Lab 1: Question 2:

Devashish Kulkarni, John-Michael Stilb, March Saper

Contents

Did Democratic voters or Republican voters report experiencing more difficulty voting in the 2020 election?

Importance and Context

From the inception of the U.S., strategies to limit the voice of the people have been used to control the course of democracy - limiting who could vote, imposing impossible literacy tests, and now in the 21st century through the passage of voter suppression laws which lead to significant burdens for eligible voters trying to exercise their constitutional right. Since 2008, states have passed increasingly strict measures to make it harder for Americans - particularly demographics which can be assumed to skew Democrat - to cast a ballot. In 2020, this ongoing crisis combined with a global pandemic which prevented large indoor gatherings of people, impacted the ability of individuals to work in-person creating an environment which further impacted voting access. In the midst of this turmoil, were both Democrats and Republicans equally impacted by voting access challenges?

Description of Data

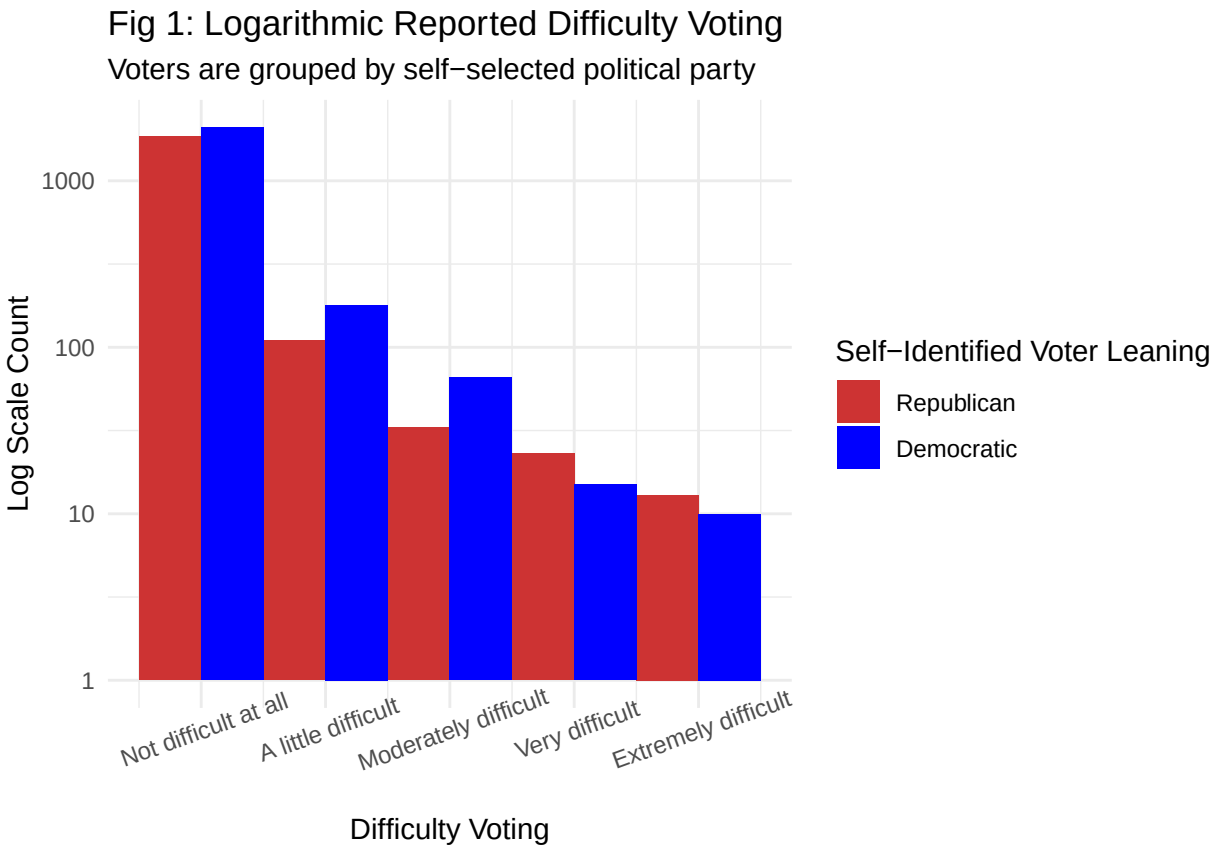
The data to answer this question comes from the ANES 2020 Time Series study. This is an observational data set comprised of pre and post election interviews administered online and through the phone. In total 8020 participants responded to the survey.

In order to examine the question of whether Democratic or Republican voters were more/less affected by difficulty voting, we need to operationalize two variables: (1) difficulty voting and (2) party leaning.

We use self-identified party to determine party since not all states require a voter to register under a specific party. We also did not want to use past voting history to determine party since we are trying to measure voting access and this might remove a critical population. Roughly 1/3 of respondents either did not identify with either the Republican or Democratic party or declined to answer. After subsetting, we were left with 5428 respondents who identified as Republican or Democrat (2864 Democrat and 2564 Republican).

The survey includes a question quantifying how difficult a respondent found it to vote on an ordinal scale from not difficult at all to difficult. We cleaned this data by removing respondents who did not complete the survey, or who answered "not applicable" to this question, perhaps due to voting by mail etc. After removing these respondents we were left with 2360 Democrat and 2028 Republican respondents to use for our testing.

To explore the data we examined a histogram of responses to the question quantifying difficulty voting by self-identified party.



For both Republican and Democratic voters there is a strong skew towards 1 - not difficult at all. This can be seen in Figure 1. (Please note the use of log scale to adequately display the data.) For both parties there is an order of magnitude greater responses to "not difficult at all" than to all other ratings combined. In addition, neither party has a noticeable skew towards more or less difficulty voting.

Most appropriate test

Since we are considering two disparate populations (unpaired) represented by ordinal data, the non-parametric Wicoxon rank-sum test will be appropriate for our null hypothesis. This test is also ideal for data with such a significant skew. Specifically, we will use the Hypothesis of Comparison because our scale is not metric.

The assumptions for the Wicoxon rank-sum hypothesis of means test are:

- Ordinal variable - The survey responses to difficulty voting are based on a Likert-style scale from least to most difficult. An ordinal variable does not require a proportional scale, only a single direction which our data satisfies.
- IID Sample - The ANES study follows survey best-practices which maximize likelihood of IID. Respondents are randomly selected and the number of samples is sufficiently large.

We will use the following null and alternate hypothesis for the test.

- Null hypothesis: There is no difference in the difficulty voting experienced by Democrats and Republicans
- Alternate hypothesis: There is a difference in the level of difficulty voting experienced by Democrats and Republicans.

We will use a two-tailed t-test to achieve robust comparison between the groups. We will look for a p-value of less than 0.05 to reject the null hypothesis.

Test, results and interpretation

```
wilcox.test(V202119 ~ is_democrat, data = dem_rep_difficulty_data)
```

The test produces a p-value of 0.005838 and we reject the null hypothesis and accept the alternate hypothesis that Democrats and Republicans experienced different levels of difficulty voting.

To evaluate the practical significance of this result, we can examine the mean difficulty voting response from Democrats and Republicans. On average, the Democratic respondents experienced a difficulty in voting of 1.167 while Republican respondents experienced a difficulty of 1.146. While these numbers are not very noticeably different, they do provide insight that confirms Democratic voters bear the greater burden of difficulty voting. Going forward this has implications for policy-makers and political groups. The ability to exercise one's right to vote is fundamental to democracy and all future elections. We must strive to promote equal and equitable access to voting for all citizens.