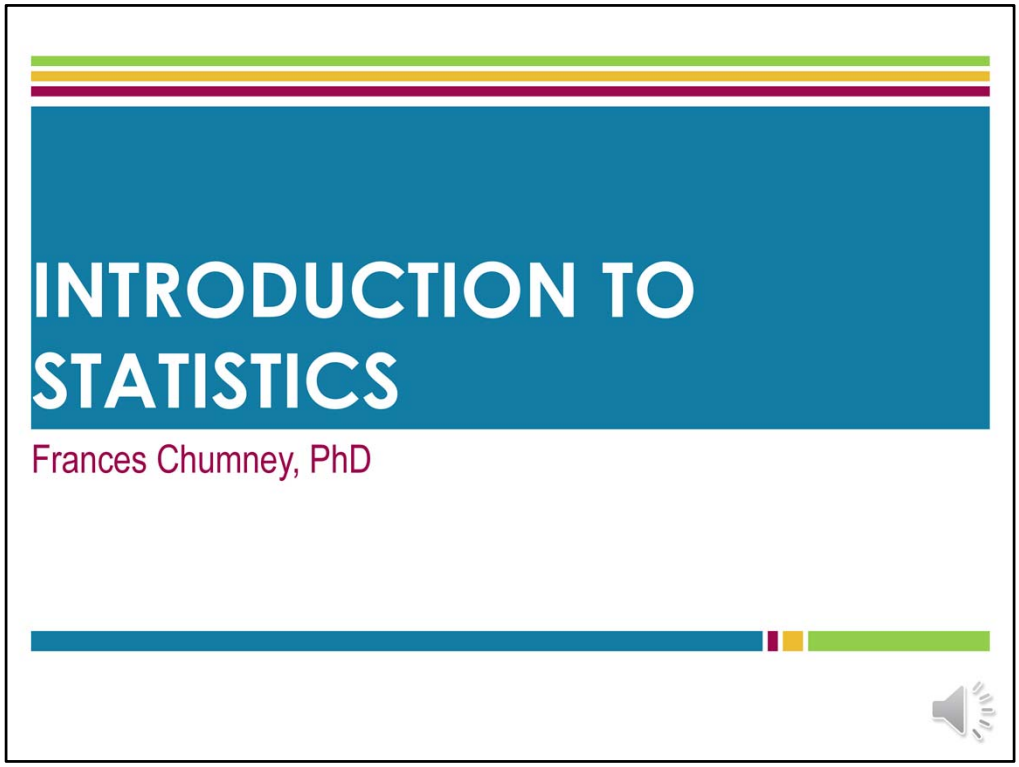
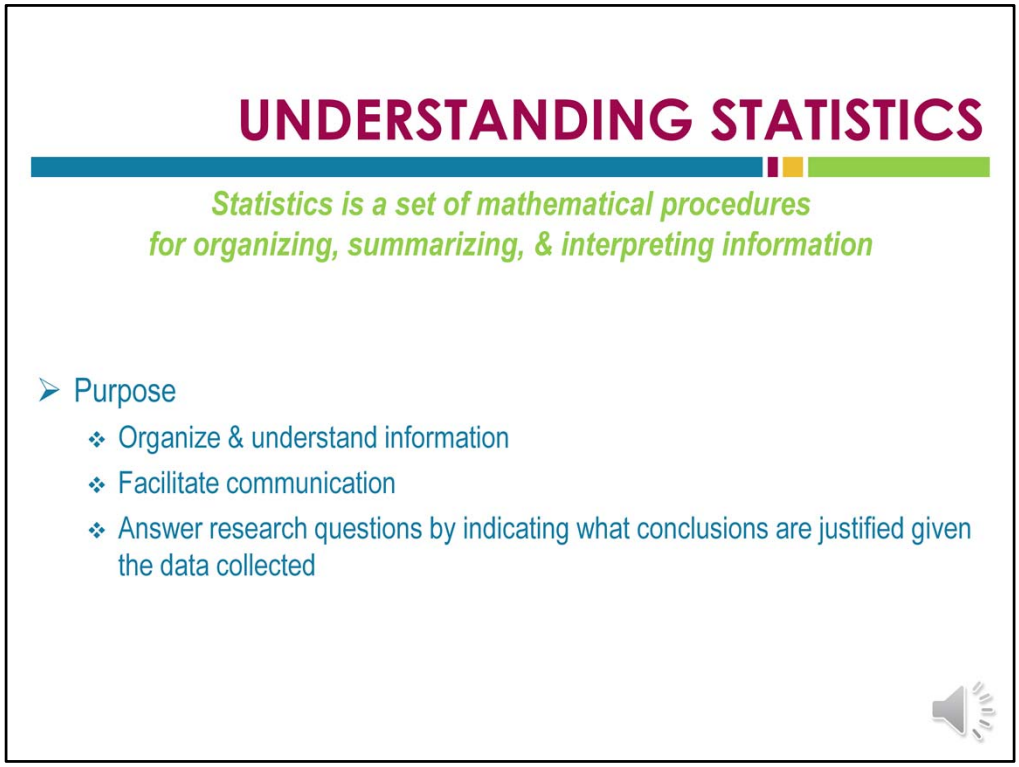


TRANSCRIPT:

This is Dr. Chumney with a brief overview of the basic concepts of statistics.

TRANSCRIPT:

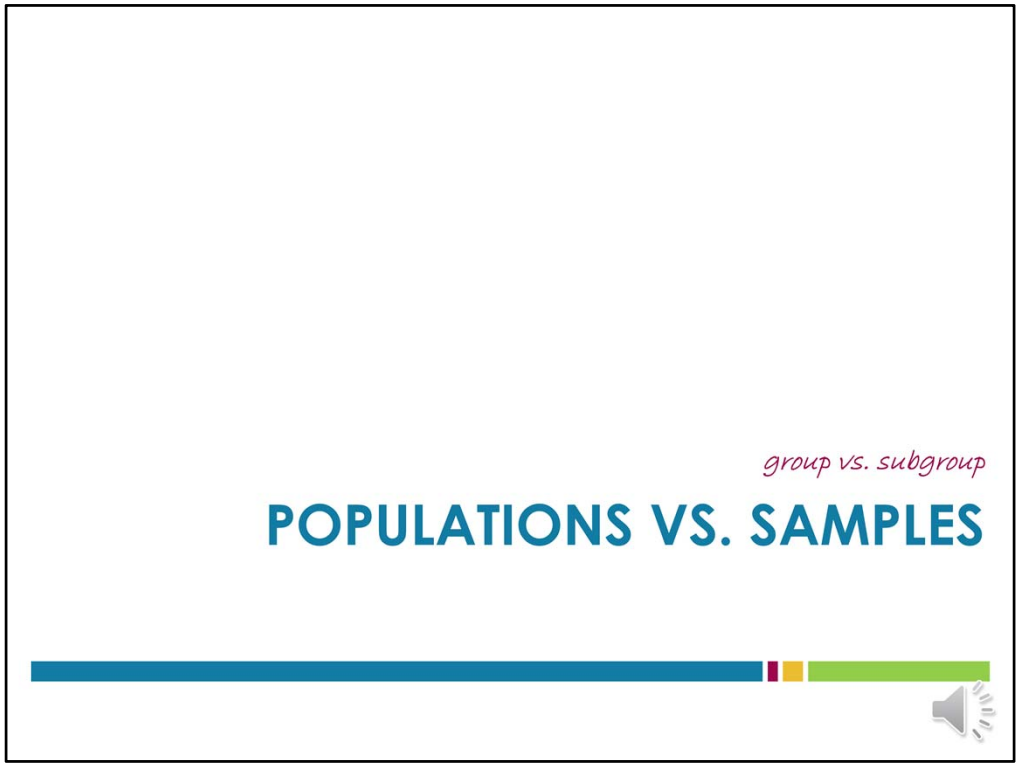
First, what are statistics?



TRANSCRIPT:

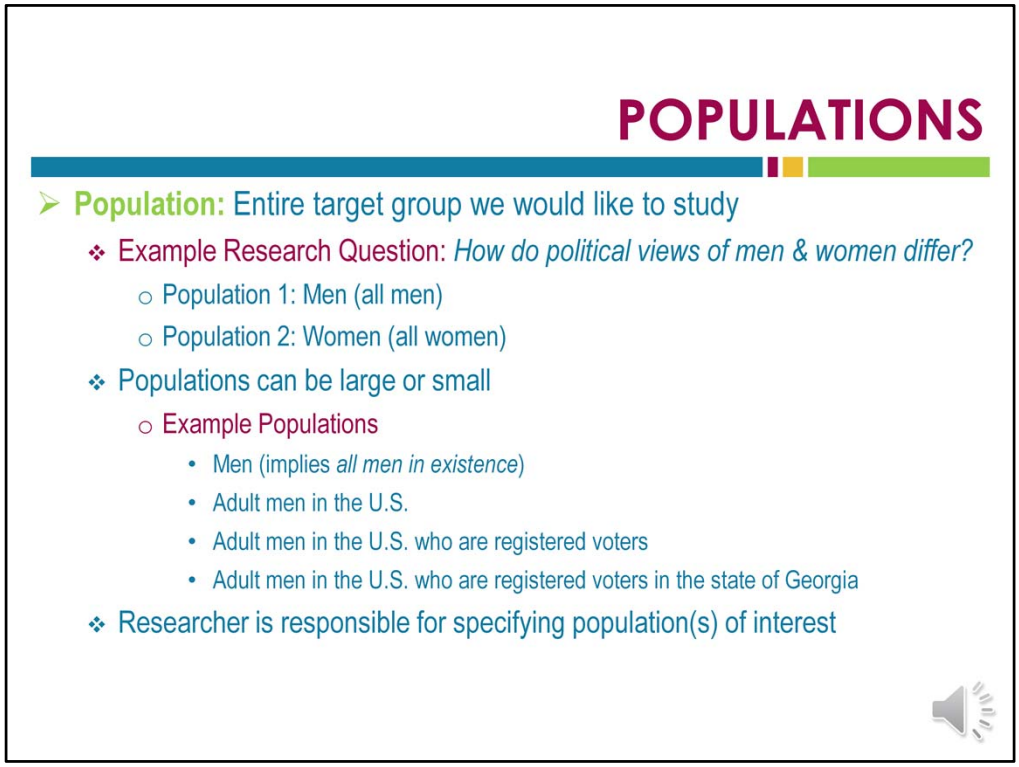
Statistics are a set of mathematical procedures used to organize, summarize, and interpret information.

Statistics are the primary form of analyzing and interpreting data collected through quantitative research studies. In this context, statistics help us to organize and understand the information contained within the data, communicate our results and characteristics of our data to others, and answer the research questions that drive the research or motivated us to collect the data in the first place.



TRANSCRIPT:

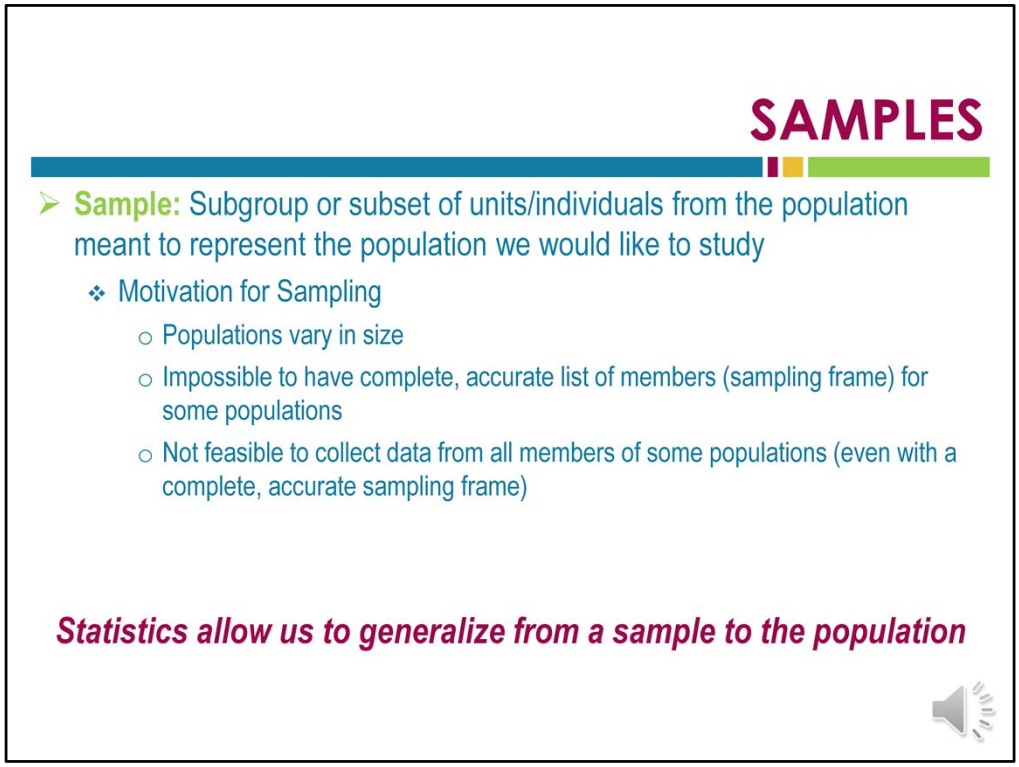
When we conduct research, we typically have to choose a sample from which to collect data, so it is important to understand how populations and samples differ.



TRANSCRIPT:

A population is the entire group that we are interested in, and it includes all of the members of whatever target group we specify. For example, if we say we are interested in studying the political views of men, we are implying that we are interested in the views of all men. Similarly, if we want to study how men and women differ in political views, then we are interested in both the entire population of men and the entire population of women.

A population can be quite large or very small. The key to remember is that the population includes all the members of the group. It is the responsibility of the researcher to specify who their target population is, and to do so in such a way as to let the reader know how general or specific that target group really is.

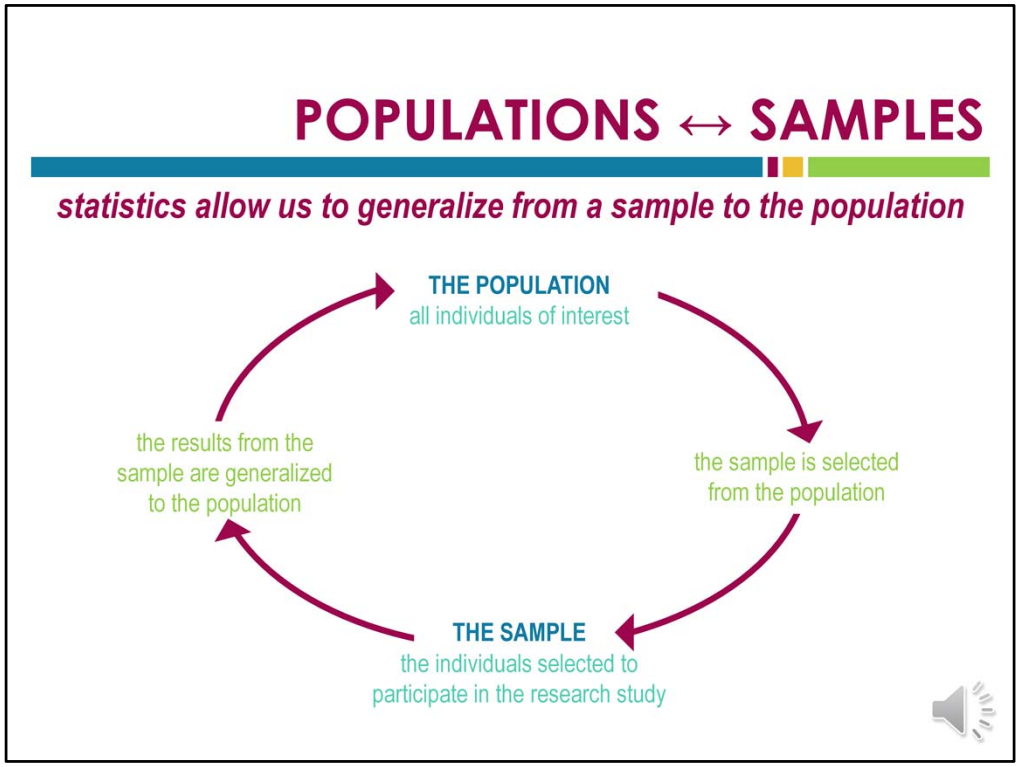


TRANSCRIPT:

Because populations can be so large, we typically rely on samples when we conduct research. A sample is a subgroup of units or individuals from the population that we want to study. This smaller group is intended to represent the larger group from which it is drawn, and this smaller group is called a sample.

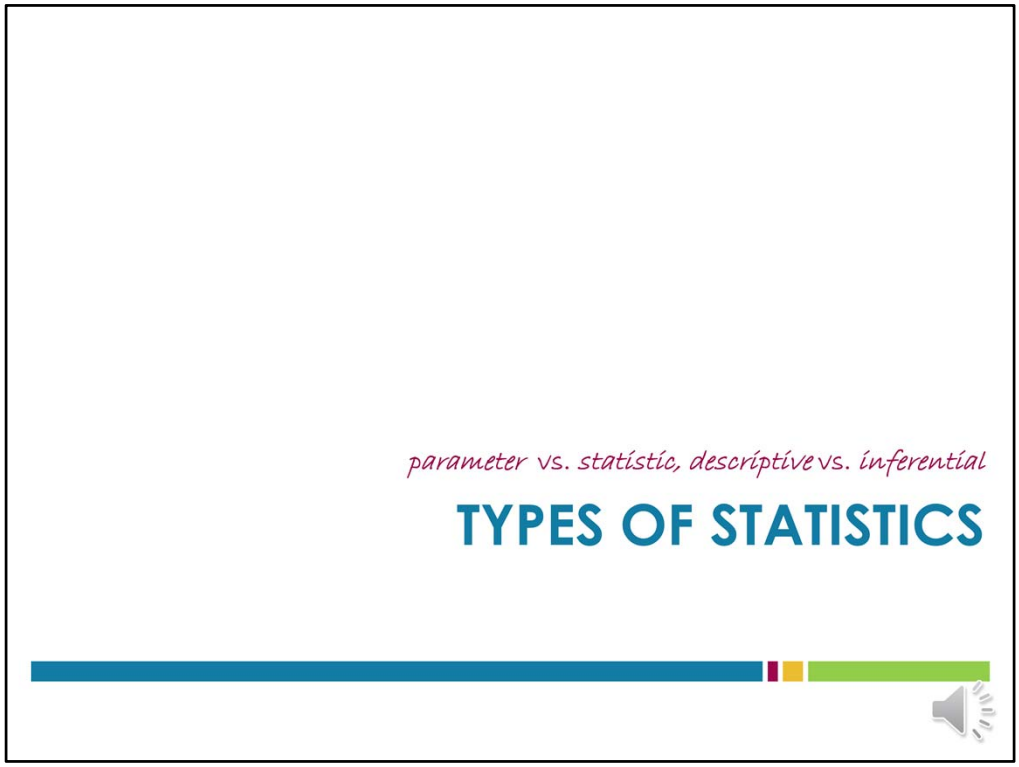
Researchers have various reasons for using samples, but they generally boil down to not having resources to collect data from an entire population.

Using statistics, a researcher can collect data from a sample, analyze those data, and then generalize those findings to the population from which the sample was selected.



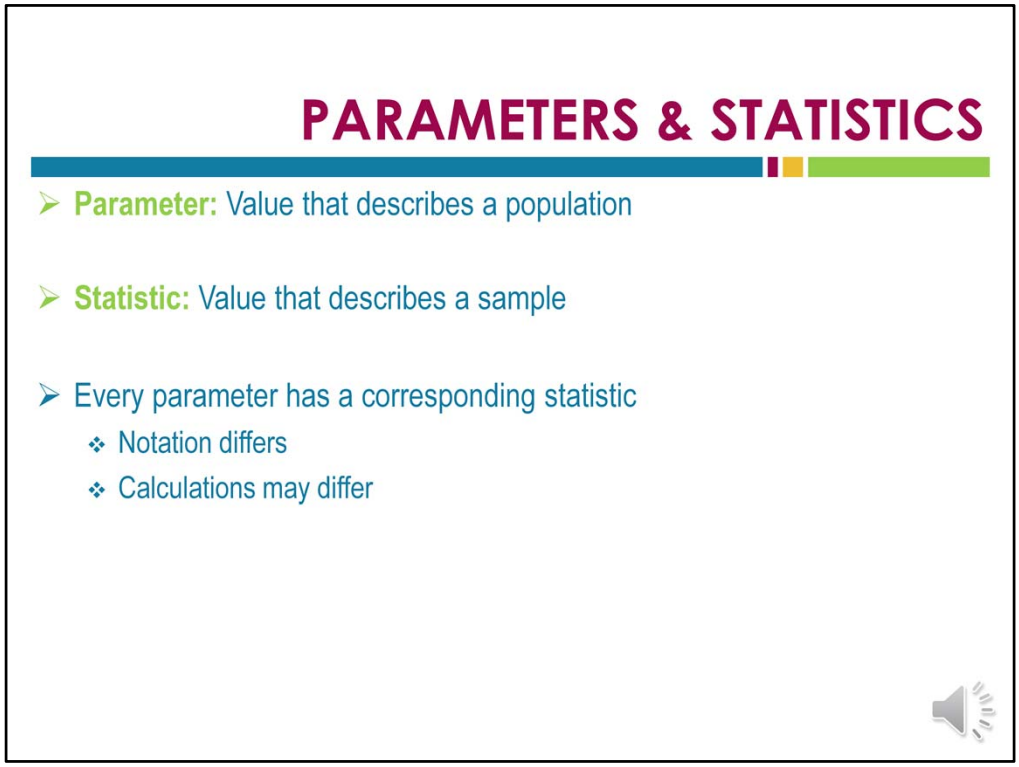
TRANSCRIPT:

This figure illustrates the relationship between populations and samples. We begin by selecting a population of interest, then we sample from that population, collect data from the sample, analyze the data, and generalize those results back to the original population.



TRANSCRIPT:

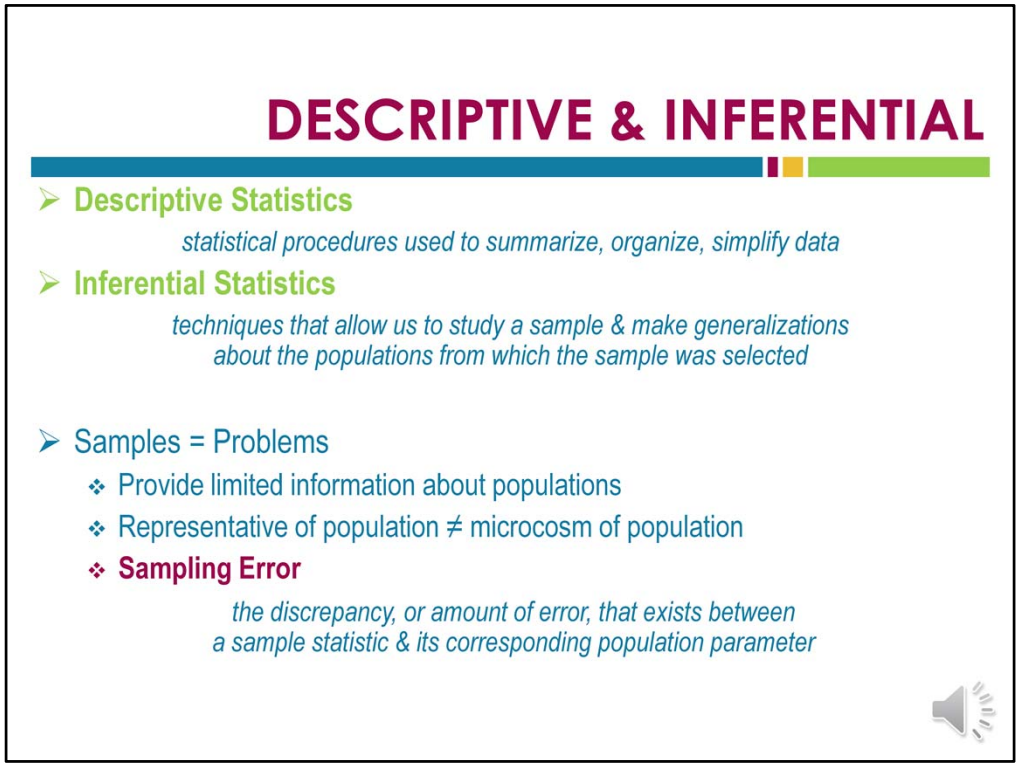
There several different categories of statistics. The most important distinctions are between parameters and statistics, and between descriptive and inferential statistics.



TRANSCRIPT:

A parameter is a calculated value that describes a population. A statistic is a calculated value that describes a sample. For every parameter (or, for every value that might be used to describe a population), a corresponding statistic can be calculated for the sample. For example, we could calculate a mean for a population if we had data from everyone in the population, and we can calculate a mean for a sample taken from that population.

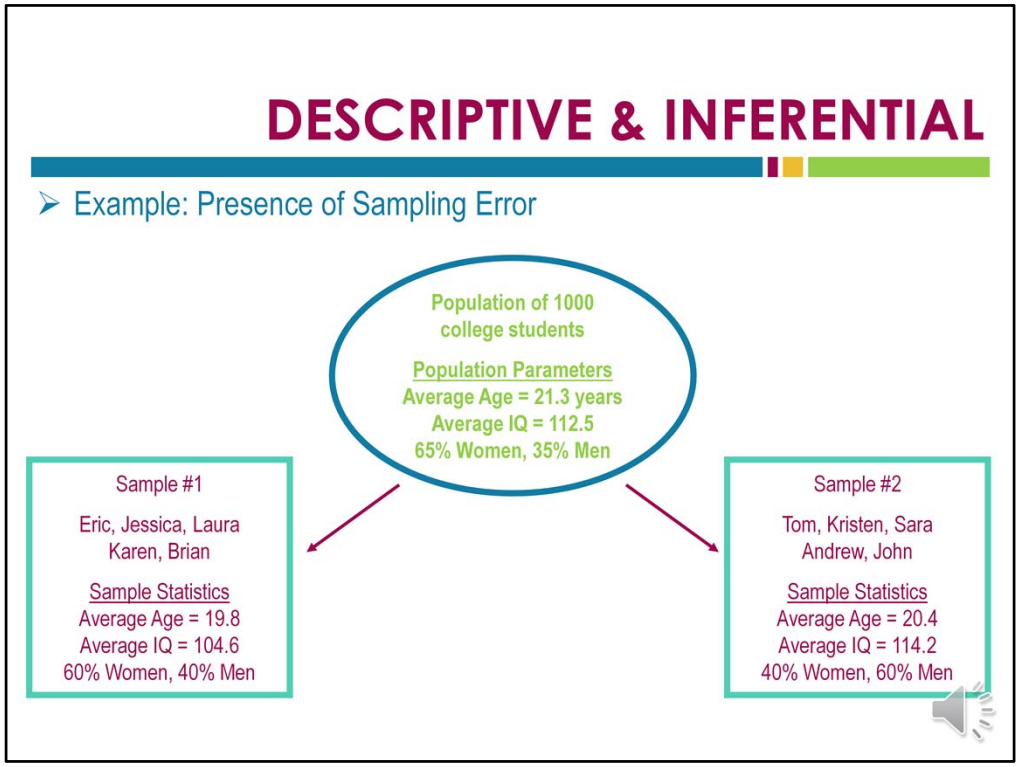
While corresponding parameters and statistics have the same meanings practically, they are technically different because they are calculated differently and different notations are used to refer to them.



TRANSCRIPT:

Another important distinction that should be made is between descriptive and inferential statistics. Descriptive statistics are procedures used to summarize, organize, or otherwise simplify data. Inferential statistics, on the other hand, are procedures that allow us to analyze data from a sample and make generalizations or draw conclusions about the population from which the sample was drawn.

There is, however, one important disadvantage when it comes to using a sample. Even when every attempt is made to draw a random sample so that it will represent the population, odds are that it does not constitute a perfect microcosm of the population. Because of this, we have to accept that every time we analyze data from a sample instead of a population, we are introducing sampling error. Sampling error is the discrepancy or amount of error that exists between the statistic that is calculated for a sample and the value of the corresponding parameter for the population from which the sample was drawn.



TRANSCRIPT:

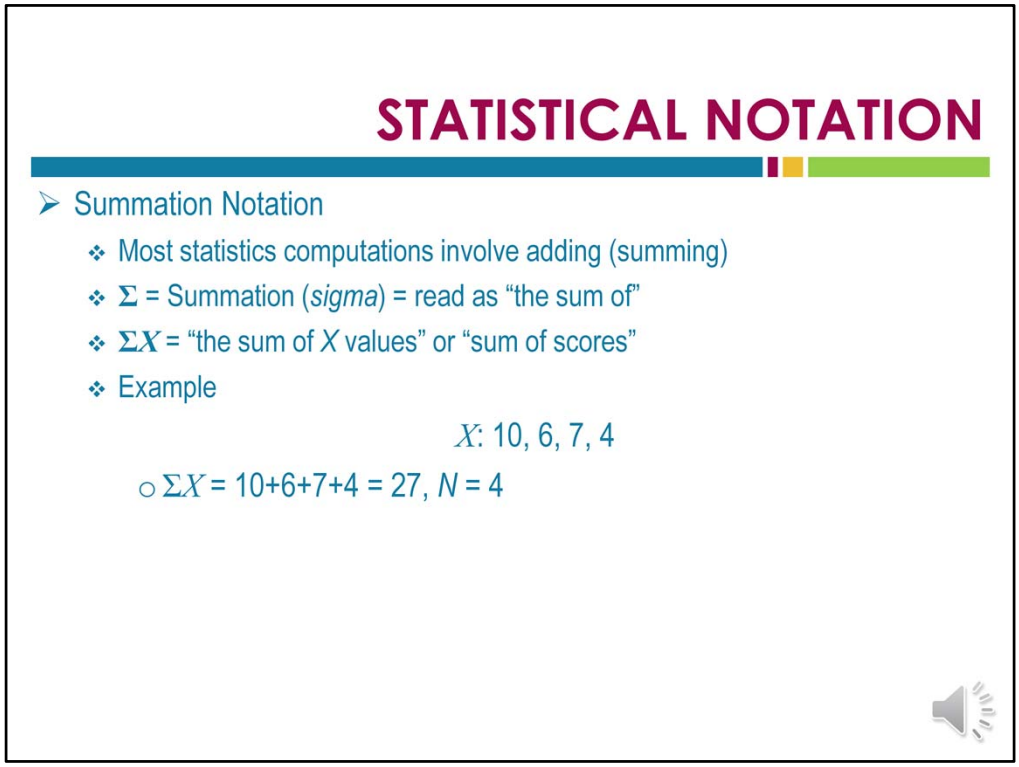
For example. Suppose we are interested in studying student satisfaction at a small university with only 1000 students. Together, those 1000 students make up the target group in which we are interested. If we do not have funding or other resources to collect data from all 1000 students, we have to sample smaller groups from that population. Typically, we would draw samples with more than 5 individuals, but to keep this example simple, let's say we draw two separate samples of 5 students from this population.

If we were to collect data on the entire population, we might find that the population has an average age of 21.3 years, and average IQ of 112.5, and is composed of 65% women and 35% men. These are all parameters because they are values that represent the population.

Let's take a look as Sample 1. Suppose these 5 students were randomly selected from the population, and we find they have an average age of 19.8 years, and average IQ of 104.6, and the sample is composed of 60% women. The sample is not a perfect representation of the population, and we know that because the values calculated to represent the sample do not exactly match the values calculated to represent the population. These differences or discrepancies between the population values and the sample statistics are sampling error.

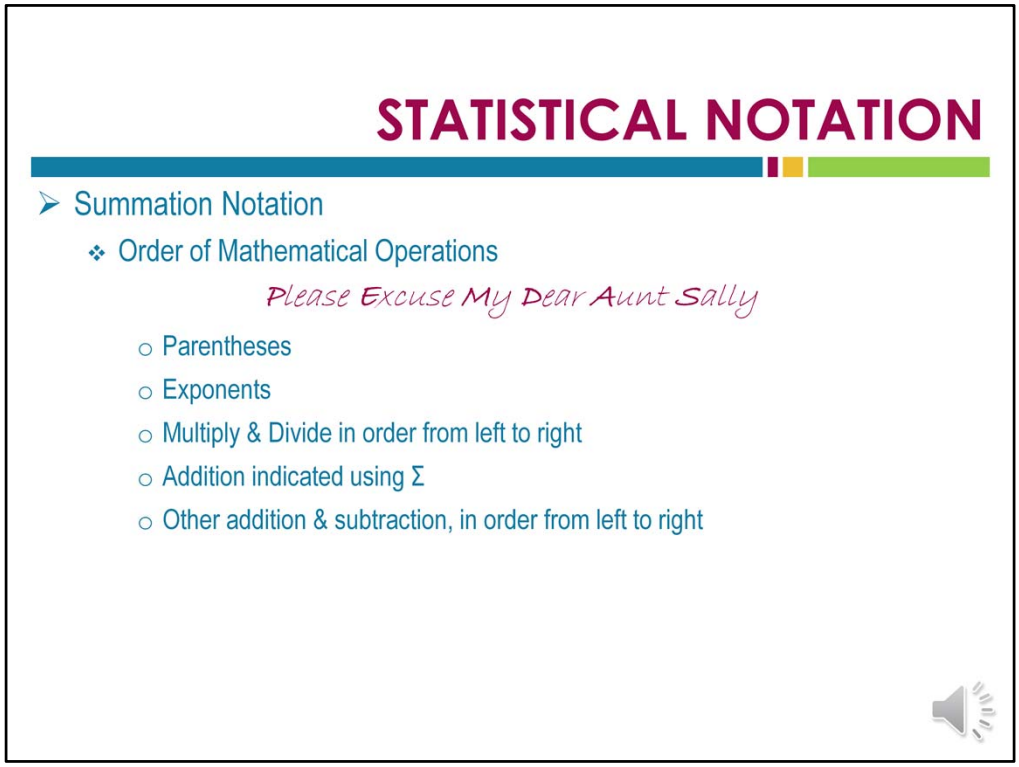
TRANSCRIPT:

There are some common symbols used in statistics, and you should be familiar with them as well.



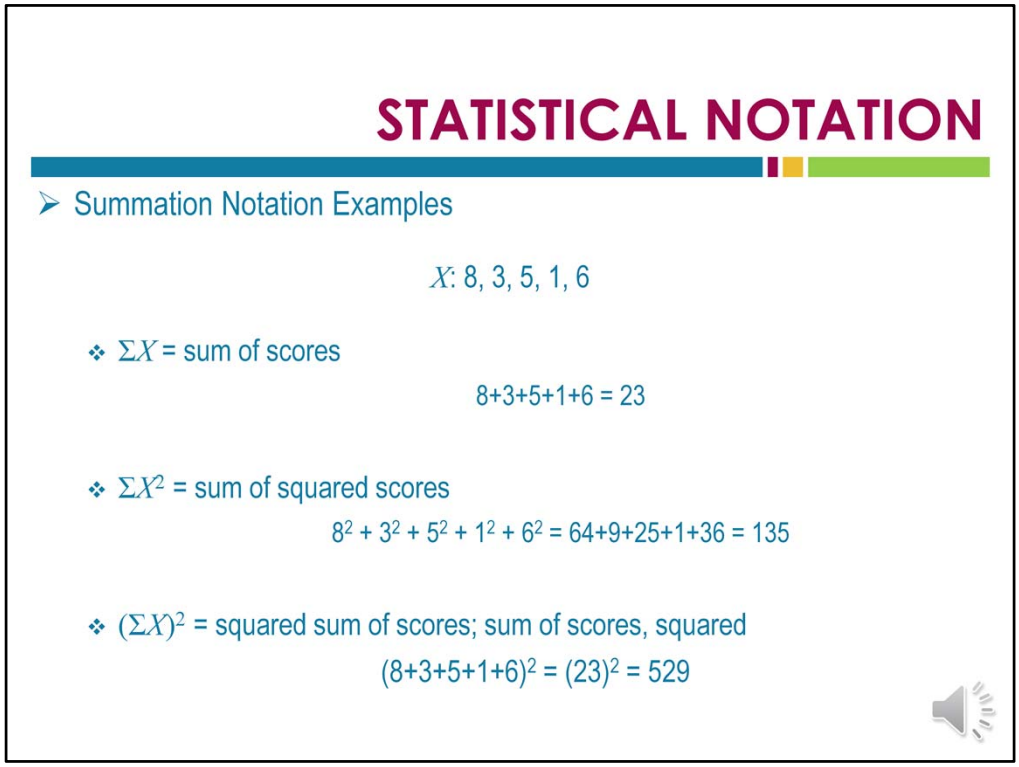
TRANSCRIPT:

In statistical notation, the symbol that indicates summation or the adding of values is capital-sigma, which is read as "the sum of." It follows, then, that sigma-X is read as "the sum of X values" or the "sum of scores" because X represents values or scores on a particular variable



TRANSCRIPT:

Some statistics formulas include multiple mathematical steps. The order in which these steps should be completed is the same as for any other mathematical expression.



TRANSCRIPT:

Here are some more examples of summation notation.

STATISTICAL NOTATION

- Summation Notation Examples

X: 8, 3, 5, 1, 6

- $\Sigma(X-1)$ = sum of X - 1 values

  (8-1) + (3-1) + (5-1) + (1-1) + (6-1) = 7+2+4+0 +5 = 18

- $\Sigma(X-1)^2$ = sum of squared X - 1 values

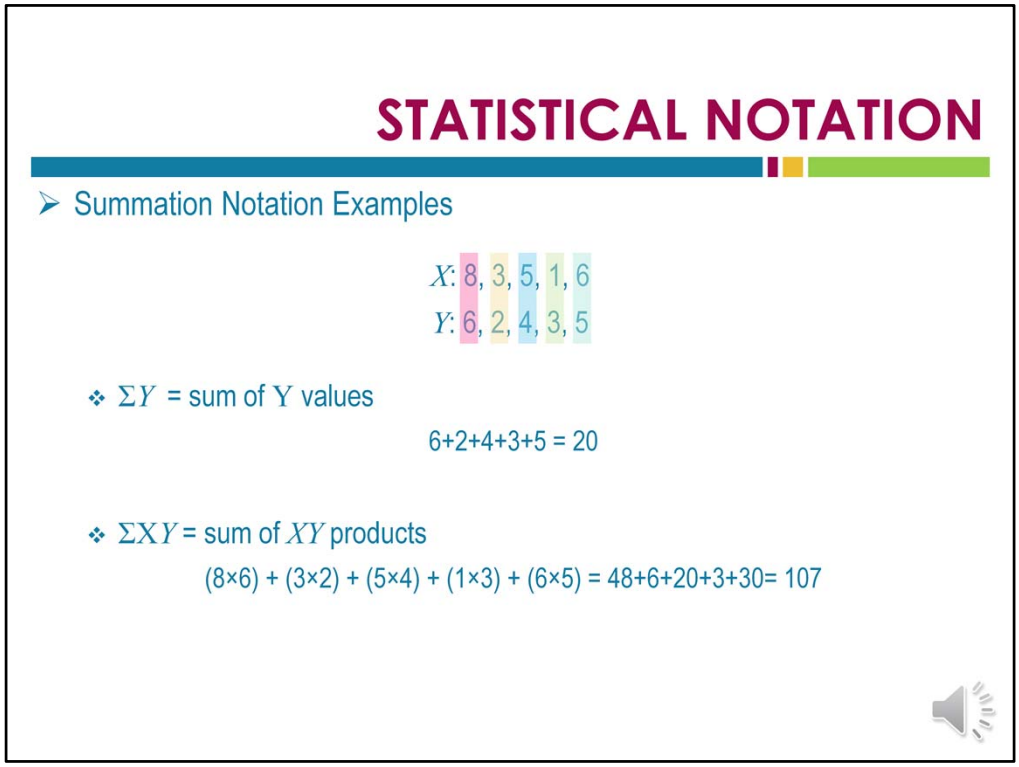
  $(8-1)^2 + (3-1)^2 + (5-1)^2 + (1-1)^2 + (6-1)^2 =$

  $7^2 + 2^2 + 4^2 + 0^2 + 5^2 =$

  49+4+16+0+25 = 94

TRANSCRIPT:

And a few more examples…



TRANSCRIPT:

And a few more examples. Notice that summation notation – and all statistical notation, really – can be written to indicate that the calculated value should include information from more than one variable.