Foundations for Inference

Submission Instructions:

- Write your responses directly into a google document in the **07 Foundations for Inference** folder in Google Drive by the due date
- Name your file `hw07_inference_username`.
- Don't forget to also turn in the interval estimation worksheet that we will work on in class to Dr. D. directly.

Parameters vs Statistics

For each question mark TRUE or FALSE, and write a justification for your answer

1. In a study of all 2857 employees at a college, it is found that 45% own a vehicle.
   TRUE OR FALSE: The number 45% is a parameter.
2. In a survey, one hundred college students are asked how many hours per week they spend on the Internet.
   TRUE OR FALSE: The parameter of interest is a proportion.
3. In a survey, smart phone users are asked how many times they used a web-based taxi service over the last year.
   TRUE OR FALSE: The parameter of interest is a proportion.
4. In a survey, one hundred college students are asked whether or not they cited information from Wikipedia in their papers.
   TRUE OR FALSE: The parameter of interest is a proportion.
5. Suppose you want to estimate the percentage of videos on YouTube that are cat videos. It is impossible for you to watch all videos on YouTube so you use a random video picker to select 1000 videos for you. You find that 2% of these videos are cat videos.
   TRUE OR FALSE: A video in your sample is a sample statistic.
6. Suppose you want to estimate the percentage of videos on YouTube that are cat videos. It is impossible for you to watch all videos on YouTube so you use a random video picker to select 1000 videos for you. You find that 2% of these videos are cat videos.
   TRUE OR FALSE: 2% is a sample proportion.

Sampling distributions

Using the results of the in class die rolling activity, answer the following questions.

1. explain the difference between a probability distribution and a sampling distribution.
2. Describe the concept of sampling variability.
3. As the sample size increases, describe what happens to the following features of the sampling distribution
   - mean:
   - standard deviation:
   - shape
4. As the number of samples increases, describe what happens to the following of the sampling distribution
   - mean:

- standard deviation:
- shape

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.

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?

Interval Estimation

In class [worksheet]. Turn this in separately, old school wise (by hand)

Explain the Central Limit Theorem

Using the continuous variable you worked with on the Foundations 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).