Party Affiliations and Voter Difficulties: A Statistical Perspective

Ryan Farhat-Sabet, Maia Kennedy, Skylar Wang, William Seward

University of California, Berkeley

Contents

Importance and Context

The right to vote is the foundation of American democracy. However many Americans face significant obstacles in exercising this fundamental right. These obstacles range from lack of election information to barriers at polling places.(1) The coronavirus pandemic further exacerbated these barriers.(2) According to the 2022 American National Election Studies (ANES) Pilot Study, 69% of registered voters reported encountering some form of difficulty while voting.(3) Furthermore, these challenges disproportionately affect certain groups, leading to voter suppression and reduced turnout.(4)(5)

Research shows that Democratic and Republican voters often face different voting difficulties. Understanding these differences is critical because voting obstacles can influence voter participation and potentially alter election results. This analysis aims to address the following research question:

Do Democratic voters or Republican voters experience more difficulty voting?

We are an analytical group that hopes the answer could reveal disparities in the voting experience between these groups and highlight whether specific challenges disproportionately impact one party. Identifying systemic biases or voter suppression linked to demographic and geographic factors can provide valuable insights for governments aiming to make voting more accessible and fair. A better understanding of these challenges can help develop targeted initiatives to reduce obstacles, increase voter turnout, and ensure a more equitable voting process for everyone.

Data and Methodology

Defining Terms & Data Cleansing

Our analysis leverages data from the 2022 ANES Pilot Study. This data comes from an incentivized online survey, where participants are eligible to receive 21 to 50 cents for each survey they complete. This means that only those who use the internet and those who know about the platform are eligible to participate, which is not a fully representative sample of the national US population, but it is a good place to begin understanding tendencies of the American voter.

Before we attempt to answer our research question, we first need to define a few of our terms: who is a voter, how do we classify them as Democrats or Republicans, and how do we define difficulty voting. Looking at our data, we have surveyed 1585 people in total. From there, we define 'voters' as individuals who are registered to vote (based on the 'reg' variable), eliminating 196 people. We believe this is the best way to define a voter using the ANES survey data, as this method allows us to include individuals who possibly did not vote because it was too difficult. Next, we tackled how to define Democrats and Republicans. First, we examined the variables 'pid1d' and 'pid1r' to separate those who

self-identify as "Democrat" and "Republican". We then examined the variable 'pidlean' to identify individuals who self-identified as 'Independent' or 'Other' but leaned toward either the Democratic or Republican parties. We categorized them accordingly to better capture a broader representation of our population. This narrows us down to 1219 voters.

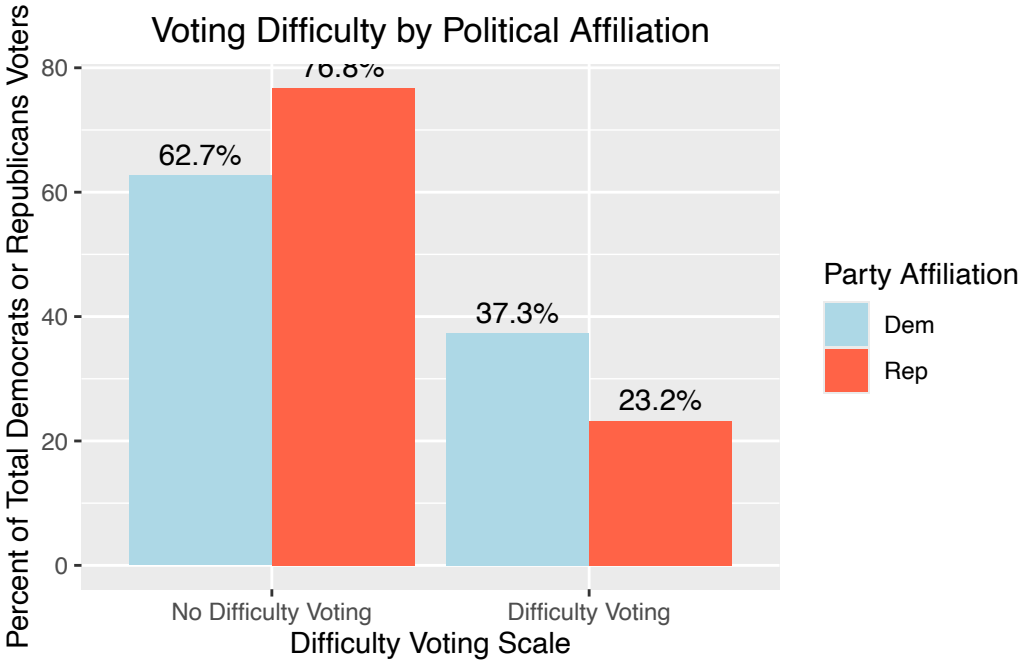
Table 1: Counting Table for Democrats & Republican Voters

Cause	Samples_Available	Removed_Samples
Start	1585	NA
Voters Only	1389	196
Democrats & Republicans	1219	170

Lastly, we had to define difficulty voting. There were a few ways we could think about this. We initially looked at the 'votehard' metric, as it seemed like a good measure of how hard people feel it is to vote. However, we realized this variable only accounts for individuals who voted in the last election, which doesn't include the entire population sampled and doesn't align with our definition of voter. We pivoted to looking at the 'vharder' variable, since it was filled out for every survey participant. The question asks people to mark reasons it might be hard for them to vote, with the final option ('vharder_12') being "none of these", indicating that it was not hard for them to vote. We used this as a proxy to measure difficulty voting. We split our population into two groups, those that had marked anything for 'vharder' for questions 1 - 11 totaling 844 and those who had marked "vharder_12" or "none of these" totalling 375. While this means we can't measure the magnitude of difficulty like we might have been able to with 'votehard', this binary variable is clearer and more standardized than the ordinal nature of the scale provided in 'votehard', as it clearly defines who has or has not experienced difficulty voting.

Putting it Together

The following figure 1 demonstrates the percentage of total Democrats or Republican Voters who did or did not experience difficulty voting, as defined by our terms above. This representation is critical in supporting our test to determine whether there is a difference in difficulty voting between the two party voters. Based on our analysis, approximately 14.1% more Democrat Voters experienced difficulty voting compared to Republicans. The inverse is true for Republicans who did not experience difficulty voting. Because of the observational nature of this dataset, we cannot directly measure whether being a Democrat or Republican will cause a higher difficulty in voting experience. We can only check whether Democrats or Republicans are more likely to claim difficulty in their voter experience.



Hypothesis Testing

Both our grouping variable (party affiliation) and our outcome variable (voting difficulty) are measured at the binary level (i.e. Democrats vs. Republicans, difficulty voting vs. no difficulty voting). Although we initially computed a two-sample t-test to demonstrate one of the tools taught in DATASCI 203, we instead decided to pursue a more appropriate test for binary data by using the chi-squared test.(6)(7) The null hypothesis of our chi-squared test can be phrased as follows:

Null Hypothesis: There is no association between political affiliation (Democrat vs. Republican) and voting difficulty ("yes, I experienced difficulty" vs. "no, I did not experience difficulty"). In other words, Democrats and Republicans are equally likely to experience difficulty or no difficulty voting.

The chi-squared test is well-suited for analyzing the relationship between categorical variables by comparing observed frequencies to expected frequencies under the proposed null hypothesis. In order for the chi-squared test to produce reliable inference, the following assumptions must be true: the variables considered must be categorical, each variable must consist of observations that are independent of each other, and there must be a large sample size.

First, each variable must be categorical. Party affiliation (Democrats vs. Republicans) and voting difficulty ("yes, I experienced difficulty" vs. "no, I did not experience difficulty") are both binary categorical variables.

Second, each variable must consist of observations that are independent of each other. Each voter's experience is independent of another, and Republicans and Democrats are independent of each other. There is no claim about needing our variables to come from the same distribution,

but we can assume that the participants from the survey are pulled from the greater US population.

Third, there must be a large sample size. The expected counts per variable must be at least 5 for the chi-squared test to be valid, which is satisfied. Smaller counts per variable would open the door for potentially using Fisher's exact test.

Results

Table 2: Difficulty Voting Between Democrats and Republicans

	No Difficulty Voting	Difficulty Voting
Dem	408	243
Rep	436	132

```
chi_test <- chisq.test(voting_data)
```

The test yields evidence that Democrats are more likely to experience difficulty voting than Republicans (X-square = 27.607, p-value = 0.0000001486). From a practical perspective, these results appear to be very highly significant. In the survey, 37.33% of Democrats reported experiencing some form of difficulty voting, compared to only 23.23% of Republicans who said the same.

Several limitations of our test affect the conclusions that may be drawn from it. These results measure a possible association, rather than any sort of causality. We highlight the fact that identifying as a Democrat or Republican does not mean that one will experience difficulty voting, nor does it mean that one party definitively experiences more difficulty voting than another. Our results may not generalize to the US population, as the ANES data is not nationally representative.

Discussion

This study found that Democrats are more likely than Republicans to deal with difficulty when voting. These results are practically important as clearly shown in Figure 1. Within the bounds of the survey, there is clearly a discrepancy between Democrats and Republicans who have dealt with some form of difficulty when voting.

The difference in difficulties experienced by Democrat and Republican voters may influence election results. Our findings may be of key interest to policymakers, public policy institutions, and election officials who are committed to ensuring fair and accessible voting for all citizens. By identifying potential difficulties in the voting process, this analysis can guide more targeted

approaches to increasing voter turnout and fight against voter suppression. While this study compares difficulties between Democratic and Republican voters, future research could explore difficulties faced by specific demographic groups, such as by race, age, disability status, and how to address these obstacles. Ultimately, a better understanding of voting difficulties will help promote fair elections, improve voter turnout, and ensure that every voice is heard.

References

1: "Executive Order on Promoting Access to Voting." (2021). https://www.whitehouse.gov/briefing-room/presidential-actions/2021/03/07/executive-order-on-promoting-access-to-voting/

2: Pew Research Center. "3. The voting experience in 2020." (2020). https://www.pewresearch.org/politics/2020/11/20/the-voting-experience-in-2020/

3: 69% is the percentage of registered voters based on the 'reg' variable that did not mark "vharder_12" or "none of these", indicating they experienced some form of difficulty (marked at least one difficulty out of variables "vharder_1-11")

4: American Progress. "Enhancing Accessibility in U.S. Elections." (2021). https://www.americanprogress.org/article/enhancing-accessibility-u-s-elections/

5: Brennan Center. "The Impact of Voter Suppression on Communities of Color." (2022). https://www.brennancenter.org/our-work/research-reports/impact-voter-suppression-communities-color

6: Stark, Philip B. "The Multinomial Distribution and the Chi-Squared Test for Goodness of Fit." University of California, Berkeley, n.d., https://www.stat.berkeley.edu/~stark/SticiGui/Text/chiSquare.ht

7: Rice, John. "Chi-Square Tests: Testing for Goodness of Fit and Independence." *University of California, Berkeley*, n.d., www.stat.berkeley.edu/~rice/Stat2/Chapt28.pdf