Homework 2 (due Wednesday, April 11th)

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

Question 0

Suppose we are calculating a 95% confidence interval for a mean based on a sample of size n.

a.

At first, the interval is just too wide to be useful. We decide to increase the sample size to get a narrower interval. What sample size would we need to get an interval that will probably be about half as wide as the original one?

b.

Suppose $n = 100$, and the confidence interval has the form $\bar{X} \pm cSE$ where SE is the sample standard error of the mean (note: this is not the same as a the sample standard deviation), and the constant c is obtained from the t-distribution with $n - 1$ degrees of freedom. For example, for a 95% confidence level, c is

```
qt(.975, 99)
```

```
## [1] 1.984217
```

What is the length of the interval? The answer should be some constant times SE. What is the length if we increase the confidence level to 99%? (You will need to use `R` to answer this second part, but can copy the answer to this page by hand).

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
```

```r
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 a highly cited paper about neonatal sex differences in "social perception." In that study, the researchers presented neonatal infants with two stimuli to look at: a picture of a face and an object. They recorded how much time each infant spent looking at either of the two stimuli, and recorded preferences as percent of total time spent looking at each. These researchers were interested in seeing if there was a difference in the average preferences between female and male infants. The numbers below are for the percent of time spent looking at the object (not the face). The study included 58 female and 44 male subjects.

```r
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)
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, 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?

Reference

Connellan, J., Baron-Cohen, S., Wheelwright, S., Batki, A., & Ahluwalia, J. (2000). Sex differences in human neonatal social perception. Infant behavior and Development, 23(1), 113-118.

https://pdfs.semanticscholar.org/50c7/f3fb0b144fb648fb4238d01d2837c57ce36b.pdf