Difficulty Voting by Party Affiliation

Datasci 203: Lab 1 Part 2

Kolby Devery, Nan Li, Matt Pitz, Ada Guan

March 3, 2022

Contents

1 Importance and Context

The 2020 election in the United States had closely contested races in several states with razor thin margins of victory. Joe Biden narrowly won the 2020 election in Georgia by 11,779 votes - a margin of only .23%. Races like this highlight how a relatively small number of votes that can sway the results of an election. Given the importance of every vote in these tight races, any obstacles to voting can have a large impact on outcomes. The ACLU has identified hundreds of 'anti-voter' bills that have been introduced in recent years. These can range from strict voter ID laws to voter purges and gerrymandering.[1] The full impact of these voter bills is crucial to understanding the ability of the United States' democratic process to fully and fairly represent the voice of US voters.

This research aims to contribute to the discussion of 'voter difficulty', by focusing on the difficulty for both Democratic and Republican voters to participate in the 2020 election. Our research is hoping to answer the following research question:

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

The answer to this question could help us better understand what impact, if any, these 'anti-voter' bills have on elections. More specifically, it could help us to investigate whether there is a disproportionate impact on voters of a specific political party. With this, researchers, commentators, and the general populace can take a more informed path forward in ensuring that all eligible voters have the same ability to vote.

2 Data and Methodology

Our report utilizes data from the 2020 American National Election Studies Time Series Study (ANES).[2] The data was collected for this study using a sequential mixed-method design that included self-administered online surveys, live video interviews conducted over the internet, and telephone interviews. Participants were offered up to $200 to complete their surveys and interviews.[3] Information on each participant was collected prior to the election and after the election. The study design contains two primary subsets of data. The first represents a new cross section of participants for the 2020 election. The second captures responses from participants in the 2016 ANES study. Our analysis focuses on the 5441 respondents from 2020 cross section. This sample draws from the 231 non-institutional U.S. citizens aged 18 or older and living in the 50 states or the District of Columbia.[4]

The study collected several fields that could be used to determine political affiliation and difficulty voting. We focus our analysis on two fields, *PRE: SUMMARY: PARTY ID* ("Party ID") and *POST: HOW DIFFICULT WAS IT FOR R TO VOTE* ("Difficulty"). Party ID utilizes a nominal scale to categorize each participant as an independent, independent leaning towards Republican or Democrat, not a strong Republican or Democrat, or a strong Republican or Democrat. Consistent with the findings in Petrocik 2009.[5], we recoded those participants that partially identify with a specific party as a member of that party Difficulty is a five point Likert scale ranging from "Not Difficult at all to"Extremely Difficult". However, this variable does not capture those who attempted to vote, but were unable to do so. To supplement this variable we add a sixth, highest result based on "POST: MAIN REASON R DID NOT VOTE" and "POST: ANY OTHER REASON R DID NOT VOTE" where the participant indicated that they were not able to vote because of "Transportation", "The line at the polls was too long", "I was not allowed to vote at the polls, even though I tried", and "I requested but did not receive an absentee ballot".

[1] ACLU. "Block the Vote: How Politicians are Trying to Block Voters from the Ballot Box" (2021).

[2] American National Election Studies. 2021. "ANES 2020 Time Series Study Full Release" [dataset and documentation]. July 19, 2021 version. www.electionstudies.org

[3] ibid

[4] ibid

[5] Petrocik J.R. "Measuring party support: Leaners are not independents". Electoral Studies 28, 562-572

Table 1: Average of difficulty voting in the 2020 election

Party Affliation	All voters	Rural voters	Urban voters
Democrat	1.23	1.22	1.23
Republican	1.21	1.17	1.25

We further plan to consider whether some level of difficulty is determined by the location of the voter. We use the "POST: DOES R CURRENTLY LIVE IN A RURAL OR URBAN AREA" variable to indicate whether a participant lives in an urban setting (defined as "Urban" or "Suburban") or in a rural setting (defined as "Rural area" or "Small town"). By splitting the participants into pools of urban and rural voters, we will be able to see whether difficulty increases or decreases in those settings for participants affiliated with each political party. After removing participants who identify as independent and removing missing observations, we are left with 3716 responses.

As we show in Table 1, the average difficulty experienced can differ between areas.[6]. When we plot the data by urban and rural voters in Figure 1, we see that the difficultly experienced by voters of each party varies

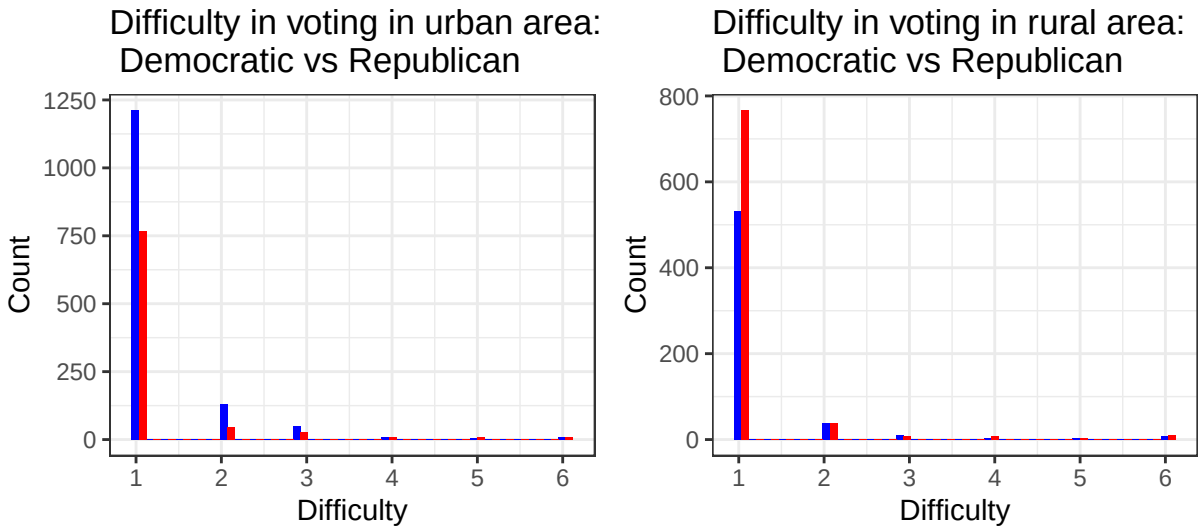

Figure 1: Voter Affiliation and location. The red series is Republicans, the blue series is Democrats

We use the Wilcoxon rank-sum test to determine whether Democratic voters experience more or less difficulty in voting than Republican voters. The Wilcoxon rank-sum test (hypothesis of comparisons version) can identify differences in Likert scale data. We group variables for the comparison test by party affiliation (Democratic or Republican). For the Wilcoxon rank-sum test to produce reliable inference, the following assumptions must be met: the data must be drawn from an i.i.d sample, Each X_i is drawn from the same distribution, each Y_i is drawn from the same distribution, and all X_i and Y_i are mutually independent. and finally measured on the ordinal scale.

The 2020 ANES survey used a mixed-mode design which included self-administered online surveys, live video interviews conducted online, and telephone interviews. Although there is a possibility that this design introduces dependencies, there is a large sample size for both parties, which suggests that links between individuals should be rare and the sample can be considered i.i.d.. Similarly of the sample observations are

[6]Since difficultly is a Likert variable, the averages do not have a clear interpretation. We provide this table only for demonstration purposes

draw from separate distributions and are independent. A participant cannot be both a Democrat and a Republican. Finally The outcome variable, level of difficulty, is a Likert variable measured on an ordinal scale satisfying the second assumption.

The null hypothesis of our Wilcoxon rank-sum test is as follows:

> **Null Hypothesis:** *Democratic voters and Republican voters experience the same level of difficulty voting in the 2020 election.*

To address the effect of the participant's location (urban vs. rural) on the level of difficulty in voting, we added two secondary null hypotheses:

> **1.** *Democratic voters and Republican voters from urban areas experience the same level of difficulty voting in the 2020 election.* **2.** *Democratic voters and Republican voters from rural areas experience the same level of difficulty voting in the 2020 election.*

3 Results

```
#Wilcoxon test for all areas
test_all <- wilcox.test(anes_filtered$difficulty_calc ~ anes_filtered$party_id_calc)
```

We see that the p-value is 0.0015, so we reject the null hypothesis that Democrats and Republicans experience the same level of difficulty voting.

Next, we run the Wilcoxon rank-sum test after dividing the participants into two groups based on whether they are in an urban or rural area. This test's whether a participant's difficulty voting may be related to their geographic location.

```
#Wilcoxon test for urban only
test_urban <- wilcox.test(anes_filtered_1$difficulty_calc ~ anes_filtered_1$party_id_calc)

#Wilcoxon test for rural only
test_rural <- wilcox.test(anes_filtered_2$difficulty_calc ~ anes_filtered_2$party_id_calc)
```

The p-values from the tests for urban areas and rural areas are 0.091 and 0.057, respectively, which are both larger than 0.05. So we fail to reject the null hypotheses that Democrats and Republicans experienced the same levels of difficulty voting in the 2020 election when considering their geographic area.

4 Discussion

The mixed results of our tests demonstrate the importance of understanding the limitations of the chosen test, how we structure our hypotheses, and how we interpret the outcomes. The Wilcoxon rank sum test does not indicate a direction to the relationships nor does it indicate causality. These results may be consistent with fact pattern that Republican lead states have erected barriers to voting in urban areas that tend to lean towards Democrats. While those barriers may effect both side equally in an urban setting, if urban areas are a major source of Democratic votes, the state wide effect would be significant. The same could also be true for rural Republicans in Democratic leaning states.

While our results are inconclusive, they may be of interest to researchers who plan to dig further into this topic. Future researchers may want to consider the use of statistical methods that can determine causal relationships and estimate the magnitude of the impact of confounding variables like geographic area. Finally, future researchers may wish to consider a time series analysis to determine whether barriers to voting have been erected or changed overtime.