

# Lab 1: Question 2

Ian Dela Cruz, Steve Hall, Fengjiao Sun

## Importance and Context

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

In early election year of 2020, Mr. Biden committed to pick a woman as his running mate. As widespread protests against racial injustice had swept the country in the aftermath of the killing of George Floyd, the demands to pick a black woman were growing louder. NPR reported that a poll in mid-June revealed 46% of Democrats said it was important for Biden to choose a candidate of color as his running mate, which was up from 36% in early April. Later in August, Biden announced Harris as his running mate.

While it was generally believed that Black women are successful in mobilizing others to support Democratic candidates, are Democratic voters more enthusiastic about Harris? And would that elevate the favoritism of Biden? To answer these questions, we would first need to understand whether Democratic voters are more enthusiastic about Biden or Harris. The answer could provide some understanding on whether Harris was a winning choice of the running mate, and may also help to guide campaign strategies on promoting the candidates.

## Description of Data

We will address this question using data from ANES 2020 Time Series Study Preliminary Release: Pre-Election Data. The 2020 ANES survey used a contactless, mixed-mode design that was created in response to challenges related to the COVID-19 pandemic. The face-to-face mode was dropped for 2020. Instead, a sequential mixed-mode design was implemented that included self-administered online surveys, live video interviews conducted online, and telephone interviews.

The study object is Democratic voters, so we focused on V201018 PRE: PARTY OF REGISTRATION and select the samples whose answers are 1 (Democratic party). We intentionally did not include “voting status” V201020 as this may result in bias if the study only include samples who had voted by the time of study. To examine the enthusiasm about Joe Biden and Kamala Harris, we focused on V201151 PRE: FEELING THERMOMETER: JOE BIDEN, DEMOCRATIC PRESIDENTIAL CANDIDATE and V201153 PRE: FEELING THERMOMETER: KAMALA HARRIS, DEMOCRATIC VICE-PRESIDENTIAL CANDIDATE. The emotion intensity is reported on a Likert scale ranging from 0 to 100 with some predefined interval. There are also apparently “abnormal data” - 998, 999, -4, -9 value to denote other answers such as “Don’t know”, “Don’t recognize”, “Technical error”, “Refused”. We removed those values so that the leftover 0 to 100 are good ordinal scale. The remaining number of total valid responses yields a large enough number of samples (1788) for analysis.

The first thing is to explore the basic stats of the paired data sets, which includes median, mean and quartile. The median of the feeling thermometer scores are the same for Biden and Harris, both being 85. The mean score of Biden is 73.89, which is 0.71 higher than that of Harris, whose mean score is 73.18.

```
summary(question2$Biden)
```

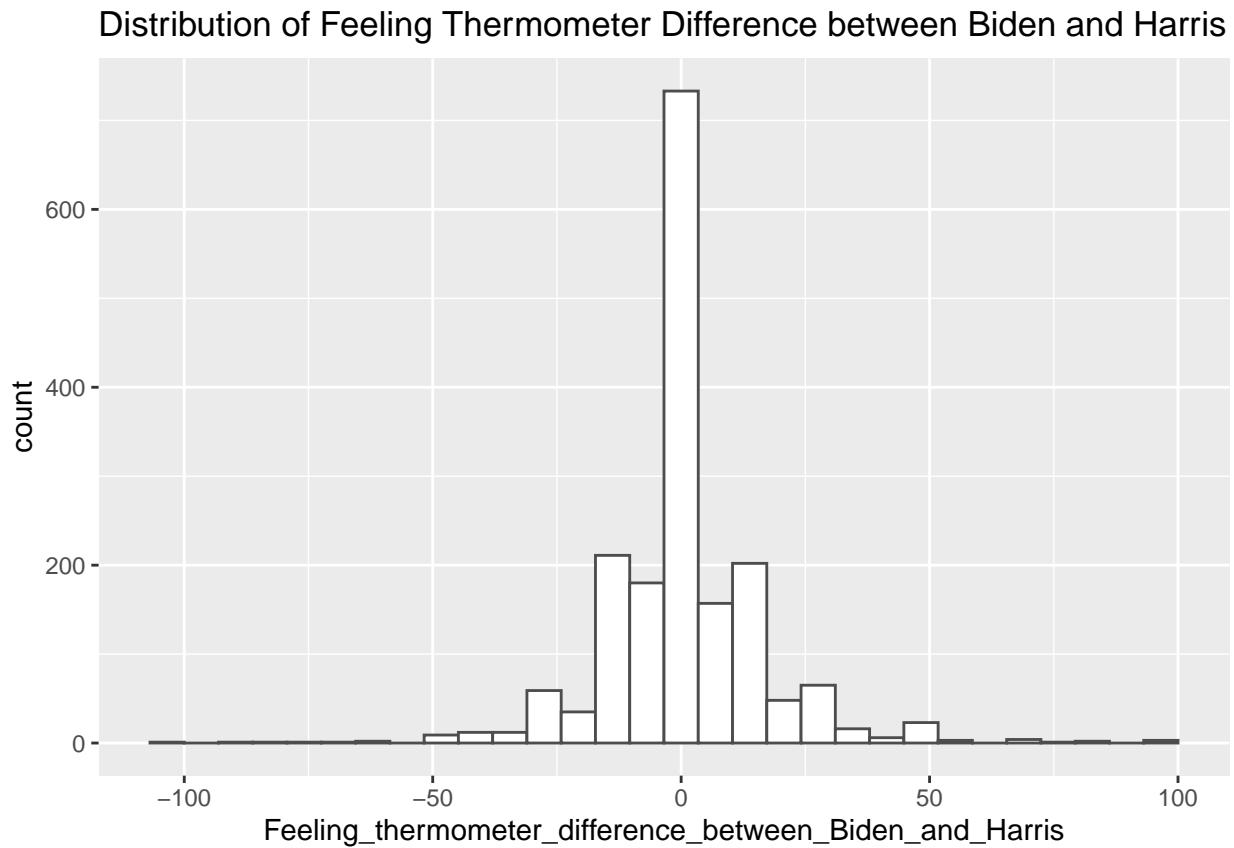
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.00	60.00	85.00	73.89	85.00	100.00

```
summary(question2$Harris)
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.00	60.00	85.00	73.18	87.75	100.00

Then we also looked into the distribution of the feeling thermometer difference between Biden and Harris, and found it is approximately symmetric.

```
ggplot(question2, aes(Feeling_thermometer_difference_between_Biden_and_Harris)) +
  geom_histogram(fill = "white", color = "grey30", bins = 30) +
  ggtitle("Distribution of Feeling Thermometer Difference between Biden and Harris")
```



### Most appropriate test

We test whether the enthusiasm thermometer is higher for Biden or for Harris. Because these variables are measured on interval scale, a non-parametric test is appropriate. In particular, the variables are both measured on the same metric scale from 0 to 100. Furthermore, the data is paired, since the same individual measures their feeling for Biden and Harris. Therefore, we decided to use Wilcoxon Signed-Rank test.

The Wilcoxon Signed-Rank test requires the following assumptions to be true:

- Independent and identically distributed data. The ANES 2020 Time Series Study pre-election sample consists of several groups, including 2016 ANES groups and freshly drawn cross-section. The fresh cross-sectional sample was a random draw from USPS computerized delivery sequence file (C-DSF) with all included residential address across 50 states and Washington DC having equal probability of selection. Those users were then randomly assigned to one of three sequential mode conditions, which further minimizes the possibility of introducing dependencies.
- Ordinal scale. The feeling thermometer showing an increase in Likert scale from 0 to 100. Ratings between 50 degrees and 100 degrees mean that the person feels favorable and warm toward the candidate. Ratings between 0 degrees and 50 degrees mean that the person does not feel favorable toward the

candidate and that he/she does not care too much for that candidate. These data are not metric because there is no standard measurement - A's scoring of 40 might be B's scoring of 60. In addition, the data are clustered on several data points instead of spreading out.

- There is natural pairing for Biden and Harris, because the person who provided answers to Biden and Harris is the same Democratic voter.
- Distribution of thermometer difference between Biden and Harris is symmetric around some mean. From the Description of Data section, we learned that the distribution of the difference is approximately symmetric around 0.

If the test were to **reject the null hypothesis**, we would conclude that among the Democratic voters who evaluated feeling thermometer values for Biden and Harris in 2020 have a measurable different enthusiasm about Joe Biden compared to Kamala Harris in 2020. If the test were to **fail to reject the null hypothesis**, we would conclude that the hypothesized effect does not exist or there is insufficient data to validate it.

## Test, results and interpretation

```
wilcox.test(question2$Biden,question2$Harris,paired = TRUE)
```

```
##  
## Wilcoxon signed rank test with continuity correction  
##  
## data: question2$Biden and question2$Harris  
## V = 296505, p-value = 0.1498  
## alternative hypothesis: true location shift is not equal to 0
```

From the test statistics, we fail to reject the null hypothesis that the Democratic voters are more enthusiastic about Biden or Harris. The p-value for the test is 0.1498, which is outside of the rejection range.

We also take a brief look at the effect size by calculating normal approximation in R. The effect size turns out to be 0.034, which is typically considered as a small effect. Given the difference of the mean is only 0.71 between Biden and Harris over a scale of 100, the small effect makes sense.

```
test <- wilcox.test(question2$Biden,question2$Harris,paired = TRUE)  
zstat <- qnorm(test$p.value/2)  
abs(zstat)/sqrt(1788)
```

```
## [1] 0.03405975
```

From a practical significance perspective, there is no evidence that Democratic voters are more enthusiastic about Biden or Harris. People seemed to be similarly excited about both candidates.