Analyzing Voting Difficulty

Datasci 203: Lab 1 Part 2

Moonsoo Kim, Ryan Beaty, and David Dorfman

Introduction

The question of whether either US political party faces more difficulty voting has heightened importance in today's politically charged climate. Disparities in voting experiences can signify systemic issues that undermine democratic integrity.

Utilizing the 2022 Pilot Study from the American National Election Studies (ANES), this analysis aims to explore the voting difficulties encountered by individuals aligned with each party. We then apply statistical tests to quantify the significance of our findings in answer to the question:

Do Democratic voters or Republican voters experience more difficulty voting?

Conceptualization

1. Who or what is a voter?

In the United States, a voter is a citizen who is eligible to participate in federal, state, or local elections by casting a ballot. The basic eligibility criteria for voting in federal elections are: U.S. citizenship, at least 18 years old on election day, residency in the state where one is registering to vote, and not being disqualified due to mental incompetence, criminal convictions, or other specific reasons.

2. Who is a "Republican" and who is a "Democrat"?

In the United States, Republicans and Democrats are members or supporters of the two major political parties. Republicans generally lean conservative, favoring lower taxes, less regulation, and a strong national defense. Democrats typically lean liberal, advocating for social justice, higher social spending, and international diplomacy. These are generalizations, and individuals within each party may have diverse views.

For purposes of this analysis, we aim to classify voters into one of these two groups. This simplifying assumption means that varying degrees of support including leaning, identifying, or strongly identifying with a party will be classified as a party supporter (Democrat or Republican).

3. What is difficulty voting?

Difficulty voting refers to challenges or obstacles individuals may face when attempting to participate in an electoral process. These can vary from logistical issues to systemic barriers, and may include:

- Voter ID Laws: jurisdictions may require specific identification which not all eligible voters have.
- Accessibility: Physical or digital barriers can make it challenging for disabled individuals to vote.
- Long Wait Times: Insufficient polling places or resources can result in long lines.
- Voter Suppression: Tactics such as gerrymandering aimed to discourage specific groups from voting.
- Registration Barriers: Complicated or inconvenient voter registration processes can deter participation.

Operationalization

The best **possible** method of measuring difficulty in voting among Republicans and Democrats would be to get complete information on each eligible voter immediately after the election on how difficult they found it to vote, and what party they support. This might be feasible using extreme mass surveillance and mandatory surveying of all eligible voters.

However, the best **available** method of measuring this concept is to use existing ANES survey data that measures respondent's party alignment along with reported difficulty voting.

The ANES 2022 Pilot Study is a cross-sectional survey based on a sample of respondents drawn from YouGov, a platform where respondents receive rewards for completing surveys. The YouGov panel is not necessarily

representative of the U.S. population, for which correction factors have been included but not factored in to our analysis.

The ANES data includes the following key variables of interest to answer the question at hand:

1. caseid: unique identifier

2. pid_x: summary of Party ID (1 strong democrat, 2 democrat, 3 lean democrat, 4 independent, 5 lean republican, 6 republican, 7 strong republican)

3. votehard: How difficult was it for you to vote? (-1 inapplicable, legitimate skip; 1 Not difficult at all; 2 A little difficult; 3 Moderately difficult; 4 Very difficult; 5 Extremely difficult)

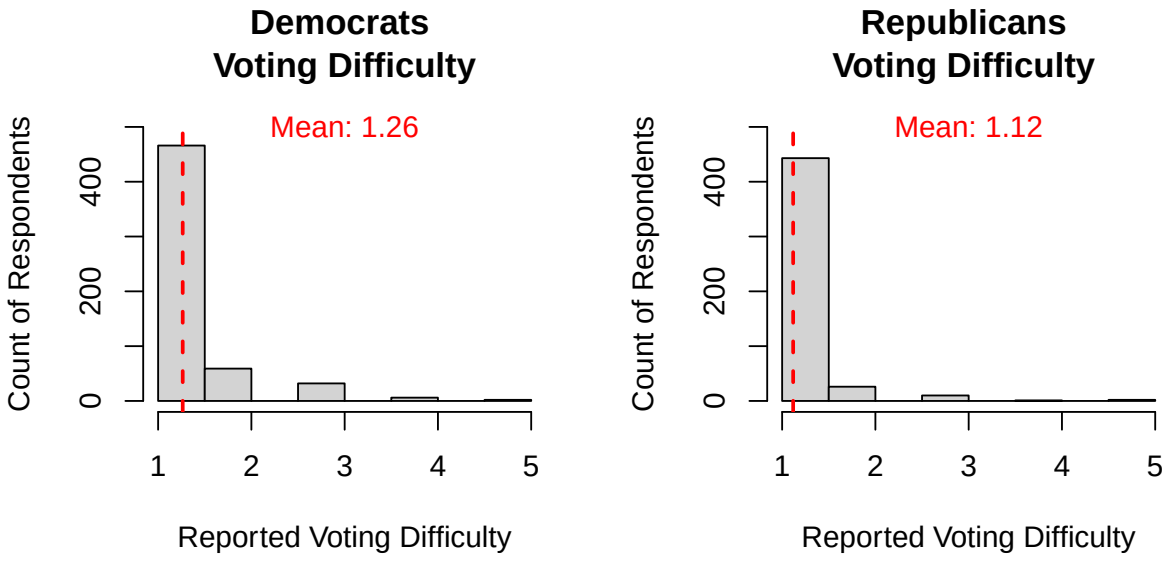
Data wrangling

We begin by reducing the dataset to the above three variables which will look at the relationship between party affiliation and difficulty voting.

Next, we eliminate independents (273 '4' responses). This enables a direct comparison of the two parties.

Lastly, we remove rows with voting difficulty response of 'inapplicable / legitimate skip' (263 '-1' responses).

We then summarize the remaining data using two histograms - one for Democrats, and one for Republicans. Additionally, we layer on the mean of each group to act as a simple summary statistic.



A preliminary review of the histograms and means of the data confirms the two groups share a non-normal distribution with similar variances.

Hypothesis testing

Due to the ordinal nature of the difficulty ratings in the ANES data, we must use the Wilcoxon Rank-Sum Test as the most appropriate way to answer the research question.

Additionally, the data compares independent groups, thus an unpaired test is appropriate.

The null hypothesis for this test is that there is equal probability that the difficulty values for Democrats are greater than Republicans, as the reverse (Republicans greater than Democrats).

Wilcoxon Rank-Sum Test Assumptions:

1. **Independence:** This assumption assumption is met as one respondent's answer should not influence or inform the difficulty of voting for another. Geographic clustering effects are a potential risk should respondents from the same region experience the same logistical voting challenges.
2. **Continuous or Ordinal Data:** Met, as the difficulty ratings are ordinal values from 1 to 5.

```
wilcox.test(voting2$voting.votehard ~ voting2$party, paired = F)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  voting2$voting.votehard by voting2$party
## W = 149130, p-value = 6.011e-06
## alternative hypothesis: true location shift is not equal to 0
```

Test, results and interpretation

As the p-value is less than 0.05 at ~.000006, it suggests that the probability(x < y) <> probability(x > y). Therefore we reject the null hypothesis that there is no difference in difficulty voting reported by Democrats vs Republicans.

The mean value is not necessarily descriptive in and of itself with ordinal data. However, given the reported difficulty range of (1 - no difficulty to 5 - very difficult), we can use the mean to describe the average experience relative to 1 (no difficulty).

Using the histograms and means shown above, we visually and quantitatively confirm the greater number of Democrats reporting difficulty voting.

In conclusion, the data suggests a statistically significant increase in difficulty voting for Democrat vs Republican voters.

Limitations

Based on ANES data field turnout22 (whether someone voted on November 8), we can confirm that a 223 of 1585 respondents did not vote. This confirms that the difficulty voting was captured even for respondents that ultimately experienced such difficulty that they did not vote.

Removing the non-voters results in a Wilcoxon Rank-Sum test that still rejects the null, with a P-value of ~0.002. For the purposes of this analysis, we chose to include both actual (voters) and perceived (non-voters) difficulty voting.

Practical Significance

If one group consistently finds it harder to vote, their voices may be underrepresented in electoral outcomes. Democrats' materially higher mean of 1.26 vs Republicans' 1.12 suggests a disparity in voting difficulty.

As the sample size of this survey is very small relative to the U.S. population at 1,585 respondents of a total population of ~330M [https://www.census.gov/popclock/], we hesitate to draw conclusions on systemic issues that undermine democratic integrity.

However, the high degree of significance in the Wilcoxon Rank-Sum test suggest that a more expansive study encompassing a larger and more diverse sample may prove informative.
