

# Question 2

Jyoti Kumari

## Contents

<b>1</b>	<b>Are Democratic voters more enthusiastic about Joe Biden or Kamala Harris?</b>	<b>2</b>
1.1	Importance and Context . . . . .	2
1.2	Description of Data . . . . .	2
1.3	Most appropriate test . . . . .	3
1.4	Test, results and interpretation . . . . .	4
1.5	Conclusion . . . . .	5
1.6	References . . . . .	5

Table 1: Democratic Voters

	Registered_Voters	Presidential_Voters	House_Voters	Senate_Voters
Count of Democratic voters	1788	3974	2980	1592

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

## 1.1 Importance and Context

The research question here is a very important to the readers as it will help measure the enthusiasm of democrats voters for not only the presidential candidate, Joe Biden but also for the vice presidential candidate, Kamala Harris. It would be interesting to analyze what impact can Harris have over Biden’s win amongst the democratic voters.

Depending on the enthusiasm, it can be inferred or predicted that voters who are more enthusiastic about Harris would more likely vote for Democrats, eventually bringing in more votes for Biden. Harris can also help pull in voters from certain communities, including Women and Liberals, that could increase the chance of Biden winning the presidential election.

## 1.2 Description of Data

We will address this question using data from the 2020 American National Election Studies (ANES). This is an observational dataset, based on a sample of respondents.

As we report in Table 1, 1788 respondents are registered democratic voters, 3974 respondents are presidential voters, 2980 are house voters and 1592 are senate voters.

To examine factors behind Democratic voters, we limit the data to individuals who are registered with democratic party(V201018). In addition to this, we also considered the voters who either selected to vote, intended to vote or preferred to vote for a democratic candidate in a presidential(V201075x), House(V201076x), Senate(V201077x), or Gubernatorial(V201078x) election. Furthermore, since count of respondents for Gubernatorial(V201078x) election is very less, it has been dropped from displaying on the table 1. After subsetting, we have table 1 observations.

The survey includes questions about how enthusiastic the respondents are for Joe Biden or Kamala Harris. Data is reported on a scale of 0-100. We have converted the given scale to a 5-point Likert scale that ranges from 1 (“least favorable”) to 5 (“most favorable”). These questions are about enthusiasm in general, not necessarily emotions relevant to voting, and the survey does not provide more information about whether these emotions *cause* voting behavior to change.

Because our research question is fundamentally causal, answering it in a convincing way would require us to conduct an experiment that might take the following form. However, the ANES data is observational rather than experimental. Consequently, we recommend care in reasoning about whatever relationship this analysis might find.

There is a strong positive relationship between a respondent’s answer to questions about feeling thermometer for Joe Biden and Kamala Harris. In Figure 1.1 we plot, the enthusiasm is almost equal for both the candidates. Clearly, there is a strong positive relationship between these measures.

The next figure 1.2 plots the relative prevalence of ANES respondent’s enthusiasm for Joe Biden and Kamala Harris. Overall, there are large number of respondents with higher enthusiasm as can be seen with taller bars on right side of figure. And, there are fewer respondents who have shown lower enthusiasm for Biden or Harris as seen on the left side of the figure. While there are about the same proportion of respondents who report being the lowest category in both, there are more who report being somewhat more enthusiastic about

Joe Biden than for Kamala Harris. This is offset by fewer individuals reporting high levels of enthusiasm for Kamala Harris.

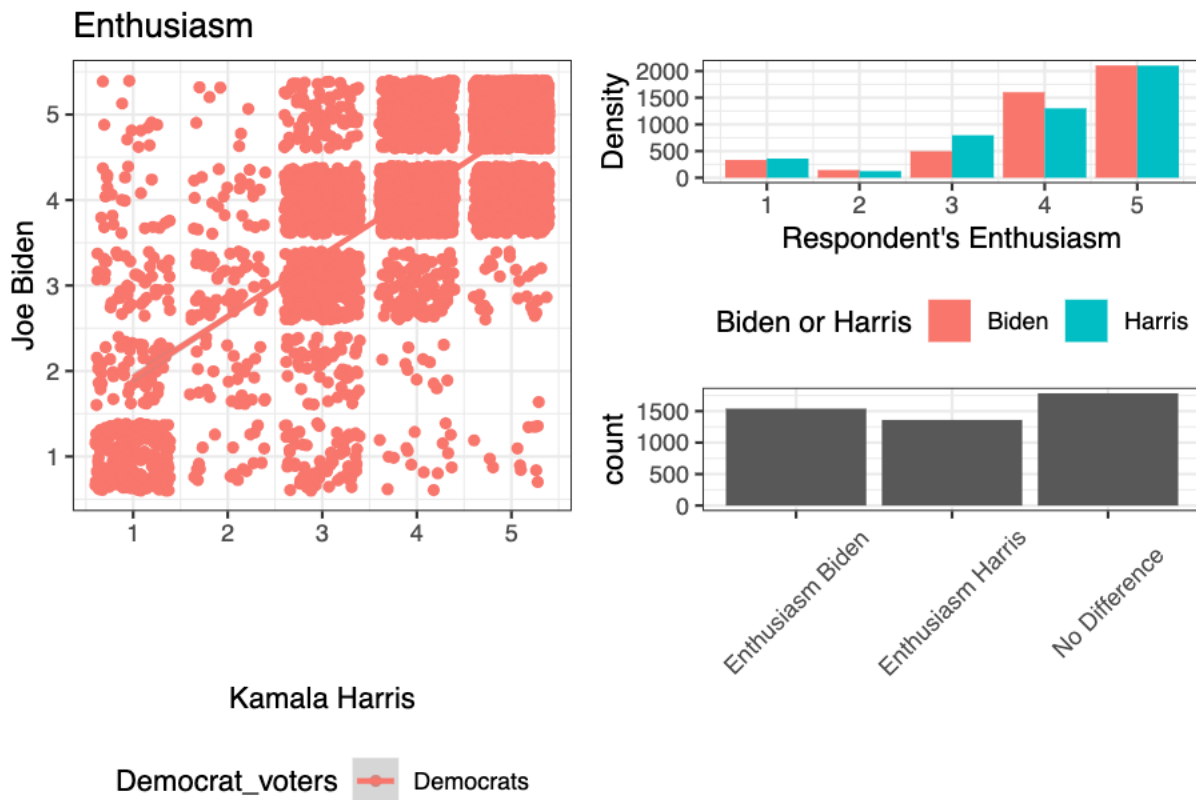


Figure 1: Democrats enthusiasm for Harris and Biden

### 1.3 Most appropriate test

We then test whether this subgroup reports being more enthusiastic about Joe Biden or Kamala Harris. Because these variables are measured on ordinal scales, a non-parametric test is appropriate. Furthermore, the data is paired, since each individual has a pair of measurements. Some researchers may use a signed-rank test in these circumstances, but that test, while technically non-parametric, requires data on a metric scale. Instead, we will use a sign test, implemented in R using `binom.test`.

The sign test uses the `'binom.test'` function to decide if a binomial distribution has an equal chance of success and failure (In this case, an equal enthusiasm from democratic voters for Joe Biden and Kamala Harris).

There is also a pairing for this data. The respondent themselves is the unit of observation, and there are two measurements for each unit. This suggests that the sign test should be the most appropriate test. The required assumptions are:

- i.i.d. data. The ANES 2020 uses respondents from a fresh cross-section of respondents, and a collaboration to interview a subset of respondents from the General Social Survey (GSS). This data collection process also rewards individuals for filling out surveys. There is a possibility that this introduces dependencies. For example, participants may tell friends or family members about this survey, resulting in a cluster of individuals that give similar responses. Nevertheless, since the data collection process claims to have millions of users, which suggests that links between individuals should be rare.

As a modern survey, the GSS looks for each respondent beginning from scratch to avoid dependencies. You might worry that some people are more likely to respond than others. But this isn't a problem for iid, it just means that the population of respondents is not the same as the population representing all Americans. It's a problem for generalization, but the test is still valid with that qualification. Since there is only a finite number of Americans, it is impossible to have perfect independence. Once a person is drawn, they cannot be drawn again, so the distribution for the next draw has changed. However, given there are a substantial number of Americans, the change is likely to be very small, so we can safely ignore these finite-sample effects (in general something to worry about with much smaller populations). On the whole, there are issues to worry about, but modern surveys like the GSS do a pretty good job of making the iid assumption valid.

- Ordinal scale. The data levels show an increase in intensity from “least favorable” to “most favorable.” There is some question, however, as to whether participants use the same scale when they think about enthusiasm for Joe Biden or Kamala Harris. Another way of noting this is to ask whether voters hold the same anchoring points on these scales.

We note that these are ordinal variables. It is clear that the difference between, say, ‘least favorable’ to ‘most favorable’ cannot be equated to the difference between ‘particularly favorable’ and never. This means that we cannot run a parametric test.

For the binomial test:

$H_0$ (probability of equal enthusiasm for Joe Biden and Kamala Harris) = 0.5

$H_A$ (probability of equal enthusiasm for Joe Biden and Kamala Harris)  $\neq$  0.5

```
binomial_test <- anes_timeseries_data %>%
  filter(
    Democrat_voters == 'Democrats') %>%
  mutate(
    more_enthusiastic_for_biden = V201151 > V201153,
    more_enthusiastic_for_harris = V201153 > V201151) %>%
  binom.test(
    x = sum(more_enthusiastic_for_biden),
    n = sum(more_enthusiastic_for_biden, more_enthusiastic_for_harris))
binomial_test

##
## Exact binomial test
##
## data: sum(more_enthusiastic_for_biden) and sum(more_enthusiastic_for_biden, more_enthusiastic_for_harris)
## number of successes = 1539, number of trials = 2899, p-value =
## 0.0009434
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.5125106 0.5491725
## sample estimates:
## probability of success
##           0.5308727
```

The number of people who are more enthusiastic for Biden than Harris is 179 (13% of Harris enthusiasts).

## 1.4 Test, results and interpretation

We have a highly significant test. We can reject the null hypothesis that there is an equal enthusiasm for Joe Biden and Kamala Harris among the democratic voters. Indeed, the p-value for the test is 0, which is within the rejection range.

To express practical significance, the common language effect size is a nice solution: Among people that express a difference, ' $r = (\text{data in favor} - \text{data against}) / (\text{total data}) = (1539 - (2899 - 1539)) / 2899 = 0.06$ ' showed more enthusiasm for Joe Biden than Kamala Harris.

This suggests a very small effect: Consider that the number of people who showed more enthusiasm for Joe Biden is just a little over the enthusiasm for Kamala Harris.

As a general rule of thumb, a correlation this small ( $< 0.1$ ) is considered practically insignificant.

## 1.5 Conclusion

As per the hypothesis testing, democratic voters are a bit more enthusiastic for Joe Biden than Kamala Harris but that difference is still small and practically insignificant.

## 1.6 References

American National Election Studies. 2021. ANES 2020 Time Series Study Preliminary Release: Pre-Election Data [dataset and documentation]. February 11, 2021 version. [www.electionstudies.org](http://www.electionstudies.org)