

# A Statistical Analysis of Voting Difficulty in Democrat and Republican Voters

Thomas Lee, Victoria Brendel, Viola Qiu

July 1, 2024

## 1 Introduction

Voting difficulty is a critical issue in the study of electoral behavior and democratic participation. Having difficulty deciding who to vote for can impact voter turnout, election outcomes, and the overall health of a democracy. Our research question is: **do Democrat voters or Republican voters experience more difficulty voting?**

We wish to determine if, on average, Democrat voters or Republican voters find it more difficult to vote for their party's presidential candidate. This has implications for campaign strategies for political parties. The ANES 2022 survey is used to conduct a Two-Sample t-Test to provide a rigorous statistical approach to answering our research question.

## 2 Data Source and Description

The data used in this analysis comes from the ANES 2022 Pilot Study following the 2022 midterm elections. This dataset includes a wide range of variables related to political attitudes, behaviors, and demographic characteristics of U.S. votes. It is publicly available through the ANES website and comes with a codebook, which is crucial for understanding definition and relationship among variables.

In this analysis, we are interested in variables that are relevant to political identification, registration and voting behav-

iors, and voting difficulties. Some key variables include `[reg]` indicating whether the respondent is registered to vote, **turnout22**, **turnout22ns**, **pipevote22a** indicating whether the respondent voted in the 2022 elections, **pid\_x** combining responses from several questions to categorize respondents as Democrats, Republicans or Independents, and **vote24dt** indicating the respondent's vote choice in 2024.

## 3 Conceptualization & Operationalization

The detailed data pipeline implementation with comments to provide documentation is included in a separate .R file within the project root directory.

### 3.1 Valid samples

According to the ANES codebook [1], out of the 1585 total samples, 1500 of them passed a quality control check. These samples can be identified as having a non-NULL **weight** feature. We removed all 85 samples which had a NULL value for **weight**, leaving 1500 samples.

### 3.2 Who is a voter?

We decided to use participation in the 2022 election as a metric to characterize who is and who is not a voter. Using the **reg** feature, we removed 192 non-registered

individuals, leaving 1308 samples. Participants were randomly given variations of the question asking about participation in the 2022 election. **pipevote22a** directly asks participants whether they did or did not vote during the 2022 election, and we kept all samples with "yes" responses. **turnout22** asks participants if they voted, did not vote, or were unsure. They could clarify whether they think they probably did or did not vote in **turnout22ns**. We believe that if they stated that they probably voted in **turnout22ns**, it would still show a tendency to vote. Thus, we decided to treat both individuals who stated that they definitely did vote and individuals who stated that they probably voted as voters. We labeled 857 of our 1308 samples as being voters and removed the 451 non-voters from the data.

### 3.3 Who is a "Republican" and who is a "Democrat"?

We used the **pid\_x** feature to classify whether a participant is a Democrat or a Republican. After removing missing values, we assigned all individuals who chose options 1 through 3 (Strong Democrat, Democrat, and Lean Democrat) to be a Democrat. We assigned all individuals who chose options 5 through 7 (Lean Republican, Republican, and Strong Republican) to be a Republican. Using this scheme, we labeled 759 of the remaining samples as being either Democrat or Republican and removed the other 98 samples.

### 3.4 What is difficulty voting?

We further operationalized the definition of "difficulty voting" as "difficulty with their party's candidate." We can label this difficulty by seeing which participants would not vote for their party's nominee in the upcoming 2024 election. Using **pid\_x** to determine political identity and **vote24dt** to determine intent, we assign a value of 0

to individuals who plan on voting for their party's nomination and a value of 1 to individuals who plan on voting for someone other than their party's nomination. This binary variable is a valid metric to use in our tests because the interpretation of 0 and 1 is the same for every individual for both parties. In data cleaning, we only had to remove one NULL sample from this feature.

We considered using the **votehard** feature as a measure of difficulty voting. However, from reading through the description of the survey, we realized that the magnitude of the **votehard** value is subjective for each participant, making it a non-robust measure of difficulty voting.

## 3.5 Hypothesis Statement

**Null Hypothesis:** The mean voting switching rate for Democrat voters is equal to the mean voting switching rate for Republican voters.

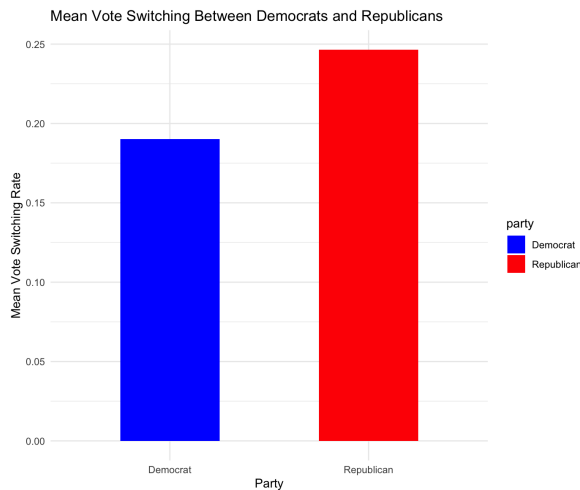
**Alternative Hypothesis:** The mean voting switching rate for Democrat voters is different from the voting switching rate for Republican voters.

## 4 Test Selection and Assumptions

An appropriate statistical test for this study is a Welch's Two-Sample t-Test. The first assumption for a two-sample t-test is that our data is independent and identically distributed. Since each voter's party affiliation, as well as their difficulty deciding who to vote for, should not have any effect on other voters' affiliation or decision, this assumption is valid for our data. The second assumption is that the random variable of interest is on a metric scale. Since our decision variable is binary, it can be treated as a metric variable, since the rule for metric variables is that the intervals between each level are equivalent. Since there are only two levels, 0 and 1, this assumption is

met. The last assumption is that there are no major deviations from normality, given the sample size. This assumption can be assessed indirectly using the Central Limit Theorem, which states that the distribution of the sample mean will be approximately normal if the sample size is large enough. Since our data is binary, we cannot test for normality, but with our large sample size of 758 data points, we can reasonably apply the central limit theorem.

## 5 Test Results and Interpretation



| Metric                         | Value               |
|--------------------------------|---------------------|
| p-value                        | 0.06206             |
| Democrat mean vote switching   | 0.19012             |
| Republican mean vote switching | 0.24646             |
| Test statistic                 | -1.8688             |
| 95% confidence interval        | [-0.11552, 0.00285] |

The p-value of 0.06206 indicates that there is not strong enough evidence to reject the null hypothesis, with a typical alpha level of 0.05. Furthermore, the confidence interval of our test statistic (Democrat - Republican mean vote switching) contains 0, so we cannot confidently state that there is a difference in mean vote switching between Democrats and Republicans. Additionally,

the sample size may not have been large enough to detect a statistically significant difference, so further research with a larger sample size may provide more insight into the difference of means, because having a larger sample size would improve the power of the test.

## 6 Conclusion

This study investigates the voting difficulty faced by Democrat and Republican voters, with the key variable being whether voters would switch their vote from their party's candidate. Using Welch's t-test to compare voting difficulty between the two groups, we found no statistically significant difference, with a p-value of 0.06206. Thus, we fail to reject the null hypothesis.

The reliance on self-reported data and a specific subset from the ANES 2022 Pilot Study may limit generalizability. To gain more accurate insights, future research with larger datasets with more objective variables would be beneficial.

## References

- [1] American National Election Studies. Anes 2022 pilot study [dataset and documentation]. December 14, 2022 version. <https://www.electionstudies.org>.