Lab 8

Taylor Blair

Math 141, Week 9

Due: Before your Week 10 lab meeting

Goals of this lab

1. Practice hypothesis testing.
2. Practice simulating a null distribution.
3. Practice computing and interpreting p-values.

Problems

```
# Insert libraries here
library(tidyverse)
library(infer)
```

Problems

Problem 1

In 2005, the researchers, Antonioli and Reveley, poised the question "Does swimming with the dolphins help depression?" To study this question, they recruited 30 subjects with clinical depression whose ages ranged from 18 to 65 years old. Each subject discontinued any other treatment four weeks prior to the experiment and were randomly assigned to either swim with dolphins (the treatment group) or to do yoga (the control group). After two weeks, each subject was categorized as "showed substantial improvement" or "did not show substantial improvement".

Taylor Here: HOW THE HECK DID THEY GET FUNDING FOR THIS AND LEGITMETLY WRITE A PAPER FOR IT https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1289317/

```
dolphins <- read_csv("/home/courses/math141f18/Data/Dolphins.csv")
```

Here's a contingency table of `improve` and `group`.

```
dolphins %>%
  count(group, improve)
```

```
## # A tibble: 4 x 3
##   group     improve     n
##   <chr>     <chr>   <int>
## 1 Control   no         12
## 2 Control   yes         3
```

```
## 3 Treatment no          5
## 4 Treatment yes        10
```

a. State the null and alternative hypotheses in terms of conjectures and in terms of parameters.

- H_0 *Null Hypothesis*
  - The Control and Treatment group have the same proportion of improvments
  - Or the $propcontrol - proptreatment = 0$
- H_a *Alternative hypothesis*
  - The proportion of Treatment group that improves will be greater than the proportion of control group

---

b. Describe what a Type I error represents in the context of the problem. Describe what a Type II error represents.

- Type 1 error
  - False positive
  - We accept a false null hypothesis. Which would mean saying that they have the same proportion of improvments, but in actuality they do not.
  - The study was bogus but called facts
- Type 2
  - False negative
  - The study was concluded with dolphings having no greater impact on individuals than the treatment group. In reality there was an impact.

---

c. Which error is worse for this example? Justify your answer.

- In general, stastcians hate false positives in studies a lot more. I have never seen someone say "*Thank god I classified this bogus study as demonstrably true*"
- Type 1
- *I was going to reach out to Christian Antonioli, a doctoral student at the time, to ask how he feels, but I cannot find his email.*

---

d. Pick a value for α based on your answer above.

- $\alpha = 0.05$
- I wanna make the person grading this dissapointed that I would use the thing that Kelly said was too mainstream.

---

e. Compute the observed statistic of interest. In your answer, label the statistic as $\hat{p}_D - \hat{p}_Y$.

```
p_hat_dolphins <- 10/15

p_hat_yoga <- 3/15

test_stat<- p_hat_dolphins - p_hat_yoga
```

$\hat{p}_D - \hat{p}_Y$ 0.4666667

---

f. Describe how to use 30 slips of paper to find one null test statistic. Make sure it is clear how the slips should be labeled and how your process assumes H_o is true.

---

1. Slips should labeled the following:
   - Control with **no** improvment: 12
   - Control with an improvment: 3
   - Treatment with **no** improvment: 5
   - Treatment with an improvment: 10
2. Get two hats
3. Put the control slips in one hat and treatment in another
   - Bootstrap a proportion for each group, take the idfference, repeat.

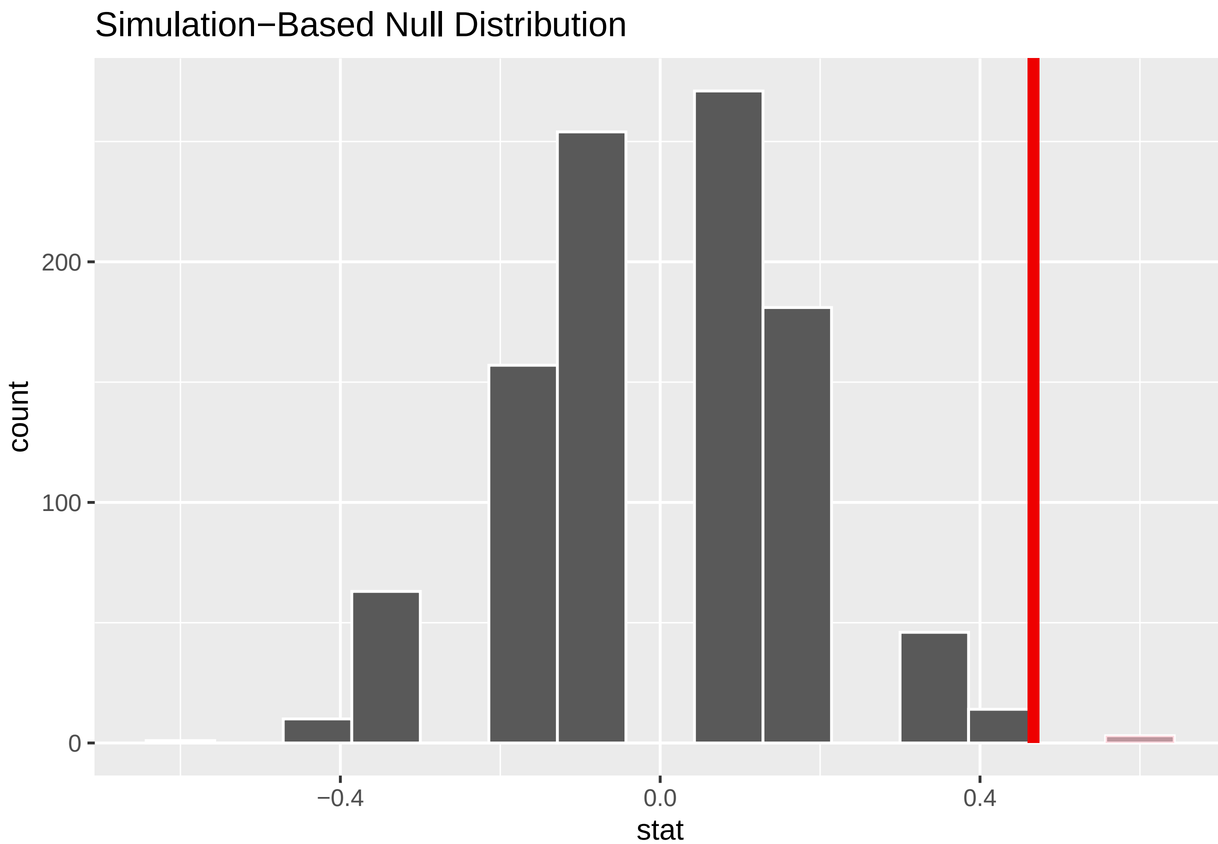
---

g. Instead of generating the null distribution with note cards, let's ask R to do the work for us. Generate a null distribution.

```
dolphins$group_factor <- factor(dolphins$group)

null_dolphins<- dolphins %>%
  specify(improve~group, success = "yes") %>%
  hypothesize(null = "independence") %>%
  generate(reps =  1000, type = "permute") %>%
  calculate(stat = "diff in props", order = c("Treatment","Control"))
```

h. Create a histogram of the null distribution. Comment on the shape and center of the plot. Does it agree with H_o?

```
null_dolphins %>%
  visualize() +
  shade_p_value(obs_stat = test_stat, direction = "right")
```



- Centered around zero.
- Wide spread
- Shows that the null should be rejected

---

i. Find the p-value. Explain what probability the p-value represents in this scenario.

```
dolphin_p <- null_dolphins %>%
  get_p_value(obs_stat = test_stat, direction = "right")

dolphin_p
```

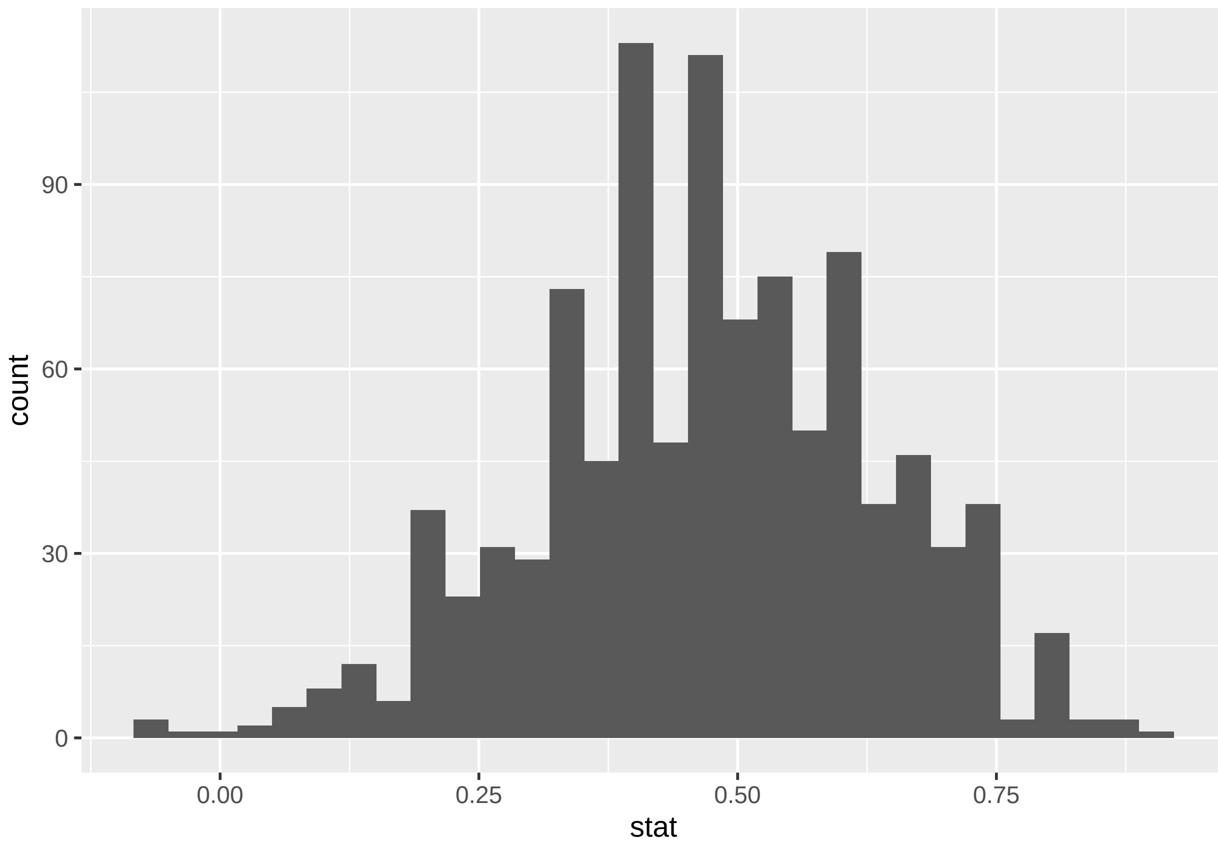
```
## # A tibble: 1 x 1
##   p_value
##     <dbl>
## 1   0.017
```

- 0.017 is less than 0.05. Therefore the null is rejected

Now let's construct a confidence interval for the parameter of interest.

j. Find the bootstrap distribution for the difference in sample proportions. Create a histogram of the bootstrap distribution.

```r
boot_dolphins<- dolphins %>%
  specify(improve~group, success = "yes") %>%
  #hypothesize(null = "independence") %>%
  generate(reps =  1000, type = "bootstrap") %>%
  calculate(stat = "diff in props", order = c("Treatment","Control"))

ggplot(boot_dolphins, aes(stat)) + geom_histogram()
```



k. Compare the null distribution and the bootstrap distribution. How are they different? Do they have the same center? Why or why not?

- Different centers. bootstrap distribution was centered around the test statistic while the null distribution was centered around zero.

---

l. Find the 90% confidence interval for the parameter of interest. Interpret the interval in the context of the problem.

```r
boot_dolphins %>%
  get_confidence_interval(point_estimate = test_stat, level = 0.9)
```

```
## # A tibble: 1 x 2
##   lower_ci upper_ci
##      <dbl>    <dbl>
## 1    0.194    0.733
```

- The median is within the values above 90% of the time.

---

m. Based on your analyses, effect size, and study design, what conclusions can we make about swimming with dolphins and depression? Address both *causation* and *generalizability* in your answer.

- Realistically, nothing. The study was flawed in both its setup and its variable.
- The study was 90% woman, and 73% white. Each of the individuals had to stop their antidepressants, take time off of work, and pay for their own flight to Honduras.
- The individuals that were in the dolphin group additionally had outdoor time, and hydrotherapy.
- Not a blinded expirement.
- What is measured at the end of the day was a rather interesting question. The study focused on a before and after depression questionare value.
- The question remains, where do ten and three come from? This quote from the study shines some light on the answer "*Because of logistical and financial limitations, we did not do a follow-up study; however, three months or more after the intervention, the 10 participants in the animal care programme and the three participants in the outdoor nature programme who had a score of no more than 7 on the Hamilton rating scale for depression at the end of treatment (clinically important improvement) provided a self report about their mental health status.* **Nine of the 10 participants in the animal care programme and all three of the outdoor nature programme reported lasting improvement and did not require treatment.**"
- In reality. We slightly skewed the results. The "treatment" group had nine individuals that imporved, and one who ended up reversing the impact treatment had on them. It is also worth noting this questionarre wasn't performed by a psychiatrist, it was a self evaluation.
- The lack of follow up also further highlights that the results were a mere snapshot.
- "Coorelation not causation"
- The control group did get to play with dolphins eventually
- The treatment and control group were aware of which group they were in, this might have impacted their mood.
  - Placebo is also one of the most powerful anti depressants. By making an individual aware that they are recieving a "treatment" as opposed to no treatment you manipulate the impact of treatment.

---

n. Which type of error, Type I or Type II, did you possibly commit for this example? Justify your answer.

- Type 1. We rejected the null hypothesis, so we accepted the study. For reasons above, I maintatin that this study could have been done better, and that the p value we generated was flawed.

Problem 2

Let's return to the movies example from Lab 7, Problem 2 and test the conjecture that audience ratings and critic ratings have a positive linear relationship. We want you to conduct a hypothesis test and write-up your results. Make sure to include:

- Useful exploratory analyses.

- The null and alternative hypotheses, the observed test statistic, and the p-value.
- Interpret the p-value in the context of the problem.
- Some conclusions about the conjecture.

```r
# Load data and drop rows that are missing variables of interest
movies <- read_csv("/home/courses/math141f19/Data/HollywoodMovies.csv") %>%
  drop_na(RottenTomatoes, AudienceScore)
```

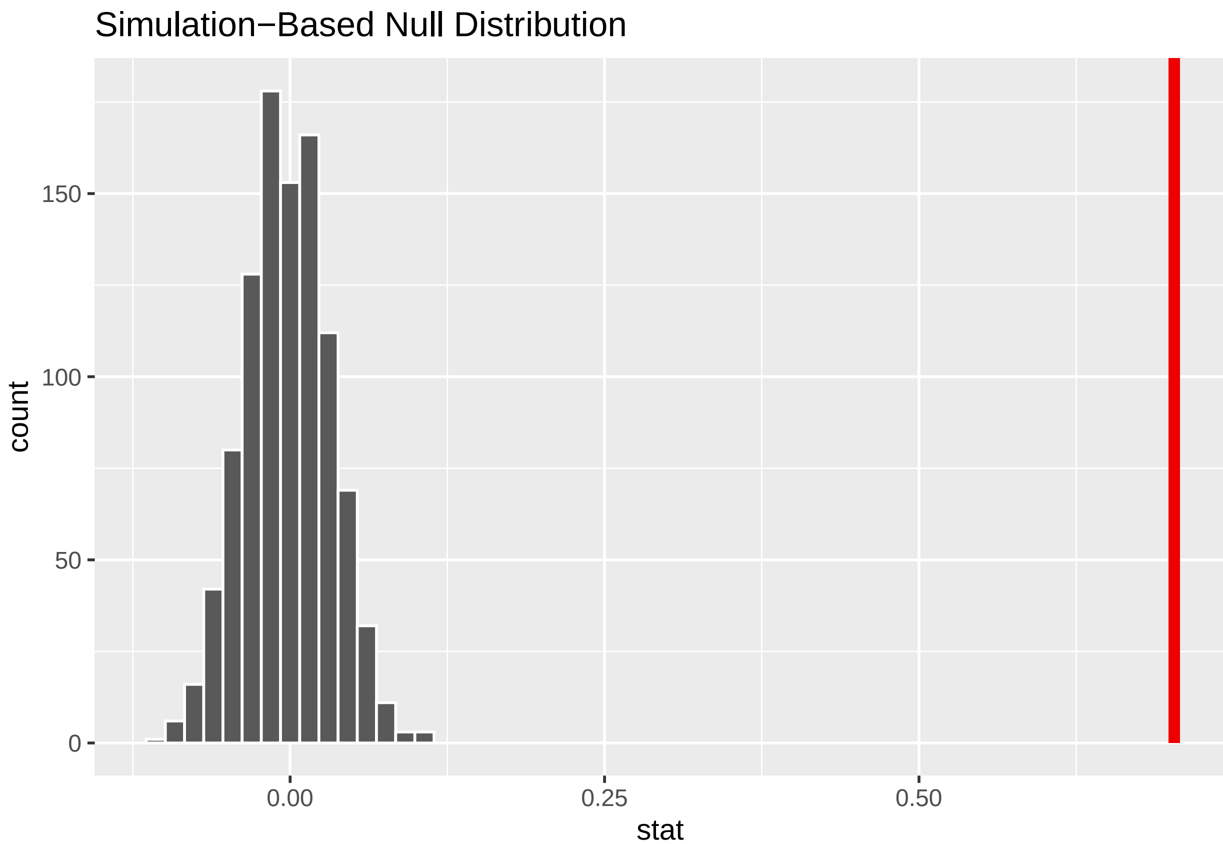
```r
test_stat_movie <- cor(movies$AudienceScore, movies$RottenTomatoes)

null_movie<- movies %>%
specify(RottenTomatoes~ AudienceScore) %>%
hypothesize(null = "independence") %>%
generate(reps =  1000, type = "permute") %>%
calculate(stat = "correlation")

movie_p <- null_movie%>%
get_p_value(obs_stat = test_stat_movie, direction = "right")

print(movie_p)
```

```
## # A tibble: 1 x 1
##   p_value
##     <dbl>
## 1       0
```

```r
boot_movie <-movies %>%
specify(RottenTomatoes~ AudienceScore) %>%
#hypothesize(null = "independence") %>%
generate(reps =  1000, type = "bootstrap") %>%
calculate(stat = "correlation")

values_less_than_zero <- boot_movie %>%
  filter(stat<0)

values_less_than_zero <- length(values_less_than_zero$stat)

print(values_less_than_zero)
```

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

```r
null_movie %>% visualize() +
  shade_p_value(obs_stat = test_stat_movie, direction = "right")
```



- H_0
  - The two categories are independent from one another.
  - *I reject the null hypothesis because the p-value is literally zero. Which should be impossible, but ehhhhhh*
- H_a
  - Greater than 0 bootstraped coorelation
  - Accept the alternative because there are 0 instances where the coorelation vlaue is less than zero.
- P value
  - 0
  - Magically zero, really shouldn't be, but the data is extreme.
- Conclusion
  - Accept the alternative, reject the null. Hypothesis is true.

Problem 3

Let's return to the anchoring example from Lab 5, Problem 2. We want you to determine to answer the following: Does the X value given serve as an anchor to how you answered question two? In other words, does the previously supplied number serve as a starting point when we estimate the percentage of UN nations that are African? (Note: We had to rename the column called `x` to `Anchor_Number` to not upset `infer`.)

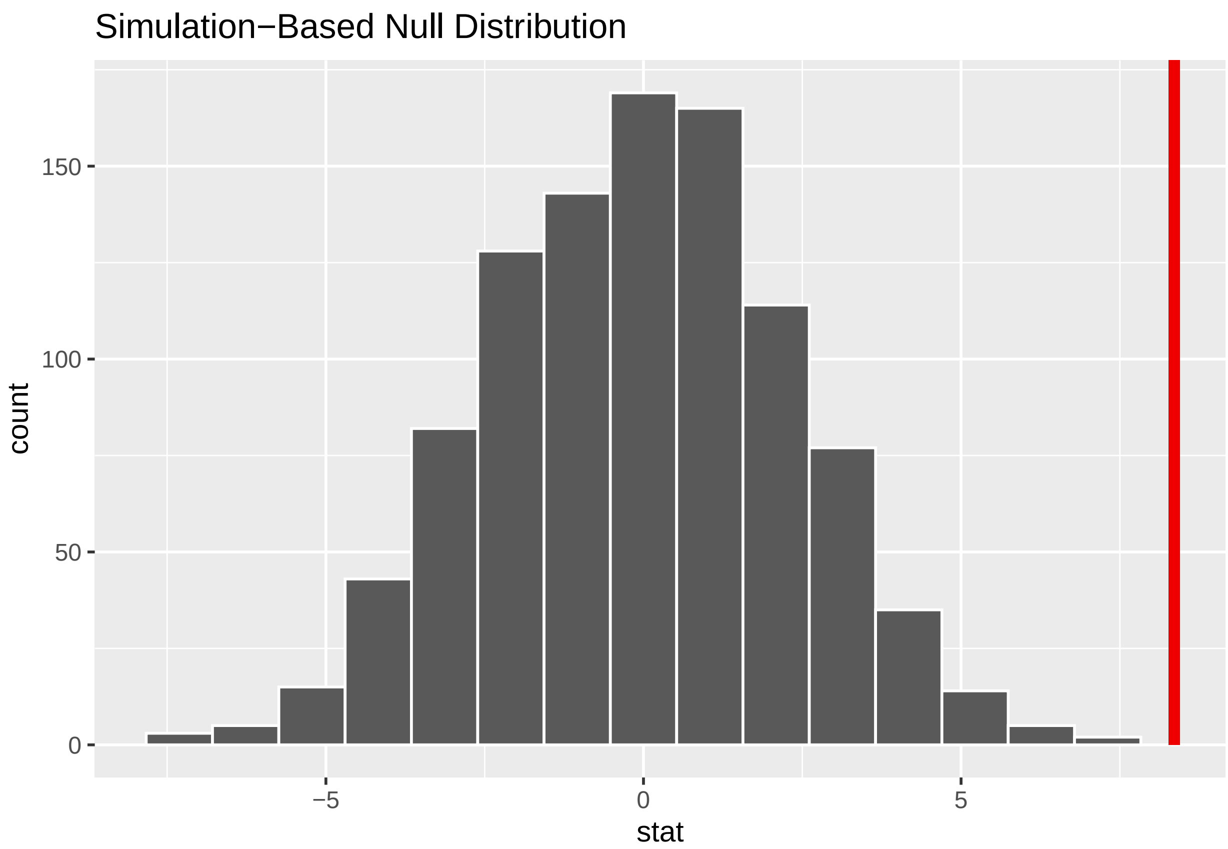
We want you to conduct a hypothesis test and write-up your results. Make sure to include:

- Useful exploratory analyses.
- The null and alternative hypotheses, the observed test statistic, and the p-value.

- Interpret the p-value in the context of the problem.
- Some conclusions about the conjecture.

```
# Load the data (also includes previous Math 141 classes)
anchor <-  read_csv("/home/courses/math141f20/Data/anchor.csv") %>%
  rename(Anchor_Number = x)
```

```
test_stat_anchor<- mean(anchor[1][anchor[2]=="sixty-five"]) -
  mean(anchor[1][anchor[2]=="ten"])


null_anchor <- anchor %>%
  specify(response=Value, explanatory = Anchor_Number) %>%
  hypothesise(null= "independence") %>%
  generate(reps=1000, type="permute") %>%
  calculate("diff in means", order= c("sixty-five", "ten"))


null_anchor %>% visualize() +
  shade_p_value(obs_stat = test_stat_anchor, direction = "right")
```




```
anchor_p <- null_anchor %>%
  get_p_value(obs_stat = test_stat_anchor, direction = "right")

print(anchor_p)
```

```
## # A tibble: 1 x 1
##   p_value
```

```
##     <dbl>
## 1       0
```

- H_0
  - The difference in means will be zero
- H_a
  - The difference in means will be greater than zero
- α
  - 0.05
- p
  - 0
  - Very itty bitty
- Conclusion
  - $p < \alpha$
  - Reject the null accept the alternative. Hypothesis is true.
