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

Importance and Context

Like many events in the year 2020, the 2020 United States general election was unprecedented. The Democrats scored a major victory in the presidential election; in particular, the election of Kamala Harris as the first female Vice President, the first Asian-American Vice President, and the first African-American Vice President was groundbreaking. The new President and Vice President have a long road ahead of them to address many of today's pressing issues, including the pandemic and economic crisis. However, we are curious if Democratic voters had differing levels of enthusiasm for Biden and Harris leading up to the election. Were Democratic voters more enthusiastic about Biden leading up to the election or Harris? Knowing the public perception of each individual could influence their public image during the course of Biden's presidency. 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 and a paired t-test.

Description of Data

The American National Election Studies (ANES) conducts surveys of voters in the United States. The pre-election results we use from this survey contain information from 8,280 pre-election interviews with U.S. citizens of voting age. The ANES survey fits well in our research question, and after careful exploration, we selected the most relevant pre-election variables:

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

that operationalize the concepts in the research question best.

We use `V201018` to select Democratic voters. `V201151` and `V201153` represent voter enthusiasm for the respective candidates and are measured on a scale from 0 to 100, with 100 being the most enthusiastic.

As no data could identify who a party registrant voted for, we assume that the registrant voted for the candidate from their registered party (i.e. A voter registered as a Democrat likely voted for the Democratic candidate). We considered using the
`FEELING THERMOMETER` variables to predict which candidate an individual would vote for. However, we felt this would involve too many ambiguous assumptions, and we opted to use `PARTY OF REGISTRATION` instead.

For both of the `FEELING THERMOMETER` variables, there were non-answer values, which we removed from the data. Out of 1,861 voters registered as Democrats, 1,788 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 distributions only represent Democratic voters, the scores are both fairly skewed, and each histogram in Figure 1 shows a left tail.

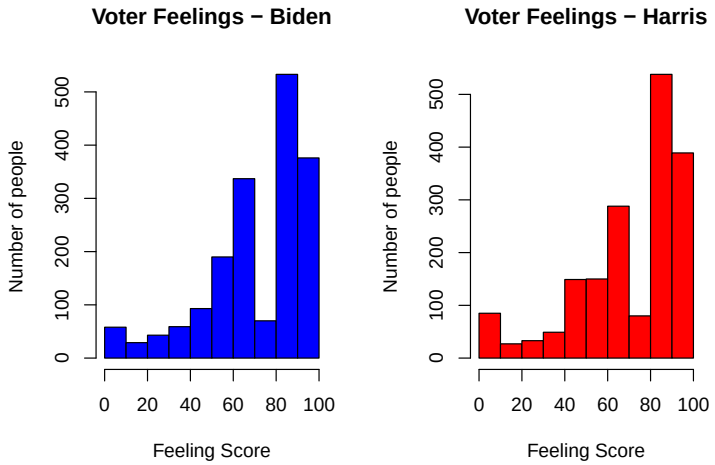

Figure 1: Histograms of Democrats' feelings towards Biden and Harris. Since our sample only includes Democrats, it makes sense that the distributions have a left skew.

Most appropriate test

As the same voter is expressing her/his feelings towards both candidates and the feelings are recorded as non-ordinal values, the correct test is a paired t-test.

We evaluate the assumptions for paired t-test:

1. The dependent variable (feelings rating) is metric since it is on a scale from 0 to 100.
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 in Figure 1, the data is fairly skewed and far from symmetric, so it is a concern here. However given the sample size (1,788), the CLT should apply and make the sampling distribution of the mean reasonably normal-like.

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 (μ_B) and Harris (μ_H) are the same. The *alternative hypothesis* is that they are not equal. Given we have no strong initial inclination in either direction, we run a two tailed test using the standard 5% significance level.

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

```
##
##  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 fail 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 Democratic voters' 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). Because the value 0 is included in this confidence interval, we cannot reject the null hypothesis at this significance level. We are 95% confident that this interval contains the true difference in levels of support.

On a practical level, we can see that the absolute the mean difference is fairly small (0.71), less than 1 point on the $0 - 100$ scale. Furthermore, as can be seen in Figure 2 below, when we plot the all absolute differences (with 20 bins, each representing a 5 points difference), we see that the vast majority of Democrat voters' feeling scores for Biden and Harris are close to each other, so the effect size is indeed small. Although curiously, there are a few people who gave Biden and Harris very different feeling scores, as much as 100!

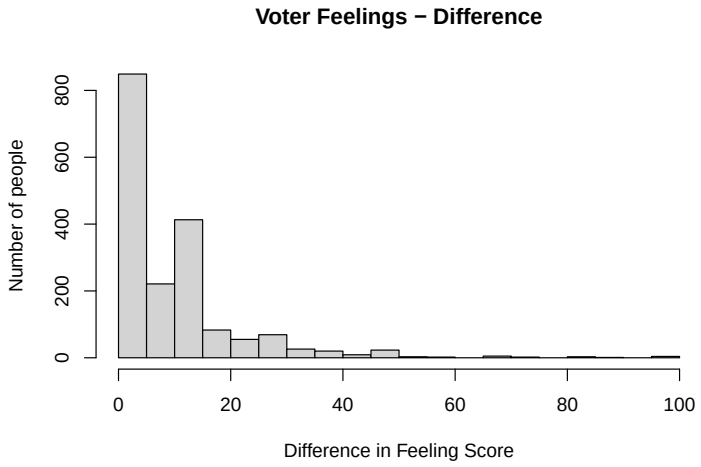

Figure 2: Histogram of absolute differences between individuals' scores for Biden and for Harris. Most individuals have a small difference in feelings between the two.

```
## [1] "40.6 % of Democratic voters gave Biden and Harris the same score."
```

```
## [1] "47.48 % of Democratic voters gave Biden and Harris scores within 5 points."
```

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

While the results from this pre-election survey are not statistically significant, having this knowledge can establish a baseline for how Democrats view Biden versus Harris. This could influence how the Democratic Party uses each individual to influence Party activities. In particular, knowing how the feelings for each individual changes over time could impact which candidate the Democratic Party runs in the 2024 election. We now have a baseline for understanding how Democratic voters feel towards the two individuals, and keeping track of these feelings will be important going into the future.