Correlation - Its Role in Research - by Sue Roe, DPA, MS, BSN, RN

Correlation is used in a variety of contexts and is a term heard in many casual conversations. Someone might say, "Did you know there is a correlation between tall people and the type of car they drive" or "There is a correlation between popcorn lovers and the number of movies they watch". While used often, correlation is a term with special meaning.

The word correlation has an interesting background. It has been part of the English language since the 16th century. The Latin origin for correlation is the root "co" which means with, together, or join. By the 19th century, correlation became a term many disciplines adopted for their own uses. Therefore, it is important to be aware that when one employs correlation in research and statistics, it has its own unique use.

The best way to define correlation is to say that it is a relationship existing between phenomena, sets of data, or variables, which tend to vary, be associated, or occur together and are not expected to do so by chance alone.

The purpose of correlation in research and statistics is to provide information about the relationship between sets of data or variables. As a statistical measure, correlation indicates the extent to which two or more sets of data or variables may or may not fluctuate together. When this happens, it does not imply "causation" but rather in correlation, the reason for fluctuation is that the sets of data or variables occur at the same time. For example, when there is a strong relationship between the number of people who smoke and who also drink alcohol, it does not mean there is a causal relationship between the two but rather one is associated with the other in a particular way. Therefore, in correlation the focus is not on cause but on the strength and also the direction of the relationship.

Possible correlations can range from +1 to -1. A zero correlation indicates there is no relationship between the data or variables. A correlation of -1 points to a perfect negative correlation, meaning as one goes up, the other goes down. For example, the more miles a person drives, the less fuel there is in the car. A correlation of +1 implies a perfect positive correlation, meaning both move together in the same direction. For example, as people age, their height increases.

These scatterplots of data below show different correlational directions. Also noted are the strength of each. The closer to +1 or -1, the greater the strength of the relationship.

---

Statistically, correlation is appropriate for examining relationships between quantifiable data (e.g., temperature, weight) rather than categorical data such as gender or a person's favorite sport and it is measured using a coefficient of correlation. The statistical tool most familiar is Pearson Product Moment Correlation where a linear relationship (as above) is examined. Here interval level data are required. When measured in a population Pearson Product Moment Correlation is labeled by use of rho (\( \rho \)). When in a sample, by the letter "r"; sometimes called Pearson's r. For example, a resulting correlation coefficient r might be +.85 which would be interpreted as a strong positive correlation.

There are limitations to using correlation.

- While helpful in analysis, it does not completely speak to all one wants to know about the data and results.
- Outlying data/variables can strongly influence the correlation coefficient.
- It does not work as well with curvilinear relationships, those that do not follow a straight line. For example, if the relationship being examined is age and the use of complementary therapies, while they may be related, the relationship may not follow a straight line. Younger adults may use a variety of different therapies than those who are older.
- If the sample size is small, correlation should not be used alone. Here, significance level is important because it will provide information about how likely the correlation may be due to chance because of a random sampling error.

While, Pearson Product Moment tends to be the most used statistical tool when examining correlational relationships, there are others that are appropriate for different levels of data. If there are two ordinal level variables, Spearman Rank Correlation Coefficient (rho) or the Kendall Rank Correlation Coefficient (tau) can be selected.

Correlation is one of many ways of reporting research results. While, it determines the strength and direction of a relationship that is most helpful, it can also offer a launching point for more detailed analyses in later studies.