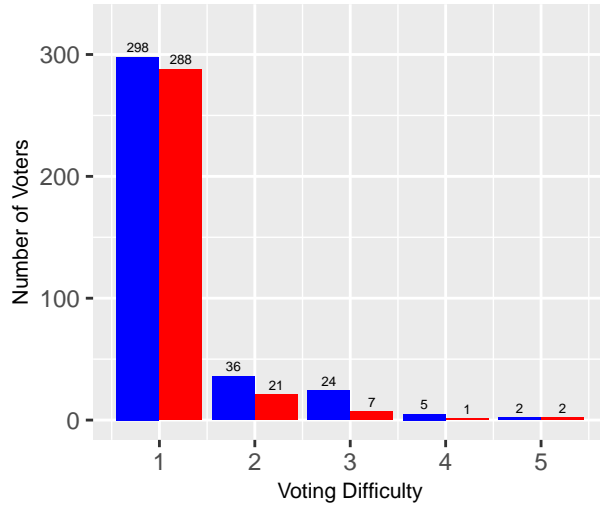
Based on the distributions above, it appears that a higher proportion of Republicans compared to Democrats report both the lowest and highest difficulty voting (Voting Difficulty in (1,5)), with the difference between more pronounced for Voting Difficulty = 1 (8.7% difference) compared to Voting Difficulty = 5 (0.1% difference). A comparatively larger proportion of Democrats compared to republicans report the middle levels of difficulty voting (Voting Difficulty in (2,3,4)). While these differences between the distributions are worthy of interest, a statistical test is required to indicate which party experiences more overall difficulty voting.

Given that Voting Difficulty represents non-metric data (Likert Scales are ordinal in nature) and given that the distribution of Voting Difficulty is not normally distributed for either political party, Voting Difficulty will be compared between parties using a Wilcoxon Rank-Sum test with a critical $\alpha = 0.05$. The null hypothesis for of our Wilcoxon Rank-Sum test is as follows:

Null Hypothesis: *The probability that a Democrat has a high difficulty voting ranks higher than a Republican with a high difficulty voting is equal to the probability that a Republican has a high difficulty voting ranks higher than a Democrat with a high difficulty voting.*

In order for the Wilcoxon Rank-Sum test to produce reliable results, we must have: data measured on an ordinal scale; sample data that is independent and identically distributed (i.i.d). We address each of these requirements below.

First, the data must be measured on an ordinal scale. Ordinal data describes data that has ordered categories, but the distance between these categories is not known. A Likert Scale test, as seen in the dataset, is a categorical measurement (from a scale of 1 to 5) to measure the perceived voting difficulty by a voter that has no standardized distance between each value.

Second, the data must be generated via an i.i.d process. The 2022 ANES pilot survey uses data from the YouGov platform. The use of the this platform, along with the reward system YouGov uses to motivate respondents, could introduce income-level, and education level clustering to the dataset. However, YouGov boasts millions of users, which decreases the likelihood of these dependencies to an acceptable amount.

3 Results

statistic	63376.5
p.value	0.00101187649907185
method	Wilcoxon rank sum test with continuity correction
alternative	two.sided

The test yields evidence that the null should be rejected based on the current data (W = 63376, p-value $\approx$ 0.001012). Based on the available data, these results support the claim that Democrats have more difficulty voting than Republicans. As Figure 2 illustrates, 81.64% Democratic party voters indicated that they had no difficulty voting, compared to 90.28% Republican party voters that indicated they had no difficulty voting, which may be driving the observed difference based on the available data.

Several limitations on the sample data exist that limit the conclusions that may be drawn from it. As mentioned above, a Likert Scale cannot be quantified in any meaningful metric form, and the levels of voting difficulty may change depending on the survey users interpretation of each question used to measure voting difficulty. Additionally, the ANES survey may not be generalizable to the entire United States population, as it is based in the assumption that participation in the survey is independent of all other survey factors. We have previously discussed that clustering related to income or education level, may be related to survey responses.

4 Discussion

Based on the available data, this study found evidence that voters who identify as Democrat may have more difficulty voting than those who identify as Republican. Furthermore, the effect appears practically significant particularly when comparing the proportion of voters from each party that report no difficulty voting, with the proportion of Republican voters that report experiencing no difficulty voting being 8.7% higher than that of Democrats. This effect size and our results more broadly may be of interest to politicians that want to introduce policies to sway elections towards or away from Democratic or Republican opponents. Our results may also be of interest to political activist groups that attempt to increase voter turnout for all parties. This study also opens the door for more discussion around the topic of why certain groups face difficulty voting, and how political representatives may be able to influence the level of voting difficulty different groups face. Finally, we hope these results will be leveraged to increase participation in our democracy.