

Lab 1: Question 2

Ian Dela Cruz, Steve Hall, Fengjiao Sun

```
library(haven)
library(tidyverse)
library(dplyr)
raw_data <- read_sav("anes_timeseries_2020_spss_20210211.sav")
```

Importance and Context

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

Running mate has always been a critical decision during the election. To some extent, the decision would shift the force of the campaign away from the competitor's presidency onto Biden. It is also believed that the choice would offer Biden an opportunity to unify the party. At a minimum, the running mate should not drag the Democratic votes' excitement behind Biden's candidacy.

Having a popular running partner plays a key role in election. Understanding whether the candidate or the running mate is more favorable among its party voters would help to provide guidance for the running mate decision. It may also help to provide guidance to future political campaigns.

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 mean of the paired datasets.

```
question2 <- raw_data %>% select(starts_with("V201018"),
                                starts_with("V201151"),
                                starts_with("V201153"))
colnames(question2) <- c("Party", "partyz", "Biden", "Harris")
question2 <- filter(question2, Party == 1,
                    Biden %in% (0:100),
```

```
Harris %in% (0:100))
question2 <- within(question2, thermo_difference <- Biden - Harris)

summary(question2$Biden)
```

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

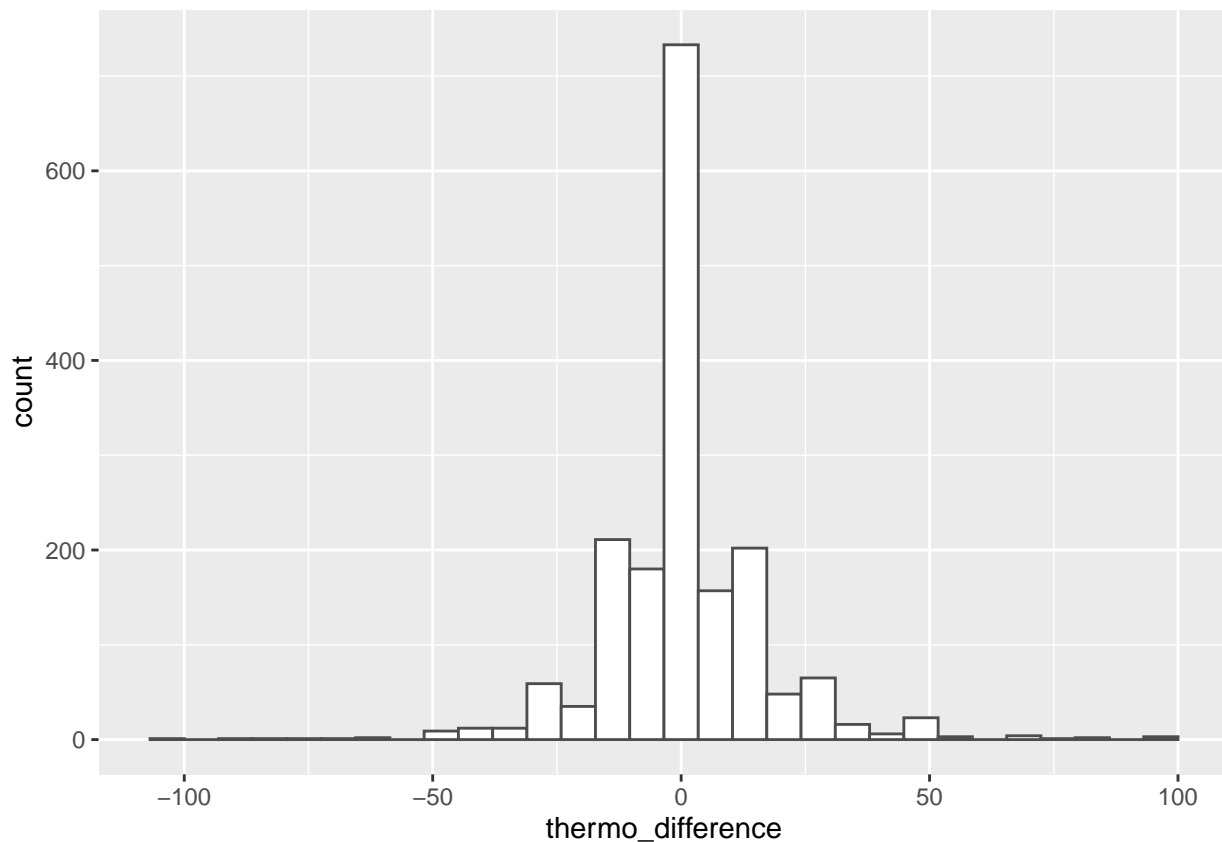
```
summary(question2$Harris)
```

```
##      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 difference and found it is approximately symmetric.

```
ggplot(question2, aes(thermo_difference)) +
  geom_histogram(fill = "white", color = "grey30")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with
## `binwidth`.
```



Most appropriate test

We test whether the enthusiasm thermometer is higher for Biden or for Harris. Because these variables are measured on ordinal 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 ranking from 0 to 100. 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”. Once we removed those values, the leftover 0 to 100 are good metric scale for analysis.
- 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.

Test, results and interpretation

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.

```
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. This means that the difference of feeling thermometer between Biden and Harris is trivial.

```
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 fairly excited about both candidates of the party.