The ultimate purpose of any scientific enquiry is to establish relationships among variables. The term **variable** has a slightly different meaning in the context of mathematics than in the context of empirical science. In mathematics, a variable is a symbol that can stand for a defined range of values. For example, the symbol $X$ could be a variable associated with numbers that might range in value from 1 to 1000. In the context of empirical science, a variable is something that can be measured in some way and associated with more than one value. The term is often used interchangeable with the term **factor**. For example, the type of family, single parent or two parent, could be a variable. More commonly, variables are associated with numbers; examples include things like body weight, grade point average, number of beers consumed each day, hours of sleep, etc. The term variable does not necessarily refer to something that can change. Your biological sex is a variable but you cannot change your sex chromosomes. Similarly, date of birth is a variable; once established it cannot be changed. (At least, you are not supposed to change it.) Other variables often do change; body weight tends to fluctuate from day to day. Some variables are constantly changing ... for example, time of day. When doing science, psychologists select a behavioral variable of interest, for example grade point average and try to find other variables that might be related to the target variable; for example, grade point average might be related to study time. Both grade point average and study time are variables. Science is about these types of relationships, relationships among variables.

Variables can be sorted into two general types: a) qualitative variables and b) quantitative variables. You can create a **qualitative variable** by sorting people (or behaviors or events) into groups or categories in such a way that all of the people in one group are similar to one another and differ in some well-defined way from all of the people in another group. For example, it is common practice to sort people into groups on the basis of biological sex, male and female. In this case, the people in one group share one common characteristic that is not shared by the people in the other group; the groups differ from one another by this one well-defined characteristic, biological sex. If the groups differ from one another on the basis of kind rather than amount or magnitude then the variable would be considered a qualitative variable. For example, usually males are thought to differ from females in kind rather than with regard to an amount of some characteristic. Another example of a qualitative variable might be political affiliation such as “republican”, “democrat”, “independent”, “other”. The people in these categories are not thought to differ from each other in an amount of something but rather we think of them as differing in kind. Notice that qualitative variables must include two or more categories. These categories must also be **mutually exclusive**; an observation must fit into one and only one category. For example, suppose that you sorted people into categories based on whether or not they are employed full time. An individual could not both be employed full time and not employed full time. In this case, the categories would be mutually exclusive.

**Quantitative variables** are created when differences among people (or behaviors or events) are measured with respect to an amount or magnitude of something rather than in kind. For example, suppose you wanted to study reaction time to the presentation of a target, such as a particular letter, presented on a computer screen. Participants in such a study might be seated at a table with a response key. They are told to press the key as quickly as possible when the target appears on the screen. The researcher collects response times to the event in milliseconds. A person who responds in 230 milliseconds responds faster than someone who responds in 235 milliseconds.
The difference in response times represents an amount of time; response time is a quantitative variable.

When doing psychological research, psychologists chose one or more behaviors to study. For example, a psychologist might be interested in aggressive behavior in children. The behavioral variable that the psychologist chooses to study is usually referred to as the dependent variable although sometimes the terms outcome variable or criterion variable are used in its place. In the context of thinking about research, psychologists look for other variables that they think might be related to differences that they observe in the dependent variable. For example, if you were interested in aggressive behavior of children, you might ask yourself “why are some children more aggressive than other children” and then try to imagine possible answers to that question: what other variables might be related to differences in aggression? The possible answers that you might come up with are referred to as hypotheses; hypotheses are simply statements about possible relationships that might exist between variables. Using our aggression example, maybe it occurs to you that differences in children’s aggression might be related to differences in some feature of their home environments. Maybe children from single parent homes are likely to be more aggressive than children from two parent homes. In this case you might formulate a hypothesis that the variable “family structure” is related in some way to the variable aggression. When we do the science of psychology we attempt to collect data that might provide information concerning this type of hypothesis. The statistical procedures that we will be discussing in this class are simply tools that psychologists use to mine data for information concerning our hypotheses; the procedures are nothing more than tools that scientists use to describe variables and identify and describe relationships among variables. When statistical procedures are used to describe variables and relationships among variables the procedures are often referred to as descriptive statistics: when used to make statistical inferences about relationships among variables the procedures are usually referred to as inferential statistics.

In research we also have the idea of an independent variable: a term usually associated with a variable that is manipulated in the context of an experiment. The term, however, is often used to describe any variable in a study other than the dependent variable and is often used in place of the term predictor variable. In the statistics literature the term has a slightly different meaning. By definition, the value of the independent variable does not depend on the value of the dependent variable while the value of dependent variable does depend on the value of the independent variable. It is this asymmetry that determines the role of variables in a relationship and not by the research methods used to empirically investigate the relationship between variables. When discussing statistical procedures, unless otherwise specifically noted, we are going to use the term “independent variable” in the statistical sense rather than in the sense commonly encountered when describing types of variables in psychological research.

Individual Differences

By implication, if not by definition, in psychological research a dependent variable is an aspect of behavior that varies. [Link to Variability] It is this variability of the behavior that we study when we do psychological research. It may surprise you that psychology is about behavioral differences. Psychology as a science attempts to characterize, account for, explain or predict behavioral differences. These differences are referred to as individual differences. Individual
differences come in two forms; a) inter-individual differences and b) intra-individual differences. Inter-individual differences are those behaviors that differ between people. For example, some people smoke more cigarettes per day than other people. This difference is an inter-individual difference. Intra-individual differences generally refer to changes in the behavior of the same people at different points in time. Suppose you recorded smoking behavior at different times of the day. You might find that a person smokes more in the evening than in the morning. Such a difference would be an intra-individual difference. As another example, it is common in psychology to examine the behavior of children as they age. If we examine the same behavior of the same children as they get older, differences in their behavior as they age would represent intra-individual differences. When we do psychological research, we are trying to account for, explain, or predict individual differences. Psychological research will always focus on individual differences; we will always be interested in the extent to which individuals vary from one another.

Imagine that you want to study cigarette smoking by college freshmen. You might want to find out why some students smoke more than other students. Before you can find answers to this question, you must come up with some way to measure cigarette smoking. One way to do this might be to ask students how many cigarettes they smoke each day. For the moment, assume that this is an acceptable measure of cigarette smoking. Imagine that you collect this information from 10 freshman students attending a particular university. You now have data (see Table 1.1) that includes numbers or scores that might range from 0 to, let’s say, 12 and you have 10 of these values. Number of cigarettes smoked per day is a variable and there are individual differences in the number of cigarettes smoked per day by the participants in our study.

What is the psychologist going to do with these numbers? For starters, the psychologist is going to want to come up with some way to describe the number of cigarettes smoked per day as a set of numbers: the psychologist is going to try to find a way to organize the numbers so that they can be easily examined and evaluated. When we have a set of numbers, we refer to the set as a distribution of numbers, values or scores. The word distribution implies that not all of the numbers, values or scores in the set are equal to one another. There are a wide variety of descriptive statistics procedures that might be used to characterize or describe a distribution of numbers. We will discuss some of these techniques in this class.

In addition to describing the numbers in the cigarette smoking distribution, it will also be necessary to determine if number of cigarette smoked per day is related to other situations, events or conditions that the researcher thinks might be related to these differences; you want to establish a relationship between cigarette smoking and some other variable that you imagine might be related to cigarette smoking. For example, suppose that you have a hypothesis that cigarette smoking is related to anxiety. To determine if such a relationship exists, it is obvious that you will need information from the students who are participating in your study about their levels of anxiety. Of course, this is going to require that you measure each student’s level of anxiety in some way. Assume for the moment that you have a way to do this; for example, giving them an anxiety test of some kind and recording their scores on the test. Notice that once each student’s level of anxiety has been recorded you are going to have two distributions of numbers: a) one distribution composed of the number of cigarettes smoked per day and b) one composed of anxiety level values (see Table 1.1). It is extremely important that each distribution of scores
must include a range or variety of values; if all of your study participants happen to smoke the same number of cigarettes per day or if each has the same level of anxiety, then it will be impossible to determine if there is a relationship of any kind between cigarette smoking and anxiety.

This outline of our cigarette smoking study illustrates the fundamental features of any psychological research study. First, we identified the variables that we wished to study, in this case smoking behavior and anxiety. Then we established some way to measure these variables. We found that some students smoked more than other students; there were individual differences in smoking behavior. We suggested that there might be a way to characterize these differences using descriptive statistical methods. We then hypothesized that cigarette smoking might be related to another variable, in this case anxiety. In our attempt to establish such a link we measured anxiety and found a distribution of anxiety scores: some students were more anxious than others. We then suggested that there might be a way to determine the extent to which differences in smoking can be linked or associated with differences in anxiety. There are different inferential statistical procedures that can be used to identify and to characterize the strength of the links between variables. In this class we will discuss a few of these statistical methods in detail. For now, however, the important point is that there are statistical tools that we can use to establish links or associations between variables.

Notice that when describing this study we did not identify one of the variables as the independent variable and one as the dependent variable. There are two reasons for this omission. First, from an empirical research perspective, it is not clear which variable, cigarettes smoked or anxiety, is the dependent/independent variable. One hypothesis about this relationship might be that increases in anxiety in some way lead to increases in cigarette smoking. In this case, cigarette smoking would be the dependent variable. But it is also possible to imagine that people who smoke become anxious when they hear about the adverse health effects associated with smoking. Perhaps the more they smoke the more anxious they become. In this scenario anxiety becomes the dependent variable. Which variable you choose to treat as the independent/dependent variable will depend on theory and on your understanding of the nature of the relationship. There are no statistical procedures that can sort this out for you.

The second reason that I did not identify the either of the variables in our hypothetical study with the dependent/independent variable is related to the fact that in some cases variables are not functionally related to each other: they are simply associated. In fact, in the case of the relationship between cigarette smoking and anxiety, you might be hard pressed to settle on either of the purposed hypotheses. When you find yourself in this situation, you will hypothesize that the variables are related to each other but you will have to admit that you are not sure how they are related. As a result you will not be able to establish which of the variables functions as the independent/dependent variable. There are descriptive and inferential statistical procedures specifically designed to treat data of this sort.