

The Statistics Behind Voting Difficulties

Carla Cortez, Redwan Hussain, Samuel Omosuyi, Jorge Bonilla

March 2, 2022

Contents

1	Importance and Context	1
2	Data and Operationalization	1
2.1	Preliminary Analysis	2
2.2	Methodology	2
3	Results	2
4	Discussion	3

1 Importance and Context

The right to vote is the hallmark of a free and democratic society. In recent years, new laws have been passed to expand the ability for citizens to vote. As a result, the motivation behind this paper is to analyze the data published by the American National Election Studies (ANES) to provide additional insights into the voting experience between Democratic and Republican voters. This analysis presents the statistical results derived from the surveys conducted by the ANES on the experiences that voters had during the 2020 elections.

The primary research question for this analysis is stated as follows:

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

The contributions from this analysis provide concrete evidence on whether a given political party experienced a measured level of difficulty. This information helps policy makers to improve the voting experience for their targeted voters so that everyone’s needs and concerns are addressed in our society.

2 Data and Operationalization

Our study focuses on the experiences of voters during the 2020 Presidential Election to determine if there is a difference in difficulty voting from individuals in either the Democratic or Republican party. To conduct this analysis, we defined three unique characteristics in this study as it relates to who the voters are, what it means to vote, and in which election the voting took place. The ANES data set is composed of 8,280 observations.

First, we defined voting as the action of casting a vote for the 2020 presidential election. In our data cleaning process, we used variable V202072 in which the survey asks, “How about the election for President? Did you vote for a candidate for President?” The ordinal values for these answers are 1 for “Yes, voted for President” and 2 for “No, didn’t vote for President.” We excluded a total of 2,328 observations consisting of subjects who did not vote in the 2020 presidential election, observations with issues during the data collection, and observations where the subject did not answer the question, defining them as non-voters.

Second, we defined a voter as an individual who performed the action of casting a vote while registered as a Democrat or Republican. This analysis did not take into account other political parties. To determine pre-party registration we used the variable V201231x which specifies the pre-election party identification. The ordinal values to answer this question are: 1) Strong Democrat, 2) Not very strong Democrat, 3) Independent-Democrat, 4) Independent, 5) Independent-Republican, 6) Not very strong Republican, and 7) Strong Republican. Based on this variable, we created a new derived variable titled “party_registration_from_party_id” that re-classifies answers 1 through 3 as Democratic and answers 5 through 7 as Republican. Since the research question posits if either Democrats or Republicans experienced difficulty voting, we analyzed the voters who remained with the same party post-election. We did this to eliminate the potential bias of having individuals who may switch political parties to vote against undesirable candidates within other parties. The post-election party identification is provided by variable V202065x and we kept the answers that state 1 for “Democratic Party” and 2 for “Republican Party.” We created a new column using these relabeled values and named it as “post_party_id.” As a result, we excluded voters who had changed party affiliation between the pre- and post-election interviews and those who were independents/non-affiliates, leading to a removal of 3,629 observations (see Table 1 below).

Table 1. Party Identification

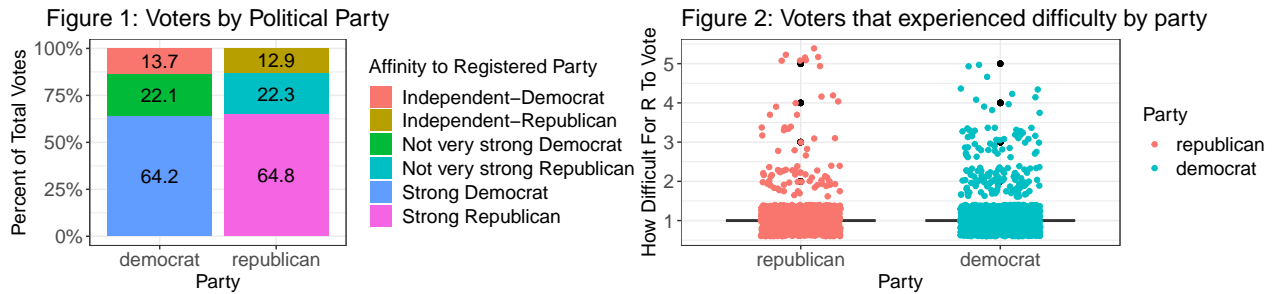
PARTY REGISTRATION FROM PARTY ID	PARTY OF REGISTRATION SUMMARY (PRE-POST)	
	Democrat	Republican
Democrat	1,328	50
Republican	102	995

Lastly, the concept of voting difficulty is defined using variable V202119. This asks the respondents “how difficult was it for you to vote in this election?” This variable lists difficulty as a Likert scale in the following

levels: 1) Not difficult at all, 2) A little difficult, 3) Moderately difficult, 4) Very difficult, and 5) Extremely difficult. The observations with answers that are not in this Likert scale list were removed as they were related to either incomplete interviews or subjects that chose not to respond. It is important to choose this post-election variable since we can safely assume that the action of voting has taken place. The final data set contained 2,323 observations and there were no missing or null values.

2.1 Preliminary Analysis

After the data wrangling process, we conducted a preliminary analysis to understand the characteristics and distribution of voters for our hypothesis test. We compared the voter turnout of Democrats and Republicans based on affinity to registered party (i.e. “lean” and “strong” party members). In Figure 1, we can see that the voter turnout was consistent for all segments. We also visualized the density of difficulty (Figure 2) and noticed that Democrats had a slightly higher frequency for difficulty scores ranging between 2-4.



2.2 Methodology

After evaluating the below assumptions required for the Wilcoxon Rank-Sum Test to use the Hypothesis of Comparisons, we determined that it is the most appropriate hypothesis test to conduct on our dataset to answer our primary research question. We meet the following assumptions. Ordinal Scale: The data for the voting difficulty variable is measured on an ordinal scale. The data is reported in a 5-point Likert scale with a 1 to 5 ordered range to describe the level of voting difficulty. IID Data: The sample responses are all independent and identically distributed events because the response of one voter does not influence the response of the other and each subject was asked the same questions. The selection probabilities for the interview subjects were the same ¹. Unpaired Data: A voter can be only part of one group, Democratic or Republican.

The null hypothesis and two-sided alternative hypothesis of our Wilcoxon Rank-Sum Test are defined as follows:

Null Hypothesis: *The probability that a voter from the Republican party experienced difficulty voting is the same as the probability that a voter from the Democratic party experienced difficulty voting.*

Alternative Hypothesis: *The probability that a voter from the Republican party experienced difficulty voting is not the same as the probability that a voter from the Democratic party experienced difficulty voting.*

One limitation of this type of test is that it has a lower power than a parametric test – this needs to be taken into consideration when analyzing the results.

3 Results

```
test_result<-wilcox.test(df$how_difficult_for_r_to_vote~df$pre_post_party_registration,
                        alt="two.sided",correct=TRUE, paired=FALSE, conf.int=TRUE)
```

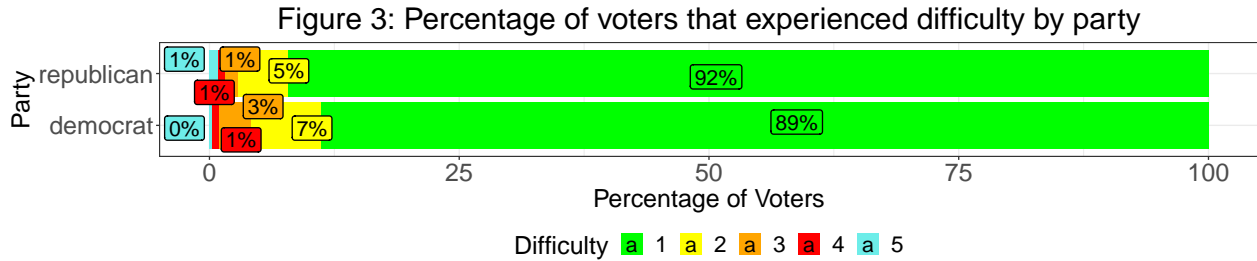
¹American National Election Studies. ANES 2020 Time Series Study Full Release, User Guide and Codebook (2022). 11.

The Wilcoxon Rank-Sum Test showed the p-value is **0.00848** (statistically significant) which is less than 0.05 and indicates we have enough evidence to reject the null hypothesis.

The test gave an estimated location-shift of 0.000032 with a 95% confidence interval associated with the test (-0.000020, 0.000040).

The measure of effect, meaning, the magnitude of the difference between the groups, is small given the effect size is calculated to be **0.0546** ($Z = -2.63$, $N = 2,323$). Effect size is computed based on the equation $\frac{zscore}{\sqrt{N}}$

Figure 3 below shows a notable difference in the percentage of Democratic voters who experienced no difficulty compared to their Republican counterparts.



Finally, we conducted a test to see if there is a correlation between difficulties experienced by the two parties. Considering the difficulty experienced response is ordinal data that have at least three categories and the categories have a natural order, we used Spearman's correlation test. The Spearman correlation can evaluate a monotonic relationship between two variables. Spearman's correlation coefficient ranges from -1 (perfect negative correlation) to +1 (perfect positive correlation) with 0 indicating no correlation.

The spearman's correlation test yielded **0.018**; although this is positive correlation, the value is close to zero which denotes no relationship between the Democratic and Republican difficulty experience.

In summary, based on the Wilcoxon Rank-Sum Test, we have statistical significance (p-value=0.00848) to reject H_0 and accept the alternate hypothesis that the probability that a voter from the Republican party experienced difficulty is not the same as the probability that a voter from the Democratic party experienced difficulty. We saw no correlation (0.018) between difficulties experienced by voters in both parties, and the magnitude of the effect size (0.0546) between the two groups was small.

4 Discussion

While there is enough evidence to reject the null hypothesis and the Spearman's correlation test shows a weak positive correlation, it is important to consider the difference between statistical and practical significance.

The study does not go into detail with the types of difficulties that are experienced during the voting process. For example, variable V202120e tells us whether the participant mentioned any difficulty getting to the polling place. However, we do not know if it is due to distance, mode of transportation, scheduling issues, or any other reason. Future studies should include more targeted questions to capture better data on difficult voting experiences based on location and demographics (e.g. age, ethnicity, socio-economic status).

Also, our current scope and hypothesis test only included those who had voted in the 2020 presidential election. A future study should analyze those who had both experienced difficulties and did not cast a ballot to determine if there is a causal effect.