Difficulty Voting in the 2020 Presidential Election

Team03: Savita Chari, Denny Lehman, Tymon Silva

10/19/2021

1 Introduction

The 2020 election, held on November 3, recorded the highest voter turnout in US election history as well as the highest number of mail-in voters.[1] And, although more participation is generally seen as a good thing, the 2020 voting experience left many concerned and troubled, casting a shadow on the safety, reliability, and structure of the voting process. Videos of citizens showing up armed with weapons at the voting booths, stories of locals intimidating and harassing voters standing in line, and reports of issues with mail-in ballot boxes, seemed all too common this last general election[2]. Is this just the media hyping the drama around the voting process or did certain groups of voters really experience increased difficulty voting compared to others? The answer seems to change based on who is asked and how they lean politically [3]. The situation has spurred many discussions around the voting process and how it will look in the future.

In this report, we answer a basic, but relevant question:

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

We focus our analysis on the frequency at which Democrats and Republicans experienced difficulties voting and omit analysis on the severity or type of voting challenges. In answering this question, we can contribute to the conversation and help guide strategies that reform and/or improve the voting process. We must protect our voters to ensure a fair election takes place.

2 Data and Methodology

Our data comes from the 2020 Time Series Study performed by the survey firm, Westat, and generated by the American National Election Studies (ANES). The data consists of responses from both a pre and post election survey, in which participants were interviewed prior to election date and contacted again to capture their experiences and opinions shortly after the election had concluded. Addresses across the US were randomly drawn from a USPS computerized delivery sequence file and the selected addresses were sent a series of letters to recruit a household member. All mailers offered financial incentives to participate in the survey. The study involved three modes of response: web -only, mixed web, and mixed video (live video interviewing via Zoom). The pre-election survey was conducted between August 18, 2020 and November 3, 2020. The study re-interviewed the same respondents in the post-election survey between November 8, 2020 and January 4, 2021. The overall response rate for the pre-election survey and the post-election survey was 40.9% and 87.9%, respectively.

Additionally, all sample members who completed the post-election survey from the ANES 2016 Time Series Study were invited to participate in the 2020 survey online. This survey data was merged using the case ID variable to join the two datasets. We created a data pipeline function to process the combined data set. The function aided in preprocessing - removing invalid entries such as participants who did not complete both

---

[1]https://www.newsweek.com/highest-voter-turnouts-history-us-elections-1545235
[2]https://www.pewtrusts.org/en/research-and-analysis/blogs/stateline/2020/10/15/election-officials-prepare-for-voter-intimidation-threat
[3]https://apnews.com/article/election-2020-ap-fact-check-joe-biden-donald-trump-technology-49a24edd6d10888dbad61689c24b05a5

Table 1: Distribution of Political Party

Party	Frequency	Proportion
Democrat	3031	0.54
Republican	2618	0.46

pre and post surveys, respondents with eligibility concerns, and participants with questionable identities. Additionally, Independents were removed from the analysis.

In order to operationalize the concept of difficulty voting, we first identify the Democratic and Republican voters. Per guidance from an article written by political scientists[4], we group voters into political parties by the direction they "lean." In other words, a voter that indicates they "lean democratic" is classified as a Democrat and vice versa. The table below describe these distributions.

Table 1 shows the frequencies of Democrats and Republicans in our dataset and the proportion of each group, which is roughly 54% Democrat and 46% Republican.

The key variable for our analysis is the question pertaining to how difficult it was for the respondent to vote. Responses were collected on a 5-point Likert scale that ranged from 1 ("Not difficult at all") to 5 ("Extremely difficult"). In order to operationalize the frequency in which Democrats had experienced difficulty voting compared to Republicans that experienced difficulty, we grouped voters in a binary fashion; as either "had difficulty voting" or "had no difficulty at all" and compared their means.
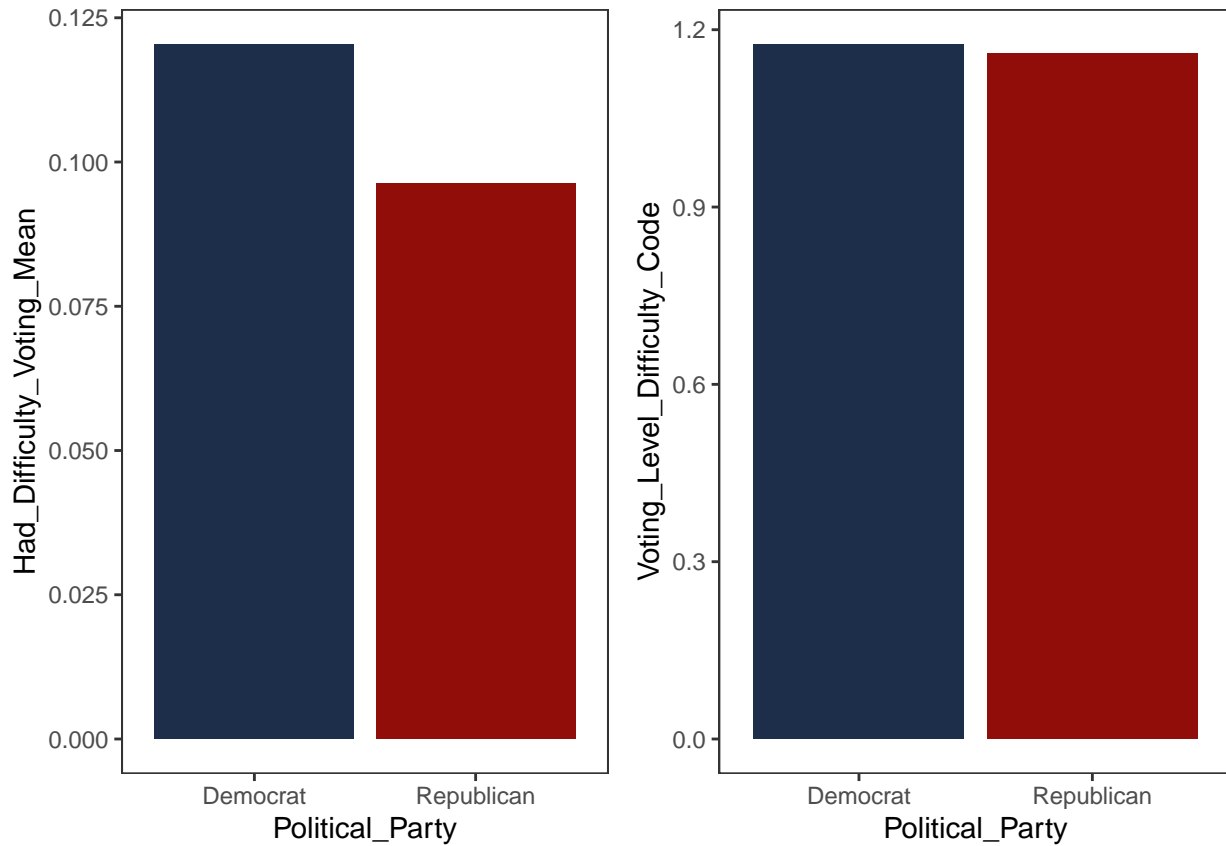
We will test if one party is more likely to experience difficulty voting than the other. To test this, we can use a two-sample t-test. The null hypothesis for this t-test can be expressed as follows:

---

[4]https://www.sciencedirect.com/journal/electoral-studies

Null Hypothesis: *The probability that a Democratic voter experiences some difficulty during their voting process is equal to the probability that a Republican voter experiences some difficulty during their voting process.*

For this t-test, the following assumptions must be true:

1. **i.i.d. data** First, data must be generated via an iid process (Independent and Identically Distributed). The ANES 2020 survey merged data from a panel, which was comprised of participants in the ANES 2016 survey. This is a concern when validating that the data be identically distributed, since the year in which these surveys were administered differs greatly. However, for the purposes of this study, we will assume the impact of this possible violation is negligible.

2. **Metric scale** Even though our raw data includes ordinal variables, we have transformed them to binary form. A binary variable qualifies as metric data because there is only a single interval, which goes from zero to 1.

3. **Sufficient normality** Even after removing ineligible respondents, the sample contains sufficient size (n≈5600). As a general rule of thumb, when $n > 30$ the distribution of the sampling statistic approaches the normal distribution per the Central Limit Theorem.

3 Results

```
t.test(ANES_mean$Voting_Level_Difficulty_Code ~ ANES_mean$Political_Party)
```

The result of the Two Sample t-test returned a t-statistic of 2.92, giving a p value of 0.0034. Since the p-value is less than 0.05 (our false positive rate), we reject the null hypothesis with 95% confidence; there is a difference in voting difficulty between Democrats and Republicans. 12% of Democrats faced some difficulty voting while 9.6% of Republicans experienced issues. Comparing means, Democrats are 25% more likely to experience incidents of voting difficulty compared to their Republican counterparts. The results are statistically significant.

4 Discussion

This study found that Democrats experience 25% higher frequency of voting difficulty than Republicans in the 2020 election. This is a surprising result, considering the influx of mail in voting and historically large election turnout. In a country that views voting as a right for all citizens, we find that challenges to voting are not evenly spread among Americans. The difference in voting difficulty based on political party is especially concerning because lifting restrictions often requires new laws. We suspect that many politicians would prefer their opposition have harsher restrictions to the polls, not more access. As Republican State Legislators pass stricter voting laws making it harder to vote in many states [5], we recommend further analysis on voting difficulty and party. This election was the largest yet, but there is still work to be done enabling all to participate in democracy, regardless of political affiliation.

---

[5]https://www.usnews.com/news/best-states/articles/2021-07-02/17-states-have-passed-restrictive-voting-laws-this-year-report-says