Homework 8 (due Thursday, April 12th)

By signing my name, I agree to abide by the Stern Code of Conduct ______________________

Question 1

This question is about whether or not a certain coin is fair. We model the coin as a Bernoulli random variable $C \sim \text{Ber}(p)$, with $P(C = \text{heads}) = p$ and $P(C = \text{tails}) = 1 - p$. We toss the coin n times and record the outcomes as $C_1, C_2, \ldots, C_n$.

a.

We want to know if the coin is fair or not. What are the null and alternative hypotheses?

b.

If we toss the coin n times and record how many times it lands on heads, we can use that number as a test statistic. What is the distribution of this test statistic under the null hypothesis?

c.

Suppose the observed value of this test statistic is x. How could you calculate a p-value using the distribution under the null hypothesis?

d.

Suppose $n = 10$ and the number of heads is 9. For the *two-sided alternative*, which outcomes are at least as extreme as 9? Use `pbinom` to calculate the probability, under the null distribution, of seeing an outcome at least as extreme as 9, and write the answer below. Based on this, would you reject the null hypothesis at the 5% significance level? What about at the 1% significance level?

e.

Suppose $n = 2$. In this case, what is the smallest possible p-value for a one-sided test?

Question 2

In 403 episodes of The Joy of Painting, Bob Ross often painted landscapes that involved "happy trees" and "majestic mountains." See below for some output from `R` of a "cross tabulation" counting how many of the paintings had no trees or mountains, no trees but mountains, no mountains but trees, and both mountains and trees.

```
counts <- table(bob_ross$trees, bob_ross$mountains)
```

```
##
##             no mountains mountains
##   no trees            61         5
##   trees              243        94
```

```
chisq.test(counts)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  counts
## X-squared = 11.222, df = 1, p-value = 0.0008081
```

What is the null hypothesis for this test? What do you conclude based on the test output?

Question 3

a.

The numbers below are from the paper discussed in class about neonatal sex differences in "social perception." These particular numbers are for the percent of time spent looking at an object.

```
x1 <- 51.9
s1 <- 23.3
n1 <- 44
x2 <- 40.6
s2 <- 25.0
n2 <- 58
# This next line is just degrees of freedom
# it's a complicated formula, ignore it
df <- (s1^2/n1 + s2^2/n2)^2/((s1^2/n1)^2/(n1-1) + (s2^2/n2)^2/(n2-1))
c <- qt(.975, df)
```

```r
SE <- sqrt(s1^2/n1 + s2^2/n2)
c((x1-x2) - c*SE, (x1-x2) + c*SE)
```

```
## [1]  1.756463 20.843537
```

What did this code calculate, what is the meaning of the output? Based on this output, do we reject the null hypothesis that there is no difference in percent time looking at an object at the 5% significance level? Do we reject the null hypothesis at the 1% significance level? For both of these, answer yes, no, or not enough information.

b.

Below is a 95% confidence interval for the difference (male - female) in percent of time spent looking at a face.

```
## [1] -12.692188   5.092188
```

Based on this interval, do we reject the null hypothesis that male babies spent 5% more time looking at a face than female babies at the 5% significance level? (Yes, no, not enough information)

c.

If we have two hypothesis tests computed at the 5% significance level, each one individually has a 5% probability of type 1 error (false rejection). Assuming the tests are independent, and both null hypotheses are true, what is the probability of at least one false rejection? The answer should be a number, but show your work.

Question 4

What is p-hacking? Give one example of how it might be done in practice. What are some recommendations on how to avoid it?