

## What is Statistics?

Statistics is the science of reasoning from data, so a natural place to begin your study is by examining what is meant by the term **data**. You will find that data *vary*, and variability abounds in everyday life and in academic study. Indeed, the most fundamental principle in statistics is that of variability. If the world were perfectly predictable and showed no variability, you would not need to study statistics. Thus, you will learn about variables and consider their different classifications. You will also begin to experience the interesting research questions that you can investigate by collecting data and conducting statistical analyses.

### Vocabulary Chapter 1

1. **Statistics** is a collection of methods for planning experiments, obtaining data, and then organizing, summarizing, presenting, analyzing, interpreting and drawing conclusions based on the data.
2. A **population** is the complete collection of all elements (scores, people, measurements, and so on) to be studied. The collection is complete in the sense that it includes all subjects to be studied.
3. A **sample** is a subcollection of elements selected from a population.

Example: In a study of household incomes in a small town of 1000 households, one might conceivably obtain the income of every household. However, it is probably very expensive and time consuming to do this. Therefore, a better approach would be to obtain the data from a portion of the households (let's say 125 households). In this scenario, the 1000 households are referred to as the population and the 125 households are referred to as a sample.

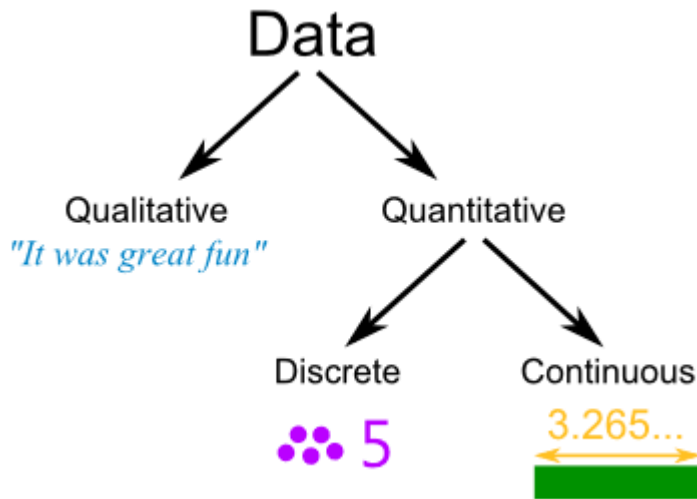
4. A **parameter** is a numerical measurement describing some characteristic of a *population* and computed from all of the population measurements.
5. A **statistic** is a numerical measurement describing some characteristic of a *sample* drawn from the population.

Example: In the household incomes example from above, the average (mean) income of all 1000 households is a parameter, whereas the average (mean) income of the 125 households is a statistic.

6. Data can be **qualitative or quantitative**.

**Qualitative (categorical) data** is descriptive information (it describes something)

**Quantitative data**, is numerical information (numbers that represent counts or measurements).



7. **Discrete data** result when the number of possible values is either a finite number or a countable number. (That is, the number of possible values is 0 or 1 or 2 and so on.)

Example: The numbers of fatal automobile accidents last month in the 10 largest US cities

8. **Continuous data** result from infinitely many possible values that correspond to some continuous scale that covers a range of values without gaps, interruptions, or jumps.

Example: The finishing times of a marathon

9. A **variable** is a *characteristic* that varies from person to person or from thing to thing. The person or thing is called an **observational unit**.

10. Variables can be classified as **categorical** or **quantitative**, depending on whether the characteristic is a categorical designation (such as gender) or a numerical value (such as height).

### Four different levels of measurement

The four levels of measurement in ascending order of precision are: **nominal**, **ordinal**, **interval** and **ratio**.

#### Nominal

At the first level of measurement, numbers are used to classify data. In fact words or letters would be equally appropriate. Say you wanted to classify a football team into left footed and right footed players, you could put all the left footed players into a group classified as 1 and all the right footed players into a group classified as 2. The numbers 1 and 2 are used for convenience, you could equally use the letters L and R, or the words LEFT and RIGHT to label the groups of players. Numbers are often preferred because text takes longer to type out and takes up more space. Another example is blood groups where the letter A, B, O and AB represent the different classes

#### Ordinal

In ordinal scales, values given to measurements can be ordered. One example is shoe size. Shoes are assigned a number to represent the size, larger numbers mean bigger shoes so unlike the nominal scale that just reflects a category or class, the numbers of an ordinal scale show an ordered relationship between numbered items – we know that a shoe size of 8 is bigger than a shoe size of 4. What you can't say though

is that a shoe size of 8 is twice as big as a shoe size of 4. So numbers on an ordinal scale represent a rough and ready ordering of measurements but the difference or ratios between any two measurements represented along the scale will not be the same.

As for the nominal scale, with ordinal scales you can use textual labels instead of numbers to represent the categories. So, for example, a scale for the measurement of patient satisfaction with the care they received in hospital might look like this: | Not satisfied | Fairly satisfied | Satisfied | Very satisfied |

There are many everyday examples of measurements assigned to ordinal scales: social class grading I, II, III, IV; grades A, B, C, D; house numbers 1,3,5...2,4,6, etc.

### **Interval**

On an interval scale, measurements are not only classified and ordered therefore having the properties of the two previous scales, but the distances between each interval on the scale are equal right along the scale from the low end to the high end. Two points next to each other on the scale, no matter whether they are high or low, are separated by the same distance. So when you measure temperature in centigrade the distance between 96 and 98<sup>o</sup>, for example, is the same as between 100 and 102<sup>o</sup> C. Remember though is that for interval scales, a measurement of 100<sup>o</sup>C does not mean that the temperature is 10 times hotter than something measuring 10<sup>o</sup>C even though the value given on the scale IS 10 times as large. That's because there is no absolute zero: the zero is arbitrary. On the centigrade scale, the zero value is taken as the point at which water freezes and the 100<sup>o</sup>C value when water begins to boil and between these extreme values the scale is divided into a hundred equal divisions. Temperatures below 0<sup>o</sup> on the centigrade scale are designated negative numbers. So the arbitrary 0<sup>o</sup>C does not mean 'no temperature'. But when expressed on the Kelvin scale, a ratio scale, a measure of 0<sup>o</sup> K equivalent to -273<sup>o</sup>C does indeed mean no temperature! Other examples of interval measurements are rare, but there's one you will be familiar with. Calendar years are an interval scale. The arbitrary 0 was assigned when Christ was born and time before this is labeled 'BC'.

### **Ratio**

Measurements expressed on a ratio scale can have an actual zero. Apart from this difference, ratio scales have the same properties as interval scales. The divisions between the points on the scale have the same distance between them and numbers on the scale are ranked according to size. There are many examples of ratio scale measurements, length, weight, temperature on the Kelvin scale, speed and counted values like numbers of people, exam marks – a score of zero really does mean no marks!! Returning to the Kelvin scale of temperatures, at the temperature of 0 K<sup>o</sup> the lowest temperature possible, it is so cold that all molecules have stopped moving.

## **Example Problems**

### **a) Determine whether the given value is a statistic or a parameter.**

1. A sample of students is selected from FIU and their average age in years is 23.7.
2. In a study of all current major league baseball players, it was found that 78% batted exclusively right-handed.

### **b) Determine whether the given value is from a discrete or continuous data set.**

1. A research poll of 1015 people shows that 752 of them have internet access at work.
2. Josh Becket's fastball was clocked at 98 mph during the World Series.
3. A student spent \$86.53 on her calculator for class.

The \_\_\_\_\_ is the set of *all* measurements of interest to the investigator. (**population**)

A \_\_\_\_\_ is a subset of measurements selected from the population of interest. (**Sample**)

**Some more examples (answers are given at the end):**

Which branch of statistics deals with organizing and summarizing data?  
(Descriptive or Inferential)

If I would like to predict the average fuel economy for my new car, in what branch of statistics will I find prediction methods?  
(Descriptive or Inferential)

A researcher at FIU is studying the effects of anti-anxiety drugs on memory. She wants to know if people suffering from PTSD due to combat stress will have a reduction in symptoms while taking an anti-anxiety medication. She plans to put 30 PTSD patients on the new drug. What is the population for this study? What are the experimental units?

Are the 30 patients described above a sample or the population?

Height and weight are both examples of \_\_\_\_\_ variables, while eye color is an example of a \_\_\_\_\_ variable.

**Answers:** Descriptive;  
Inferential ;  
All people suffering from combat stress related PTSD; the individual patients who will receive the drug are the experimental units.  
Sample  
Quantitative; Qualitative

**Examples:**

**1. Now consider the students in your class as observational units. Classify each of the following variables as categorical or quantitative.**

- How many hours you have slept in the past 24 hours
- Whether or not you have slept for at least 7 hours in the past 24 hours
- How many states you have visited
- Handedness (which hand you write with)
- Day of the week on which you were born
- Whether or not you have used a cell phone today
- Whether you prefer baths or showers
- How much time you spent on your most recent bath or shower

**Variables of State**

**2. Suppose that the observational units of interest are the fifty states. Identify which of the following are variables and which are not. Also classify the variables as categorical or quantitative.**

- a. Gender of the state's current governor
- b. Number of states that have a female governor
- c. Percentage of the state's residents older than 65 years of age
- d. Highest speed limit in the state
- e. Whether or not the state's name contains one word
- f. Average income of the adult residents of the state
- g. How many states were settled before 1865

**3. The wife of a farmer in a small rural town of 1050 people wishes to open a small video rental store. Before doing so, she would like to estimate the number of people in that town who would be interested in renting videos. Over the course of one week, she decides to ask 50 people randomly at a local post office whether or not they would rent videos.**

- a. What is the population of interest?
- b. What is the population sampled from?
- c. Identify the variable of interest.

**4. We are interested in the TV viewing habits of the country. A sample of Americans is surveyed and the average amount of time spent watching TV is found to be 4.6 hours per day.**

Is this average a population parameter or a sample statistic?