Democrats or Republicans. Who had more difficulty voting?

W203: Statistical Analysis

Ivy Chan, Jonathan Hodges, Dipika Kumar, Christian Lee

Research Report for a Think Tank from a Consulting Group

Contents

1. Importance and Context

Voter engagement for the 2020 United States presidential election was at record levels, with more voters saying “it really matters” who wins the presidency than at any point over the last twenty years. At the same time, challenges such as the pandemic and social unrest led to half of the registered voters indicating it would be difficult to vote in the election. There has been a significant change since October 2018, shortly before that year’s midterm election, when 85% of registered voters said it would be easy to vote in the midterm elections[1]. A better understanding of the various influencers of voter turnout is useful for your organization for party strategists and campaign managers.

This study will focus on the 2020 election and better understand the difficulty levels in voting between parties, which is one of many factors impacting voter turnout. Specifically, the goal of the analysis is to address the following research question:

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

As consultants we are here to answer this question in order to provide guidance and a foundation for your future research. This includes determining if voter difficulty is a major bellwether of turnout and, if so, further decomposing the factors that lead to these difficulty levels, such as registration, absentee versus in-person voting, long waits at polling places, or bad weather, which can be analyzed in the future. In addition, this analysis will allow us to identify if the difference in difficulty voting between Republicans vs. Democrats was statistically significant in the recent election. If the results are statistically significant, the reasons for difficulty in voting can be further assessed and mitigated.

2. Data and Methodology

The study utilized data from the 2020 Times Series Study conducted by the American National Election Studies (ANES). The Times Series Study interviewed 8280 individuals and comprised pre-election and post-election interviews from August 2020 through the end of December 2020. The sample we created from a subset of the ANES Times Series Study is limited in terms of generalizing to the US voter population across all demographic groups. This is due to the fact that we didn't leverage the weighting provided by ANES that is based on the US census.

Before answering the question of which political party had more difficulty voting, we need to operationalize the concepts, including who is a voter, their political affiliation, and the type of difficulty the individual had. Having this information provides context to exhibit who had more difficulty voting statistically.

To classify a respondent as a voter, we look at those who have already registered to vote (at their current address, another address, or without an address) or are planning on registering to vote. As registering is a prerequisite for voting, we believe this variable is a strong indicator of being a voter. For V201008, values 1, 2, and 3 gave information about their registration address, while the other values did not give additional information. V201009 determined if voters were registered to vote, and value 1 gave applicable information for the study.

We use the following fields for voter identification:

Variable Name	Description	Value
V201008	PRE: WHERE IS R REGISTERED TO VOTE (PRE-ELECTION)	1, 2, 3
V201009	PRE: WEB ONLY: IS R WITHOUT ADDRESS REGISTERED TO VOTE (PRE-ELECTION)	1

There are multiple ways to identify a respondent's respective party, for instance, voting behavior in past elections, voting in the primaries, the party they are currently registered to, etc. The pre-election self-identified affiliation variable was the best way to measure a respondent's political stance due to the quality and quantity of the data. Their political stance before and during the act of voting was taken into consideration. Values 1 and 2 correspond to the political party, while the other values did not give any more information.

[1]Pew Research Center -https://www.pewresearch.org/politics/2020/08/13/election-2020-voters-are-highly-engaged-but-nearly-half-expect-to-have-difficulties-voting/ August 13, 2020

We use the following field for party affiliation:

Variable Name	Description	Value
V201228	PRE: PARTY ID: DOES R THINK OF SELF AS DEMOCRAT, REPUBLICAN, OR INDEPENDENT	1, 2

How difficult it was for respondents to vote and the main reason respondents did not vote were used to determine the difficulty in voting. The combination of the two factors encompassed how hard it was for voters to cast their vote and why they found voting difficult. The values (2,3,4,5) of having difficulty voting were used in the study to determine which party had more difficulty for V202119. We utilized values that were not in the respondents control that induced difficulty in voting for V202123.

We use the following field for difficulty:

Variable Name	Description	Value
V202119	POST: HOW DIFFICULT WAS IT FOR R TO VOTE	2, 3, 4, 5
V202123	POST: MAIN REASON R DID NOT VOTE	9, 10, 11, 12, 13, 14, 15

After assigning true or false values based on if it was difficult for party members to vote, we observed how many democrats and republicans had difficulty voting.

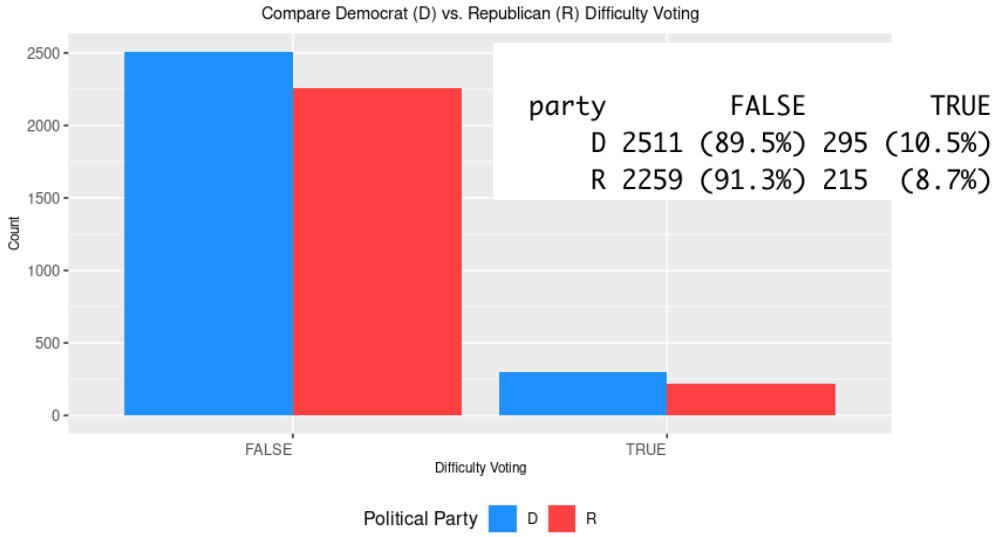

Figure 1: Difficulty Voting Democrats(D) vs Republicans(R)

Democrats had more true and false values, which aligned with the population of democrats being higher than republicans. However, we looked into the true and false cases as a percentage of the respective party population samples. The percentage difference between true and false was not notable between parties. Utilizing the party as a grouping variable and defining the response variable as proportional count of difficulties per group we can conduct some statistical tests.

We evaluated various reputable hypothesis tests to determine which is best for answering our research question. After reviewing the assumptions that must be met for each test, we were able to narrow our choices down to the two-sample proportion test and the comparison version of the Wilcoxon rank sum test. As we evaluated difficulty as a binary true or false value instead of a scale of ordered categories, the data was determined to be more appropriately wrangled for the proportion test. Additionally, the Wilcoxon rank sum test is of lesser statistical power; therefore, we determined the proportion test was best suited for our analysis from these two test options.

The proportional statistic assumptions were validated for proper use of the test statistic. The first assumption for the proportional two-group comparison test was independence and identical distributions (i.i.d). Independence can be assumed since random sampling occurred, and one sample's information cannot be inferred from other samples. However, the population of the samples changes from 8280 (pre-election interviews) to 7,782 (post-election

interviews). There is a 7% decrease in the population size, which will not heavily affect the probability distribution. The following other assumptions are correct as the sample population follows a binomial distribution, and the data are simple random values from the population.

While evaluating for practical significance, Cohen's d is not well suited as it requires normality. For our binary data, we chose the Phi coefficient, which yielded a value of -0.031. The result indicates that there is virtually no relationship between political party and voter difficulty, meaning other factors drive difficulty.

Below is the accounting table summarizing the data wrangling after the exploratory data analysis (EDA) of the dataset.

Cause	Number of Samples Available for Analysis (after removal for cause)	Removed Number Samples for cause
Start	8280	0
Non voter	7888	392
Independents (non-partisans) or no party affiliation provided	5280	2608

Note: For the response variable value that is not included in the selections, we assign a binary/boolean value to the record. Therefore, we do not have any missing or out-of-range response variable values.

3. Result

As we have two independent samples, Democratic voters and Republican voters, and difficulty being defined and organized as binary/boolean values, we derived the counts for total voters and the counts for voters with difficulty for each party. We use the proportion hypothesis test for binary data.[2]

> ***Null Hypothesis****: The proportion of Democrat voters with difficulty voting is equal to the proportion of Republican voters with difficulty voting.*
>
> H_0: $p_1=p_2$, versus select one H_A: $p_1 \neq p_2$ [3]

The proportion hypothesis test returns a p-value of 0.02847, which leads us to fail to reject the null hypothesis as it is not within the rejection region (< 0.025).The result is not statistically significant.

```
        2-sample test for equality of proportions with continuity correction

data:  c(democrat_with_difficulty$count, republican_with_difficulty$count) out of c(total_democrat, total_republican)
X-squared = 4.7996, df = 1, p-value = 0.02847
alternative hypothesis: two.sided
95 percent confidence interval:
 0.001973025 0.034483097
sample estimates:
   prop 1    prop 2
0.1051319 0.0869038
```

4. Discussion

While this is limited based on how Democrats and Republicans responded to ANES interviews, our study found evidence that voting difficulty does not have a strong relationship with the voter's political party. It is consistent with our findings that there is no significant difference in the Democrat and Republican populations we compared for voting difficulty. Additionally, we have shown that difficulty wasn't experienced at a very high rate across either party, with just over 10% of Democrats and Republicans expressing this in their survey responses.
Since the 2020 election had the highest turnout in United States history, it does not appear that difficulty had a significant role in preventing people from getting to the polls. Instead, it may make sense for future research to look at other predictors of voter turnout, such as education or socioeconomic status.

---

[2] Comparing Hypothesis Testing for Continuous, Binary, and Count Data "https://statisticsbyjim.com/hypothesis-testing/comparing-hypothesis-tests-data-types/" Date: NA.
[3] Comparing Two Proportions "https://online.stat.psu.edu/stat415/lesson/9/9.4" Date: NA.