STA130H1S – Fall 2022

Hypothesis Testing ABCD Guessing

() and STA130 Professors

This exercise worksheet is accompanied by an Quercus quiz here. It is recommended that you work through this document and the quiz at the same time.

There are multiple studies exploring the relationship between gratitude journaling and life satisfaction, mental health, adjusting to university and more. An example of gratitude journaling is writing down three things that you are thankful for at the end of the day. We might all be able to benefit from this science-backed way to live more satisfying lives. To learn more, check out this 10-minute video An antidote to dissatisfaction from the Kurzgesagt youtube channel. To hear a little bit more about mental health more generally, check out this 20-minte TED talk This could be why you're depressed or anxious by Johann Hari.

The data this week is synthetic, but is loosely based on a study of students in their first year of a five-year college in Turkey. Specifically, 61 Turkish students in their first year of college were randomly assigned to either practice gratitude journaling for three weeks (treatment group, n=30) or to write down a factual account of their day for three weeks (control group, n=31). At the end of the three week period, their life satisfaction (on a scale from 0 to 100) was measured as was their adjustment to university life (on a scale from 48 to 336, where scores over 144 are considered to indicate successful adjustment while scores lower than this indicate difficulties with adjusting). To read more about the study, you can read the original article.

Question 1

Adjust the code below to create a new variable called `adjusted` that has the value `TRUE` if the student has a score greater than 144 and FALSE if the student has score of 144 or less. Then create an object called `group_props` using `summarise()` to find out what proportion of students have "successfully adjusted to university life" in each group.

```
library(tidyverse)
gratitude <- read_csv("gratitude.csv")
gratitude_new <- gratitude %>%
  mutate(adjusted = NA)  # this line needs to changed

group_props <- gratitude_new %>%
  group_by(treatment) %>%
  summarise(prop_adj = mean(adjusted))

group_props
```

```
## # A tibble: 2 x 2
##   treatment prop_adj
##   <chr>        <dbl>
## 1 control         NA
## 2 treatment       NA
```

What percentage of students who were part of the gratitude journaling treatment group successfully adjusted to university life?

A. 13%

B. 30%

C. 47%

D. 100%

Question 2

Suppose one of your peers is very skeptical about "this whole gratitude nonsense". They think that there will be no real difference between the treatment and control groups with regard to the proportion who successfully adjust to university life.

Which of the following statements is TRUE?

A. Your peer's opinion is like the null hypothesis for a two sample hypothesis test.

B. Your peer's opinion is like the alternative hypothesis for a two sample hypothesis test.

C. Your peer's opinion is like the null hypothesis for a hypothesis test on one proportion.

D. Your peer's opinion is like the alternative hypothesis for a hypothesis test on one proportion.

Question 3

Suppose you want to test whether there is a difference between the proportions of students in the treatment and control groups that successfully adjust to university (i.e. have a score of more than 144 on the adjust to university life scale).

Which of the following would be the correct null hypothesis for this test?

A. $H_0 : \bar{x}_{treatment} - \bar{x}_{control} = 144$

B. $H_0 : p_{treatment} - p_{control} = 0$

C. $H_0 : p_{adjust.to.univeristy} = 0$

D. $H_0 : \mu_{treatment} - \mu_{control} = 0$

Question 4

Which of the following would be the BEST alternative hypothesis for this test?

A. The average adjustment to university score for first-year Turkish college students who practice gratitude journaling for three weeks is not the same as the average adjustment to university score for first-year Turkish college students who did not practice gratitude journaling for three weeks.

B. In our sample, the proportion of students who successfully adjust to university life in the treatment group is not the same as the proportion in the control group.

C. It is not the case that the proportion of first-year Turkish college students who successfully adjust to university life among those who practice gratitude journaling for three weeks is the same as the proportion of first-year Turkish college students who successfully adjust to university life among those who did not practice gratitude journaling for three weeks.

D. The proportion of first-year Turkish college students who practice gratitude journaling for three weeks is the same as the proportion of first-year Turkish college students who did not practice gratitude journaling for three weeks.

Question 5

Building off of the code in Question 1, update the code below to calculate the test statistic for this investigation.

```
group_props
```

```
## # A tibble: 2 x 2
##   treatment prop_adj
##   <chr>        <dbl>
## 1 control         NA
## 2 treatment       NA
```

```
test_stat <- group_props %>% #prop_adj
  summarise(test_stat = diff(NA)) # this is the line you need to update, replace the NA
test_stat
```

```
## # A tibble: 0 x 1
## # ... with 1 variable: test_stat <lgl>
```

```
test_stat_num <- as.numeric(test_stat$test_stat)
test_stat_num
```

```
## numeric(0)
```

What is the test statistic for this investigation?

A. 13%

B. 0.338

C. 0.467

D. 0

Question 6

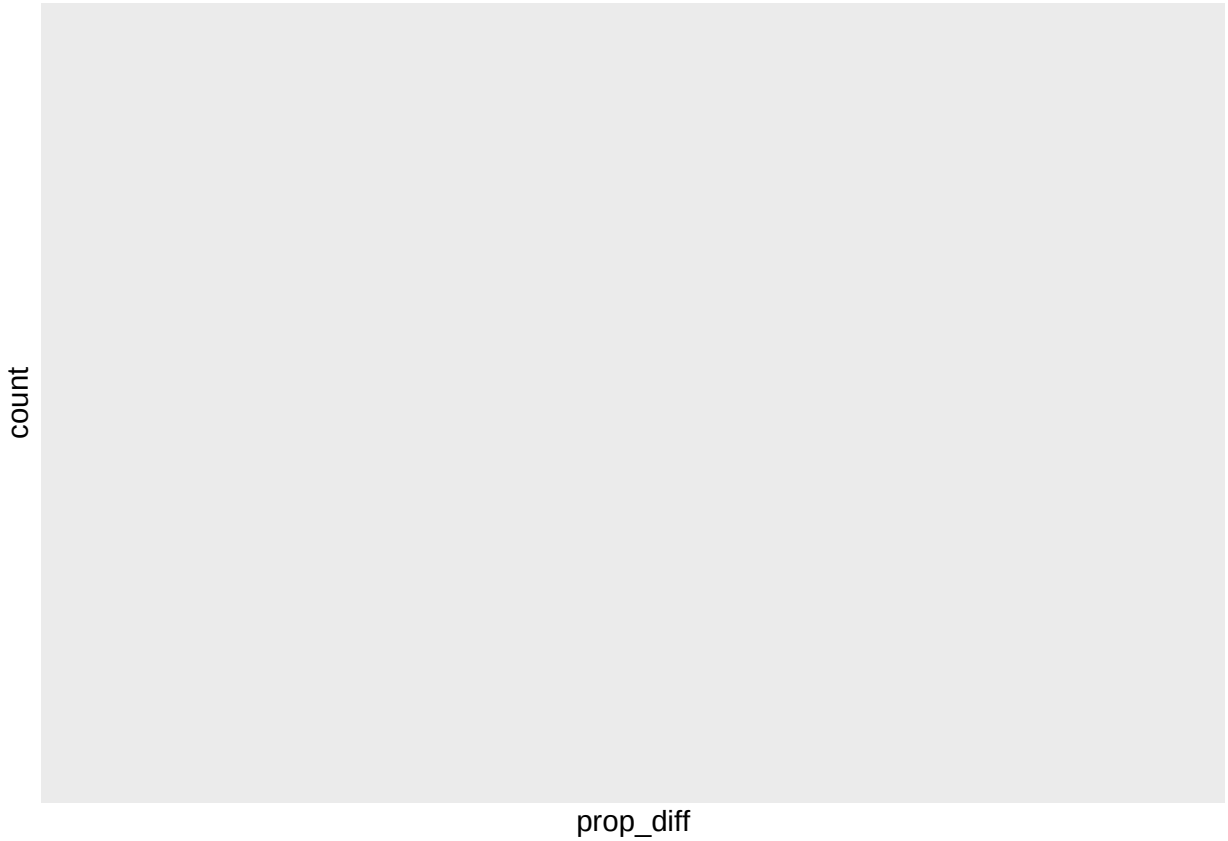
```
set.seed(42)
repetitions <- 1000
simulated_values <- rep(NA, repetitions)
test_stat <- test_stat_num # this line uses the work from Q5

for(i in 1:repetitions){
  simdata <- gratitude_new %>%
    mutate(treatment = sample(treatment)) %>%
    group_by(treatment) %>%
    summarise(prop = mean(adjusted)) %>%
    summarise(value = diff(prop))

  simulated_values[i] <- as.numeric(simdata)
}

sim <- tibble(prop_diff = simulated_values)
```

```r
sim %>% ggplot(aes(x=prop_diff)) +
  geom_histogram(binwidth = 0.06, color="black", fill="gray") +
  geom_vline(xintercept = abs(test_stat), colour = "red") +
  geom_vline(xintercept = -abs(test_stat), colour = "red")
```



Based on the plot alone, which of the following would you estimate the p-value to be closest to?

A. 0 (exactly)

B. 0.005

C. 0.10

D. 0.30

Question 7

Based on the p-value from Question 6, what kind of error might we be making, assuming we are using an α of 0.01?

A. Type I

B. Type II

C. Selection bias

D. Coercion error

Question 8

Suppose the question "What is your favourite number between 0 and 100?" had been asked of each student when they signed up for experiment (and before they were randomly assigned into the gratitude journal group or not). Consider the *median* favourite numbers between the treatment and control groups.

Which ONE of the following statements about this new investigation is FALSE?

A. Any differences between the median favourite numbers for the two groups is just due to chance.

B. As this variable was measured for each student before the experiment was conducted, any difference between the group medians is just due to the random allocation of students to treatment groups and not due to the treatments themselves.

C. Differences between the median favourite numbers for the treatment and control groups should be due to one of two reasons: chance acting alone or chance and something else acting to produce a difference.

D. We would expect to get a large p-value at the end of this investigation.

Question 9

Choose the correct hypotheses to investigate if there is a difference in median favourite numbers between the treatment and control groups.

A. $H_0 : median_{treatment} = median_{control}$ and $H_1 : median_{treatment} \neq median_{control}$

B. $H_0 : \tilde{x}_{treatment} = \tilde{x}_{control}$ and $H_1 : \tilde{x}_{treatment} \neq \tilde{x}_{control}$

C. $H_0 : sample.median_{treatment} = sample.median_{control}$ and $H_1 : sample.median_{treatment} \neq sample.median_{control}$

D. $H_0 : median_{treatment} - median_{control} = 0$ and $H_1 : median_{treatment} - median_{control} > 0$

Question 10

Study participant students favorite numbers before being assigned to do gratitude journaling or not are recorded next to their other measures in the dataset `gratitude_v2.csv`. Replace the `sum()` function below with the correct function to calculate the test statistic for this investigation into whether there is a difference in median favourite numbers between the treatment and control groups.

```
gratitude_v2 <- read_csv("gratitude_v2.csv")

# Calculate your test statistic
test_stat_num <- gratitude_v2 %>%
  group_by(treatment) %>%
  summarise(medians = sum(fav_num)) %>% # this line needs to be updated
  summarise(diff_med = diff(medians)) %>%
  as.numeric()

test_stat_num
```

```
## [1] -97
```

What is the test statistic for this investigation?

A. -97

B. -11

C. -1.46

D. 0.00

Question 11

The code below relies on your completion of the previous question and you don't need to make any changes to the code.

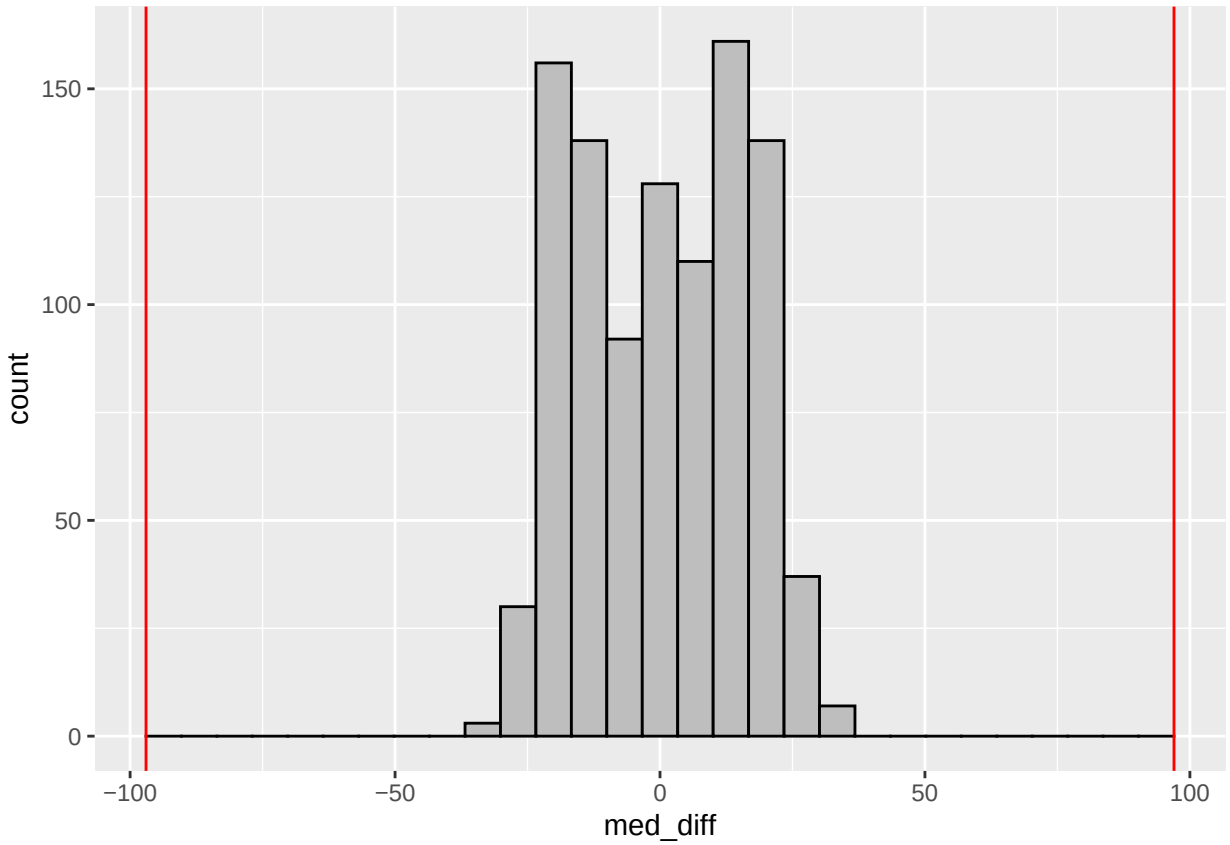
```
set.seed(123456789)
repetitions <- 1000
simulated_values <- rep(NA, repetitions)
test_stat <- test_stat_num # this line uses the work from the previous question

for(i in 1:repetitions){
  simdata <- gratitude_v2 %>%
    mutate(treatment = sample(treatment)) %>%
    group_by(treatment) %>%
    summarise(medians = median(fav_num)) %>%
    summarise(value = diff(medians))

  simulated_values[i] <- as.numeric(simdata)
}

sim <- tibble(med_diff = simulated_values)

sim %>% ggplot(aes(x=med_diff)) +
  geom_histogram(color="black", fill="gray") +
  geom_vline(xintercept = abs(test_stat), colour = "red") +
  geom_vline(xintercept = -abs(test_stat), colour = "red")
```




```
# Calculate p-value
num_more_extreme <- sim %>%
  filter(abs(med_diff) >= abs(test_stat_num)) %>%
  summarise(n())

p_value <- as.numeric(num_more_extreme / repetitions)
p_value
```

```
## [1] 0
```

Which of the following is the BEST interpretation of the p-value calculated above?

A. There is a 67% chance that there is no difference between the median favourite numbers of the two groups.

B. We have strength of evidence against the null hypothesis of $\alpha = 0.665$.

C. We have no evidence against the hypothesis that there is no difference in the median favourite numbers between the treatment and control groups.

D. At the 50% significance level we fail to reject the alternative hypothesis.