# PS 9: A Matter of Taste II

## Data Analysis

```
library(tidyverse)
```

```
-- Attaching packages --------------------------------------- tidyverse 1.3.2 --
v ggplot2 3.3.6      v purrr   0.3.4
v tibble  3.1.8      v dplyr   1.0.10
v tidyr   1.2.1      v stringr 1.4.1
v readr   2.1.2      v forcats 0.5.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
```

```
library(infer)
```

Be sure to tag your group mates when you submit this to Gradescope!

1. List any changes that you made to your experimental protocol.

   Our taster tried the two flavors first from the cans as a control before doing the blind taste test. The materials handler also shuffled the cups on the table before giving them to the taster. We only did 10 trials as we had 10 cups to use.

2. Create a data frame based on the data you collected and print it out into your pdf. You can print all rows your data frame using `print(my_df, n = nrow(my_df))`. Consult the notes "A Tool for Computing with Data" for a refresher of how to make a dataframe.

```
trial_num <- c(1,2,3,4,5,6,7,8,9,10)
actual_flavor <- c("Passionfruit", "Passionfruit", "Passionfruit", "Lime",
                   "Passionfruit", "Lime", "Lime", "Passionfruit",
                   "Lime", "Lime")
```

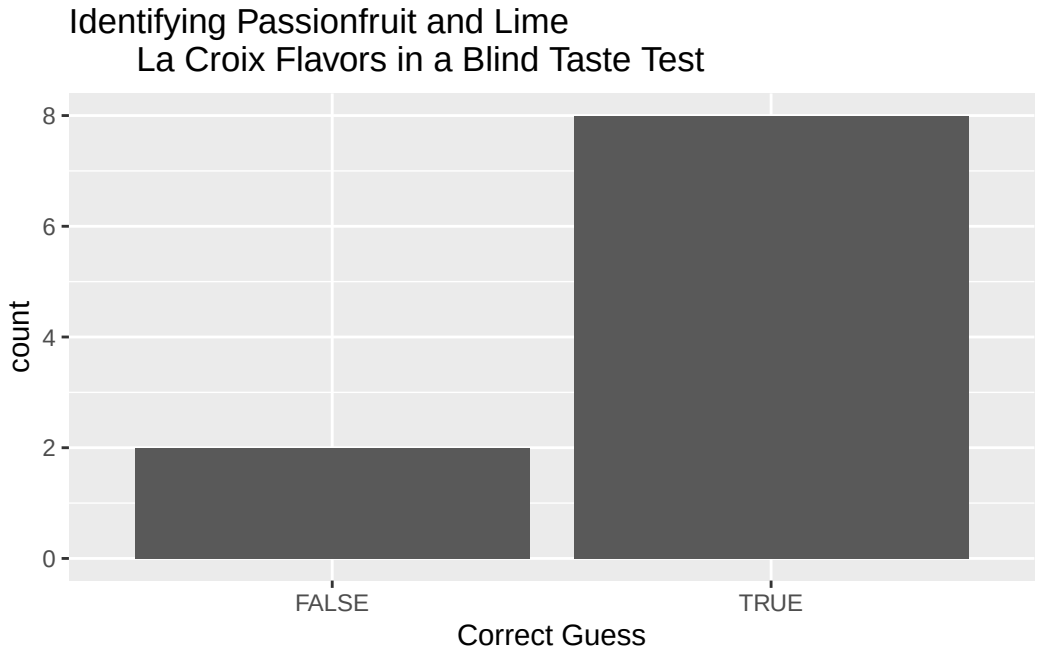
```r
guessed_flavor <- c("Passionfruit", "Passionfruit", "Passionfruit", "Lime",
                    "Lime", "Lime","Lime", "Passionfruit", "Lime",
                    "Passionfruit" )
correct <-c(TRUE, TRUE, TRUE, TRUE, FALSE,
            TRUE, TRUE, TRUE, TRUE, FALSE)

taste_test_df <- data.frame(trial_num,
                            actual_flavor,
                            guessed_flavor,
                            correct)
#print(taste_test_df, n = nrow(taste_test_df))
taste_test_df
```

```
   trial_num actual_flavor guessed_flavor correct
1          1  Passionfruit   Passionfruit    TRUE
2          2  Passionfruit   Passionfruit    TRUE
3          3  Passionfruit   Passionfruit    TRUE
4          4          Lime           Lime    TRUE
5          5  Passionfruit           Lime   FALSE
6          6          Lime           Lime    TRUE
7          7          Lime           Lime    TRUE
8          8  Passionfruit   Passionfruit    TRUE
9          9          Lime           Lime    TRUE
10        10          Lime   Passionfruit   FALSE
```

3. Create a visualization of the data you collected (not the null distribution) similar to the one you sketched in the handout. Does it look clearly in support of your claim or contrary to your claim or somewhere in between?

The data does clearly support our claim as 80% of the trials resulted in a correct identification of the La Croix flavor. The difference between the proportion correct (0.8) and the proportion wrong (0.2) is not 0, so therefore the alternative hypothesis appears to be true.

```r
taste_test_df %>%
  ggplot(aes(x=correct)) +
  geom_bar() +
  labs(title="Identifying Passionfruit and Lime
       La Croix Flavors in a Blind Taste Test",
       x="Correct Guess")
```




4. Conduct a hypothesis test to determine whether your data is consistent with the null hypothesis. Be sure to provide.

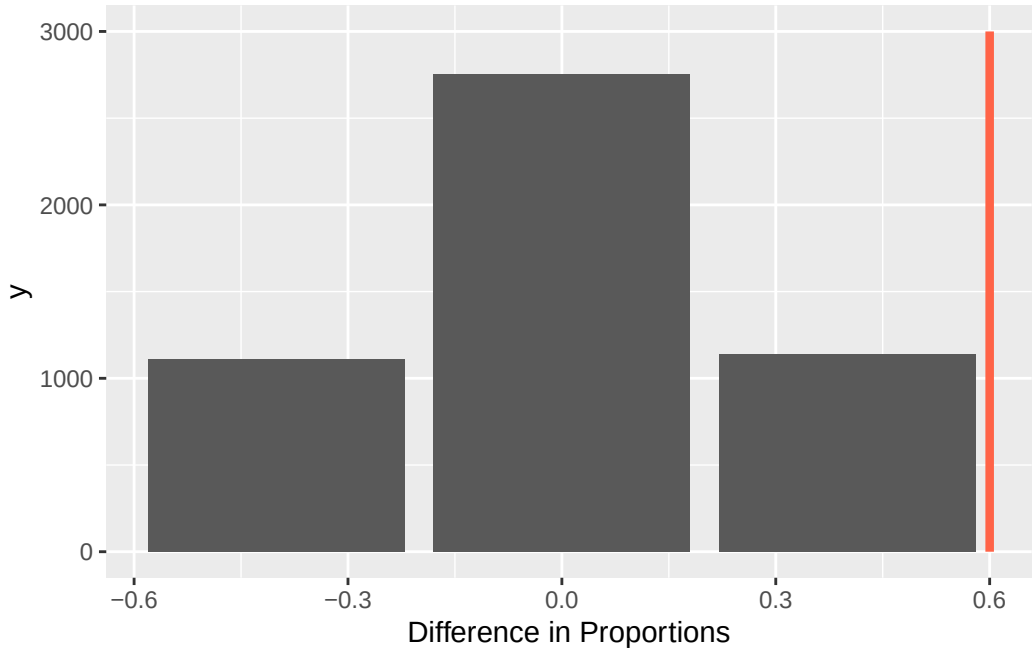
   - The null and alternative hypotheses.
     - Null: The difference between the proportion identified correctly and proportion identified incorrectly is 0.
     - Alternative: The difference between the proportion identified correctly and proportion identified incorrectly is not 0.
   - The value of the observed test statistic.
     - Proportion identified correctly - Proportion identified incorrectly: 0.8-0.2= 0.6
   - A visualization of the null distribution and observed test statistic with the p-value shaded in.

```
null <- taste_test_df %>%
  specify(response = correct,
          explanatory = actual_flavor,
          success = "TRUE") %>%
  hypothesize(null = "independence") %>%
  generate(reps = 5000, type = "permute") %>%
  calculate(stat = "diff in props")
```

```
Warning: The statistic is based on a difference or ratio; by default, for
```

```
difference-based statistics, the explanatory variable is subtracted in the
order "Lime" - "Passionfruit", or divided in the order "Lime" / "Passionfruit"
for ratio-based statistics. To specify this order yourself, supply `order =
c("Lime", "Passionfruit")` to the calculate() function.
```

```r
null %>%
  ggplot(aes(x = stat)) +
  geom_bar() +
  labs(x = "Difference in Proportions")+
  annotate("segment", x = .6, xend = .6,
           y = 0, yend = 3000,
           color = "tomato", lwd = 1.5)
```




d. The p-value and your conclusion (use $\alpha = .05$) regarding null hypothesis and original claim.

The p-value is 0 and since p<0.05, then we can reject the null hypothesis and our claim of "Our taster can correctly identify two different flavors of La Croix" is accepted.