

Disparities in Difficulty Voting Between U.S. Democrats and Republicans

DATASCI 203: Lab 1 Part II Report

Gary Kong, Yeshwanth Somu, Maiya Caldwell, and Clara Zhu

Contents

1	Importance and Context	1
2	Data and Methodology	1
3	Results	3
4	Discussion	3

1 Importance and Context

After current Democratic President Joe Biden displaced former Republican President Donald Trump in the 2020 presidential election, there was much speculation about voting suppression and fraud. Republican voters angrily attributed Trump’s loss to such suppression, in many cases altogether dismissing the validity of the election. While voter suppression has always been a feature of U.S. political history, unprecedented controversy surrounding the validity of the 2020 election is contributing to even more division to the U.S. political climate.

Voter suppression is a multifaceted topic, encompassing multiple barriers to voting, including, but not limited to legal, economic, educational, and process barriers. This analysis provides a starting point towards understanding the extent of voter suppression in the 2020 elections by addressing the following question:

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

The answer to this question could help quantify the degree of voting disparity between parties, potentially paving the way towards creating a more equitable voting process. By providing insight into which of the two major U.S. parties were faced with more difficulty in voting, it could provide greater clarity to help remediate controversy surrounding the 2020 election.

2 Data and Methodology

Our analysis leverages data from the 2020 American National Election Studies (ANES). The dataset comprises re-interviewed respondents from the 2016 ANES Time Series Study and a freshly-drawn 2020 cross-section, which comprised three survey mode sequence groups: a) Web, b) Phone/Web, and c) Video/Phone/Web. To ensure identical distribution of Democrat and Republican groups in our analysis, we chose to only use the 2020 cross-section data.

The 2020 cross-section sample’s target population was the 231 million non-institutional U.S. citizens aged 18 or older living in the 50 US States or the District of Columbia. The sample was a random draw from the USPS computerized delivery sequence file, with all included residential addresses across the 50 states and Washington DC having equal probability of selection. Respondents had to reside at the sampled address and be a U.S. citizen aged 18 or older. Only one adult member of each household could complete the survey.

The 2020 cross-section sample includes 5,441 respondents/observations who completed pre-election interviews, with 4,783 also completing post-election re-interviews. As we used variables from both pre- and post-election questionnaires, we chose to only include observations with complete data for both pre- and post-election questionnaires. This leaves 4,783 observations, of which 2,012, 2,136 and 635 are in the Web-only, Phone/Web, and Video/Phone/Web groups, respectively.

We note that the sample may not be nationally representative due to selection and response bias. Some demographic groups may be more likely to be ineligible for not residing at the sampled address. The sample may be underweight on subpopulations without internet access and the elderly who may not be digitally literate. Within each household, individuals of certain characteristics may be more likely to participate in the survey than others (e.g., busy, working professionals may have less time / incentive to participate vs. the unemployed).

ANES provides survey weights to correct for the biases outlined above. However, we chose not to use these for our analysis. ANES calculated weights for each of the three survey groups. Meanwhile, our analysis involves dividing the sample into Democrat and Republican groups and testing for differences in difficulty voting between these groups. The unit of weighting is thus inconsistent with the unit of our analysis, which could lead to inaccurate results and difficulty in interpretation if weights were to be applied.

The dataset includes questions about voter registration status in both the pre- and post-election surveys, as well as a question about whether a respondent actually voted in the November 2020 election. We defined a voter as a respondent that is either registered to vote and/or voted in the November 2020 election (irrespective of stated voter registration status). We included all registered respondents to account for the potential for such respondents to not vote due to difficulties in voting. This results in 4,507 valid observations.

To operationalize party affiliation, we categorized voters registered in the Democratic Party or Republican Party in the post-election survey as affiliated to that party. For those unregistered as either, we used a 7-point Likert Scale summary variable of party identification from the pre-election survey, whereby voters leaning Democrat and Republican are classified as affiliated to that party. We excluded voters who do not classify as either Democrat or Republican, resulting in 4,098 total voter observations, of which 2,233 are Democrat, and 1,865 are Republican.

We considered several alternative methods of operationalizing party affiliation. We considered using which party the voter actually voted for, but concluded that this was sub-optimal versus our chosen method for several reasons. Firstly, voters

may not always vote according to their political identification. Voters may also voted for conflicting parties in the different elections. Finally, there may be Democrat or Republican ‘voters’ who did participate in the elections. We also considered using a liberal-conservative self-placement scale, but concluded that this is a less direct mapping to party membership.

We assumed that difficulty voting refers to encountering difficulty in the voting process, as opposed to psychological indecision in voting. We operationalized difficulty voting (the outcome variable) using a survey question (“How difficult was it for you to vote in the election?”) which most directly approximates this concept. Data is reported on a 5-point Likert scale that ranges from 1 (“Not difficult at all”) to 5 (“Extremely difficult”). Excluding voters with invalid ‘difficulty voting’ responses results in 3,772 total valid voters, of which 2,066 are Democrat, and 1,706 are Republican.

The survey includes a question (“Did you encounter any of the following problems in trying to vote in this election”) whereby respondents can select multiple answers from a list of 10 possible problems voting. We considered measuring difficulty voting by summing the number of problems voting. Nonetheless, we chose the 5-point Likert scale over this as the number of problems encountered does not necessarily map to how difficult the overall process was (i.e., a voter could encounter one critical problem voting vs. encountering multiple, minor problems). As a follow-up, we would consider using a composite outcome variable based on both the difficulty voting Likert scale and the number of problems encountered.

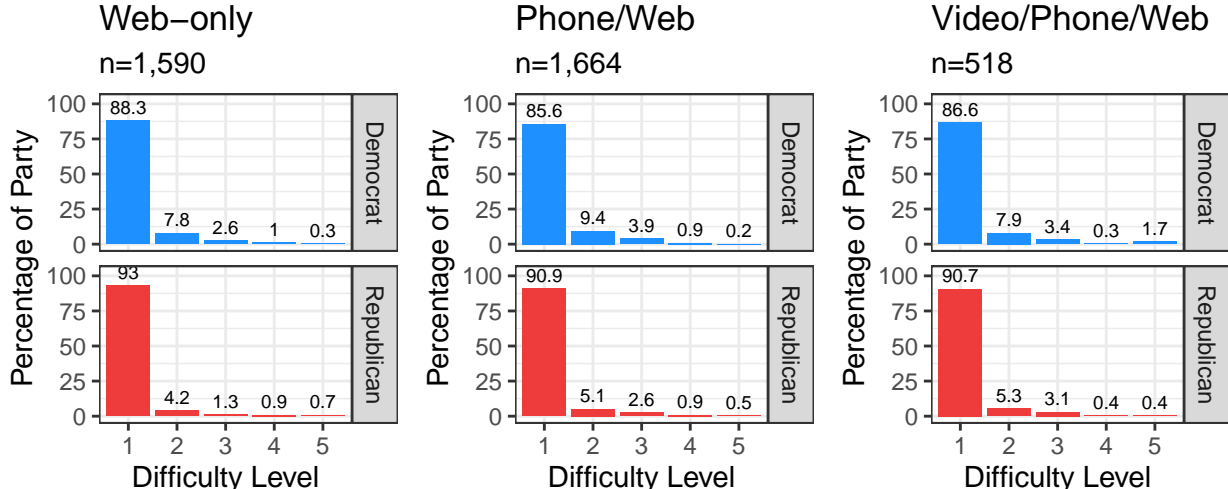


Figure 1: Distribution of difficulty voting scores, by party and survey method

Our grouping variable is categorical, while our outcome variable is ordinal. Common tests could include a two-sample T-test (in certain scenarios) or a Wilcoxon Rank Sum (Hypothesis of comparison) test. A two-sample T-test would require treating a 5-point Likert scale as interval data, which may not be appropriate as differences between categories may not be equal. Additionally, as seen in the above histograms, the distribution of data is highly skewed, necessitating a non-parametric test. As such, we chose a Wilcoxon Rank Sum (Hypothesis of comparison) test.

The null hypothesis of our Wilcoxon Rank Sum test can be phrased as follows:

Null Hypothesis: *The probability that difficulty voting is higher for Democrats than Republicans is equal to the probability that difficulty in voting is higher for Republican than Democrats*

For a Wilcoxon Rank Sum test to produce reliable results, the data must be measured on an ordinal scale, and sample data must be generated via an I.I.D. process, where data from Democratic and Republican voters alike must have been drawn from the same distribution and must be mutually independent from one another. We address each of these requirements in turn.

First, the outcome variable must be measured on an ordinal scale. In our case, the 5-point Likert scale outcome variable is ordinal as the intervals between categories are monotonic but not so well-defined as to be numerically uniform increments.

Second, the sample data must be generated via an I.I.D. process. The 2020 cross-section data consists of three survey method groups. Notably, the distribution of the survey method groups differ between Democrat and Republican voters, as shown in Table 1. The survey method groups cannot be considered identical as the survey method used may have impacted responses. As such, to ensure identically distributed data, we stratified our analysis by survey method. Each observation in the dataset is also mutually independent. Each survey respondent can only provide one set of responses and are classed as either Democrat or Republican. Additionally, 2020 cross-section was sampled from a random draw, and only one member of each household could participate in the survey.

Table 1: Distribution of valid responses, by party and survey method

	Democrat	Republican
Web-only	43.3	40.8
Phone/Web	42.6	46.0
Video/Phone/Web	14.1	13.2

To aid interpretation of results, we created an additional binary variable, based on whether respondents rated at least “Moderately Difficult” voting (i.e., 3-5) on the Likert scale used as our outcome variable.

3 Results

```
test1 = wilcox.test(x = Dem_w, y = Rep_w, alternative = "two.sided", correct=FALSE)
test2 = wilcox.test(x = Dem_pw, y = Rep_pw, alternative = "two.sided", correct = FALSE)
test3 = wilcox.test(x = Dem_vpw, y = Rep_vpw, alternative = "two.sided", correct = FALSE)
```

For two of the three survey method groups, the test yields evidence that there is a significant difference in difficulty voting between Democrats and Republicans (Web-only: $W=3.25 \times 10^5$, $p=0.002^*$, $n=1,590$; Phone/Web: $W=3.63 \times 10^5$, $p=0.001^*$, $n=1,664$; Video/Phone/Web: $W=3.4 \times 10^4$, $p=0.151$, $n=518$). As seen in Table 2, compared to Republicans, a greater proportion of Democrats rated at least “Moderately Difficult” voting in each of the three survey method groups. This difference is most pronounced in the Video/Phone/Web group. As such, it is plausible that the lack of statistical significance in the Video/Phone/Web group is due to smaller sample size in this group.

Table 2: Percent of voters rating at least "Moderately Difficult" voting, by party

Survey Method	Democrat	Republican	Delta
1. Web-only	3.91	2.87	1.04
2. Phone/Web	5	3.95	1.05
3. Video/Phone/Web	5.48	3.98	1.50

From a practical perspective, our results appears potentially important. As seen in Table 2, the percentage of Democrats who considering voting at least “Moderately Difficult” exceeds that of Republicans by 1.04-1.5 percentage points. This could swing elections in certain states if it translates to comparable discrepancies in electoral participation rates.

Several limitations of our analysis may affect the conclusions that could be drawn from it. As mentioned previously, the ANES is not nationally representative, suggesting that results may not generalize to the U.S. population. Additionally, the 5-point Likert scale used as our outcome variable measures perception of difficulty voting, as opposed to being an objective measure of it.

4 Discussion

This study found evidence that Democratic voters find it more difficult to vote than Republican voters. The effect appears practically significant, which challenges arguments about Republican voter suppression leading to Trump’s loss in 2020. Though the difference in voting difficulty between Democratic and Republican voters may appear modest, in such a polarized political climate, even small discrepancies may be enough to misrepresent the collective choice for president in certain states.

Our results may be of particular interest to political strategists, commentators, researchers, and policymakers who can use this information to inform electoral reform. While this study focuses on quantifying disparities between Democratic and Republican voters at a high-level, a broader objective is to rectify them. Future studies could therefore focus on uncovering specific aspects of the voting process that contribute to voting difficulty, and how voting difficulty varies demographically. Finally, we hope that a general understanding of differences in voting difficulty between the major U.S. political parties can benefit society more broadly by raising awareness of the issue.