Central Limit Theorem:

Goals of the lab:
To familiarize students with the Central Limit Theorem

Background:

The Central Limit Theorem:
Suppose X is a RV with mean \( \mu \) and standard deviation \( \sigma \). If we take a sample of size \( n \geq 20 \) from this distribution and \( n \geq 20 \), then the average of those \( n \) observations \( \bar{X} \) will be a Normal random variable with mean \( \bar{X} = \mu \), and standard deviation \( \sigma_{\bar{X}} = \sigma / \sqrt{n} \). The standard deviation of the sampling distribution of \( \bar{X} \) is also called the standard error of \( \bar{X} \) and is sometimes denoted \( \text{SE}(\bar{X}) \).

The Central Limit Theorem is one of the most fundamental and important results in statistics. However, it is also a very complicated one. The great difficulty that occurs with discussing this result is that so many concepts meet in this one theorem that it is hard for intelligent people (not just Stat 211 students to grasp). First, the distribution of \( X \) can be of any kind. It can be a discrete distribution or a continuous distribution. It can have a symmetric distribution or a heavily skewed distribution. The only things we need know about the distribution of \( X \) are its mean \( \mu \) and standard deviation \( \sigma \). Second, we sample \( n \) observations from the distribution of \( X \). Third, the distribution of the sample mean, \( \bar{X} \), will have the same mean as the distribution of \( X \), \( \bar{X} = \mu \). Fourth, the standard deviation of the distribution of the \( \bar{X} \) will be \( \text{SE}(\bar{X}) = \sigma / \sqrt{n} \). This implies that the standard deviation of the distribution of the sample mean will be smaller than the standard deviation of the distribution of \( X \) by a factor of \( \sqrt{n} \). Finally, when \( n \geq 20 \), the shape of the distribution of the sample means will be (approximately) that of a Normal distribution.

So it is clear that there are at least 5 distinct concepts involved in the central limit theorem. By having two labs on this, we can separate out some of these concepts. The first lab involves the last concept which is the idea that the distribution of the sample averages becomes more and more like a Normal distribution as \( n \) increases. The second lab deals with the concept of the mean and standard deviation of the sampling distribution of \( \bar{X} \).

Examples:

- Suppose \( X \) is a RV with mean \( \mu = 7 \) and standard deviation \( \sigma = 3 \). If we were to take a sample of 65 observations from this distribution and average them, what would be the distribution of the sample mean?

The mean of the sampling distribution of \( \bar{X} \) would be \( \mu_{\bar{X}} = 7 \) and the standard deviation of the sampling distribution of \( \bar{X} \) would be \( \text{SE}(\bar{X}) = \sigma / \sqrt{n} = 3 / \sqrt{65} = 0.372 \). Since \( n \geq 20 \), the distribution of \( \bar{X} \)'s possible values would have a Normal distribution.

- Suppose that \( Y \) is a Normal RV with mean 0.85 and standard deviation 0.15. If we were to take a sample of 12 observations from this distribution and average them, the distribution of that average would be what?

The mean of the sampling distribution of \( \bar{Y} \) would be 0.85 and the standard deviation of the sampling distribution of \( \bar{Y} \) would be 0.15 / \( \sqrt{12} = 0.043 \). Since the distribution of \( Y \) is Normal, then the sampling distribution of \( \bar{Y} \) will be a Normal distribution.
• Suppose H is a RV with mean 702 and standard deviation 61. If we were to take a sample of 15 observations from this distribution and average them, the distribution of that average would be what?

The mean of $\overline{H}$ would be 702 and the standard deviation of $\overline{H}$ would be $61 / \sqrt{15} = 15.75$. However, neither $n \geq 20$ nor the distribution of H is Normal, so we cannot say anything about the shape of the sampling distribution of $\overline{H}$. 