Lab 1: Question 2

Jun Qian, Lucas Schroyer, Ryan Mitchell, Oliver Chang

Contents

0.1 Importance and Context

Research question: "Are Democratic voters more enthusiastic about Joe Biden or Kamala Harris?"

Voting is the heart and basis of the American democracy. Every single vote is counted equally and it represents the will and voice of the people. Throughout American history, a few number of votes have determined the outcome of key elections. For example in the 2020 presidential election, the elected candidates secured only slight more votes in the swinging states than their opponents by sometime less than 1%. This tiny lead eventually determined who was going to run the Office. Under such circumstances, high levels of enthusiasm of the voters towards the presidential candidates is accompanied by high rates of voter turnout, which can affect the results of the election. As a matter of fact, understanding which candidate is more attractive/likable to the voters is crucial for the parties to adjust their election strategies and campaigning decisions.

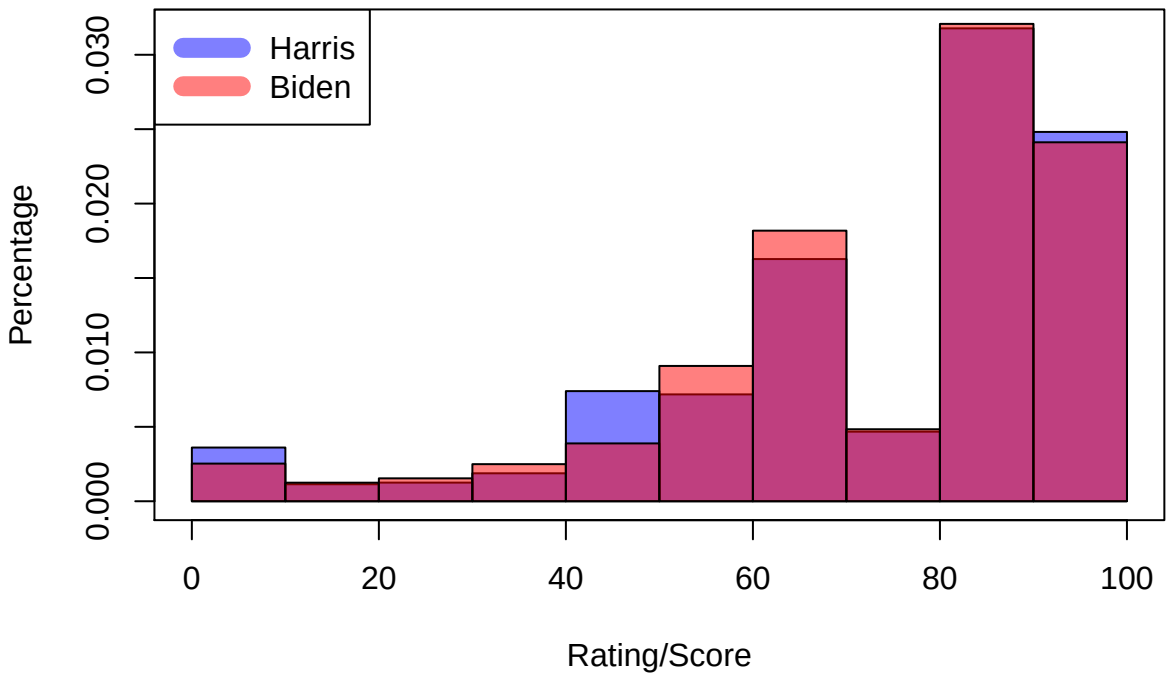
0.2 Description of Data

To answer our research question, we downloaded the pre-election data from the 2020 American National Election Studies (ANES) site. We then extracted the following variables related to this question for further processing and exploration:

- biden_rating(V201151)
- harris_rating(V201153)
- party_of_registration(V201018)
- democrat_party_rating(V201156)
- republican_party_rating(V201157)
- voted_early_in_gen_election(V201022)
- plan_to_vote_in_gen_election(V201032)
- registered_to_vote_loc(V201008)

The two variables, biden_rating (V201151) and harris_rating (V201153), were the main variables we used to infer the voter's enthusiastic rating in this study. These two variables were metric and ranged in integer value from 0 to 100. We applied filters excluding missing or inappropriate responses. For the purposes of this study, we defined "voters" as survey respondents who: (1) registered to vote (V201008) and either (2) voted early in general election (V201022) or (3) planned to vote (V201032). We also filtered the data down to Democratic voters only, using the following variables: (1) party_of_registration (V201018), (2) democrat_party_rating (V201156), and (3) republican_party_rating (V201157). In this study, we defined Democratic voters as voters who were registered as Democrats or either intended to vote or voted early for Democratic candidates. In addition, we noticed 48% of the voters didn't have applicable registered party information. However, we did have their ratings for both parties. Hence, we inferred their partisan preferences using democrat_party_rating and republican_party_rating. If the voters showed a stronger preference towards one party (rating one party 50 higher than the other party), then we assigned the voters to their preferred party. We believed including the partisan rating data into the testing samples allowed us to better capture the actual supporter population for the two parties. The more samples we had, the more statistical power we had to test our hypothesis. The total voter counts were 2691 with all the filters applied. Fig. 1 illustrates the normalized histogram of the Democratic voters' enthusiastic ratings of Biden vs. Harris. The distribution of the ratings were very similar to each other. The ratings both peak between 80 and 90, with skewed distributions to the left, due to the constraint of the rating (capped at 100).

Fig. 1. Biden vs. Harris Voter Enthusiastic Rating



0.3 Most appropriate test

To compare the relative voter's enthusiastic ratings for Biden and Harris, We thought the Sign Test was the best test to conduct for the following reasons:

(1) The data was paired.

(2) We wanted to compare the relative enthusiastic level within each voter, not with other voters. That is, we only wanted to know if a voter liked Biden or Harris more, not if a voter liked Biden or Harris more than another voter. For example, if one voter rated the two candidates at 80-90, and another voter rated at 10-20, the ratings from the former voters would account for more weights in a parametric test and could lead to wrong test results and conclusions.

Hence, we needed to convert the parametric data into non-parametric. We took the signs using the two main variables, biden_rating and harris rating. If one candidate had a higher rating, then a + was assigned to the candidate. Once the signs were marked, we calculated the sum of total + for Biden and Harris (Table 1). There were 779 + for Biden, 765 + for Harris, and 1147 draws. In the Sign Test (Binomial Test), the draws were excluded because they didn't tell if one favored either candidate. Hence, the total number to trails were $779 + 765 = 1544$.

There were some assumptions for the test to be valid.

(1) The data extracted from the population had to be I.I.D.. Clustering effect could skew the data during sampling process. In addition, the weights that accounted for the population size of each sampling location were not applied in this study for simplicity reasons.

(2) The data had to be non-metric. We had converted the parametric data to non-parametric, binary in this case, in the previous step. This method also didn't require a strict minimal sample size. However, we had more than 1000 samples, which strengthened the credibility of the test.

The goal of this test was to exam if one candidate were statistically more favored by the voters than the other. The Null Hypothesis for the test was p = 0.5, which meant the two candidates were equally favored by the voters. We used a two-tailed test in this study, since we didn't know which candidate had more + signs at the beginning.

Table 1: Counts of Signs of Voters' Enthusiastic Ratings

Biden_signs	Harris_signs	Total_Trials
779	765	1544

```
biden_harris <- anes_timeseries_q2 %>% select(biden_rating, harris_rating) %>%
      filter(harris_rating >= 0 & harris_rating <= 100
             & biden_rating >= 0 & biden_rating <= 100)
biden <- biden_harris %>% select(biden_rating)
harris <- biden_harris %>% select(harris_rating)
biden_higher <- ifelse((biden - harris) > 0, 1, 0)
total <- ifelse((biden - harris) != 0, 1, 0)
binom.test(sum(biden_higher), sum(total), p = 0.5,
         alternative = c("two.sided"),
         conf.level = 0.95)

harris_higher = sum(total)- sum(biden_higher)
r = (sum(biden_higher)-harris_higher)/sum(total)
```

0.4 Test, results and interpretation

The test results indicated that the Null Hypothesis was not rejected. The reported p-value was 0.74 for a two-tailed Sign Test. This meant, statistically, the Democratic voters weren't more enthusiastic for either Biden or Harris in the 2020 presidential election. The results suggested that there was no single Democratic candidate that attracted statistically more votes than the other. The results could also suggest that the Democratic voters were equally satisfied (both had high ratings in average) with the two candidates. In addition, the effect size (r) using proportion approach was 0.01. The effect size was really small that the difference didn't have practical significance either. Therefore, no further major adjustment for the campaign strategy regarding the two candidate was required. Practically speaking, if the results came back to be significant, then the Democratic Party might consider strategies to enhance the enthusiastic rating for the candidate with lower ratings or consider other candidates.