Practice Session 6

Introduction to Hypothesis Testing:

Generally speaking, hypothesis testing is a way for researchers to answer questions in a meaningful way. The methodology leverages statistical reasoning and probability to either "reject" or "fail to reject" a statement about the population as a whole. Hypothesis testing in statistics often focuses on questions about population parameters (e.g., μ, π). We do not write hypotheses in terms of the sample statistics (e.g., $\bar{x}$, $\hat{p}$). This aligns with the general theme of the course, which focuses on using samples to better understand the population.

Hypothesis testing can be used to answer questions about a single sample, or to compare samples from multiple groups. Although the calculations and assumptions may differ, the overarching framework is roughly the same in each case.

Hypothesis Testing Terms and Definitions:

Terms:

- Significance Level (α)
- Null Hypothesis
- Alternative Hypothesis
- P-value

Definitions:

- The probability of observing results at least as extreme as what we observed, assuming the null hypothesis is true.
- A statement of interest that represents the *status quo*, or that there is no effect or difference.
- The probability of rejecting the null hypothesis assuming the null hypothesis is true (i.e., the probability we incorrectly reject the null hypothesis).
- A statement of interest that proposes there is an effect or difference.

Hypothesis Testing for the Proportion

Example 1: AP Multiple Choice

Is the letter "C" the most common answer choice? Answer this question with a hypothesis test. Use the `APMultipleChoice` data set from the `Lock5Data` library. Make sure to follow all 5 steps of hypothesis testing. *Hint:* when defining the null hypothesis, consider how likely each answer choice would be if answers were selected at random.

1.) State the null and alternative hypothesis in words:

- [Null]
- [Alternative]

2.) Calculate the observed statistic of interest (obs_stat)

```
# Your code here
```

3.) Create a null distribution in R using `do_it()` and `rflip_count()`. Plot the null distribution, and add a vertical red line at your point estimate.

```
# Your code here
```

4.) Calculate the p-value (probability of obtaining a result as or more extreme than what we observed) using the `pnull()` function.

```
# Your code here
```

5.) Report the p-value. Make a decision on whether the results are statistically significant, and state your conclusion.

Example 2: Compassionate Rats

Would a rat attempt to free a trapped rat? In a recent study, some rats showed compassion by freeing another trapped rat, even when chocolate served as a distraction and even when the rats would then have to share the chocolate with their freed companion.

Conduct a hypothesis test to see if female rates are more compassionate than their male counterparts.

1.) State the null and alternative hypothesis in words:

- [Null]

- [Alternative]

2.) Calculate the observed statistic of interest (obs_stat)

```
# Your code here
```

3.) Create a null distribution in R using `do_it()` and `rflip_count()`. Plot the null distribution, and add a vertical red line at your point estimate.

```
# Your code here
```

4.) Calculate the p-value (probability of obtaining a result as or more extreme than what we observed) using the `pnull()` function.

```
# Your code here
```

5.) Report the p-value. Make a decision on whether the results are statistically significant, and state your conclusion.

Exploring the `pnull()` Function

In this exercise, we will practice using some basic R coding to calculate a p-value without using the `pnull()` function.

1.) Consider null distribution and point estimate from the previous exercise. Using the `>` operator, write a single line of code that will show which values in the null distribution are greater than the point estimate. Run this code and observe what the values look like. Save this result to a vector. *Hint:* don't over think this part. Just write the inequality as you would in math class.

```
# Your code here
```

2.) Using your vector from above, count how many values from the null distribution are greater than the point estimate. Use the `sum()` function for this. *Hint:* the `==` operator can be used within the `sum()` function to compare values.

```
# Your code here
```

3.) Repeat the exercise above simply using the `sum()` function on your vector from part 1.). What do you notice? What happens when you run the code `sum(TRUE)`?

```
# Your code here
```

4.) Using your result from part 2.), find the percentage of observations in the null distributions that are greater than the point estimate. Compare to the number from `pnull()` earlier.

```
# Your code here
```

Hypothesis Testing for the Difference of Means