The objective of much research or scientific analysis is to identify the extent to which one variable relates to another variable. For example:

• Is there a relationship between a person's education level and their health?

• Is pet ownership associated with living longer?

These and other questions are exploring whether a correlation.

Correlation is a popular measure to quantify the association between two variables. Correlation describes the power and direction of a relationship. Not only can we measure this relationship but we can also use one variable to predict the other. Correlation is used when the experiment is not possible or inconvenient.

Correlational studies are quite common in psychology, particularly because some things are impossible to recreate or research in a lab setting. Instead of performing an experiment, researchers may collect data to look at possible relationships between variables. From the data they collect and its analysis, researchers then make inferences and predictions about the nature of the relationships between variables.

We can measure correlation by calculating a statistic known as a correlation coefficient. A correlation coefficient is a number from -1 to +1 that indicates the strength and direction of the relationship between variables. The correlation coefficient is usually represented by the letter r.

The number portion of the correlation coefficient indicates the strength of the relationship. The closer the number is to 1 (be it negative or positive), the more strongly related the variables are, and the more predictable changes in one variable will be as the other variable changes. The closer the number is to zero, the weaker the relationship, and the less predictable the relationships between the variables becomes. For instance, a correlation coefficient of 0.9 indicates a far stronger relationship than a

correlation coefficient of 0.3. If the variables are not related to one another at all, the correlation coefficient is 0.

The correct values for correlation coefficients depend on your study area.

The sign - positive or negative - of the correlation coefficient indicates the direction of the relationship. A positive correlation means that the variables move in the same direction. Put another way, it means that as one variable increases so does the other, and conversely, when one variable decreases so does the other. A negative correlation means that the variables move in opposite directions. If two variables are negatively correlated, a decrease in one variable is associated with an increase in the other and vice versa.

When two variables are found to be correlated, it is tempting to infer that the correlation represents a causal relationship. After all, something is causing variable Y to be correlated with variable X. However, it is not possible to know what is causing what after conducting a correlation analysis. There are several correlations involving real-world data that do not represent causal relationships. For example:

According to data from the USDA and the US Census, there is a positive correlation between the divorce rate in some US states and eating margarine. However, eating margarine does not cause divorce. Likewise, getting divorced does not cause one to eat margarine.

According to data from the National Science Foundation and the US Census, there is a positive correlation between the total revenue of arcade games and the number of doctoral degrees awarded in computer science. But, the revenue of arcade games does not cause people to pursue doctorates. And, being awarded a computer science doctorate does not cause people who own arcade games to earn more revenue.

Why might two variables be associated? Considering the following example:

A researcher finds a positive correlation between the time spent with someone and how much someone trusts that person. There are three possible ways one could make a causal inference regarding this correlation:

Spending time together causes trust between two people.

Trust between two people causes them to spend time together.

There is another variable that influences the correlation between trust and time spent together. For example, two people may come to trust one another when the time they spend together involves sharing their values and past experiences. Or, two people might establish trust when they spend time together overcoming difficulties. What about instances where two people spend a lot of time together and do not trust one another?

A correlation between variables, however, does not automatically mean that the change in one variable is the cause of the change in the values of the other variable. A correlation only shows if there is a relationship between variables.

Finding the linear correlation coefficient requires a long, difficult calculation, so most people use a calculator or software such as Excel or a statistics program.