Summary Statistics:

Statistics are seen everywhere and are an important part of understanding the world in which we live. For example, many statistics get thrown around during political elections regarding which candidate is favored. A professor might talk about the average score on the last exam you took which is another statistic. However, to truly understand what a statistic is, we must first understand the difference between a *population* and a *sample*.

*A **population** consists of all individuals that we're interested in knowing some characteristic about*. *A **parameter** is a characteristic of that population.* In the example of a political election, our population might be all voters in the United States and our parameter might be the proportion of voters who favor a given candidate. Since it is infeasible to talk to every voter in the United States, we can utilize a sample as an effective alternative.

*A **sample** is a subset of or a smaller group within the population*. *A **statistic** is a characteristic of a sample*. Rather than talk to every voter in the United States, we might talk to a few randomly selected voters and record which candidate they will vote for. From the sample, we can then compute the proportion (which is a statistic) who favor the candidate of interest. This is illustrated in Figure 1.

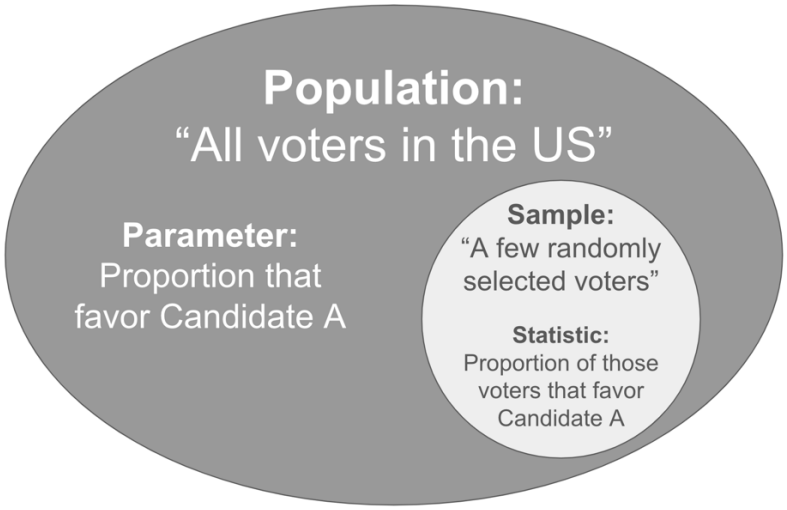

Figure 1: Population vs. Sample, Parameter vs. Statistic

We don't only utilize samples because they are more convenient to obtain. If done correctly, the statistics obtained from samples can be generalized to the corresponding population. For instance, if 30% of people in the sample favor candidate A, we could reasonably assume that about 30% of the population favors that candidate.

In statistics, there are many types of statistics and parameters. The ones we are generally concerned with measure where the center is, what is the proportion, how spread out the data is, etc. The table below gives the most common formulas to calculate these statistics/parameters. Note that capital N refers to the population size and n refers to the sample size.

Name	Symbol	Parameter/Statistic	Formula
Measures of Center			
Population Mean	μ	Parameter	$\sum_{i=1}^{N} \frac{x_i}{N}$
Sample Mean	$\bar{x}$	Statistic	$\sum_{i=1}^{n} \frac{x_i}{n}$
Median	N/A	Statistic	Order numbers from least to greatest and find the middle one. If the middle falls between two numbers, take the mean of those two numbers.
Mode	N/A	Statistic	The number that occurs the most frequently.
Proportions			
Population Proportion	p	Parameter	$\frac{X}{N}$, $X =$ successes of pop.
Sample Proportion	$\hat{p}$	Statistic	$\frac{x}{n}$, $x =$ successes of sample
Measures of Spread			
Range	N/A	Statistic	$\max - \min$
Population Variance	σ^2	Parameter	$\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}$
Sample Variance	s^2	Statistic	$\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$
Population Standard Deviation	σ	Parameter	$\sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}}$
Sample Standard Deviation	s	Statistic	$\sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}$