

Lab 1: Question 2

Tilman Bayer, Kieran Berton, Ruilan Liu, Toby Petty

Question 2: Are Democratic voters more enthusiastic about Joe Biden or Kamala Harris?

Importance and Context

The 2020 US General Presidential Election marked the first time since the 1992 US General Election that an incumbent candidate was ousted by their challenger, a feat that has only been performed 10 times previously¹. Biden's success was largely due to him receiving the highest number of votes for any candidate in US history, over 81 million, with many votes coming from ballots cast early and returned by mail². With such marked enthusiasm for a candidate, it is important to understand what were the driving factors that caused voters to prefer the Democratic presidential ticket.

President Biden's running companion, Kamala Harris, was also in her own way historic, being the first African-American, first Asian-American, and third female vice presidential nominee on a major party ticket. Civil unrest in the years prior to the election and the growth of the Black Lives Matter movement have created a polarized political landscape where identity and demographic politics have grown in importance to voters, and it would be remiss to not attribute some amount of the Democratic party's electoral success to the inclusion of a candidate of minority and mixed ethnic background. Long-time Democratic voters are likely more familiar with Joe Biden, whose political career has spanned over four decades and who most recently was Vice President under Barack Obama between 2009 and 2017. But did this familiarity translate to enthusiasm prior to the 2020 election, or were voters more enamored with the Senator from California, Vice President-Elect Kamala Harris?

Using data from the American National Election Study (ANES) 2020 Pre-Election Survey, we will compare Democratic voter enthusiasm between Joseph Biden and Kamala Harris to judge which candidate was a greater pull for voters heading into the general election.

Description of Data

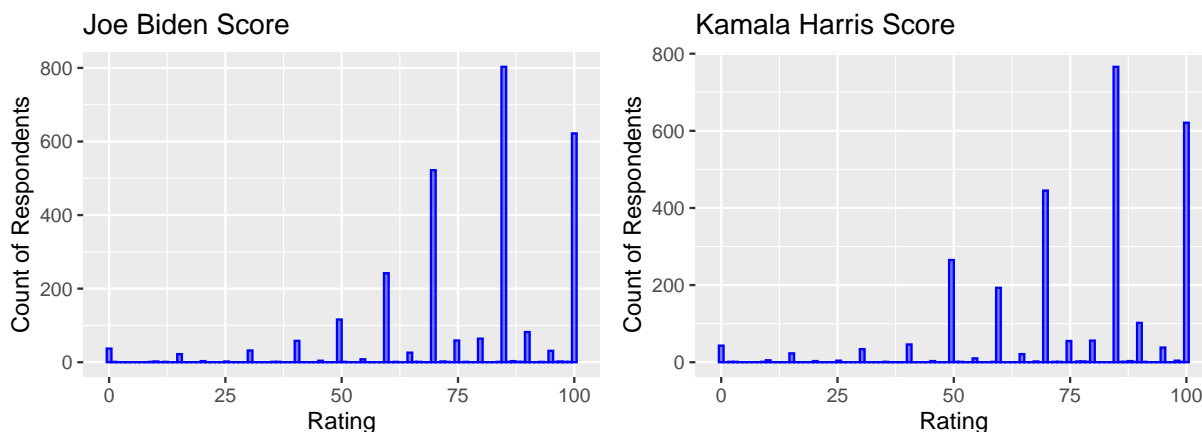
Please refer to our report for question 1 for general remarks about the ANES dataset,³ and for details about our conceptualization of "Democratic voter" based on this dataset, which yielded a subset of 2864 people, or about 34.6% of survey respondents.

Of this subset, we were interested in those respondents who gave answers to questions V201151 and V201153 on the ANES Survey, which asked respondents to rate their enthusiasm for Biden and Harris respectively on scales from 1-100. Of the 2864 reported Democrats, 2761 responded to both questions asking them to rate their enthusiasm for the two candidates. Because only a small fraction of democrats did not respond to both questions, we chose to remove these respondents (which represent 3.6% of reported democrats), as we don't think it will bias our results significantly and it allows to use the paired T-test because we can directly compare responses from the same respondents. Below is a histogram showing the distribution of responses for each candidate.

¹<https://www.cnn.com/2012/11/09/politics/btn-second-term-presidents>

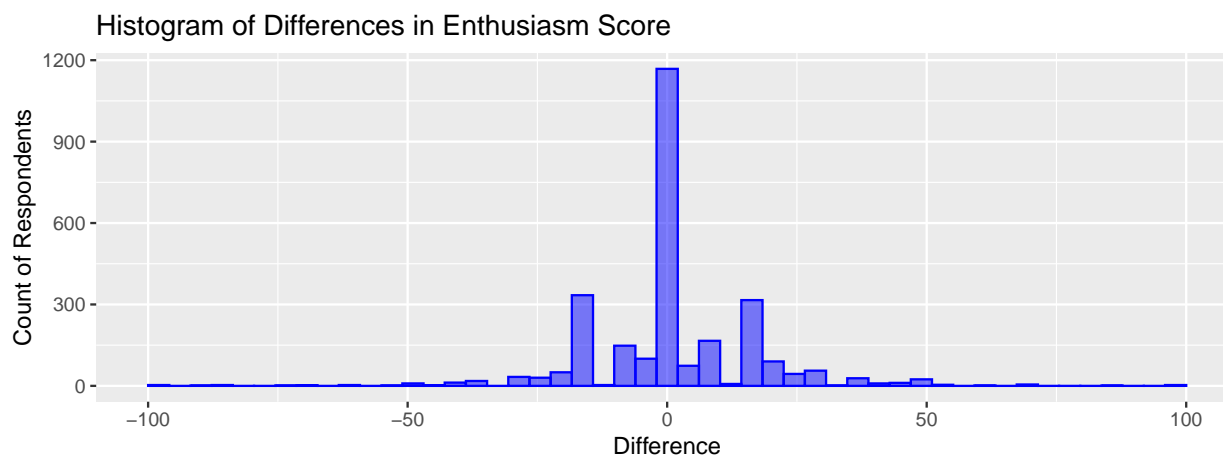
²<https://www.newsweek.com/joe-biden-crosses-80-million-votes-10-million-more-obama-got-1549612>

³To call out one aspect here: As advised by ANES, an accurate analysis based on this data would have to account for the survey weights included in the dataset. Per the instructors, we are ignoring them for the purposes of this labs project.



As we can see, responses are concentrated on “round numbers” that end in 0’s or 5’s, so that the distribution of responses is near 0 everywhere except for those spikes in response rate.

The means of the two distributions are difficult to identify by eye, and harder still is identifying which candidate has the higher average enthusiasm score. The mean average enthusiasm score for Joe Biden is about 77.9, while the mean average score for Kamala Harris was about 76.8. Given that the distributions overlap almost entirely, it is much more useful to examine the distribution of differences between responses for each candidate for each respondent.



Most appropriate test

To answer the question of whether Democratic voters are more enthusiastic about Joe Biden or Kamala Harris we choose to use the Paired T-Test because the data consists of a single set of responses in which each survey respondent was asked to separately rate their enthusiasm (via a “feeling thermometer”) for both candidates, so there is a natural pairing of responses for both candidates from every individual who answered both questions. We naturally expect there to be some dependence between individuals’ responses for each candidate; for example, those who were feeling overall really enthusiastic about the ticket’s chances in the election may give high responses for both candidates, and conversely those feeling doubtful about their party’s chances of winning may rank both candidates with low enthusiasm. Since we expect some relationships like this to exist in the data, we choose to use the Paired T-Test on the relative difference between the enthusiasm scores for the 2 candidates, which naturally lends itself to establishing if there is a difference in average rating across all Democratic voters in the sample.

The Paired T-Test assumes that responses are recorded on the same metric scale where the difference in meaning between values remains consistent with the metric distance between values. In this case, our

responses are recorded on the same scale from 0-100, and though we cannot be entirely sure that respondents have consistent feelings about what level of enthusiasm is represented by each and every number on that scale, we can reasonably assume that respondents treated the differences between numbers on this scale consistently. We assert that since a rating of 0 to 100 is analagous to a percentage, with which most people are very familiar, that this is a reasonable assumption on which to proceed.

The Paired T-Test also assumes that each pair of responses is drawn I.I.D., which in this case is satisfied because of the experimental methodology of ANES. The ANES sample is built from a pool of adults 18 years or older living in the United States and was built using mailing addresses provided by the United States Postal Service (USPS). Lastly, it must be the case for the Paired T-Test to be valid that the differences between responses be distributed normally, or be sufficiently normal given the sample size. Given our sample size of 2761, we can reasonably assume the distribution of differences to be normal, and we also observe that the distribution plotted in the histogram of differences appears approximately normal, in that it is clustered around the center and not heavily skewed in either direction.

Since we are using a two-sided Paired T-Test, the null and alternative hypotheses can be stated as:

- **Null Hypothesis:** The mean of the differences between Democratic voters' enthusiasm scores for Joe Biden and Kamala Harris is zero.
- **Alternative Hypothesis:** The mean of the differences between Democratic voters' enthusiasm scores for Joe Biden and Kamala Harris is either above or below zero.

We will assess the statistical significance of our test using an alpha value of $\alpha = 0.05$, based on this being a standard value used in many political science analyses.

We think this is the most appropriate test to use in this situation. A non-parametric test or a non-paired test would have the disadvantage of lower statistical power (i.e. being less likely to reject the null hypothesis if it is not true).

Test, results and interpretation

```
t.test(h_and_b$V201151, h_and_b$V201153, paired = TRUE, alternative = "two.sided")
```

Performing a Paired T-test, we can see that with $p < 0.001$, the mean enthusiasm for Joe Biden is greater than the mean enthusiasm for Kamala Harris, with the mean difference being about 1.09 in response. Though this p-value is very significant, a response that is only higher by 1.09 on a scale from 0 to 100 practically represents very little. For example, the computed Cohen's D statistic for the difference is 0.0673, which using the conventional interpretation qualifies it as an effect size somewhere between "Small" (0.2) and "Very Small" (0.01)⁴. It makes sense that there would be greater enthusiasm for a primary candidate rather than their running mate in any election cycle and for any political party given that voters likely consider the primary candidate to be the one holding the more important office. However, such a small difference in enthusiasm teaches us relatively little about the underlying reasons for voters being enthusiastic about either candidate and does not indicate there is any great difference in opinion about the two running mates among Democratic voters.

⁴https://en.wikipedia.org/wiki/Effect_size#Cohen's_d