Foundations for Inference

Submission Instructions:

- Write the answers to this assignment directly in Google Docs in the **07 Foundations for Inference** folder in Google Drive.
- Name this file `foundations_userid`.
- Use headers to make your assignment readable.

Sampling distributions

Review the results of the die rolling exercise from the [Hack MD Lec05 collaborative notes], and the uploaded pictures in the `simulation` folder. Answer the following questions.

1. explain the difference between a probability distribution and a sampling distribution.
2. Describe the concept of sampling variability.

Scale up

1. Open a web browser and go to: https://gallery.shinyapps.io/CLT_mean/
2. Using the slider on the left side, reduce the sample size to 2, and the number of samples to 10.
3. How to read this interactive app:
    a. The plot in the first tab is the *Population Distribution* of the x's.
    b. The smaller plots on the *Samples* tab are the results of random samples drawn from this population. Each data point (x) is shown as a dot, the vertical bar is the sample mean.
    c. The green plot at the bottom of the *Sampling Distribution* tab is a histogram and density plot for the sampling distribution. This is the distribution of the means from each of the samples drawn above.

Your job: Pick a shape (non-normal, and not uniform) for the population. If you chose a skewed distribution, play around with both a high and a low skew. If you chose a uniform distribution, play around with upper and lower bounds. Change the parameters (e.g. mean, sd, or upper and lower bound).

0. Describe the population distribution that you chose. Describe location, shape, spread.
1. Adjust the number of samples. What happens to the sampling distribution (location/shape/spread)?
2. Adjust the sample size. What happens to the sampling distribution (location/shape/spread)?
3. At what sample size does the sampling distribution of a low skewed population become approximately normal?
4. What about for a distribution with high skew?
5. Can you ever actually observe the sampling distribution? Why do we need to know why it behaves?

Interval Estimation

In class [activity] on interval estimation. Turn this in separately, old school wise (by hand)

Central Limit Theorem

Using the continuous variable you worked with on the interval estimation worksheet, explain to someone else what the Central Limit Theorem tells you about the sampling distribution. But don't use fancy technical terms. For example if you were researching a person's age, describe the distribution of the individual ages, and the distribution of the average age (complete with numbers).