Lab 1: Question 2

Jeffrey Adams, Brittany Dougall, Jerico Johns, Li Jin

```
install.packages("rstatix", repos = "http://cran.us.r-project.org")
install.packages("coin", repos = "http://cran.us.r-project.org")
library(coin)
library(rstatix)
library(dplyr)
library(ggplot2)
library(tidyverse)
```

Importance and Context

In presidential elections, the presidential and vice presidential candidates run as a joint ticket. A significant difference in public enthusiasm between members of a joint ticket may influence voting patterns. For example, a strong favorable opinion regarding one of the members of a ticket may motivate a voter to still vote for the ticket even if the voter feels somewhat negatively or ambivalent about the other ticket member. Say, if a voter feels ambivalent about Harris but strongly about Biden, the presence of Biden on the ticket may motivate the voter to still vote for this joint ticket. Conversely, if enthusiasm about one candidate is low, the presence of the other member may not be sufficient to motivate the voter to choose the ticket. For example, if a voter had little enthusiasm for Biden, the presence of Harris on the ticket may not be sufficient to inspire the voter to choose the Democratic ticket. In this scenario, voters who would otherwise vote Democrat may choose to vote for another candidate ticket or to abstain from voting at all. For this reason, it could be useful to determine if Democrats are more enthusiastic about Biden or Harris.

Description of Data

Data is collected from an ANES 2020 time series study, that uses cross-sectional random sampling on a USPS survey a single individual from each household in the sample regarding their political opinions and voting behavior in the U.S. presidential election. To operationalize our research question of whether or not Democratic voters are more enthusiastic about Biden or Harris, we chose the variables V201228, V201151, and V201153. V201018 contains which party voters identify most closely with, allowing us to select only registered Democratic voters. As seen from the output below, there are more respondents who entered a 1 (Democrat) for this field than there are registered Democrats based upon voting record (V201018).

```
##
##   -9   -8   -1    1    2    4    5
##    9    2 4010 1861 1336 1029   33
```

```
##
##   -9   -8   -4    0    1    2    3    5
##   44    4    1    7 2864 2564 2527  269
```

For this reason, we chose to use field V201228, which party voters identify as. Since we are only interested in Democrat's enthusiasm for the candidates, we filtered for those who entered a 1 (identify as Democrat). V201151 contains voter enthusiasm scores for Biden and V201153 contains voter enthusiasm scores for Harris. All 3 columns contain numeric scores. V201018 contains a numeric value that corresponds to voter party affiliation. V201151 and V201153 contain numeric scores ranging from $0 - 100$ to indicate voter enthusiasm

level, as well as the following values for both candidates: 998 (don't know how to rate), −4 (technical error), −9 (refused). V201153 contains the additional potential entry for Harris: 999 (don't know the candidate). We filtered the data frame on V201018 to select only voters who are registered Democrats and only included rows where candidates had a valid score. As seen in the below output, the majority of respondents entered a score for both Biden and Harris, when a valid score was entered. The first result represents the number of rows where a valid score was entered for Biden but not Harris, the second result the number of rows with a valid score for Harris but not Biden, and the third result the number of rows with a valid score for both candidates.

```
## [1] 49
```

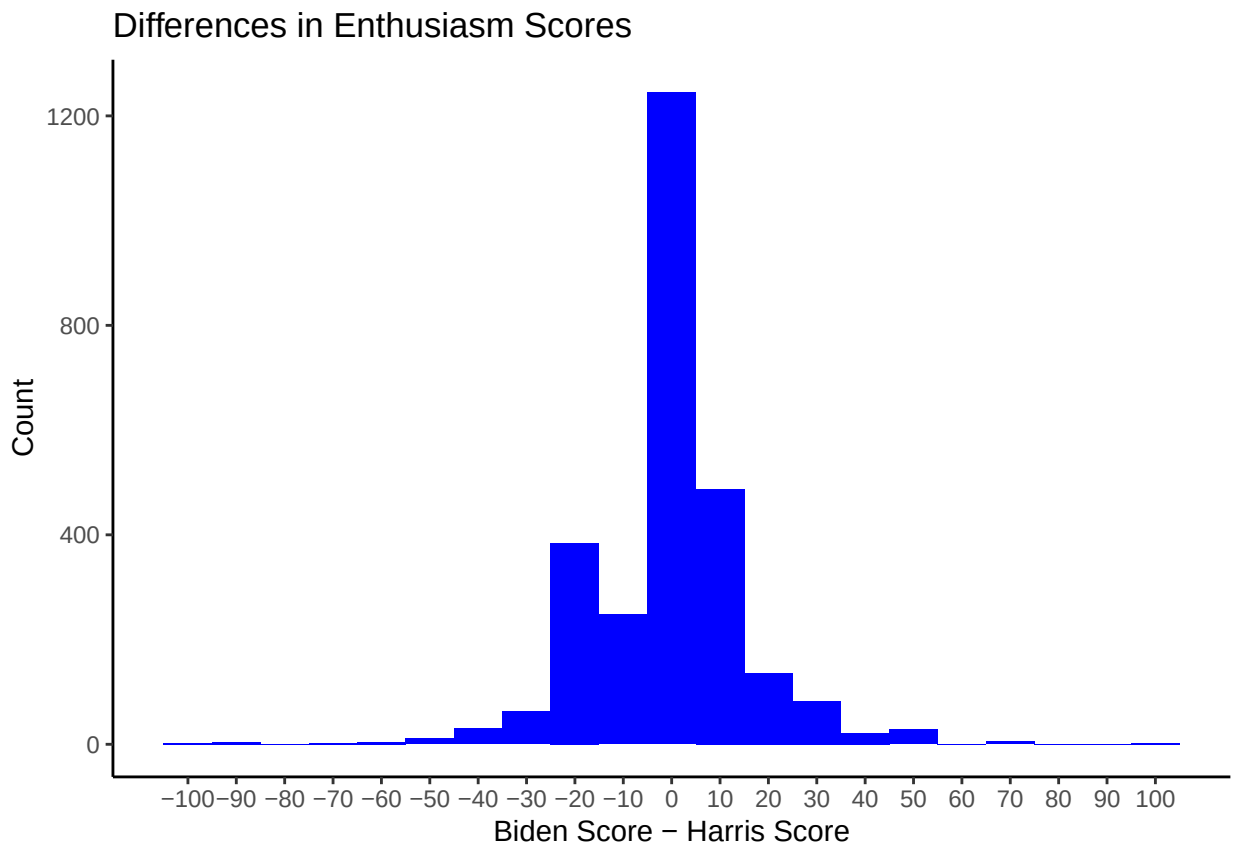
```
## [1] 26
```

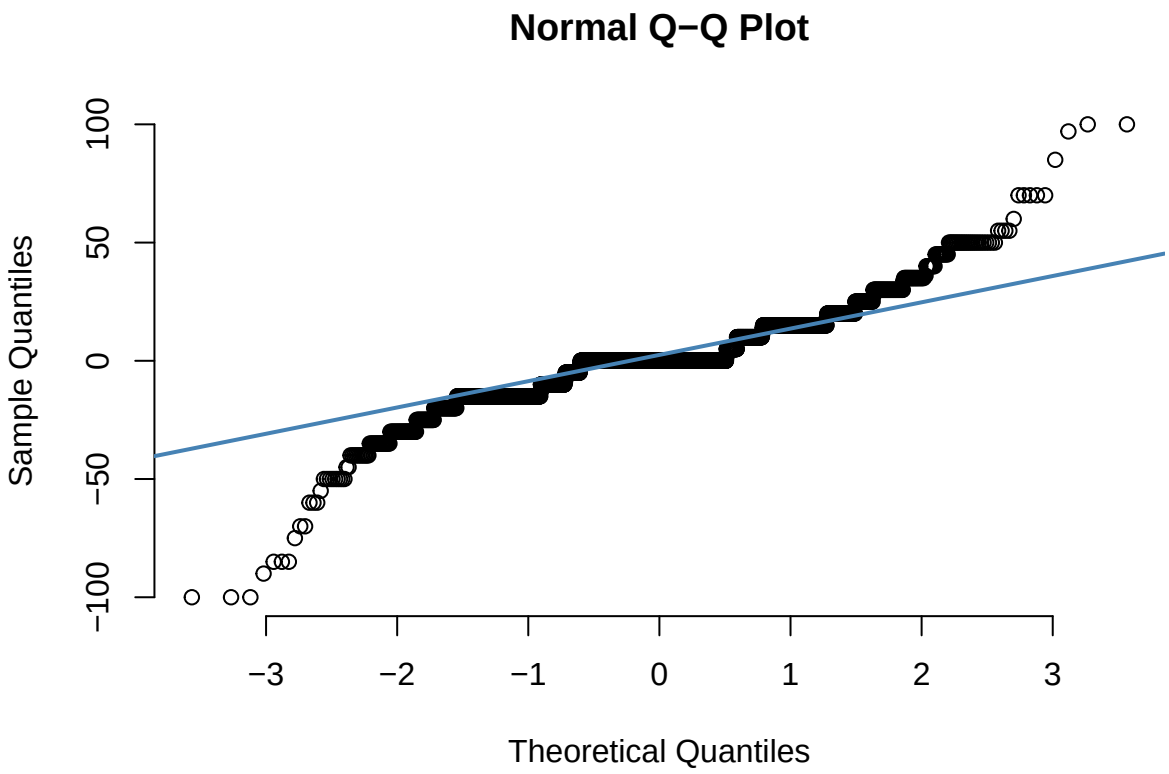
```
## [1] 2761
```

For this reason, we only included rows where a valid enthusiasm score between 0 and 100 was entered for both candidates in order to effectively compare differences in scores between these candidates among Democratic voters.

Most appropriate test

As seen above, the majority of Democrats entered a score for both candidates, indicating that we should choose a paired test. The data for our scale is ordinal numeric (0 - 100), where one cannot reasonably interpret meaning behind a single unit. However, since we are interested in testing if the mean enthusiasm score is higher for Biden or Harris among Democratic voters, and are taking the difference between scores across each respondent, we are imposing a metric scale. As seen in the below histogram of Biden enthusiasm scores minus Harris enthusiasm scores, the differences in enthusiasm scores do not appear to be normally distributed. The following qqplot of the correlation between Biden scores minus Harris scores against the norm also suggests that the data do not follow a normal distribution and that we should choose a non-parametric test. For these reasons, we chose to conduct a Wilcoxon signed-rank test to test our null hypothesis that the difference in enthusiasm scores between Biden and Harris for a given Democratic voter is 0.


Differences in Enthusiasm Scores
Count
1200
800
400
0
−100 −90 −80 −70 −60 −50 −40 −30 −20 −10 0 10 20 30 40 50 60 70 80 90 100
Biden Score − Harris Score



The assumptions behind this test are: 1) the data have a metric scale, 2) I.I.D. data - each pair of measurements is drawn from the same distribution, independently of all other pairs, and 3) the distribution of the difference between enthusiasm scores is symmetric around some mean. Per the documentation for the columns for Biden and Harris' enthusiasm scores, valid scores for both columns take on values between 0 and 100 and we impose a metric scale by taking the difference in these values - in our set up, the enthusiasm score for Biden minus the enthusiasm score for Harris (which meets the metric scale requirement). Scores outside of this range will be excluded from our analysis. Per the documentation, only 1 member of each household was surveyed and respondents were randomly selected from the USPS delivery sequence file. It is possible that individuals who chose to participate encouraged others outside their household to participate; however, given that all addresses within the 50 states and D.C. were given equal likelihood of being selected, this possibility is rare (meeting the I.I.D. requirement). From the above histogram distribution of Biden enthusiasm score minus Harris enthusiasm score, the distribution of the difference in scores is approximately symmetric around the median, meeting the 3rd requirement.

Test, results and interpretation

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  democratic_rating_df$biden_rating and democratic_rating_df$harris_rating
## V = 710762, p-value = 0.0001215
## alternative hypothesis: true location shift is not equal to 0
```

```
## Warning: attributes are not identical across measure variables;
## they will be dropped
```

```
## # A tibble: 1 x 7
##   .y.              group1       group2        effsize    n1    n2 magnitude
```

```
## * <chr>            <chr>        <chr>           <dbl> <int> <int> <ord>
## 1 enthusiasm_score biden_rating harris_rating  0.0575  2761  2761 small
```

At approximately 0.0001215, the p-value is less than our significance level of 0.05, indicating that we should reject the null hypothesis (the null hypothesis being that there is no difference in mean enthusiasm for Biden and Harris among Democratic voters). However, the effect size is small, suggesting little practical significance.