Activity 12

Coin Flip
In order to get a better sense and understanding of randomness we are going to flip coins. More accurately we are going to simulate the flipping of a coin. Understanding the role of randomness is key to understanding how to interpret/consume inferential statistics.

1. Simulate 10 flips of a fair coin. What were the number of heads? Run the simulation 5 times.

This is random so everyone will get different answers

	Sim 1	Sim 2	Sim 3	Sim 4	Sim 5
Number of Heads	7	7	5	8	4
Proportion of Heads	0.7	0.7	0.5	0.8	0.4

Describe what is happening across the simulations.

We see the number of heads fluctuate in the number and/or order they show up in 10 flips.

2. Now let's modify how many times we flip the coins (n) and see how that changes things.

This is random so everyone will get different answers

	n = 10	n = 20	n = 50	n = 100	n = 500	n = 1000
Number of Heads	6	11	24	58	237	494
Proportion of Heads	0.6	0.55	0.48	0.58	0.474	0.494

a) Describe what is happening as you increase the number of flips.

As we up the number of flips we see that that the empirical/observed proportion of heads tends to 0.5 which is the parameter (population value) we set in the simulation.

b) If you run them all again, will the results be exactly the same? Explain any differences and/or similarities

The results will not be exactly the same because we are running a random process. Meaning the values won't be exactly the same. However, the trend of the empirical/observed proportion of heads tending to 0.5, the parameter (population value), will still be evident.

3. Let's try this out with an **unfair** coin. That is, use a weighted coin that favors heads ('p = 0.6').

	n = 10	n = 20	n = 50	n = 100	n = 500	n = 1000
Number of Heads	7	8	30	61	300	614
Proportion of Heads	0.7	0.4	0.6	0.61	0.60	0.614

a) Describe what is happening. Is it surprising? Explain.

As we up the number of flips we see that that the empirical/observed proportion of heads tends to 0.6 which is the parameter (population value) we set in the simulation. No, this is not surprising. As we collect more information (flips) the observed proportion should reflect the TRUTH about the coin which is it is unfair with a higher probability of landing on heads (0.6). b) Do you think you could tell that the coin was unfair with only 10 flips? Explain.

Probably not. Even a fair coin will produce 6 heads in 10 flips quite often (by random chance). The slightly unfair coin will produce 5 heads in 10 flips quite often too. With only 10 flips it will be very hard to determine this.

Random Assignment

4. Consider the 'CDC Dataset'. We are conducting an experiment to test the effects of a vitamin supplement on health. People assigned to the treatment group will take the vitamin daily and people assigned to the control group will not take the vitamin.

Assign people to groups using Random Assignment. What do you notice about the means for each group? Consider running the simulation several times to see how this plays out with different randomizations.

They are very close to being the same for each variable. When we rerun this over and over again, the means tend to be similar. Sometimes the differences are a little more pronounced, but the values are still fairly similar.

Assign people to groups using Voluntary Assignment. (This simulation serves as an example and is based on hypothetical data.) Voluntary assignment allows for the introduction of confounding variables. What might be impacted and how might this impact the conclusions about the vitamin in our experiment?

Each variable is no longer necessarily similar between the groups.
A confounding variable for this group could be healthy lifestyle. Healthy people tend to exercise more and are more health conscious so would be more likely to take the vitamin supplement. Their overall health (bmi, heart rate, cholestoral etc) would naturally be better than the other group but not necessarily caused by the vitamin.

5. Consider the 'Survey Dataset'. We are conducting an experiment to test if there is a difference in student performance between in-person classes and online asynchronous classes. Students assigned to the treatment group will take the class on-line and people assigned to the control group will take the class in-person.

Assign students to groups using Random Assignment. Run the simulation a few times and pay attention to the means for each group. The means should be approximately the same between groups.

Now assign students to groups using Voluntary Assignment. (This simulation serves as an example and is based on hypothetical data.) How might the conclusions from the results of the experiment differ between Random and Voluntary Assignment? What could be a confounding variable in Voluntary Assignment (does not have to be in the dataset)?

Perhaps students who are more self motivated are more likely to take asynchronous classes. Self motivated students are likely to achieve better grades but that does not mean the online asynchronous delivery is improving student performance

Causation vs Correlation

Choose 2 articles from ["is it random or not?"]
http://jfmueller.faculty.noctrl.edu/100/correlation_or_causation.htm
For each article decide:
- Does the headline imply a causal or correlation claim?
- What's being identified as the treatment group?
- What might the omitted variable be (if not random)?

ARTICLE 1: More Math Helps Young Scientists: More math in high school is being identified as the treatment in the title. In the article they say it correlates (so not being randomly assigned).

The omitted variable could be the quality of math educators or the math curriculum used at each student's high school.

ARTICLE 2: