Assignment 6: Who busts the Mythbusters?

Sangwon Yum

2023-10-16

Exercise 1

- Q) Which variable is the response variable and which is the explanatory variable?
- A) (yawn), (group)
- Q) What value in the response variable is classified as a success?
- A) (yawn) success (yes)

```
summary(experimental_data)
```

```
##    subject_id        yawn              group
##  Min.   : 1.00   Length:50          Length:50
##  1st Qu.:13.25   Class :character   Class :character
##  Median :25.50   Mode  :character   Mode  :character
##  Mean   :25.50
##  3rd Qu.:37.75
##  Max.   :50.00
```

Exercise 2

- Q) Choose one of the four answer options for should be used to build the null distribution
- A) 4

Exercise 3

```
specify(response = yawn, explanatory = group, success = "yes")
```

Exercise 4

```
hypothesize(null = "independence")
```

Exercise 5

```
generate(reps = 10000, type = "permute")
```

Exercise 6

```
calculate(stat = "diff in props", order = c("Treatment", "Control"))
```

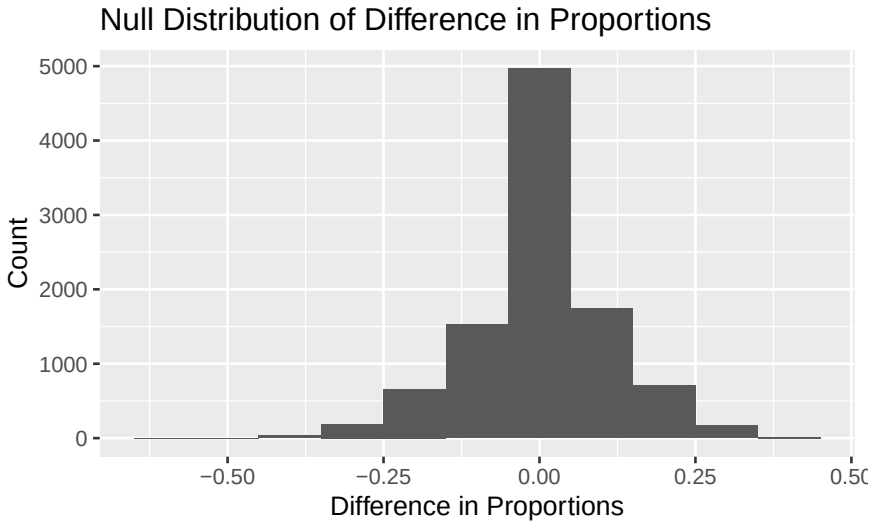
Exercise 7

```
set.seed(50)

yawn_null = experimental_data %>%
  specify(response = yawn, explanatory = group, success = "yes") %>%
  hypothesize(null = "independence") %>%
  generate(reps = 10000, type = "permute") %>%
  calculate(stat = "diff in props", order = c("Treatment", "Control"))

print(yawn_null)
```

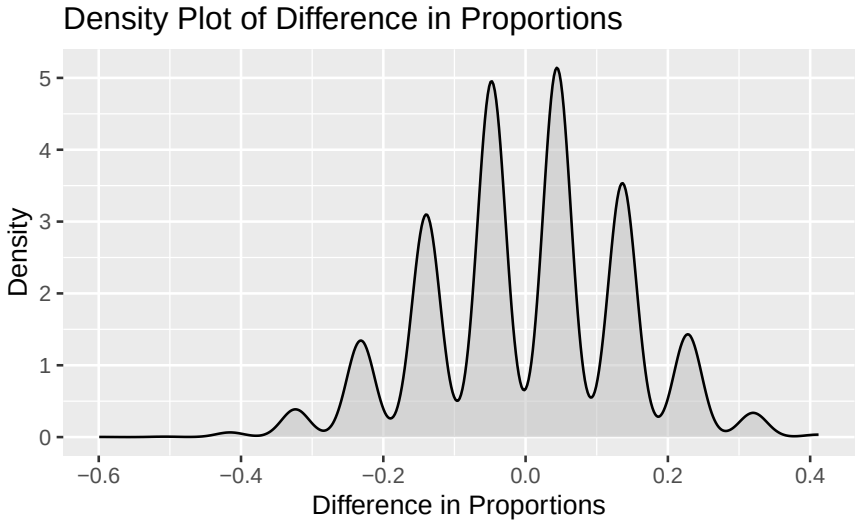
```
## Response: yawn (factor)
## Explanatory: group (factor)
## Null Hypothesis: independence
## # A tibble: 10,000 x 2
##    replicate    stat
##        <int>   <dbl>
##  1         1 -0.0478
##  2         2 -0.232
##  3         3  0.136
##  4         4  0.136
##  5         5 -0.140
##  6         6  0.136
##  7         7 -0.140
##  8         8 -0.0478
##  9         9  0.136
## 10        10  0.136
## # i 9,990 more rows
```

Exercise 8

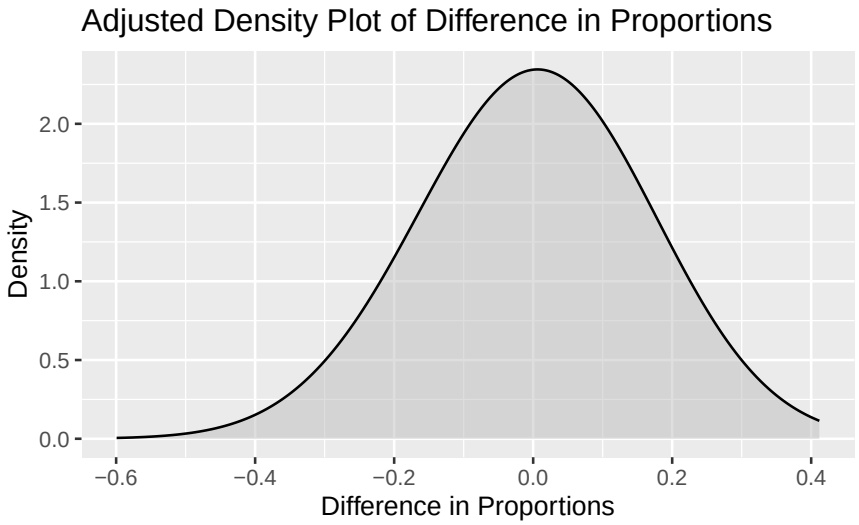
```
# The histogram
ggplot(yawn_null, aes(x = stat)) +
  geom_histogram(binwidth = 0.1) +
  labs(title = "Null Distribution of Difference in Proportions",
       x = "Difference in Proportions",
       y = "Count")
```




```
# The first density plot
ggplot(yawn_null, aes(x = stat)) +
  geom_density(fill = "grey", alpha = 0.5) +
  labs(title = "Density Plot of Difference in Proportions",
       x = "Difference in Proportions",
       y = "Density")
```

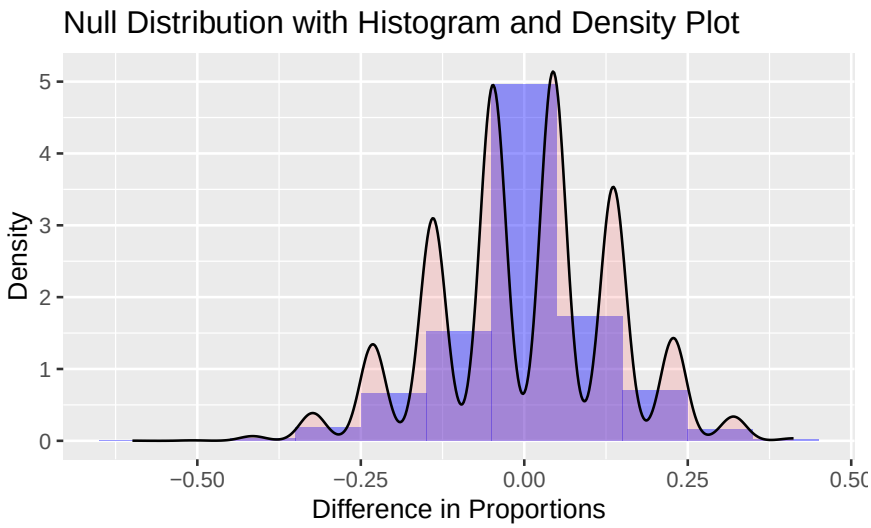


```r
# The second density plot
ggplot(yawn_null, aes(x = stat)) +
  geom_density(adjust = 5, fill = "grey", alpha = 0.5) +
  labs(title = "Adjusted Density Plot of Difference in Proportions",
       x = "Difference in Proportions",
       y = "Density")
```



Exercise 9

```r
ggplot(yawn_null, aes(x = stat)) +
  geom_histogram(aes(y = ..density..), binwidth = 0.1, fill = "blue",  alpha = 0.4) +
  geom_density(alpha = 0.2, fill = "#FF6666") +
  labs(title = "Null Distribution with Histogram and Density Plot",
       x = "Difference in Proportions",
       y = "Density")
```

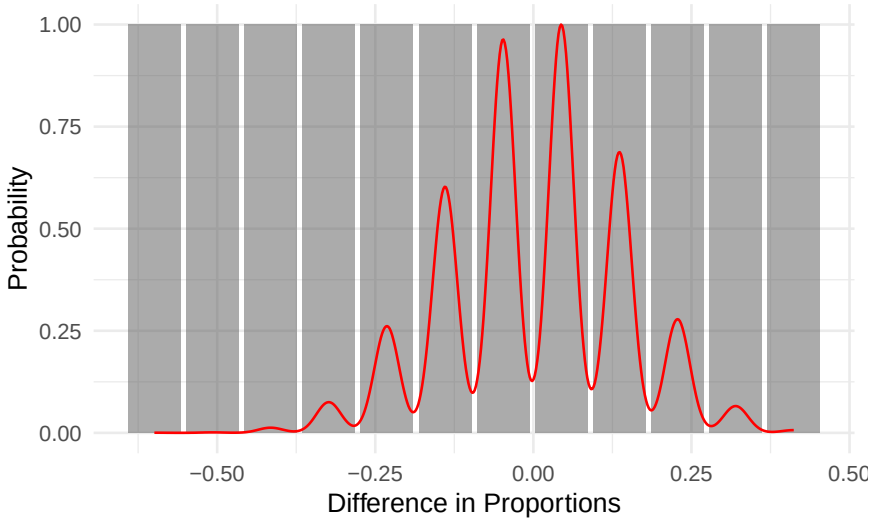


Exercise 10

```
yawn_null_pmf = yawn_null %>%
  group_by(stat) %>%
  summarise(prob = n() / sum(n()))
print(yawn_null_pmf)
```

```
## # A tibble: 12 x 2
##       stat  prob
##      <dbl> <dbl>
##  1 -0.599      1
##  2 -0.507      1
##  3 -0.415      1
##  4 -0.324      1
##  5 -0.232      1
##  6 -0.140      1
##  7 -0.0478     1
##  8  0.0441     1
##  9  0.136      1
## 10  0.228      1
## 11  0.320      1
## 12  0.412      1
```

```
ggplot() +
  geom_col(data = yawn_null_pmf, aes(x = stat, y = prob), binwidth = 0.1, alpha = 0.5) +
  geom_density(data = yawn_null, aes(x = stat, y = ..scaled..), color = "red") +
  labs(x = "Difference in Proportions", y = "Probability") +
  theme_minimal()
```



Exercise 11

- Q) Is there sufficient evidence to reject the null hypothesis?
- A) p-value 0.5094 0.05 ( ). null treatment

- Q) Explain why using examples from the previous exercises
- A) 8~10 . 11 (p-value) .

```
yawn_obs_stat = experimental_data %>%
  specify(response = yawn, explanatory = group, success = "yes") %>%
  calculate(stat = "diff in props", order = c("Treatment", "Control"))
print(yawn_obs_stat)
```

```
## Response: yawn (factor)
## Explanatory: group (factor)
## # A tibble: 1 x 1
##     stat
##    <dbl>
## 1 0.0441
```

```
p_value = yawn_null %>%
  get_p_value(obs_stat = yawn_obs_stat, direction = "greater")
print(p_value)
```

```
## # A tibble: 1 x 1
##   p_value
##     <dbl>
## 1   0.515
```

```r
yawn_null %>%
  visualize() +
  shade_p_value(obs_stat = yawn_obs_stat, direction = "greater") +
  labs(title = "Distribution of Null Hypothesis with Observed Statistic",
       x = "Difference in Proportions",
       y = "Density")
```