MA117 - Worksheet 9
Mid-Block Review
Week 2, Friday

Problem 1. (a) 1000 statistical studies rejected their null hypotheses at the 1% significance level. Roughly how many of them reached the wrong conclusion?

(b) 1000 statistical studies failed to reject their null hypotheses at the 1% significance level. Roughly how many of them reached the wrong conclusion?

(c) 1000 statistical studies constructed 99% confidence intervals for some parameter of interest. Roughly how many of those confidence intervals don't contain their parameter of interest?

Problem 2. The distant planet of Ungli is home to a race of aliens who call themselves Ungliwallahs. Half of the population of Ungliwallahs have 20 fingers and the other half have 40 fingers. Let X be the random variable modeling the number of fingers on a randomly chosen Ungliwallah. What is the expected value of X? What is the variance of X?

Problem 3. In the Kingdom of Okane, there are two types of coin in circulation called *kinka* and *ginka*. The masses of *kinka* are normally distributed with mean 2 g and standard deviation 0.1 g. The masses of *ginka* is normally distributed with mean 3 g and standard deviation 0.05 g. Okanemochi is the richest aristocrat in the Kingdom of Okane. Every morning, he fills his empty coin purse with exactly 10 *kinka* and 20 *ginka* before he heads out for his daily aristocratic errands. The coins he puts into his coin purse are chosen independently of one another.

(a) What is the expected total mass of the coins in Okanemochi's coin purse every morning?

(b) What is the standard deviation of the total mass of the coins in Okanemochi's coin purse every morning?

Problem 4. About 30% of human twins are identical, and the rest are fraternal. Identical twins are always the same gender. One-quarter of fraternal twins are both male, one-quarter are both female, and one-half are mixes (one male, one female). You have just become a parent of twins and are told that they are both girls. What is the probability that they are identical twins?

Background. FiveThirtyEight conducted a survey about singular versus plural usage of the word "data." Specifically, the survey asked respondents the following questions:

A. How would you write the following sentence?

(a) Some experts say it's important to drink milk, but the data is inconclusive.

(b) Some experts say it's important to drink milk, but the data are inconclusive.

B. When faced with using the word "data," have you ever spent time considering if the word was a singular or plural noun?
C. How much, if at all, do you care about the debate over the use of the word "data" as a singluar or plural noun?

Problem 5. Answer the above questions A–C for yourself! ☺

The results of FiveThirtyEight's survey are recorded in the following `csv` file:

`https://sagrawalx.github.io/teaching/fa21-b1_ma117/class/datasgpl.csv`

Responses are encoded as follows:

A. The `Prefer` column of the data set records respondents' responses to question A, with `Singular` indicating sentence (a) and `Plural` indicating sentence (b).
B. The `ThoughtAbout` column records respondents' responses to question B. The possible responses are `Yes` or `No`.
C. The `CareAbout` column records respondents' responses to this question B. The possible responses were `Not at all`, `Not much`, `Some`, or `A lot`.

The survey also asked respondents for some demographic information, which are included in the `csv` file in a self-explanatory way. Note that some respondents did not respond to some of the questions, so you'll want to throw out blank responses from the relevant questions when doing some of the calculations below. Answer the following using both a confidence interval and a p-value hypothesis test.

Problem 6. Are these survey results compatible with the hypothesis that the proportion of people who prefer `Singular` usages of the word "data" is 50%?

Problem 7. Does this survey provide evidence that `Prefer` is correlated with `ThoughtAbout`?
Recall. If you named your data frame `df`, typing `table(df$Prefer, df$ThoughtAbout)` into the console will show you a contingency table.

Problem 8. Does this survey provide evidence that `Prefer` is correlated with `Gender`?