The Difficulty Adjustment

Lab 1, Question 2: Did Democratic voters or Republican voters report experiencing more difficulty voting in the 2020 election?

Hannah Gross, Matt Kinkley, Jacquie Nesbitt, Huyette Spring

Contents

1 Importance and Context **2**

2 Description of Data **2**

3 Most appropriate test **3**

4 Test, Results, and Interpretation **4**

5 Test Limitations **4**

6 Bibliography **4**

1 Importance and Context

A citizen's right to vote is a pillar of democracy in the United States. However, the attainment of voting rights has been a long-standing struggle for women, minorities, and the socioeconomically disadvantaged. Today, voter suppression and election integrity have become an acute source of conflict between the two US political parties. Democrats view increasing voter restrictions to have a disproportionately negative effect on minorities and the socioeconomically disadvantaged, some of the Democrats' most reliable voters. On the other hand, Republicans have expressed deep concern over COVID-19 inspired measures to increase the use of mail-in and early voting because they believe that if methods of voting are too easy, it will encourage greater rates of voter fraud.

This dispute suggests that Democratic voters experience greater difficulty voting and Republican voters less so. But is this true? Inequity voting could have grave implications. In the aftermath of the 2020 election, at least 250 new election laws have been proposed in various levels of government that could increase this potentially disproportionate effect (Gardner, Rabinowitz and Stevens). Regardless of which side of the aisle you stand on, if certain groups of voters face greater difficulty voting than others, this indicates a potentially systemic constitutional infringement of rights that must be addressed.

2 Description of Data

We addressed this research question using data from the 2020 American National Election Studies (ANES) Time Series Study. This study is an observational dataset based on a combined sample of respondents who participated in the prior ANES 2016 Time Series Study and a freshly drawn cross-section of randomized addresses of eligible voters at the time of recruitment. The study is a preliminary release of combined pre and post-election data and, therefore, subject to future changes in cleaning, processing, documentation, data, and variables.

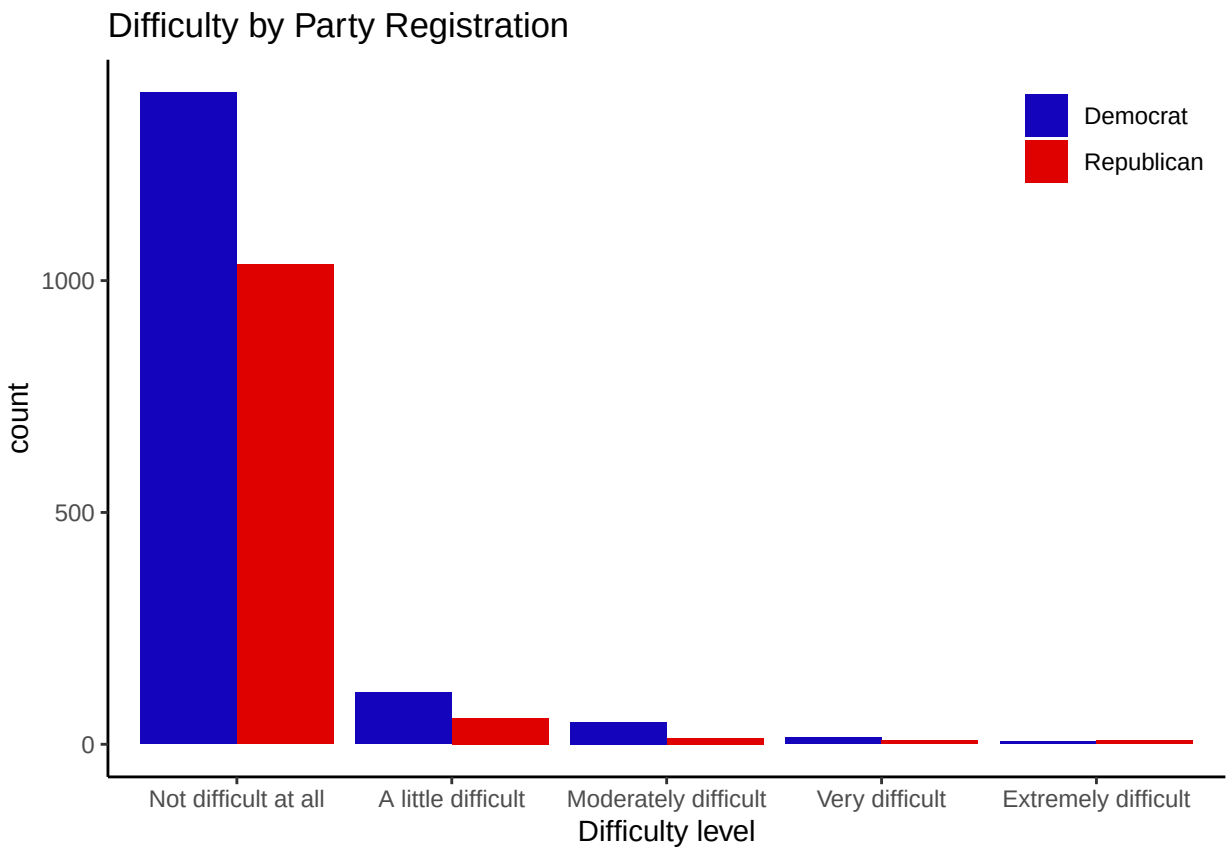
For our grouping variable, we determined whether a participant was a Democrat or a Republican using the following survey question (V201018): "What political party are you registered with, if any?" Respondents answered by selecting one out of the following four options: "Democratic party", "Republican party", "None or 'Independent' ", or "Other". This question appeared only if the respondent was registered to vote or registered to vote in a state where the party may be registered at the time of survey. We believed this is an appropriate and conservative variable to operationalize the research question because we wanted to focus on individuals who have self-declared their party and not their ideological leanings.

We considered alternative variables like the self-reported party identification variable. However, we decided against using a variable associated with party identification because of the iterative process of asking study participants to define their partisan leanings. The iterative process compels respondents to identify with one of the two major US political parties. For this research question, we did not want to use these potentially inaccurate responses. Rather we wished to identify Republicans and Democrats by how they self-identified in their voter registration, a process independent of the ANES survey.

Another alternative was to create a composite variable that would include the sum of votes across all political races. However, this would have required more time and resources available but is worthy of future exploration.

For our response variable, we used the following survey question (V202119): "How difficult was it for you to vote in this election?". The data is reported on a Likert scale from 1 - Not difficult at all to 5 - Extremely difficult. This question was presented in forward and reverse order to mitigate the effect of inattentive survey respondents that may attempt to speed through the survey or select the same answer for each question. We chose this question because it provides a range of difficulty levels and represents the total experience of the respondent. That being said, this is self reported data which is subjective.

We operationalized the party of registration variable by removing all empty values and dropping all answers not equal to the Democratic Party or Republican Party. As for the age variable, we removed all empty values and non-applicable values. As for our difficulty voting variable, we filtered for only values between "1 - Not difficult at all" to "5 - Extremely difficult". After subsetting the data to only registered Democrats and Republicans, our sample contained 2,709 observations.



3 Most appropriate test

Difficulty voting is measured on an ordinal scale, so a non-parametric test is necessary to evaluate if Democrats or Republicans are more likely to experience difficulty voting. We also determined that the test was unpaired for evaluating difficulty voting for two separate groups of respondents, Democrat/Republican party registrants. Therefore, we decided on a Wilcoxon Rank Sum Test, implemented in R using wilcox.test. Also, we decided to adopt a more conservative statistical approach by conducting a two-tailed Wilcoxon Rank Sum Test, as one-tailed Wilcoxon Rank Sum Test limits the scope of the test and may hide important statistical results.

The Wilcoxon Rank Sum Test requires the following assumptions to be true.

The data is an Independent and Identically Distributed (I.I.D.) sample:

The Wilcoxon Rank Sum Test requires that the data is generated through an I.I.D sampling process. This survey is a combined sample of respondents who participated in the ANES 2016 Time Series Study and a freshly drawn cross-section of randomized addresses across 50 states and the District of Columbia of eligible voters at the time of recruitment. For the fresh cross-section, respondents were sent a series of letters to recruit one household member to complete a survey for each randomly selected address. However, it is important to note that ANES investigators detected instances where the person who took the initial screener was not the same person who completed the ANES survey. ANES investigators flagged those instances and are expected to provide updated data in the final release. This dataset meets the I.I.D. requirement because each data entry in the set does not provide any information about any other entry in the set. However, the mix-matching of individuals is a limitation in the quality of the data and statistical methods used to analyze it.

The scale of measurement is ordinal:

Wilcoxon Rank Sum Test requires that the data is ordinal. The difficulty voting variable fits this requirement since we are working with ordinal data.

4 Test, Results, and Interpretation

Null Hypothesis: $P(X > Y) = P(X < Y)$

Alternative Hypothesis: $(X > Y) \neq P(X < Y)$ (two-tailed)

Rejection Criteria: $P - value \leq .05$

The null hypothesis for this test is that the probability that a Democrat's difficulty voting ranks higher than a Republican's difficulty voting is equal to the same probability that a Republican's difficulty voting ranks higher than a Democrat's difficulty voting. The alternative hypothesis is that the probability that a Democrat's difficulty voting ranks higher than a Republican's difficulty voting is not equal to the probability that a Republican's difficulty voting ranks higher than a Democrat's difficulty voting. We also set our rejection criteria to be less than or equal to .05.

We rejected the null hypothesis because the p-value is 0.001738, which is highly significant and less than our rejection criteria (.05). Also, the 95% confidence interval indicates... of -7.39e-07 - 4.75e-05.

From a practical standpoint, the Wilcoxon effect size is small, so even though there is a statistically significant difference in voting difficulty between Democrats and Republicans, the difference is negligible. Indicating that while there is a difference in difficulty voting for registered Democrats and Republicans, it is very small.

5 Test Limitations

The limitations of our test stem from our choice to limit our sample to only registered Republicans and Democrats from the pre-election survey as well as the research question itself. While our test was statistically significant, the raw number of respondents in our sample who reported any amount of difficulty voting was exceptionally small compared to those who had no trouble voting. It is possible that registered voters are more experienced voters, and by operationalizing the party registration variable from the pre-election survey, we may have discounted voters who registered close to or on election day. Furthermore the scope of the research question required us to group our data by Democrats and Republicans, two very broad categories. Perhaps we would see more compelling results grouping the data by age, ethnicity or geographic location. Despite the limitations of our test, we were able to find a statistically significant difference in expected difficulty voting for Democrats and Republicans. We hope that this finding will be impactful enough to inspire additional investigation.

6 Bibliography

American National Election Studies. 2021. ANES 2020 Time Series Study Preliminary Release: Combined Pre-Election and Post-Election Data [dataset and documentation]. March 24, 2021 version. www.electionstudies.org

BBC, "Georgia voting: Fact-checking claims about the new election law"

Gardner, Rabinowitz and Stevens, "How GOP-backed voting measures could create hurdles for tens of millions of voters"

Petrocik, "Measuring party support: Leaners are not independents"

Pew Research Center "Voters Rarely Switch Parties, but Recent Shifts Further Educational, Racial Divergence"