1.3: Types of Variables

All things about which data are collected are variables, however, not all variables are measured or represent the world in the same way. There are several ways variables can be defined and categorized based on the nature of the variable itself and its operationalization. To **operationalize** is to define a variable for the purposes of measurement. Therefore, when a researcher or statistician operationalizes, they are defining the way a variable will be measured to gather data for their research. The nature of variables limits how they can be operationalized and measured. The way they are measured then limits how they can be analyzed. Therefore, it is important to understand the different kinds of variables and ways they can be categorized.

Qualitative verses Quantitative

The first main way to categorize variables is by whether they are qualitative or quantitative. **Qualitative variables** are those which vary in characteristic, category, type, or kind rather than amount. Eye color is qualitative because we use categories to define the type of color each individual's eyes are. Major in college is also qualitative because names are used to differentiate among the majors. When someone says their major is Psychology and someone else says their major is Nursing, the two individuals vary in type of major rather than amount of major. We wouldn't say that one of this is more or less "majory" than the other because the names of the majors indicate differences in type of major not amount of major. Thus, eye color and major are two examples of qualitative variables.

In contrast, **quantitative variables** are those which exist and are measured in amount. Weight is a quantitative variable because values are used to represent differing amounts of weight. When a variable is quantitative, comparisons of amount are possible. For example, we can say that a cat that weighs 19 pounds is heavier than a cat that weighs 14 pounds. Here the word "heavier" indicates a comparison of the amount of weight between the two cats. Many other variables are quantitative such as sleep measured in hours, distance driven measured in miles, or math knowledge on a 10 point quiz. In each of these examples you can see what the variable is (sleep, distance driven, math knowledge) and how each is being quantitatively operationalized (as hours, miles, or points, respectively).

> *When something varies in characteristic, category, type, or kind, it varies qualitatively.*
> *When something varies in amount, it varies quantitatively.*

Discrete verses Continuous

There are a variety of ways quantitative variables can exist and, thus, be properly measured. One way they can differ is by being discrete variables or continuous variables.

Discrete variables are sometimes referred to as discontinuous variables. When a quantitative variable is **discrete** it means it exists and can be measured in counts. The data for discrete variables are often whole numbers. For example, the number of children someone has is a discrete version of a quantitative variable. Number of children is quantitative because data are being collected to represent an amount. This amount is discrete because children can only exist in whole units and are not divisible into parts of a whole. You can have 0 children, 1 child, or perhaps 12 children but you cannot have 1.50 children. Discrete variables can be thought of as thing that can only vary in specific counts rather than any amount.

In contrast, when a quantitative variable is **continuous**, it means it can exist and, thus, be measured in infinitely differentiated amounts. Hours of sleep is continuous because the amount of time someone sleeps can be exactly 7.00 hours or 8.00 hours, for example, or it could be anything between those such as 7.50 hours, 7.25 hours, or even 7.9999 hours if we have a very accurate and specific timer. When data are continuous it means that it is possible that amounts vary at or between integers.

Scales of Measurement

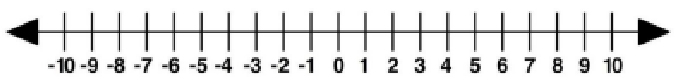
Another way to distinguish among types of variables and how they are measured is through the scales of measurement. When a variable is operationalized, one of four scales of measurement can be applied. The four scales of measurement are: ratio, interval, ordinal, and nominal. These are the categories for the four different ways things can be measured. The first two scales of measurement (ratio and interval) differentiate between the two common ways something quantitative can be measured. The fourth scale of measurement (nominal) is used when something qualitative is measured. The third scale of measurement, the ordinal scale, is best understood as existing in the tension between what is quantitative and what is qualitative.

The Ratio Scale

The ratio scale is the most specific and least limited of the quantitative scales of measurement. The **ratio scale** organizes amounts that increase by the same magnitude throughout a sequence with a true zero used to represent a complete absence of that which is being measured. Thus, ratios scales have two defining features: 1. A true zero and 2. Equally sized changes in amount between each whole number. Think of the ratio scale as referring to anytime something is measured based on where it would fall on the right side of the number line, starting from 0. On the number line 0 represents a complete absence of something. As the amount of the thing being measured increases, values move to the right on the number line. This scale of measurement is the one that is often presumed when considering real world values in math. When we are given a problem such as "If you have three apples and get two more apples, how many apples do you have?" the ratio scale is presumed used.

The word "ratio" is given to these scales to indicate that they have a feature no other scale of measurement has: the ability to make meaningful comparisons with ratios. A ratio refers to the relative size between two amounts. Another way of saying this is that a ratio tells how many times one amount is a multiple of the other. These kinds of relative comparisons can only be made when a true zero exists and is known based on the measure used. Let's consider ratios using coffee measured in ounces. The least amount of coffee one could have is 0 ounces. Someone who has 16 ounces of coffee has twice as much as someone who has 8 ounces of coffee. We can make this statement because coffee is being measured on a ratio scale. Hours of sleep also uses the ratio scale and, thus, comparisons of relative amount can be made. We can correctly state that 4 hours of sleep is twice as much as 2 hours of sleep or that 5 hours of sleep is half as much as 10 hours of sleep.

This may seem like an absurdly obvious thing to say because, of course, 4 is twice as much as 2 and 5 is half as much as 10. The reason this seems so obvious is because most of the mathematical concepts we know about the world are focused on things which exist on the ratio scale of measurement. We rarely overtly learn or think about quantitative things which are not on this scale of measurement. We are accustomed to thinking of 0 as representing a complete absence because that is its mathematical role. That is what it is meant to represent. Therefore, we often treat positive numbers as though they represent the relationship of those numbers to 0 on the number line. We may not have ever considered "what if there was no 0" or "what if 0 wasn't really known to be 0." However, sometimes 0 is not truly knowable based on the way something can be measured. In those cases, different scales of measurement are used and ratio-based statements of comparison are not appropriate.



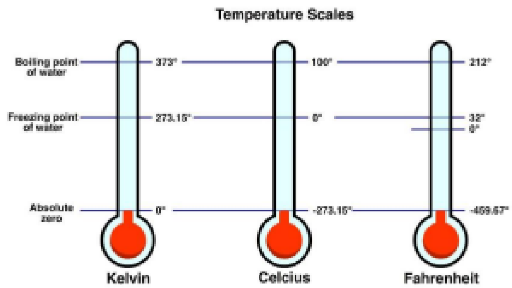
The Interval Scale

The interval scale is a quantitative scale of measurement without a true, known zero. **Interval scales** organize amounts that increase by the same magnitude throughout a sequence but do not have a true, known zero point. Thus, they are quantitative but only meet one criteria that is also met by the ratio scale: Equally sized changes in amount between each integer. Integers refer to whole numbers (such as 3) or their negatives (such as -3). An interval is an equally sized segments of change in magnitude. This scale of measurement was named for this defining feature.

Think of the interval scale as being a segment of the number line without starting at the true zero. Equally sized differences between numbers on the scale reflect equal differences in magnitude. However, the true zero is either not known or is not represented as "0." Many interval scales do not use 0 because the value chosen for 0 would be somewhat arbitrary. When 0 is used on an interval scale, the statistician must keep in mind that this value does not indicate a known and complete absence of that which is being measured. Some interval scales have a zero but also have negative numbers. Ratio scales cannot measure into the negative because you cannot physically have less than nothing (or absolute zero) of something.

Comparing the Ratio and Interval Scales

When an interval scale is used, statements of ratio are not appropriate. Let's explore why with an example. A good example of something that is measured on an interval scale is temperature in Fahrenheit. Temperature is quantitative and exists in a continuous way. However, there are lots of ways it can be operationalized. We could, for example, measure temperature using Fahrenheit, Celsius, or Kelvin. These are three different quantitative temperature scales. Fahrenheit and Celsius are interval scales of measurement. Each has a 0 but it is not a true zero because it does not represent a complete absence of heat. If it did, then it would not be possible to observe of experience a negative temperature on either scale.



When each of these two interval scales were created, a relevant point of temperature was labeled as 0°, but neither was an absolute zero. Anders Celsius developed the Celsius scale and chose to label the temperature at which water froze as 0°. Daniel Gabriel Fahrenheit developed the first version of the Fahrenheit scale and chose to label the temperature at which a saltwater mixture froze as 0°. Thus, each of these scales call different amounts of heat 0 and it is possible to experience temperatures below 0 degrees on each of these scales. This is because the point that was chosen to be labeled as 0 on each of these scales does not represent a complete absence of heat. Because of this, it is not accurate to state that 80° Celsius is twice as hot as 40° Celsius nor is it accurate to state that 80° Fahrenheit is twice as hot as 40° Fahrenheit. The only way these kinds of ratio statements can be made is if the 0 on the scale represents a true zero.

However, temperature can be measured with the ratio scale known as Kelvin. Kelvin is a way of measuring temperature that counts up from a point known as absolute zero. There are no negative values on the Kelvin scale. Absolute zero, which is 0 degrees Kelvin, represents the absence of heat; it is the lowest possible point determined for heat and the Kelvin scale counts up from this point. Because of this, the Kelvin scale meets both of the criteria to be a ratio scale of measurement.

When data are collected on a ratio scale, it is possible to make ratios. For example, 80° degrees Kelvin is twice as hot as 40° degrees Kelvin, something we could not say for these temperatures on the Celsius or Fahrenheit scales. This is because comparisons of ratio can only be made when a quantitative scale of measurement with a true zero, like the Kelvin scale, is used. Fortunately, however, there are few other limitations to using interval scales and, thus, data which were collected using interval or ratio scales of measurement can often be analyzed using the same techniques.

The Ordinal Scale

The **ordinal scale** is used when something can be organized into quantitative sequences without specifying the precise magnitude of differences in the sequence. Thus, when things can be put into ranks or orders of magnitude, they are on the ordinal scale. For example, the speed of runners can be ranked from fastest to slowest or from first place (representing fastest) to last place (representing slowest). Ordinal scales are limited in that the magnitude of change between ranks is not necessarily, or known to be, consistent. When numbers are used to specify who is the fastest (1st) to who is the slowest (5th) among five runners, for example, the numbers represent the five ranked or ordered places. However, the difference in how much faster first place was from second place is not necessarily the same as how much faster second place was from third, third was from fourth, or fourth was from fifth. Height can also be measured on the ordinal scale by organizing persons from shortest to tallest. In this way, the heights of five people could be ranked from 5 (tallest) down to 1 (shortest). However, the amount that each person's height differed from the next tallest person in the ordered sequence is not necessarily the same because the specific amount of change in height from person to person is rarely stable. The tallest person may be 1 inch taller than the second tallest person but the second tallest person might be 3 inches taller than the third tallest person.

The ordinal scale is unique in a few ways. First, it is the only scale of measurement where words or symbols are often used to represent quantity instead of numbers. Words like “first,” “tallest,” “most,” “last,” “shortest,” and “least” are used for some variables measured using ordinal scales. Second, it is the most limited of the three quantitative scales of measurement because it is the least specific in its ability to differentiate amounts. This limits the way the data can be computed and the kinds of quantitative comparisons that can be made. In fact, ordinal data should generally be analyzed using techniques appropriate for qualitative data rather than those that can be used to quantitative data. For this reason, the ordinal scale is sometimes argued to be a qualitative scale of measurement despite it being quantitative in nature (because it organizes data based on magnitudes). Fortunately, many quantitative variables that are sometimes measured on the ordinal scale, such as speed or height, can be measured with more specificity using the ratio or interval scales as long as the researcher or statistician operationalizes them that way. When using a ratio or interval scale is possible, it is often best to use that instead of an ordinal measurement.

The Nominal Scale

The nominal scale of measurement is the category used for all qualitative variables. The nominal scale is used when observations are labeled and categorized to differentiate them by characteristic. “Nom” means name and “nominal” roughly means “in name” or differentiated in name or type. Qualitative variables such as eye color and major, for example, are measured on the nominal scale. Data for variables measured using a nominal scale can be referred to as nominal data.

> **Connecting Topics**
>
> Interval and ratio data are often a good fit for parametric statistics while nominal and ordinal data often require non-parametric statistics.

Nominal data can be represented using words or names, symbols, or numbers. However, it is important to keep in mind that when numbers are used for nominal data, they function as individual or category names rather than indicators of quantity. An example of this is an ID number. An ID number is a set of digits used as a name for an individual or item but does not represent an amount. If one person has a student ID number of 80645 and another has a student ID number of 29448 it does not mean that one of them is more of a student than the other. The numbers are simply used the same way as names to differentiate between the individuals.

Independent Variables verses Dependent Variables

So far we have been considering variables one at a time. When one variable is considered or summarized at a time, we refer to the procedures and results as **univariate statistics**. “Uni” refers to one of something and “variate” means the something being referred to is a variable. However, it is also possible to consider two variables at a time and how they might be related or connected. When the associations or connections between two variables are considered, the procedures and results are referred to as **bivariate statistics**. “Bi” refers to the co-existence or consideration of two things at once. It is also possible to focus on two or more variables at which time the procedures and results can be referred to simply as **multivariate statistics**. For the first several chapters of this book, we will focus on univariate statistics (such as Chapters 2, 3, and 4). However, in later chapters we will also explore bivariate statistics, some of which (such as Chapters 8 and 9) further define variables as independent or dependent.

Statisticians sometimes use bivariate analyses to see if one variable is the cause of change or differences in another variable. When this is done, the variables are differentiated using the terms the independent variable and the dependent variable. An **independent variable** is the assumed or hypothesized cause in a cause-effect relationship. A **dependent variable** is the assumed or hypothesized thing which is affected in a cause-effect relationship. Whether something is considered an independent variable or a dependent variable is specific to what is being considered or tested at that moment. This means that it is possible for the same variable to sometimes be an independent variable and other times to be a dependent variable. Take grades as an example. If we want to test whether studying causes grades to be higher, studying is the independent variable and grade is the dependent variable. However, if we want to test whether grades cause students to be considered more desirable job candidates, grades are now the independent variable and desirability of the candidate is now the dependent variable.

It is important to note that not all variables are either independent or dependent. These designations are only used when cause-effect connections between or among variables are being considered to tested. In the absence of those scenarios, these terms should not be applied. For example, a statistician may want to test whether hours of sleep relate to amount of caffeine consumed where the statistician is either not interested in, cannot deduce, or is not considering, whether one variable is the potential cause of the other. In these situations, the terms independent variable and dependent variable are either not used or are used with the understanding that they are being theoretically positioned as cause and effect thought they cannot be fully tested as such. We will see some examples of these scenarios in later chapters (such as Chapter 12).

Reading Review 1.2

Use each description of a variable and how it was measured to categorize it. Variable names are underlined to help you identify them.

1. Ethnicity measured by having participants fill in the ethnicity with which they identify.
   a. Is the variable quantitative or qualitative?
   b. Which scale of measurement is being used?
2. Chips consumed measured in ounces.
   a. Is the variable quantitative or qualitative?
   b. Which scale of measurement is being used?
3. Shirt size measured in sizes of XS, S, M, L, XL, and XXL.
   a. Is the variable quantitative or qualitative?
   b. Which scale of measurement is being used?

Identify the independent variable and the dependent variable in each bivariate sentence below:

4. Exercise decreases stress.
   a. Independent variable:
   b. Dependent variable:
5. Popularity is impacted by behavior
   a. Independent variable:
   b. Dependent variable:
