

### *Population Versus sample*

We will encounter the terms *population* and *sample* on almost every lecture of this course. Consequently, understanding the meaning of each of these two terms and the difference between them is crucial.

Suppose a statistician is interested in knowing

1. The percentage of all voters in a city who will vote for a particular candidate in an election
2. The 2020 gross sales of all companies in New York City
3. The prices of all houses in California

In these examples, the statistician is interested in all voters, all companies, and all houses.

Each of these groups is called the population for the respective example. In statistics, a population does not necessarily mean a collection of people. It can, in fact, be a collection of people or of any kind of item such as houses, books, television sets, or cars. The population of interest is usually called the target population.

***Definition***

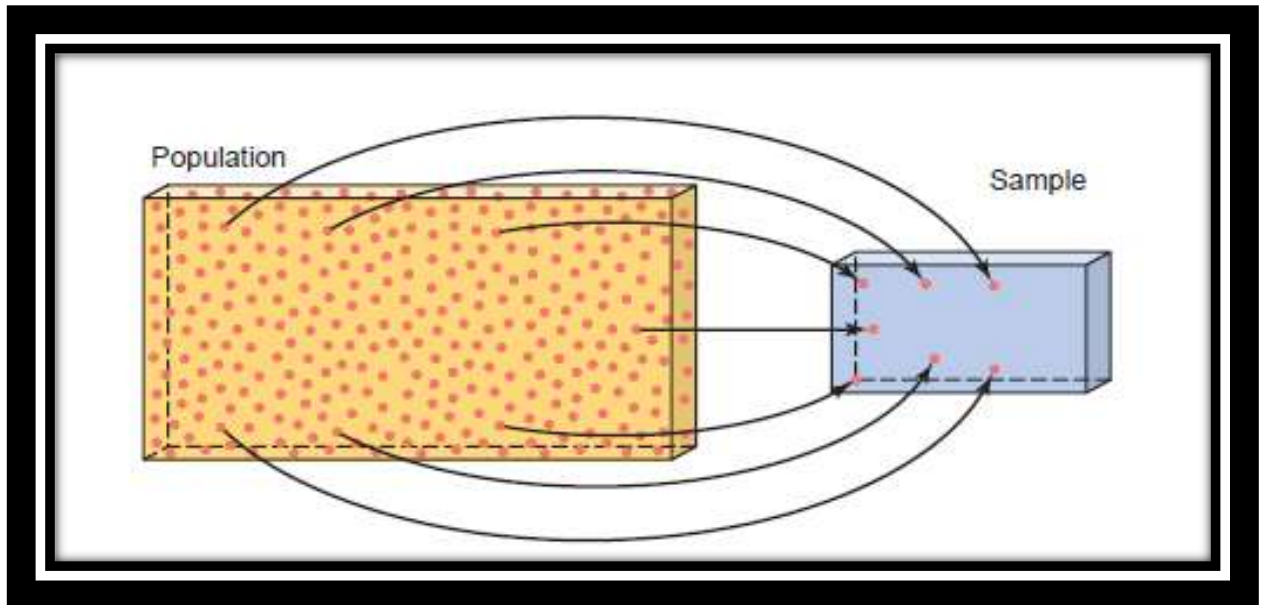
**Population or Target Population** A *population* consists of all elements—individuals, items, or objects—whose characteristics are being studied. The population that is being studied is also called the *target population*.

Most of the time, decisions are made based on portions of populations. For example, the election polls conducted in the United States to estimate the percentages of voters who favor various candidates in any presidential election are based on only a few hundred or a few thousand voters selected from across the country. In this case, the population consists of all registered voters in the United States. The sample is made up of a few hundred or few thousand voters who are included in an opinion poll. Thus, the collection of a few elements selected from a population is called a **sample**.

***Definition***

**Sample** A portion of the population selected for study is referred to as a *sample*.

Figure below illustrates the selection of a sample from a population.



The collection of information from the elements of a population or a sample is called a **survey**. A survey that includes every element of the target population is called a **census**. Often the target population is very large. Hence, in practice, a census is rarely taken because it is expensive and time-consuming. In many cases, it is even impossible to identify each element of the target population. Usually, to conduct a survey, we select a sample and collect the required information from the elements included in that sample. We then make decisions based on this sample information. Such a survey conducted on a sample is called a **sample survey**.

As an example, if we collect information on the 2020 incomes of all families in Khartoum, it will be referred to as a census. On the other

hand, if we collect information on the 2009 incomes of 50 families from Khartoum, it will be called a sample survey.

### ***Definition***

**Census and Sample Survey** A survey that includes every member of the population is called a *census*. The technique of collecting information from a portion of the population is called a *sample survey*.

The purpose of conducting a sample survey is to make decisions about the corresponding population. It is important that the results obtained from a sample survey closely match the results that we would obtain by conducting a census. Otherwise, any decision based on a sample survey will not apply to the corresponding population. As an example, to find the average income of families living in Khartoum City by conducting a sample survey, the sample must contain families who belong to different income groups in almost the same proportion as they exist in the population. Such a sample is called a representative sample. Inferences derived from a representative sample will be more reliable.

***Definition***

**Representative Sample** A sample that represents the characteristics of the population as closely as possible is called a *representative sample*.

A sample may be random or nonrandom. In a **random sample**, each element of the population has a chance of being included in the sample. However, in a nonrandom sample this may not be the case.

***Definition***

**Random Sample** A sample drawn in such a way that each element of the population has a chance of being selected is called a *random sample*. If all samples of the same size selected from a population have the same chance of being selected, we call it **simple random sampling**. Such a sample is called a **simple random sample**.

One way to select a random sample is by lottery or draw. For example, if we are to select 5 students from a class of 50, we write each of the 50 names on a separate piece of paper. Then we place all 50 slips in a box and mix them thoroughly. Finally, we randomly draw 5 slips from the box. The 5 names drawn give a random sample. On the other hand, if we arrange all 50 names alphabetically and then select the first 5 names on the list, it is a nonrandom sample because the students listed 6th to 50th have no chance of being included in the sample.

A sample may be selected with or without replacement. In sampling with replacement, each time we select an element from the population, we put it back in the population before we select the next element. Thus, in sampling with replacement, the population contains the same number of items each time a selection is made. As a result, we may select the same item more than once in such a sample. Consider a box that contains 25 marbles of different colors.

Suppose we draw a marble, record its color, and put it back in the box before drawing the next marble. Every time we draw a marble from this box, the box contains 25 marbles. This is an example of sampling with replacement. The experiment of rolling a die many times is another example of sampling with replacement because every roll has the same six possible outcomes.

Sampling without replacement occurs when the selected element is not replaced in the population. In this case, each time we select an item, the size of the population is reduced by one element. Thus, we cannot select the same item more than once in this type of sampling.

Most of the time, samples taken in statistics are without replacement. Consider an opinion poll based on a certain number of voters selected from the population of all eligible voters. In this case, the same voter is not selected more than once. Therefore, this is an example of sampling without replacement.