

The Big Picture

CO-1: Describe the roles biostatistics serves in the discipline of public health.

Throughout the course, we will add to our understanding of the definitions, concepts, and processes which are introduced here. You are not expected to gain a full understanding of this process until much later in the course!

To really understand how this process works, we need to put it in a context. We will do that by introducing one of the central ideas of this course, the **Big Picture of Statistics**.

We will introduce the Big Picture by building it gradually and explaining each component.

At the end of the introductory explanation, once you have the full Big Picture in front of you, we will show it again using a concrete example.

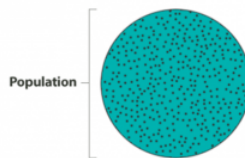
Learning Objectives

LO 1.3: Identify and differentiate between the components of the Big Picture of Statistics

Video

[The Big Picture of Statistics](#) (4:59)

The process of statistics starts when we identify what group we want to study or learn something about. We call this group the **population**.



Note that the word “population” here (and in the entire course) is not just used to refer to people; it is used in the more broad statistical sense, where population can refer not only to people, but also to animals, things etc. For example, we might be interested in:

- the opinions of the population of U.S. adults about the death penalty; or
- how the population of mice react to a certain chemical; or
- the average price of the population of all one-bedroom apartments in a certain city.

Note

The **population**, then, is the entire group that is the target of our interest.

In most cases, the population is so large that as much as we might want to, there is absolutely no way that we can study all of it (imagine trying to get the opinions of all U.S. adults about the death penalty...).

A more practical approach would be to examine and collect data only from a sub-group of the population, which we call a **sample**. We call this first component, which involves choosing a sample and collecting data from it, **Producing Data**.



Note

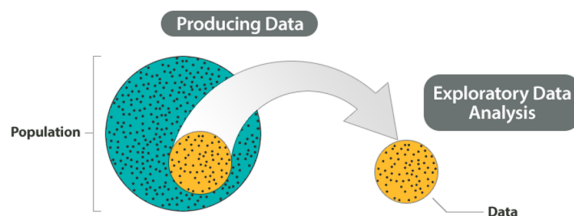
A **sample** is a subset of the population from which we collect data.

It should be noted that since, for practical reasons, we need to compromise and examine only a sub-group of the population rather than the whole population, we should make an effort to choose a sample in such a way that it will represent the population well.

For example, if we choose a sample from the population of U.S. adults, and ask their opinions about a particular federal health care program, we do not want our sample to consist of only Republicans or only Democrats.

Once the data have been collected, what we have is a long list of answers to questions, or numbers, and in order to explore and make sense of the data, we need to summarize that list in a meaningful way.

This second component, which consists of summarizing the collected data, is called **Exploratory Data Analysis** or **Descriptive Statistics**.



Now we've obtained the sample results and summarized them, but we are not done. Remember that our goal is to study the population, so what we want is to be able to draw conclusions about the population based on the sample results.

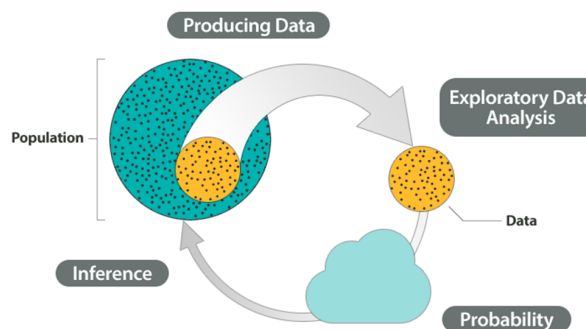
Before we can do so, we need to look at how the sample we're using may differ from the population as a whole, so that we can factor that into our analysis. To examine this difference, we use **Probability** which is the third component in the big picture.

The third component in the Big Picture of Statistics, **probability** is in essence the “machinery” that allows us to draw conclusions about the population based on the data collected in the sample.



Finally, we can use what we've discovered about our sample to draw conclusions about our population.

We call this final component in the process **Inference**.



This is the **Big Picture of Statistics**.

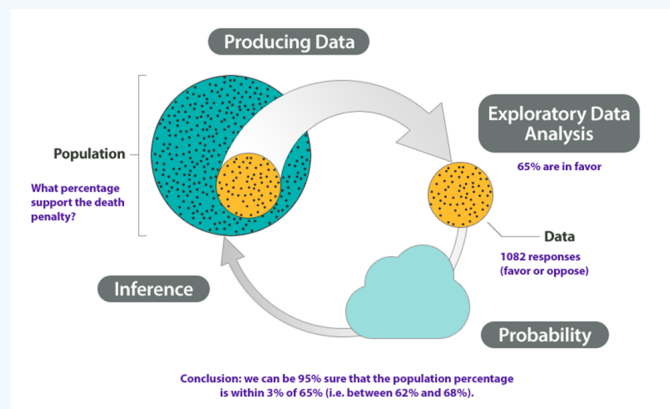
✓ EXAMPLE: Polling Public Opinion

At the end of April 2005, a poll was conducted (by ABC News and the Washington Post), for the purpose of learning the opinions of U.S. adults about the death penalty.

1. Producing Data: A (representative) sample of 1,082 U.S. adults was chosen, and each adult was asked whether he or she favored or opposed the death penalty.

2. Exploratory Data Analysis (EDA): The collected data were summarized, and it was found that 65% of the sampled adults favor the death penalty for persons convicted of murder.

3 and 4. Probability and Inference: Based on the sample result (of 65% favoring the death penalty) and our knowledge of probability, it was concluded (with 95% confidence) that the percentage of those who favor the death penalty in the population is within 3% of what was obtained in the sample (i.e., between 62% and 68%). The following figure summarizes the example:



Course Structure

The structure of this entire course is based on the big picture.

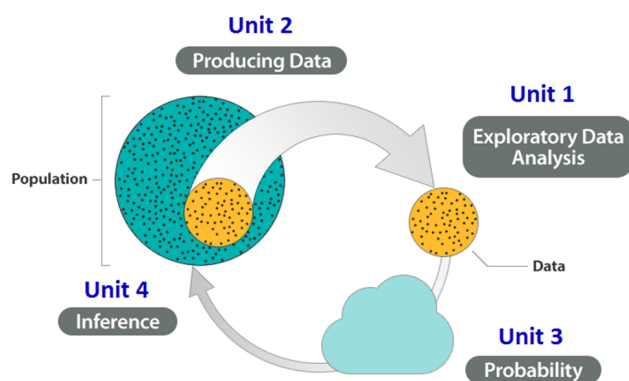
The course will have 4 units; one for each of the components in the big picture.

As the figure below shows, even though it is second in the process of statistics, we will start this course with exploratory data analysis (EDA), continue to discuss producing data, then go on to probability, so that at the end we will be able to discuss inference.

The main reasons we begin with EDA is that we need to understand enough about what we want to do with our data before we can discuss the issues related to how to collect it!!

This also allows us to introduce many important concepts early in the course so that you will have ample time to master them before we return to inference at the end of the course.

The following figure summarizes the structure of the course.



As you will see, the Big Picture is the basis upon which the entire course is built, both conceptually and structurally. We will refer to it often, and having it in mind will help you as you go through the course.

The Big Picture is shared under a [CC BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/) license and was authored, remixed, and/or curated by LibreTexts.