

## 1.1: What is Statistics?

You are exposed to statistics regularly. If you are a sports fan, then you have the statistics for your favorite player. If you are interested in politics, then you look at the polls to see how people feel about certain issues or candidates. If you are an environmentalist, then you research arsenic levels in the water of a town or analyze the global temperatures. If you are in the business profession, then you may track the monthly sales of a store or use quality control processes to monitor the number of defective parts manufactured. If you are in the health profession, then you may look at how successful a procedure is or the percentage of people infected with a disease. There are many other examples from other areas. To understand how to collect data and analyze it, you need to understand what the field of statistics is and the basic definitions.

### Definition 1.1.1

**Statistics** is the study of how to collect, organize, analyze, and interpret data collected from a group.

There are two branches of statistics. One is called descriptive statistics, which is where you collect and organize data. The other is called inferential statistics, which is where you analyze and interpret data. First you need to look at descriptive statistics since you will use the descriptive statistics when making inferences.

To understand how to create descriptive statistics and then conduct inferences, there are a few definitions that you need to look at. Note, many of the words that are defined have common definitions that are used in non-statistical terminology. In statistics, some have slightly different definitions. It is important that you notice the difference and utilize the statistical definitions.

The first thing to decide in a statistical study is whom you want to measure and what you want to measure. You always want to make sure that you can answer the question of whom you measured and what you measured. The who is known as the individual and the what is the variable.

### Definition 1.1.2

**Individual** – a person or object that you are interested in finding out information about.

### Definition 1.1.3

**Variable** – the measurement or observation of the individual.

If you put the individual and the variable into one statement, then you obtain a population.

### Definition 1.1.4

**Population** – set of all values of the variable for the entire group of individuals.

Notice, the population answers who you want to measure and what you want to measure. Make sure that your population always answers both of these questions. If it doesn't, then you haven't given someone who is reading your study the entire picture. As an example, if you just say that you are going to collect data from the senators in the U.S. Congress, you haven't told your reader what you are going to collect. Do you want to know their income, their highest degree earned, their voting record, their age, their political party, their gender, their marital status, or how they feel about a particular issue? Without telling what you want to measure, your reader has no idea what your study is actually about.

Sometimes the population is very easy to collect. Such as if you are interested in finding the average age of all of the current senators in the U.S. Congress, there are only 100 senators. This wouldn't be hard to find. However, if instead you were interested in knowing the average age that a senator in the U.S. Congress first took office for all senators that ever served in the U.S. Congress, then this would be a bit more work. It is still doable, but it would take a bit of time to collect. But what if you are interested in finding the average diameter of breast height of all of the Ponderosa Pine trees in the Coconino National Forest? This would be impossible to actually collect. What do you do in these cases? Instead of collecting the entire population, you take a smaller group of the population, kind of a snap shot of the population. This smaller group is called a sample.

**Definition 1.1.5**

**Sample** – a subset from the population. It looks just like the population, but contains less data

How you collect your sample can determine how accurate the results of your study are. There are many ways to collect samples. Some of them create better samples than others. No sampling method is perfect, but some are better than others. Sampling techniques will be discussed later. For now, realize that every time you take a sample you will find different data values. The sample is a snapshot of the population, and there is more information than is in the picture. The idea is to try to collect a sample that gives you an accurate picture, but you will never know for sure if your picture is the correct picture. Unlike previous mathematics classes where there was always one right answer, in statistics there can be many answers, and you don't know which are right.

Once you have your data, either from a population or a sample, you need to know how you want to summarize the data. As an example, suppose you are interested in finding the proportion of people who like a candidate, the average height a plant grows to using a new fertilizer, or the variability of the test scores. Understanding how you want to summarize the data helps to determine the type of data you want to collect. Since the population is what we are interested in, then you want to calculate a number from the population. This is known as a parameter. As mentioned already, you can't really collect the entire population. Even though this is the number you are interested in, you can't really calculate it. Instead you use the number calculated from the sample, called a statistic, to estimate the parameter. Since no sample is exactly the same, the statistic values are going to be different from sample to sample. They estimate the value of the parameter, but again, you do not know for sure if your answer is correct.

**Definition 1.1.6**

**Parameter** – a number calculated from the population. Usually denoted with a Greek letter. This number is a fixed, unknown number that you want to find.

**Definition 1.1.7**

**Statistic** – a number calculated from the sample. Usually denoted with letters from the Latin alphabet, though sometimes there is a Greek letter with a  $\wedge$  (called a hat) above it. Since you can find samples, it is readily known, though it changes depending on the sample taken. It is used to estimate the parameter value.

One last concept to mention is that there are two different types of variables – qualitative and quantitative. Each type of variable has different parameters and statistics that you find. It is important to know the difference between them.

**Definition 1.1.8**

**Qualitative or categorical variable** – answer is a word or name that describes a quality of the individual.

**Definition 1.1.9**

**Quantitative or numerical variable** – answer is a number, something that can be counted or measured from the individual.

**Example 1.1.1 stating definitions for qualitative variable**

In 2010, the Pew Research Center questioned 1500 adults in the U.S. to estimate the proportion of the population favoring marijuana use for medical purposes. It was found that 73% are in favor of using marijuana for medical purposes. State the individual, variable, population, and sample.

**Solution**

Individual – a U.S. adult

Variable – the response to the question “should marijuana be used for medical purposes?” This is qualitative data since you are recording a person's response – yes or no.

Population – set of all responses of adults in the U.S.

Sample – set of 1500 responses of U.S. adults who are questioned.

Parameter – proportion of those who favor marijuana for medical purposes calculated from population

Statistic – proportion of those who favor marijuana for medical purposes calculated from sample

#### Example 1.1.2 stating definitions for qualitative variable

A parking control officer records the manufacturer of every 5<sup>th</sup> car in the college parking lot in order to guess the most common manufacturer.

##### **Solution**

Individual – a car in the college parking lot

Variable – the name of the manufacturer. This is qualitative data since you are recording a car type.

Population – set of all names of the manufacturer of cars in the college parking lot.

Sample – set of recorded names of the manufacturer of the cars in college parking lot

Parameter – proportion of each car type calculated from population

Statistic – proportion of each car type calculated from sample

#### Example 1.1.3 stating definitions for quantitative variable

A biologist wants to estimate the average height of a plant that is given a new plant food. She gives 10 plants the new plant food. State the individual, variable, population, and sample.

##### **Solution**

Individual – a plant given the new plant food

Variable – the height of the plant (Note: it is not the average height since you cannot measure an average – it is calculated from data.) This is quantitative data since you will have a number.

Population – set of all the heights of plants when the new plant food is used

Sample – set of 10 heights of plants when the new plant food is used

Parameter – average height of all plants

Statistic – average height of 10 plants

#### Example 1.1.4 stating definitions for quantitative variable

A doctor wants to see if a new treatment for cancer extends the life expectancy of a patient versus the old treatment. She gives one group of 25 cancer patients the new treatment and another group of 25 the old treatment. She then measures the life expectancy of each of the patients. State the individuals, variables, populations, and samples.

##### **Solution**

In this example there are two individuals, two variables, two populations, and two samples.

Individual 1: cancer patient given new treatment

Individual 2: cancer patient given old treatment

Variable 1: life expectancy when given new treatment. This is quantitative data since you will have a number.

Variable 2: life expectancy when given old treatment. This is quantitative data since you will have a number.

Population 1: set of all life expectancies of cancer patients given new treatment

Population 2: set of all life expectancies of cancer patients given old treatment

Sample 1: set of 25 life expectancies of cancer patients given new treatment

Sample 2: set of 25 life expectancies of cancer patients given old treatment

Parameter 1 – average life expectancy of all cancer patients given new treatment

Parameter 2 – average life expectancy of all cancer patients given old treatment

Statistic 1 – average life expectancy of 25 cancer patients given new treatment

Statistic 2 – average life expectancy of 25 cancer patients given old treatment

There are different types of quantitative variables, called discrete or continuous. The difference is in how many values can the data have. If you can actually count the number of data values (even if you are counting to infinity), then the variable is called discrete. If it is not possible to count the number of data values, then the variable is called continuous.

#### Definition 1.1.10

**Discrete** data can only take on particular values like integers. Discrete data are usually things you count.

#### Definition 1.1.11

**Continuous** data can take on any value. Continuous data are usually things you measure.

#### Example 1.1.5 discrete or continuous

Classify the quantitative variable as discrete or continuous,

- The weight of a cat.
- The number of fleas on a cat.
- The size of a shoe.

#### Solution

- This is continuous since it is something you measure.
- This is discrete since it is something you count.
- This is discrete since you can only be certain values, such as 7, 7.5, 8, 8.5, 9 You can't buy a 9.73 shoe.

There are also are four measurement scales for different types of data with each building on the ones below it. They are:

### Measurement Scales:

#### Definition 1.1.12

**Nominal** – data is just a name or category. There is no order to any data and since there are no numbers, you cannot do any arithmetic on this level of data. Examples of this are gender, car name, ethnicity, and race.

#### Definition 1.1.13

**Ordinal** – data that is nominal, but you can now put the data in order, since one value is more or less than another value. You cannot do arithmetic on this data, but you can now put data values in order. Examples of this are grades (A, B, C, D, F), place value in a race (1st, 2nd, 3rd), and size of a drink (small, medium, large).

#### Definition 1.1.14

**Interval** – data that is ordinal, but you can now subtract one value from another and that subtraction makes sense. You can do arithmetic on this data, but only addition and subtraction. Examples of this are temperature and time on a clock.

## Definition 1.1.15

**Ratio** – data that is interval, but you can now divide one value by another and that ratio makes sense. You can now do all arithmetic on this data. Examples of this are height, weight, distance, and time.

Nominal and ordinal data come from qualitative variables. Interval and ratio data come from quantitative variables.

Most people have a hard time deciding if the data are nominal, ordinal, interval, or ratio. First, if the variable is qualitative (words instead of numbers) then it is either nominal or ordinal. Now ask yourself if you can put the data in a particular order. If you can it is ordinal. Otherwise, it is nominal. If the variable is quantitative (numbers), then it is either interval or ratio. For ratio data, a value of 0 means there is no measurement. This is known as the absolute zero. If there is an absolute zero in the data, then it means it is ratio. If there is no absolute zero, then the data are interval. An example of an absolute zero is if you have \$0 in your bank account, then you are without money. The amount of money in your bank account is ratio data. Word of caution, sometimes ordinal data is displayed using numbers, such as 5 being strongly agree, and 1 being strongly disagree. These numbers are not really numbers. Instead they are used to assign numerical values to ordinal data. In reality you should not perform any computations on this data, though many people do. If there are numbers, make sure the numbers are inherent numbers, and not numbers that were assigned.

## Example 1.1.6 measurement scale

State which measurement scale each is.

- a. Time of first class
- b. Hair color
- c. Length of time to take a test
- d. Age groupings (baby, toddler, adolescent, teenager, adult, elderly)

**Solution**

- a. This is interval since it is a number, but 0 o'clock means midnight and not the absence of time.
- b. This is nominal since it is not a number, and there is no specific order for hair color.
- c. This is ratio since it is a number, and if you take 0 minutes to take a test, it means you didn't take any time to complete it.
- d. This is ordinal since it is not a number, but you could put the data in order from youngest to oldest or the other way around.

## Homework

1. Suppose you want to know how Arizona workers age 16 or older travel to work. To estimate the percentage of people who use the different modes of travel, you take a sample containing 500 Arizona workers age 16 or older. State the individual, variable, population, sample, parameter, and statistic.
2. You wish to estimate the mean cholesterol levels of patients two days after they had a heart attack. To estimate the mean you collect data from 28 heart patients. State the individual, variable, population, sample, parameter, and statistic.
3. Print-O-Matic would like to estimate their mean salary of all employees. To accomplish this they collect the salary of 19 employees. State the individual, variable, population, sample, parameter, and statistic.
4. To estimate the percentage of households in Connecticut which use fuel oil as a heating source, a researcher collects information from 1000 Connecticut households about what fuel is their heating source. State the individual, variable, population, sample, parameter, and statistic.
5. The U.S. Census Bureau needs to estimate the median income of males in the U.S., they collect incomes from 2500 males. State the individual, variable, population, sample, parameter, and statistic.
6. The U.S. Census Bureau needs to estimate the median income of females in the U.S., they collect incomes from 3500 females. State the individual, variable, population, sample, parameter, and statistic.
7. Eyeglassmatic manufactures eyeglasses and they would like to know the percentage of each defect type made. They review 25,891 defects and classify each defect that is made. State the individual, variable, population, sample, parameter, and statistic.
8. The World Health Organization wishes to estimate the mean density of people per square kilometer, they collect data on 56 countries. State the individual, variable, population, sample, parameter, and statistic.
9. State the measurement scale for each.
  - a. Cholesterol level

- b. Defect type
  - c. Time of first class
  - d. Opinion on a 5 point scale, with 5 being strongly agree and 1 being strongly disagree
10. State the measurement scale for each.
- a. Temperature in degrees Celsius
  - b. Ice cream flavors available
  - c. Pain levels on a scale from 1 to 10, 10 being the worst pain ever
  - d. Salary of employees

#### Answer

- 1. See solutions
- 3. See solutions
- 5. See solutions
- 7. See solutions
- 9.
  - a. ratio
  - b. nominal
  - c. interval
  - d. ordinal

---

This page titled [1.1: What is Statistics?](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Kathryn Kozak](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.