Discovering the Central Limit Theorem

This exercise follows the information in the course packet, section 4.1-4.4. Write your answers on a separate piece of paper. This will be turned in for credit. It can be hand-written and turned in during the next class period or typed up and submitted to me via Slack.

Guided Setup

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. Orientation to this interactive app:
   1. The blue plot at the top is the population distribution of the x’s.
   2. The smaller plots below are the results of random samples drawn from this population.
      1. Each data point (x) is shown as a dot, the vertical bar is the sample mean.
   3. The green plot at the bottom is a histogram and density plot for the **sampling distribution**.
      1. This is the **distribution of the means** from each of the samples drawn above.

Initial questions

1. Are the numbers in the top right of the blue plot population parameters or sample statistics?
2. Using an English sentence, describe the distribution (location/shape/spread) of the population. Be sure to use numbers in your description.
3. Are the numbers in the top right of the small white plots population parameters or sample statistics?
4. Why do these numbers differ across samples?
5. What is meant by *sampling variation*?
6. How many x’s were drawn to create each sample? (What is the sample size?)
7. How many samples did you generate?
8. How many sample means were calculated?
9. Compare the location (mean), shape, and spread (standard deviation) of the sampling distribution to the population

Change parameters

Increase the **number of samples** to 1,000.

1. How many x’s were drawn from the population to create each sample?
2. How many samples did you generate?
3. Compare the location (mean), shape, and spread (standard deviation) of the sampling distribution to the population
4. How does increasing the number of samples affect the mean and standard deviation of the sampling distribution? (Compare the mean and standard deviation from this sampling distribution to the mean and standard deviation from the previous sampling distribution from the initial setup).

Increase the **sample size** to 50

1. How many x’s were drawn from the population to create each sample?
2. How many samples did you generate?
3. Compare the location (mean), shape, and spread (standard deviation) of the sampling distribution to the population
4. How does increasing the sample size affect the mean and standard deviation of the sampling distribution? (Compare the mean and standard deviation from this sampling distribution to the mean and standard deviation from the sampling distribution from the initial setup).

Change the population parameters

1. Change the mean and standard deviation to your favorite numbers. You are not allowed to use 0 for the mean or 1 for the standard deviation.
2. Change the sample size to 100.
3. Report the mean and standard deviation of both the population, and the sampling distribution.
4. What is the mathematical relationship between the standard deviation of the population and the standard deviation of the sampling distribution.

Create your own distribution

Pick a different shape (non-normal) 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?

**Explain the Central Limit Theorem**

In your own words, based on what you have explored through simulation and information in the textbook, what does the Central Limit Theorem say?