Democrats vs Republican Voting Difficulty

Ahmad Azizi, Don Irwin, Parastoo Javadi

Importance and Context

United States' 2020 general election was unlike prior elections. More than 158 million votes were cast, setting a record. [1]. During a raging pandemic, a divided nation headed to the polls. [2]. Prior to November 3rd, 50% of surveyed registered voters said that it would be very or somewhat easy to vote in the elections; 49% expected to have difficulty casting a ballot. Adding to the sense of high-stakes, "more voters [said] it 'really matters' who wins the presidency than at any point in the last 20 years." [3].

Against this backdrop, our team set out to understand if one group faced more difficulties in casting a ballot than other groups. Our analysis seeks to answer the following question:

1. **D**id Democratic voters or Republican voters experience more difficulty voting in the 2020 election?

Our data-driven analysis will show that Democrats and Republicans did not face the same level of difficulty casting a ballot. That is to say, one group experienced more difficulty and the other.

Additionally, we will show that while non-white GOP voters did not face different difficulty than non-white Democratic voters, or vice versa; non-white voters on the whole, regardless of party affiliation, faced different difficulty casting a ballot than white voters.

Data and Filtering and Initial Analysis

In this analysis we are using a time-series data generated by American National Election Studies (ANES) based on the data received from Westat, the survey firm that conducted the 2020 Time Series. A full description of the data collection methodology may be found at : https://www.westat.com/project/conducting-american-national-election-studies-anes-premier-survey-voting-us

The unfiltered data set contains 8280 survey responses and 1771 variables.

Our Data Filtering, Cleansing and Categorization:

1. Party Affiliation (Variable V201231x):
   (a) Democrats: For values (1,2,3) – ("Strong Democrat","Not very strong Democrat",and "Independent-Democrat"). We categorized those survey respondents as "Democrats".
   (b) Republicans: For values (5,6,7) – ("Independent Republican", "Not very strong Republican", and "Strong Republican") We categorized those survey respondents as "Republicans".

   All other responses were omitted.
2. Ethnicity (Variable V201549x): We included only those survey responses where the respondent indicated their ethnicity. And other responses were omitted.
3. When the response was administered (Variable V200004): We included only those survey responses where the survey was responded to "Before and After" the election.

Our data pipeline, cleansing, testing, and rendering libraries are includes within the enclosed file "source_files1.R'.

Our library returns objects with summaries for different samples of interest. – Some sample code below.

[1] Pew Research Center. "Turnout soared in 2020 as nearly two-thirds of eligible U.S. voters cast ballots for president." (2021).
[2] Pew Research Center. "U.S. Media Polarization and the 2020 Election: A Nation Divided." (2020).
[3] Pew Research Center. "Election 2020: Voters Are Highly Engaged, but Nearly Half Expect To Have Difficulties Voting." (2020).

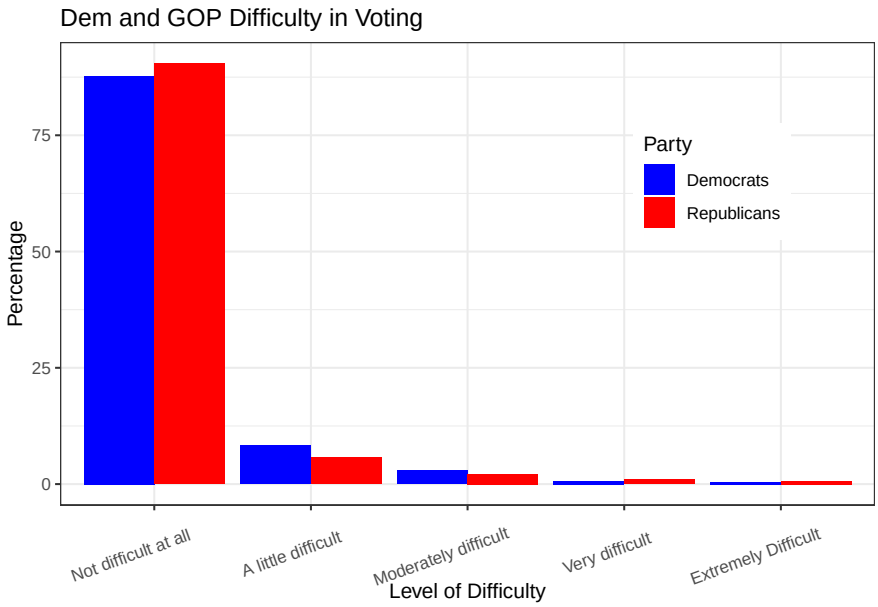
```
#source_file <- file.path(getwd(),"source_files1.R")
#source(source_file)
#young voters.
young_voters_data_set <- get_young_voters_18_to_32_data_set()
#swing state voters.
swing_state_voters_set = get_swing_state_voters_data_set()

#any test can be accessed like so.
wilcox_young_voter = young_voters_data_set$wilcox_results
```

Applying the filters listed above we were left with 5784 sample respondents, 3105, classified as Democrats, and 2679 as Republicans.

After undergoing the filtering above we processed the data to derive what percentage of sampled voters fall into voting difficulty categories.

Table 1: GOP and Dem Voting Difficulty

Voting Difficulty	Democrats	Republicans
Not difficult at all	87.98059	90.48152
A little difficult	8.025383	5.636432
Moderately difficult	2.948862	2.164987
Very difficult	0.6718925	1.0451661
Extremely Difficult	0.3732736	0.6718925



Test Selection

1. Wilcox Ranked Sum Test – Requirements Met:
   (a) **i.i.d. data** In the 2020 Time Series Study code guide book, it is explained that the sample was a random draw from the USPS computerized delivery sequence file (C-DSF), with all included residential addresses across the 50 states and Washington DC having equal probability of selection. It is safe to assume that i.i.d assumption has been fulfilled.
   (b) **Ordinal Scale** the answer to the question of how difficult was it for the voter to vote is a 5-point Likert scale ranges from ("Not Difficult") to ("Extremely Difficult").
2. Two Sampled T Test – Data can be "binarized" to meet requirement.
   (a) **i.i.d. data** In the 2020 Time Series Study code guide book, it is explained that the sample was a random draw from the USPS computerized delivery sequence file (C-DSF), with all included residential addresses across the 50 states and Washington DC having equal probability of selection. It is safe to assume that i.i.d assumption has been fulfilled.

(b) **Metric Scale** If we put respondents into two categories "no difficulty" and "any difficulty" we the data can be made metric and could be made valid for a Two Sample T Test.

Hypothesis Statement

Null Hypothesis: *The probability that Democratic voters and Republican voters experienced the same difficulty voting in the 2020 election to the probability that Democratic voters or Republican voters experienced more difficulty voting in the 2020 election.*

Test Results

```
#access the wilcox test results
wilcox_summary <- summary_data_set$wilcox_results
t_test_result <- summary_data_set$two_sample_t_test
pvalue <- round(as.numeric(wilcox_summary$Test_Value[4]), digits = 3)
W_stat <- round(as.numeric(wilcox_summary$Test_Value[3]), digits = 3)
```

Table 2: Wilcox Ranked Sum Test Results Democrats and Republicans

Test_Attribute	Test_Value
Method	Wilcoxon rank sum test with continuity correction
Alternative	two.sided
Statistic	3675848.5
P-Value	0.00415509724943191

Table 3: Two Sample T Test with 'Binarized' Difficulty.

Test_Attribute	Test_Value
Method	Welch Two Sample t-test
Alternative	two.sided
Statistic	3.03833408297036
P-Value	0.00239049774953922
CI From	0.00913751851196779
CI To	0.0423742396067332
Mean of X	0.120940649496081
Mean of Y	0.0951847704367301

The two-sided comparison result of democrats and republicans shows that Democrats and Republican voters are not experiencing the same voting difficulty. As the Table 2 shows, the Rank Sum test between the groups show a meaningful difference with a W statistic of 3.6758485×10^6 and p-value of 0.004. The outcome of the results is very important for either of the parties. We tried to dive deeper in to data set and compare Non-white and white voters voting difficulty.

Discussion

Our analysis shows the level of difficulty between the two parties is not the same. Although, this is very important result, further analysis is required to study the difference in number of minorities, first-time voters, etc in each party that might result in voting difficulty difference between parties. For instance, we performed a two-tailed t-test on white and non white voters data, and we found that non-white voters and white voters experienced different amount of difficulty voting in 2020 election.

Also, we believe that this result would be very important for swing state elections, because a small difference in number of voters can change the election result as we observed in 2020 election. We couldn't find enough amount of data for each swing state in this data set to perform an analysis; however, if there were more data available we would be interested in looking at the difficulty voting of each party in the swing states.

Supplementary Analysis

We were interested in testing if white or non-white voters experienced different amount of difficulty in voting in 2020 election. The difference in voting difficulty between white and non-white voters can contribute to difference in democrats and repulicans voting difficulty if they have different proportions of white and non-white voters.

To define voter ethnicity, we looked at an attribute that contained self-identified ethnicity of the voters. We binarized our voters into white if they identified themselves as white, and non-white if identified themselves as any other ethnicity than white. Our result showed that there are ? white voters and ? non-white voters. We performed a Wilcoxon Rank Sum test on each group to answer the following null hypothesis:

> **Null Hypothesis:** *The probability that white voters and non-white voters experienced the same difficulty voting in the 2020 election.*

The Rank Sum test, comparison version, requires the following assumptions to be true:

1. **i.i.d. data** In the 2020 Time Series Study code guide book, it is explained that the sample was a random draw from the USPS computerized delivery sequence file (C-DSF), with all included residential addresses across the 50 states and Washington DC having equal probability of selection. It is safe to assume that i.i.d assumption has been fulfilled.
2. **Ordinal Scale** the answer to the question of how difficult was it for the voter to vote is a 5-point Likert scale ranges from ("Not Difficult") to ("Extremely Difficult").

The wilcox rank sum comparison result of white and non-white voters shows that they did not experience the same voting difficulty. As the Table 3 shows, the Rank Sum test between the groups show a meaningful difference with a W statistic of 955250.5 and p-value of 0.01.

Table 4: Wilcox Ranked Sum Test Difficulty in Voting Between White and Non-Wite, regardless of Party

Test_Attribute	Test_Value
Method	Wilcoxon rank sum test with continuity correction
Alternative	two.sided
Statistic	955250.5
P-Value	0.0104768914385633

Table 5: Wilcox Ranked Sum Difficulty in GOP and Dem non-white Voters

Test_Attribute	Test_Value
Method	Wilcoxon rank sum test with continuity correction
Alternative	two.sided
Statistic	56078.5
P-Value	0.51291708854802

Conclusion and Final Questions

Our analysis indicates that one party faces a different level of voting difficulty than the other.

Non white Democrats did not experience different voting difficulty than non-white Republicans. However, people of color, did experience different level of difficulty voting that white voters – regarless of party.

This leaves us with the following, a as of yet, unanswered question: Does racial composition of a party advantage or disadvantage the party in voting?