

# Lab 1 Part 2: Analyzing Voting Difficulty

DATASCI203 Section 1 Team 5

# 1 Importance and Context

While every U.S. citizen over the age of 18 has the right to vote, many factors can impede individuals from exercising their voting rights. In recent elections, political parties have argued over voting accessibility including mail-in ballots and whether to close polling places while voters are still in line. Generally, Democrats [have argued for more accessibility and fewer voting restrictions](#), while Republicans have argued for the opposite. This alignment flows from [how the parties see voting access benefiting them electorally](#). Since the 2020 election, when many jurisdictions [increased access](#) to things like mail-in ballots, the debate has intensified. And many states have since [rolled back those changes](#). In this context, our central question is,

**Do Democratic voters or Republican voters experience more difficulty voting?**

By researching this question, we hope to add to the discussion surrounding discrepancies in the voting process. We hope that highlighting access issues in elections will help enable equal voting rights for all U.S citizens.

## 2 Data and Methodology

Our analysis is based on the 2022 data from the American National Election Studies (ANES), a widely respected national survey of voters conducted before and after every presidential election.

The data cleaning process included defining variables of interest including “voter,” a voter’s political party, and “difficulty” in voting. From the original 1585 entries of 18+ citizens, we narrowed down the result to a sample of 1308 by focusing on weighted entries intended to represent the population. We defined voters as registered voters in the ‘reg’ column . We did not use historical voting turnout to define a voter, because some voters may face obstacles preventing them from voting, which is relevant to our research.

To determine political alignment, we relied on survey questions captured under the ‘pid1d’, ‘pid2d’, ‘pidlr’, ‘pidlean’, and ‘pid2r’ columns. We classified political affiliation by any voter who identified as belonging to a party or as leaning closer to the Republican or Democratic party. Anyone who said they leaned towards “Neither” was excluded from the sample. From Table 1 below, we recorded a total of 567 Democratic, 577 Republican voters, respectively representing 44% and 45% of the sample. The 164 non-affiliated members made up approximately 13% of the sample.

Table 1: Self-Reported Number of Registered voters based on Party Affiliation

Party	Registered Voters
Democrat	567
Neither	164
Republican	577

To measure difficulty voting and capture voters who did not cast a ballot, we developed a binary variable (1 = did experience difficulty; 0 = did not experience difficulty) using a set of ANES survey questions. All voters who cast a ballot were asked: “How difficult was it for you to vote?”. Anyone who experienced any difficulty (2-5) was coded as a 1. Moreover, other respondents including non-voters that experienced any of the 11 potential voting difficulties (vharder\_0 through 11) were coded as 1. We used two additional questions (waittime, triptime) to code anyone who said they had to wait in line longer than 30 minutes or travel more than 15 minutes each way to reach their polling place as 1/‘did experience difficulty,’ as the time commitment needed to vote exceeded a working person’s usual lunch break. All other respondents were coded as a 0. From Figure 1, we observed that Democrats had fewer voters falling into the “did not experience difficulty” compared to Republicans.

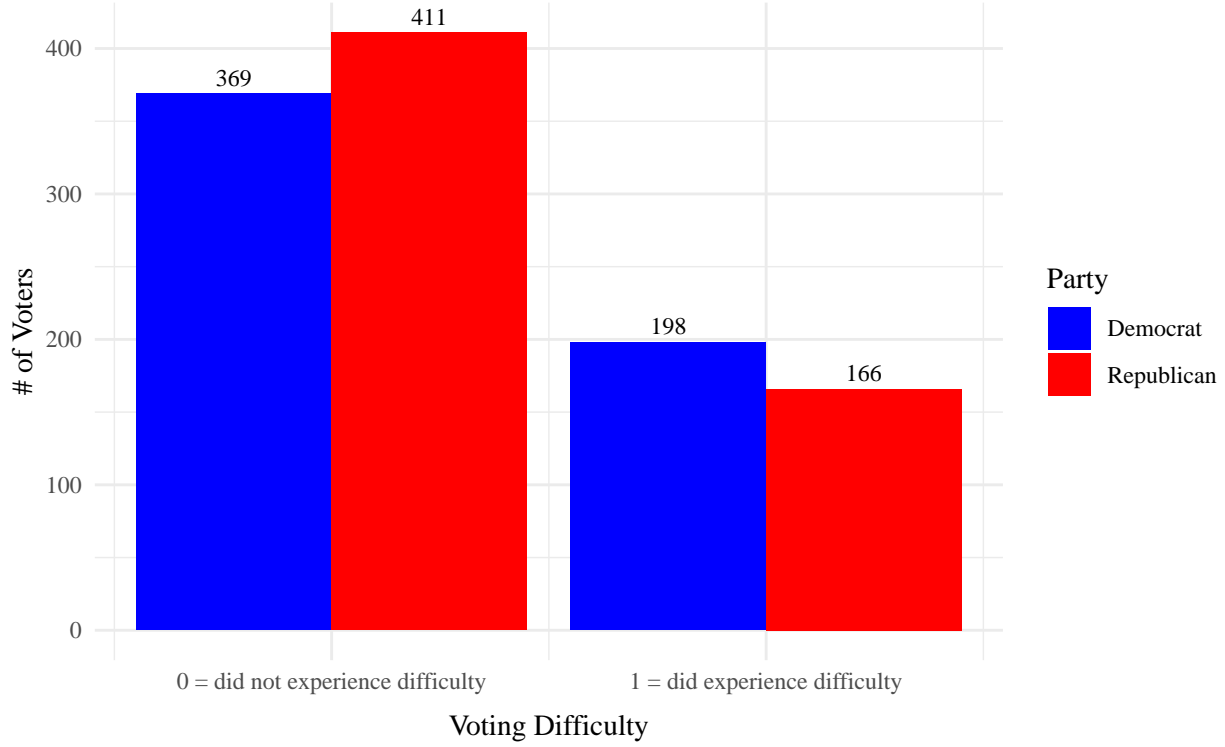


Figure 1: Bar Chart of Voting Difficulty by Party Affiliations

To further evaluate the significance of the size difference between Democratic and Republican voters' difficulty, we used the two-sample t-test, a robust statistical test for metric data.

Our null hypothesis was that there was no difference in voting difficulty between Democratic and Republican voters. In the context of the two-sample t-test, our null hypothesis was that:

*For  $R$  being the number of Republicans who experienced difficulty voting and  $D$  being the number of Democrats who experienced difficulty voting,  $E[R] = E[D]$*

To produce reliable inference from the two-sample t-test, we must meet three assumptions.

**Assumption 1:** Our variables must be measured on a metric scale. In this case, we have created a Bernoulli indicator variable to represent whether or not a single person in our sample experienced difficulty voting. Our resulting binomial variables ( $R$ ,  $D$ ) consist of counts, which are measured on a metric scale, so we meet this assumption.

**Assumption 2:** Our data must be IID. The ANES 2022 pilot uses a panel of individuals from the YouGov platform which could lead to the possibility of introducing dependencies. For example, participants may tell relatives or friends about YouGov, resulting in a cluster of individuals that give similar responses in specific areas. YouGov users presumably have internet access, which means people who are unhoused or otherwise experience internet access issues may not be represented. Nevertheless, YouGov claims to have millions of users, and the ANES weights its responses to better reflect the voter population, so we believe performing this test on this data will be useful for research purposes.

**Assumption 3:** There must be no major deviations from normality, considering our sample size. Since we have binomial variables, our theoretical density curves for  $E[R]$  and  $E[D]$  approach the normal distribution as the sample size increases. We have a large sample size (567 Democrats and 577 Republicans), so the Central Limit Theorem allows us to presume our sample values for  $E[R]$  and  $E[D]$  come from approximately normal distributions.

### 3 Results

```
# Perform a two-sample t-test
result <- t.test(republican_data$difficulty_binary, democratic_data$difficulty_binary)
result

##
## Welch Two Sample t-test
##
## data:  republican_data$difficulty_binary and democratic_data$difficulty_binary
## t = -2.2352, df = 1136.6, p-value = 0.0256
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.115505181 -0.007517569
## sample estimates:
## mean of x mean of y
## 0.2876950 0.3492063
```

The Welch Two Sample t-test on our data yielded a p-value of .025, well below the .05 threshold for statistical significance. Our t-score of -2.24 with 1136 degrees of freedom suggests that the difference in means is fairly large, though not huge, relative to the variability within the groups and its negative nature suggests that Democratic voters experience more difficulty in voting compared to Republican voters.

In practical terms, approximately 29 percent of Republicans in our sample experienced some difficulty voting, compared to 35 percent of Democrats, a 6-percent difference. Considering how close elections can be, a 6-percent difference could impact the outcome of an election if even a modest percentage of voters who experience difficulty voting end up not casting a ballot.

Another practical significance of our findings is that what could be considered a large percentage of the electorate—nearly one third of Republicans and over a third of Democrats in our sample—experienced barriers to casting their votes. Those numbers may indicate the need for further study to address cross-partisan voting access issues.

### 4 Discussion

In this study, we sought to answer the research question: "Do Democratic voters or Republican voters experience more difficulty voting?" To address this question, we conducted a two-sample T-test to determine the significance of the differences in difficulty voting between Democratic and Republican voters. Our work has limitations: the research question is complex and there are many ways to evaluate the data, though we believe we have chosen a fair and appropriate methodology. Various factors could also have affected our sample. A possible limitation is if respondents within the same community have similar experiences, it may motivate similar individuals to participate potentially creating a clustering of similar opinions. Moreover, regional differences, voting laws, and individual experiences can all impact people's perceptions of their voting experience. We cannot account for all the difficulties voters may have experienced that were not captured by the ANES survey. Nor can we account for people who had similar voting experiences, but answered survey questions about them differently. And finally, our chosen variables account only for whether or not a person experienced any difficulty voting, not the degree to which they experienced difficulty voting, so we cannot determine from our test whether one group tended to experience more extreme difficulties than the other.