Lab 1, Part 2: Hypothesis Testing
Voting Difficulty in the 2020 Election

Bailey Kuehl, Erik Sambrailo, Jenna Sparks, Roshni Tajnekar

2023-03-07

Contents

1 Importance and Context

One of the foundations of a democracy is the ability for individuals to participate in voting. This democratic process allows an individual to vote for leaders that best represent them and their interests. In the United States (US), every four years voters choose a presidential candidate that has historically been identified under two major political parties: Democrats and Republicans. Since there are a variety of methods to vote and these can vary between districts. Every voter's journey to cast their ballot is unique and requires varied levels of effort. Voters can encounter a variety of obstacles as well (i.e. travel time required, long waiting times) which can increase their difficulty to vote. Research from the Pew Research Center suggests that people of different age groups, races, and education levels experienced varying levels of difficulty voting in the 2016 US Midterm Elections.[1] The research in this article suggests that difficulty when voting was disproportionately felt by individuals under the age of 30, black, or having completed less education. The Pew Research Center has also conducted research that shows party affiliation and demographics are often associated with each other.[2] Combining these two studies indicates a potential discrepancy in the effort necessary for individuals of a certain party to vote. As this threatens the validity of the democratic voting process, we want to know if there is a measurable statistical difference between the difficulty experienced by Democrats vs. Republicans when voting in the 2020 US Presidential Election.

Our company, Statistical Answers Inc. has been contracted by US Voting Justice LLC. to answer the following question: **Did Democratic voters or Republican voters experience more difficulty voting in the 2020 election?** Our analysis will contribute to the existing body of work on voter accessibility by examining partisan differences in barriers to voting. This information will have important implications for campaign strategies and policy-making.

2 Data and Methodology

In our analysis, we use 2020 time series data from the American National Election Studies (ANES) [3]. The dataset includes responses from pre- and post-election surveys for 8,280 individuals.

2.1 Definitions of Terms

Before we begin our analysis, we must first consider and define the following relevant terms: What is a "voter"? Who is considered a "Democrat"? A "Republican?" What is meant by "voting difficulty"?

In defining each of these terms, we consider the data tables available in the ANES Codebook. We define a **voter** as someone who indicated that they voted in the 2020 election, regardless of when or how they voted. Note that we considered including those who were registered to vote but could not vote due to extreme difficulties in this definition. See discussion section for details.

To define one's **political party**, we started with a review of self-identified affiliation. In the pre-election survey, individuals were asked a series of questions about their political affiliation and its strength. The results of these questions were used to map individuals by the strength of their affiliation. Every individual showing any level of lean towards Democrat or Republican was categorized into those respective parties. To further validate this categorization, we reviewed post-election survey questions about which political party they felt close to and represented their views. We also looked at their official vote for the presidential candidate. We omitted individuals where these responses contradicted the initial political grouping. While there were few individuals that met this contradiction, we felt this secondary check was necessary for validating our political affiliation definition.

Finally, we define a voter's **voting difficulty** based on a rating of how difficult it was for them to vote on a scale of 1 to 5 (5 being the most difficult). Although we considered counting the number of difficulties reported per individual (e.g., work conflicts, long wait times), we found this approach to be impractical

[1] Pew Research Center. "In their own words: Why some people find voting difficult." (2018).

[2] Pew Research Center. "Wide Gender Gap, Growing Educational Divide in Voters' Party Identification" (2018).

[3] American National Election Studies. "ANES 2020 Time Series Study Full Release [dataset and documentation]" (2021).

because the impact of each difficulty varies and comparing them across individuals in a sample would have been challenging and potentially unreliable.

A summary of term definitions and the ANES Codebook tables from which they were derived from are shown below in Table 1.

Table 1: Defining Voter, Political Party, and Difficulty

	ANES Table	Description
Voter	V202109x	How did you vote? Did you vote? Where are you registered to vote?
Political	V201231x (primary)	Would you call your affiliation strong? Do you lean either direction?
Party	V202443, V202429,	Which political party do you feel closest to? Which best represents your views?
	V202073 (secondary)	Which presidential candidate did you vote for?
Difficulty	V202119	How difficult was it for you to vote in this election?
Voting		

Table 2 demonstrates the modifications made to our dataset in order to filter it down to a finalized, testable set.

Table 2: Accounting Table: Sample Removal

| Reason for Removal	# Samples Removed	# Samples Remaining
Start	0	8280
Select Relevant Columns	0	8280
Non-voters	1830	6450
Neutral Political Affiliation	576	5874
Binary Political Affiliation	0	5874
Validate Political Affiliation	61	5813
Non-responsive Difficulty Results	46	5767

2.2 Wilcoxon Rank Sum

We are now ready to determine which statistical methodology to use in order to test our hypothesis. To conduct a hypothesis test, we must first satisfy the assumption that the data used is independent and identically distributed (IID). We can reasonably assume this because the sample population is approximately random. Based on how we have defined "Democrat" and "Republican", a single individual can be grouped as either a Democrat or a Republican. To measure the difference in voting difficulty between these two groups, we can reasonably consider an unpaired test. Because we have chosen a Likert-scale variable to measure our dependent variable, we cannot utilize an unpaired t-test, which requires metric data. Instead, we will use the Wilcoxon Rank Sum Test to test our hypothesis, which allows for ordinal data. We assert that the data is ordinal since it is qualitative.

We choose our null hypothesis as follows: The probability of Democratic voters (D) experiencing more difficulty voting in the 2020 election than Republican voters (R) is the same as the probability that Democratic voters experiencing less difficulty voting in the 2020 election than Republican voters. Our alternative hypothesis is that the probability of Democratic voters experiencing more difficulty in the 2020 election than Republican voters is not the same as the probability that Democratic voters experiencing less difficulty voting in the 2020 election than Republican voters.

H_0: $P(D > R) = P(D < R)$ and H_1: $P(X > Y) \neq P(X < Y)$

Before beginning our analysis, we wanted to visualize the distributions of difficulty experienced by the two defined groups of Democrats and Republicans. We find the distributions to be fairly similar, and that both groups skewed towards "Not at all Difficult" responses. These distributions are shown below in Figure 1.

Figure 1: Distribution of Voting Difficulty: Democrats vs Republicans
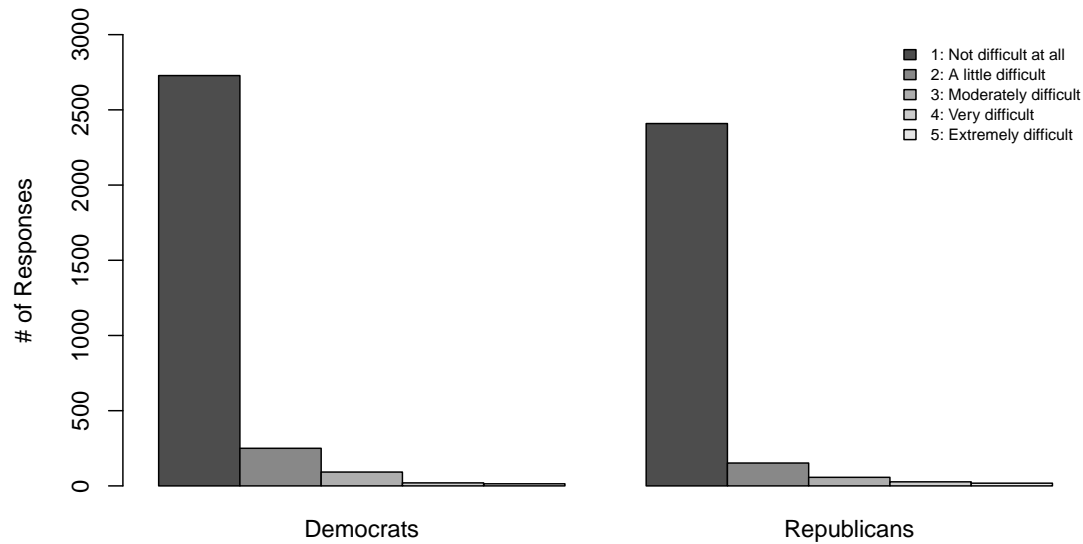
3 Results

Below are our results for the Wilcoxon Rank Sum test.

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  value by group
## W = 4235146, p-value = 0.002743
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
##  -4.705949e-05  3.942171e-05
## sample estimates:
## difference in location
##           3.828745e-05
```

We find that our test resulted in a p-value that is less than 0.05. This indicates that the likelihood of obtaining a difference as large as, or larger than, what we observed by chance is less than 5%. Thus, we reject the null hypothesis that there is equal likelihood of difficulty voting among Republicans versus Democrats, and conclude that the result is statistically significant.

4 Discussion

Our study found that, statistically, Democrats experienced more difficulty when voting than Republicans during the 2020 Presidential Election. However, we must also consider the practical significance of this result, as well as the caveats of our analysis.

First, it is essential to consider that the dataset we analyzed only contained 5,813 voters out of the 154.6 million voters who participated in the 2020 election [4] This means that our survey only represents 0.004% of voters in the election. Thus, our dataset does not represent a substantial portion of the voting population. Additionally, in deciding to keep our definition of "voter" simple (Table 1), we excluded individuals who identified in the survey that they did not vote, or did not respond to that specific question. As a result, 1,830 potential voters were removed from our sample.

While we assumed IID in our sample, we recognize that the nature of our data may pose concerns to this assumption. One concern is that all participants who filled out this survey are likely similar in that they are not easily deterred from difficult tasks, given that they voluntarily responded to this lengthy survey. As our test attempts to compare difficulties among voters, survey respondents may not be completely independent of one another. Another potential violation is that voter party affiliation is associated with geographical regions and there is no verification that the sample was taken in a systematic way to not have influence between participants.

In the future, our analysis can contribute to a broader effort to help political campaigns better understand and address the specific obstacles faced by different groups of voters. Additionally, the insights from our analysis could inform the development of more effective laws and programs aimed at improving voter accessibility.

---

[4]United States Census Bureau. "Census Bureau Releases 2020 Presidential Election Voting Report" (2022).

5 References

American National Election Studies. ANES 2020 Time Series Study Full Release [dataset and documentation]. July 19, 2021 version

Pew Research Center. In their own words: Why some people find voting difficult". November 5, 2018

Pew Research Center. "Wide Gender Gap, Growing Educational Divide in Voters' Party Identification". March 20, 2018.

United States Census Bureau. "Census Bureau Releases 2020 Presidential Election Voting Report". February 17, 2022