Democrats or Republicans: Who has more difficulty voting?

Blake Bormes, Evan Fjeld, Jujhar Bedi

Introduction

Long lines, inability to produce a government ID, and difficulty getting to polling booths are all reasons voting can be difficult on election day. For decades there have been discussions about the most just and easiest way to allow US citizens to vote. From volunteers who bus people to polling booths and a plethora of resources to assist with casting an absentee ballot, both Democrats and Republicans have assisted Americans in the voting process.

Which party needs more assistance? Are Democrats or Republicans having a more difficult time voting? Answering this question will give us a better understanding of party demographics and which voters require more assistance on election day.

With data from the American National Election Studies (ANES), a federally funded organization, we will analyze this question with statistical techniques to determine if Republicans or Democrats have more difficulty voting. ANES produces high quality voting data through surveys to inform explanations of election outcomes. In this experiment, we will be using ANES's 2020 Time Series Study data to understand Democrats' and Republicans' difficulty voting. The variables or survey questions being analyzed are "How difficult was it for [you] to vote?" and "Party ID" and the statistical test we will use to determine if an individual's party affiliation is significant when determining how difficult it is for one to vote is the Wilcoxon Rank Sum Hypothesis of Comparison Test.

Determining the best variable to answer the question

The two variables chosen from the data set were "Party ID: V201231x" and "How difficult was it for [you] to vote: V202119," which we will refer to as "Voting Difficulty." Party ID is a range from 1 (Democrat) to 7 (Republican) of how left or right leaning a survey respondent is, with 4 meaning an individual is independent.

According to John Richard Petrocik's "Measuring Party Support: Learners are not Independents," 40% of citizens are registered as independent but many of them tend to vote in line with a specific party. Using a measure of people's behavior and beliefs instead of their voter registration is a much more accurate way to determine someone's political affiliation. Additionally, for an indicator of voting difficulty we chose the variable that asked respondents on a scale from 1 (not difficult at all) to 5 (extremely difficult) how difficult it was to vote. This likert scale variable is a very clear ordinal measure of voting difficulty. Although there were other variables within the data that measured if weather, registration, identification card, or other problems occurred while voting, we determined that the variable clearly asking respondents about voting difficulty was the best measure.

General Assumptions

- Voters are truthful when the respond as to which party they learn towards.
- We will remove any respondents that refused to answer the to two questions used for this research and assume that those who refused are not signifincant in numbers are similarity to skew our results.
- We will use the general question of 'How difficult was it to vote?' as a catch-all for the various reasons why a person found it more or less difficult to vote. This assumes that the general question captures all reasons and the respondent was able to weigh all potential blockers and distill that down to one response.

- 1,879 people chose to ignore the question about how difficult it was to vote. We will assume these people did not have difficulty voting. It is not unreasonable to assume that those who did have difficulty voting are very likely to answer this question while those who had no trouble voting would pass it by.
- All survey respondents will be included regardless of whether they voted in 2020. This is due to the fact that we need to capture voters who voted but were unable to due to blockers in place that prevented them from participating in the electoral process.

Test and Hypothoses

The Wilcoxon Rank-Sum Hypothesis of Comparison Test will be used to measure if whether an individual is Democrat or Republican has a statistically significant effect on how difficult it is for an individual to vote. This test will allow us to measure the ordinal variable of difficulty voting and the unpaired sample sets of Democrat and Republican voters. Our null hypothesis will be that the probability of Democrats having a harder time voting than Republicans is equal to the probability of Republicans having a harder time voting than Democrats.

Given Democrat voting difficulty is D_D and Republican voting difficulty is represented R_D

$H_0 : D_D = D_R$

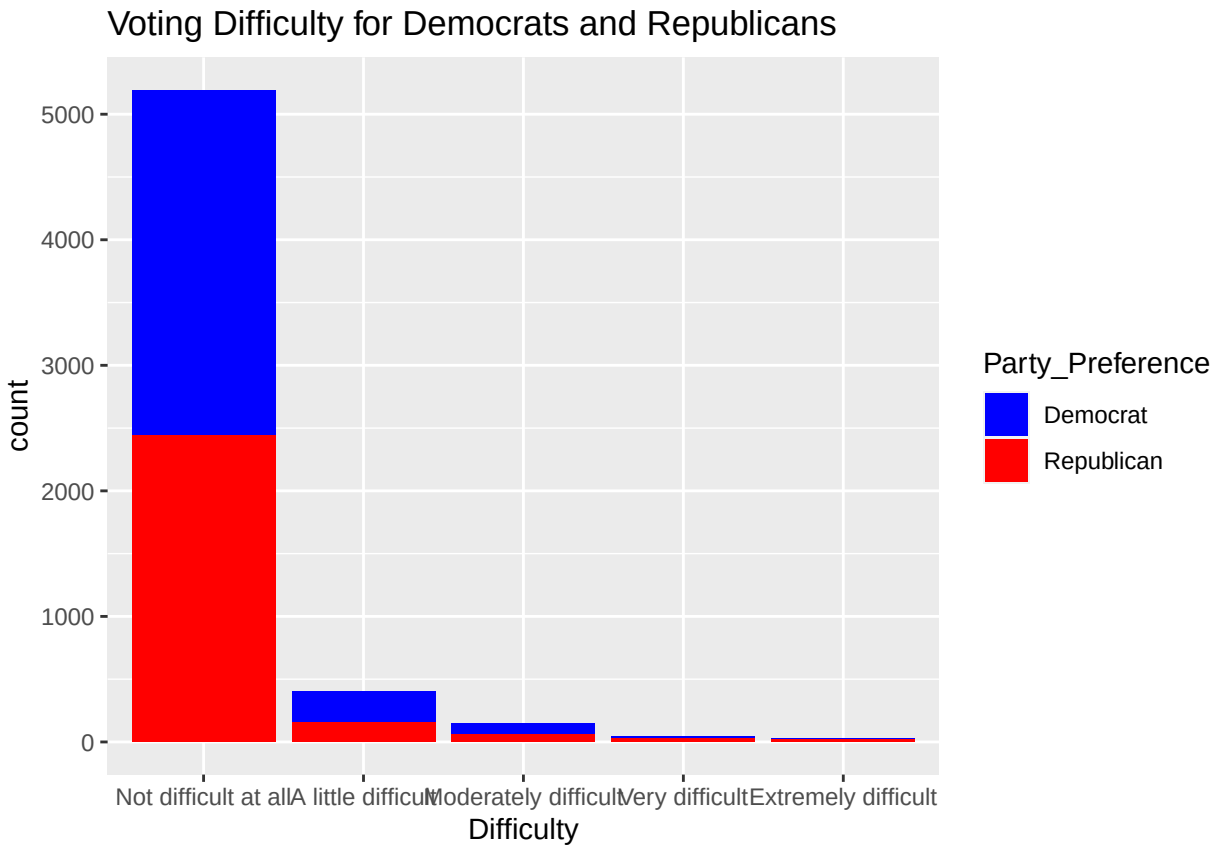
The alternative hypothesis will be that the probability of Democrats having a harder time voting than Republicans is different than the probability of Republicans having a harder time voting than Democrats.

$H_a : D_D \neq D_R$

Test Assumptions

The test assumptions required for a Wilcoxon Rank-Sum Hypothesis of Comparison Test are listed below with explanations on whether the assumptions are met and any limitations.

- Ordinal data is being analyzed.
  - The difficulty voting variable is an ordinal variable that survey respondents used a likert scale to indicate level of voting difficulty on a scale from 1 (not difficult at all) to 5 (extremely difficult)
  - The Republican or Democrat variable is a binary indicator of a survey respondent's party affiliation used to represent the two sample populations that will be used in the test.
- The data sample is unpaired.
  - The data is unpaired because Democrat survey respondents and Republican survey respondents are two separate populations of individuals.
- Data is Independent and Identically Distributed (IID).
  - The ANES data is gathered by video interviews before and after participants vote. The participants sign up for the survey in exchange for rewards. Selecting only people who register for a survey in exchange for rewards is not a random sample. There is some bias within the survey population. This means that whether a respondent is a Democrat or Republican and a respondent's voting difficulty could make the variables not independent.
  - As seen below, the distributions of difficulty voting for both Republican and Democrat populations are identical.
- The data is not parametric
  - As seen below, the distributions of difficulty voting is not parametrically distributed, it is a skewed decreasing distribution.




```
wilcox.test(as.numeric(Difficulty) ~ Party_Preference,
            data=df_new)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  as.numeric(Difficulty) by Party_Preference
## W = 4323774, p-value = 0.003541
## alternative hypothesis: true location shift is not equal to 0
```

Conclusion

Our study concluded that voting difficulty between republican and democratic voters was not equivalent. We reject the null hypothesis. The voting difficulty is not equal for Democrats and Republicans. The p-value of 0.003541, obtained from our statistical analysis, was significantly less than the significance level of 0.05. This result aligns with our expectation since the 2020 election was highly contested and marred by a global pandemic. Voting via mail-in ballots was highly politicized and it is possible that voters that did not have this option might have found it more difficult to vote due to the pandemic.

With that being said, there are several limitations in our test that need to be mentioned. First of all, the data collected from ANES is not necessarily representative of the entire population of voters in the US due to the way it was generated. Additionally, voting difficulty is subjective and individuals may quantify the same hardship differently. It is important to consider these factors as they may have impacted the results of our study. Future studies may focus more heavily on the specific difficulties that plague voters in order to simplify the voting process.