## Lecture #18 - Learning about the world through surveys

**Some important definitions:**

1. **Population** - A group of objects or people we wish to study.
2. **Parameter** - A numerical value that characterizes some aspect of this population.
3. **Census** - A survey in which EVERY member of the population is measured.
4. **Sample** - A collection of people or objects taken from the population of interest.
5. **Statistic** - A numerical characteristic of a sample data. Statistics are used to estimate parameters. Statistics are sometimes called estimators and the numbers that result are called estimates.
6. **Bias** is measured using the center of the sampling distribution: It is the distance between the center and the population parameter value.
7. **Precision** is measured using the standard deviation of the sampling distribution, which is called the **standard error**. When the standard error is small, we say the estimator is **precise**.
8. **Sampling Distribution** - the special name for the probability distribution of a statistic. Used to make inferences about a population.

**Facts:**

1. No matter how many different samples we take, the value of $\mu$ (the population mean) is always the same, but the value of $\bar{x}$ changes from sample to sample.
2. The precision of an estimator does NOT depend on the size of a population; it depends only on the sample size.
3. Surveys based on larger sample sizes have smaller standard error and therefore better precision. Increasing sample size improves precision.

**Keeping track of parameters and statistics:**

| **Parameters (typically unknown)** | **Statistics (based on data)** |
|---|---|
| 1. $\mu$ - population mean | 5. $\bar{x}$ - sample mean |
| 2. $\sigma$ - population standard deviation | 6. $s$ - sample standard deviation |
| 3. $\sigma^2$ - population variance | 7. $s^2$ - sample variance |
| 4. $p$ - population proportion | 8. $\hat{p}$ - sample proportion |

**THE CENTRAL LIMIT THEOREM - Three ways**

1. The Central Limit Theorem for a Sample **PROPORTION** tells us that if we take a random sample from a population, and if the sample size $n$ is large and the population size is much larger than the sample size, then the sampling distribution of the sample proportion $\hat{p}$ is approximately normal with mean $p$ and standard deviation

   $$\sqrt{\frac{p(1-p)}{n}}$$

   (If you don't know the value of $p$, then you can substitute the value of $\hat{p}$ to calculate the estimated standard error.)

2. The Central Limit Theorem for Sample **SUM** tells us that if we take a random sample $X_1, X_2, \ldots, X_n$ from a population, and if the sample size $n$ is large and the population size is much larger than the sample size, then the sampling distribution of the sum $X_1 + X_2 + \cdots + X_n$ is approximately normal with mean $n\mu$ and standard deviation $\sigma\sqrt{n}$.

3. The Central Limit Theorem for Sample **MEAN** tells us that if we take a random sample from a population, and if the sample size $n$ is large and the population size is much larger than the sample size, then the sampling distribution of the mean $\bar{X}$ is approximately normal with mean $\mu$ and standard deviation $\frac{\sigma}{\sqrt{n}}$.