

Lab 1: Question 2

Tal Segal, Xander Snyder, Patrick Old

Importance and Context

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

As Joe Biden’s campaign for President of the United States entered August of the most dreadful year in recent memory, it still lacked one crucial element: a running mate. There was a lot riding on this decision, and his eventual choice on August 11th, 2020 of Senator Kamala Harris was one that made history. After Biden defeated Trump by a narrow margin in November, Harris became both the first black woman and South Asian American to be elected second in command.

It is a common practice for a presidential candidate to choose a running mate that appeals to voters that they themselves lack a connection with. Harris and Biden are similar enough to work together toward shared goals, but different in many readily apparent ways. Kamala, 56, is young compared to her counterpart, who received a large amount of press for his elderly status of 78 years old at the time of taking office. Furthermore, Harris’ rich ethnic background and skin color drew much attention during the campaign, compared to Biden, whose background in these regards resembles that of almost every former president and vice president. These differences combine to reach a broader audience, but present a question; who are Democratic voters more enthusiastic about?

The answer to this question may help explain the degree to which Biden and Harris, individually, contributed to the Biden-Harris ticket’s win. Furthermore, the results could give some indication of how open Democrats are to a younger candidate with a more diverse background, compared to that of their more traditional counterparts. Additionally, they may also provide insight into how Democrats could respond to policies implemented by the Biden-Harris administration, and whether they will be excited, or merely lukewarm, about the prospect of a Harris presidency should President Biden pass while in office.

Description of Data

We will address this question using pre-election data from the 2020 American National Election Studies (ANES). We break down the research question into three key parts: 1) defining which observations qualify as being “Democratic”, 2) defining which observations qualify as a “voter”, and 3) measuring enthusiasm for each observation.

The concept of determining if an individual is “Democratic” or a “voter” is simpler in post-election data than pre-election data. Following the election, it is clear whether an individual voted and for whom. However, pre-election data is more complicated as respondents have not yet cast a vote. Instead, we must rely upon whether respondents indicated their intent to vote in the upcoming election.

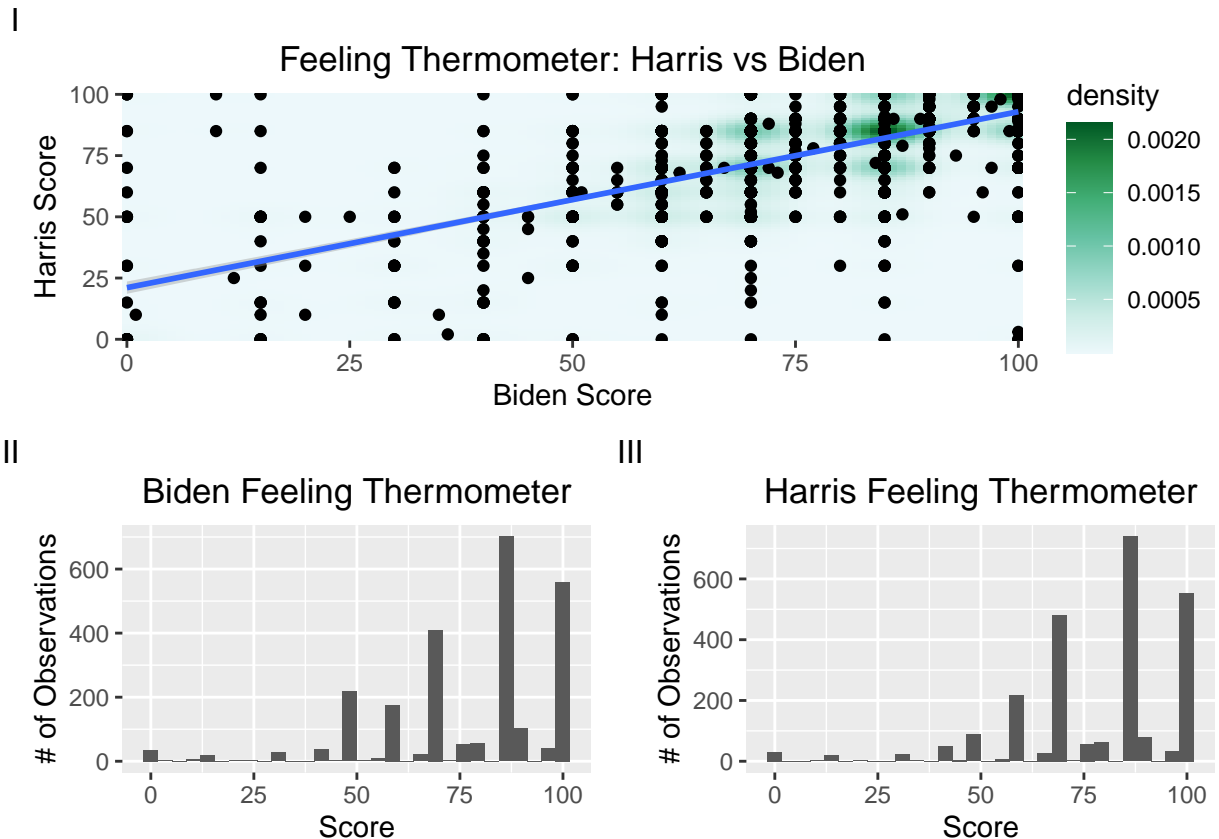
34.59% of respondents answered “democrat” when asked the following question: “Generally speaking, do you usually think of yourself as a Democrat, a Republican, or an independent?” We defined these individuals as democrats. Those who answered affirmatively to the question “Do you intend to vote in the November election for President?” were defined as voters. Voters represented 87.82% of the total sample.

Determining voter enthusiasm for specific candidates was a somewhat more complicated task. The only metric provided in this study that measured both Joe Biden and Kamala Harris separately on the same scale was what the ANES refers to as a “feeling thermometer”¹. This “feeling thermometer” is meant to measure a

¹See the ANES 2020 User Guide and Codebook page 87 for a full description

respondent's favorability and warmth toward a candidate on a scale of 0 - 100. Responses of greater than 50 indicate a warmth or favorability toward the candidate, where as anything below 50 indicates the opposite. A score of 50 is considered neutral. After filtering the feeling thermometer for valid responses (those that fall in the zero to 100 range), and applying the two other filters mentioned above, 2483 responses (29.99% of the total sample) are left to analyze.

Both distributions of Harris' and Biden's feeling thermometer have a left skew (figure 2,3), which results in a strong positive correlation when both are plotted on each axis of a scatter plot (figure 1). The plots are of no surprise considering each member is of the democratic party and those are the only voters in the filtered sample for this study. However, it is notable that there is a dip toward the upper end of each candidate's histogram, where the mode and median are both equal to 85. This indicates that while Democratic voters score Biden and Harris with high feeling thermometer scores, there still is a hesitation to back them with full score. Furthermore, respondents tended to round to the nearest 5th, with the most common responses coming in at 85, 100, and 75 for both candidates in that order. This created clearly marked density zones in the scatter plot (figure 1). Here, all points are plotted, but due to the overlap seen in a few regions that would not be visible with points alone, a density gradient is added to assist in visualizing the data.



Most appropriate test

I then test whether this subgroup reports having greater enthusiasm for Kamala Harris or Joe Biden. I initially considered three statistical tests for this study: the paired t-test, the Wilcoxon Signed-Rank Test, and the Sign Test. A two-sample paired t-test would assume that 1) the data is metric, 2) the data is paired, and 3) the sample size is large and/or follows a normal distribution. While the data appears quantitative, it is in fact ordinal, not metric. Metric data requires that the magnitude between two equally-spaced data points be equal. (i.e. the difference between 0 and 10 is the same as that between 90 and 100). However, in this case, it's unclear what a "feeling thermometer" degree even means, aside from providing for somewhat greater granularity in responses. It's certainly impossible to know if someone with a score of 70 is equally

more excited than someone with a score of 60, as that person with a score of 60 is when compared to someone with a score of 50. Therefore, we should treat the “feeling thermometer” scale as ordinal, or ranked, data.

The second possibility is the Wilcoxon Signed-Rank Test. The assumptions for this test are that 1) the data is of an interval scale, 2) the data is paired and drawn IID, and 3) the data is symmetric (as it is a paired test in this case, the symmetry is based upon the paired differences). While this data meets the second assumption clearly and largely meets the 3rd assumption (a box plot showing near symmetry is omitted due to space concerns), the first requirement of interval data fails as the data is ordinal.

This leaves us with the Sign Test. The assumptions for this test are that 1) the data is ordinal, and 2) the data is paired and drawn IID. The ordinal assumption is met based on the above discussion. The IID assumption is also met, as the ANES 2020 survey sampled from both a fresh cross-section of new respondents as well as a number of prior respondents to the ANES 2016 survey. The collection method for the fresh sample was mixed method, comprised of web surveys, phone surveys, and video conference surveys. While reaching out to people who were willing to respond to the 2016 survey could create some dependencies in that subset, the sample frame, sample size, and sampling method largely fit the i.i.d. assumption.

This simple test relies upon finding the sign of the difference between pairs, but doesn’t worry about the magnitude of that difference. While this makes the Sign Test an extremely flexible tool, it also often requires large amounts of data to find significant results since the statistic it uses relies on such basic information. Furthermore, much of the dimensionality of the survey data is discarded for this test, which is why it is not as commonly-used as others tests. Nonetheless, it is applicable for this study.

Test, results and interpretation

```
num_equal <- as.integer(as.logical(jk$kamala == jk$joe)) # find observations with no sign
kamala_greater <- as.integer(as.logical(jk$kamala > jk$joe))
binom.test(sum(kamala_greater), length(kamala_greater) - sum(num_equal), p = 0.5,
           alternative = c("two.sided"),
           conf.level = 0.95)
```

I fail to find evidence that there is a difference in enthusiasm for Joe Biden and Kamala Harris in Democratic voters. The p-value for the sign test is 0.0742. Generally speaking, 2483 data points is enough data to use a sign test in a significant manner. However, because 41.56% (1032 observations) of the sample recorded an equal level of enthusiasm for both politicians, these data points resulted in no additional data being contributed to the test statistic. In other words, they were essentially dropped from the test. This relatively high percentage of ties is not surprising given the similar distributions visible in the histograms above, as well as the clear correlation between points visible in the density scatter plot. It’s worth noting, but not wholly relevant to our test, that if twice as many samples were collected that had the same ratio as that which we calculated, the significance of the test would go up (p-value of 0.011).

Although the test failed to reject the null hypothesis, we briefly consider the effect size. Despite the large number of democratic voters that felt equally enthusiastic about Biden as Harris, 27.83% of respondents reported a higher feeling thermometer score for Kamala Harris than for Joe Biden. 30.61% reported the reverse. If tied feeling scores are ruled out as they are by the Sign Test, this results in 47.62% of Democratic voters having greater enthusiasm for Kamala Harris than Joe Biden, and 52.38% vice versa. This indicates no significant difference in sentiment for either party member. Furthermore, it is notable that while Biden and Harris have many readily apparent differences, Democratic voters across the nation appear to view them with largely equal enthusiasm.