

# Lab 1 Question 2: Are Democratic voters more enthusiastic about Joe Biden or Kamala Harris?

Yao Chen, Jenny Conde, Satheesh Joseph, Paco Valdez, Yi Zhang

## Contents

|          |                                         |          |
|----------|-----------------------------------------|----------|
| <b>1</b> | <b>Importance and Context</b>           | <b>2</b> |
| <b>2</b> | <b>Description of Data</b>              | <b>2</b> |
| <b>3</b> | <b>Most appropriate test</b>            | <b>3</b> |
| <b>4</b> | <b>Test, results and interpretation</b> | <b>3</b> |

# 1 Importance and Context

Like many events in the year 2020, the 2020 United States general election was unprecedented. For Democrats, it was a groundbreaking win. They succeeded in making Kamala Harris the first female vice president, the first Asian-American vice president, and the first African-American vice president all at the same time.

But is it Kamala that Democratic voters were more enthusiastic about in the Biden-Harris combination, or is it Biden? In this report, we analyze the relative level of support from Democratic voters between Biden and Harris using the data from the American National Election Studies (ANES) 2020 Time Series Study.

# 2 Description of Data

The American National Election Studies (ANES) conducts surveys of voters in the United States, with a flagship survey occurring immediately before and after each presidential election. ANES survey fits well in our research question. The variables tracked in the survey gives a multi-dimensional view about voters. After careful exploration, our team picked the most relevant pre-election variables to be:

- V201018 – PARTY OF REGISTRATION
- V201151 – FEELING THERMOMETER: JOE BIDEN, (D) PRESIDENTIAL CANDIDATE
- V201153 – FEELING THERMOMETER: KAMALA HARRIS, (D) VICE-PRESIDENTIAL CANDIDATE

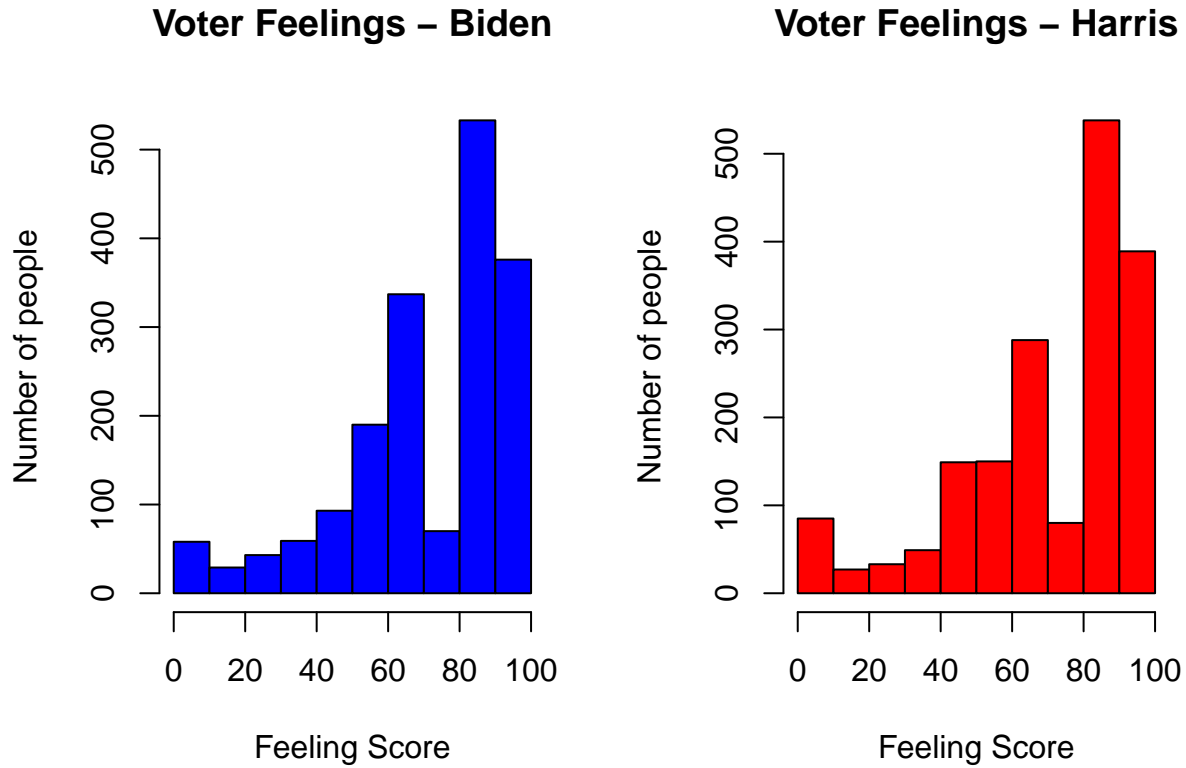
We'll use V201018 to select Democratic voters. We'll use V201151 and V201153 to represent voter enthusiasm for the respective candidates.

For both of the `enthusiasm` variables, there were potentially non-answer values, which we removed from the data. And out of 1861 voters registered as Democrats, 1788 have valid answers for both questions.

The data are summarized below:

| ## | party     | Biden          | Harris         |
|----|-----------|----------------|----------------|
| ## | Min. :1   | Min. : 0.00    | Min. : 0.00    |
| ## | 1st Qu.:1 | 1st Qu.: 60.00 | 1st Qu.: 60.00 |
| ## | Median :1 | Median : 85.00 | Median : 85.00 |
| ## | Mean :1   | Mean : 73.89   | Mean : 73.18   |
| ## | 3rd Qu.:1 | 3rd Qu.: 85.00 | 3rd Qu.: 87.75 |
| ## | Max. :1   | Max. :100.00   | Max. :100.00   |

As can be seen, the mean feeling score for both are fairly similar, and because these are Democrat voters, the scores are both fairly skewed towards the high side.



### 3 Most appropriate test

As the same voter is expressing her/his feeling towards both the candidates, the correct test for the data in non-ordinal value is paired t-test.

The assumptions for paired t-test are:

1. The dependent variable must be metric.
2. The observations are independent of one another and identically distributed.
3. The dependent variable should be approximately normally distributed given the sample size.

Going through them for this situation:

1. The feelings rating is from 0 to 100 so it is a metric scale.
2. Given the sampling frame based on a cross-section of registered addresses across 50 states and the District of Columbia, we feel the data are sufficiently close to be i.i.d.
3. As can be seen in the histograms, the data is fairly skewed and far from symmetric, so it is a concern here. However given the sample size (1788), the CLT should apply and make the sampling distribution of the mean reasonably normal-like.

### 4 Test, results and interpretation

For the test itself, we establish the *null hypothesis* to be that the average level of support of Democrats for Biden ( $\mu_B$ ) and Harris ( $\mu_H$ ) are the same.

And the *alternative hypothesis* is that they're not.

Given we have no strong initial inclination in either direction, this should be a two tailed test.

We'll be using the standard 5% significance level.

$$H_0 : \mu_B = \mu_H \qquad H_a : \mu_B \neq \mu_H \qquad \alpha = 0.05$$

```
##
## Paired t-test
##
## data: democraticDf$Harris and democraticDf$Biden
## t = -1.7858, df = 1787, p-value = 0.0743
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -1.4903780 0.0697963
## sample estimates:
## mean of the differences
## -0.7102908
```

From the test, we can see that the p-value is 0.0743, which is greater than the significance level of  $\alpha = 0.05$ , therefore we failed to reject the null hypothesis.

However given the p-value is fairly close to 0.05, it could be considered as a marginally statistically significant result for the two tailed test. Even though we do not have sufficiently strong evidence to believe Democrat voter's level of support for Biden and Harris are different given the data at a 5 significance level, there does seem to be some level of disparity.

In addition, a 95% confidence interval for the difference of voter feelings' mean between Biden and Harris  $\mu_B - \mu_H$  is (-1.4904, 0.0698). Meaning out of all confidence intervals constructed, 95% of the time the interval will contain the true mean.

On a practical level, we can see that the mean of the differences is fairly small (0.71), less than 1 point on the 0 – 100 scale.

Furthermore, if we plot the difference across the board, we can see that vast majority of Democrat voters' feeling score for Biden and Harris are within 5 points of each other, so the effect size is indeed small.

```
## [1] "40.6 % of Democratic voters gave Biden and Harris the same score."
## [1] "87.58 % of Democratic voters gave Biden and Harris scores within 20 points."
```

Although curiously, there are a few people who gave Biden and Harris very different feeling scores, as much as 100!

### Voter Feelings – Difference

