Introduction to Statistics

Statistics are used everyday to help us look at information and make decisions. For example, medical studies use statistics to determine possible causes of a disease. Engineers use statistics to predict the life of a part or device. Manufacturers use statistics to monitor the production of goods and materials.

Goals

1. To recognize the difference between a population and a sample.
2. To understand the difference between a parameter and a statistic.
3. To recognize when descriptive or inferential statistics should be used.

Important Terms and Definitions

Statistics (singular) is the science concerned with the collection, organization, and analysis of information.

A **unit** is a single entity (person or object) whose characteristics are of interest

A **population of units** is the complete collection of units about which information is sought.

A **population** is a set of all measurements corresponding to each unit in the entire collection of units about which information is sought.

A **sample** is a subset of measurements selected from the population of interest.

A **representative sample** exhibits characteristics typical of those possessed by the target population.

A **random sample** ensures that every subset of fixed size in the population has the same chance of being included in the sample.

A **variable** is a characteristic of an individual or object that may vary for different observations.

A **quantitative variable** measures a variable on a numerical scale.

A **qualitative variable** categorizes the values of the variable.

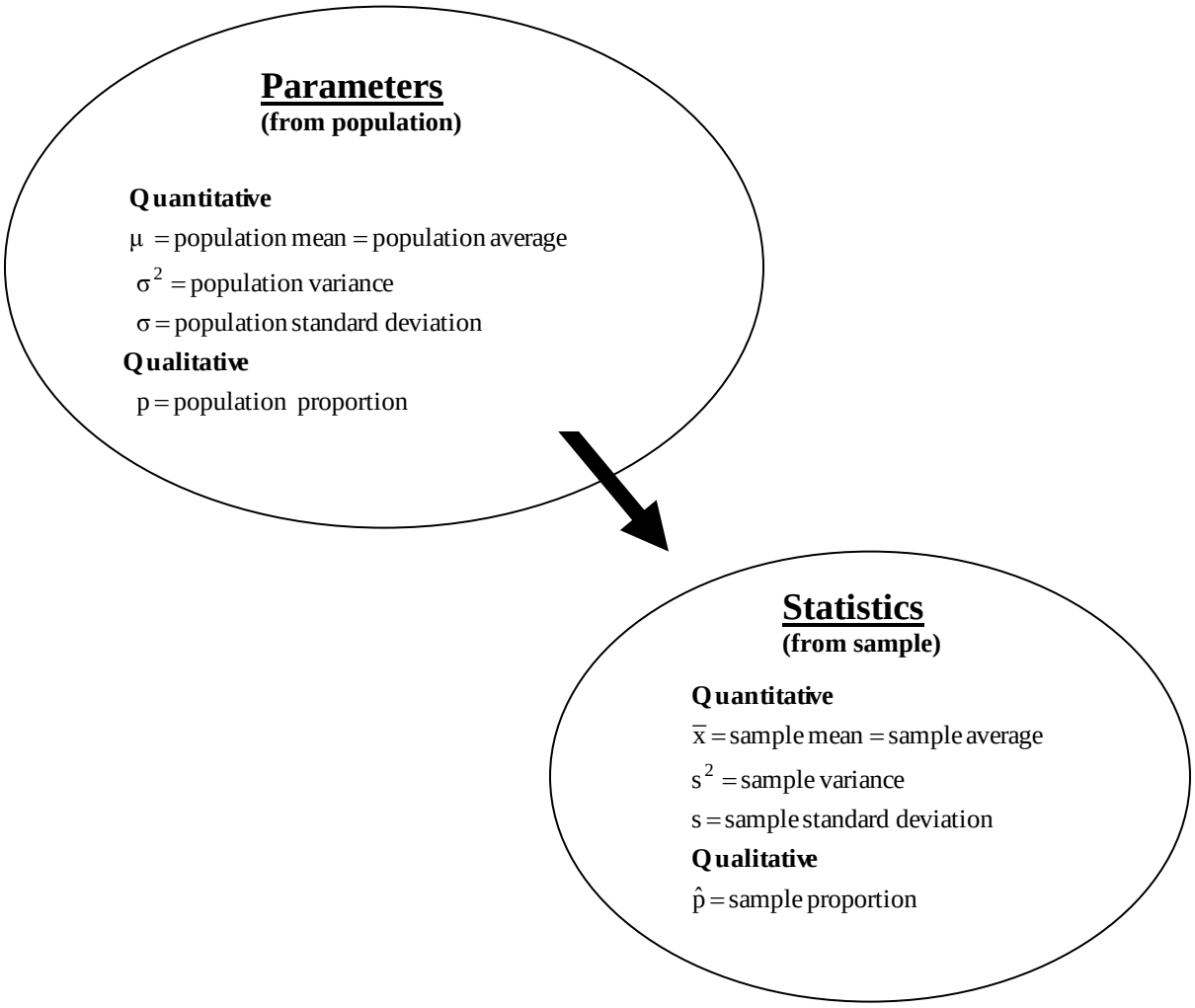
Statistics (plural) are numerical characteristics calculated from a sample.

A **parameter** is a numerical characteristic of a population. Usually a parameter is an unknown quantity.

Descriptive Statistics is concerned with organizing and describing sample information. (Descriptive statistics describe how things are.)

Inferential Statistics is concerned with making inferences (predictions or decisions) about a population based on the information contained in a sample. (Inferential statistics describes how things probably will be.) A **measure of reliability** is a statement about the degree of uncertainty associated with a statistical inference.

Some Frequently used Parameters and Statistics



n = sample size = the number of observations in a sample

The mean (average) is often used to determine the central location of the observations.
The variance and standard deviation are used to determine the variation of the observations.
The proportion is used to determine the proportion of observations, which have a particular qualitative value.

Use statistics to make inferences about parameters.

$\bar{x}$ estimates μ

$\hat{p}$ estimates p

s^2 estimates σ^2

s estimates σ

Example 1: For a class project, you are asked to determine the average credit hours taken by Clemson students this semester.

Identify the unit.

Identify the population of units.

Identify the population.

Suppose 100 Clemson students were randomly selected and polled.
Identify the sample.

Identify the variable.

Is the variable quantitative or qualitative?

Identify the symbol for the parameter of interest.

Identify the symbol for the statistic.

Example 2: A cereal company claims that the average amount of raisins in its boxes of raisin bran is two scoops. Five boxes were randomly selected and taken off of the production line. An analysis revealed an average of 1.9 scoops of raisins per box.

Identify the unit.

Identify the population of units.

Identify the population.

Identify the sample.

Identify the symbol for the parameter of interest.

Identify the symbol for the statistic.

Identify the variable.

Is the variable quantitative or qualitative?

Example 3: A survey was conducted to determine the proportion of registered voters in the U.S. that favor greater restrictions on purchasing handguns. One thousand registered voters in the U.S. were randomly selected and polled. Of those polled, 550 were in favor of greater restrictions on purchasing handguns.

Identify the unit.

Identify the population of units.

Identify the population.

Identify the sample.

What is the sample size?

Identify the symbol for the parameter of interest.

Identify the symbol for the statistic.

Identify the variable.

Is the variable quantitative or qualitative?

Example 4: For a class project you are asked to determine the average credit hours taken by Clemson students. One hundred CU students were randomly selected, and the average credit hours were calculated. Determine if the following statements are concerned with descriptive or inferential statistics.

The average semester credit hours taken by the 100 Clemson students who were polled were 16 hours.

The average semester credit hours taken by Clemson students are more than 15 hours.