UNIT 7, Text Lecture: Correlation

This unit does NOT cover a hypothesis test. Rather, it covers the Pearson Correlation. To this point in the course, we have covered some experimental methodologies that allow researchers to conclude a **cause and effect relationship between an independent and dependent variable.**

For example, if a researcher wishes to test whether caffeine **CAUSES** some effect on concentration, then the researcher would need an **EXPERIMENTAL** design, such as the between-subjects or within-subjects designs, which, again, we have previously covered. The researcher would collect data from different “groups” in these experimental designs and analyze the data with the appropriate hypothesis test in order to be able to conclude that the IV has **caused** an effect on the DV.

However, researchers aren’t always interested in **cause and effect relationships** between variables. As a matter of fact, oftentimes, the necessary experiment to establish a cause and effect relationship between two variables **is impossible, especially when the participants are human beings**.

Think about it for a moment: What if a researcher wanted to establish a cause and effect relationship between eating red meat and cancer. This experiment would clearly be unethical because we cannot set up an experiment in a way that **causes** people to get cancer. Or, imagine that a researcher hypothesizes that hippocampal damage will cause changes in memory. Clearly, a researcher can’t damage a person’s hippocampus (a structure in the brain) and then measure its effect on memory: Ethics would not allow that experiment because it is highly unethical to intentionally subject research participants to experimental manipulations that would harm them.

When experiments are impossible

When ethics precludes an experiment or when a researcher is not interested in demonstrating a causal relationship between 2 variables, but rather ONLY wants to know if there is a relationship between two variables (X and Y), the researcher can conduct a correlational study.

For example, a researcher may wish to know whether the years spent in post-high-school education is correlated with IQ; that is, the researcher wants to know if there is a relationship between the number of years of post-high-school education and IQ. In other words, are these two variables “co-related” or “correlated”? Are the changes in one variable (IQ) changing systematically with the other variable (formal education)? In this case, a researcher may, for example, select a sample of n=100 adults, maybe between 30 and 50 years of age, and collect **two** measures on each person:

X = number of post high-school education years and **Y**= IQ

The researcher would wind up **with an X and a Y measure for each person in her sample**. A common **descriptive statistical** method to analyze data such as these is:

Pearson-Product-Moment Correlation or, commonly called, the Pearson r

Using Pearson r (**r** is the statistical notation), the researcher can analyze the data to find out:

1. Whether there is a linear relationship between X and Y
2. The DIRECTION of the relationship, and
3. The DEGREE of the relationship

Let's discuss this a bit further

Let's use an example to illustrate what is meant by DIRECTION and DEGREE of a relationship between two variables, X and Y.

Imagine a researcher was interested in knowing whether there is a relationship between S.A.T. scores and college G.P.A. The researcher would get a random sample of graduating college students and measure each student on both variables: X=SAT and Y=GPA. Notice that each student would be measured twice; that is, one measure would be his or her SAT score and the other measure would be his or her final college GPA.

Let us assume the following data:

X (SAT)	Y (GPA)
900	2.00
1000	2.5
950	2.1
1200	3.0
1100	2.7
1150	3.1
1400	3.6
1550	3.7

The researcher can run a Pearson r on SPSS, and if she did, the analysis would result in a **signed number that can only range between -1 and 1.** Let us be clear that everyone can "see" where this range of numbers is located on a number line:

-1 -.9 -.8 -.7 -.6 -.5 -.4 -.3 -.2 -.1 0 .1 .2 .3 .4 .5 .6 .7 .8 .9 1

Correlation values (Pearson r's) CANNOT be outside the limit of the above range of numbers. Notice that the smallest correlation you can get is ZERO and the largest is 1. If you get a correlation value (r) bigger than 1, then you have made an error.

The DEGREE of the Relationship Between X and Y

How strong a correlation (the relationship between X and Y) is, its **DEGREE,** its **STRENGTH,** is indicated by the **numerical value** of Pearson r. The higher the numerical value (the closer to 1), the STRONGER THE DEGREE of relationship between X and Y.

Importantly, the DEGREE of the relationship (how strong or weak) is **independent of the sign** of the correlation:

A correlation of **-.5 (negative .5)** is of **equal** strength to a correlation of **.5 (positive .5)**

Again, the numerical value provides the degree (how strong) the relationship is **no matter if it is a negative or a positive numerical value**.

Correlational values of **0** (r = 0) means that there is no relationship at all between X and Y. Correlational values near 0 indicate very weak correlations. As the correlational values move away from zero toward 1—either in the positive direction or negative direction—the strength or DEGREE of the relationship between X and Y gets stronger.

Examples

A Pearson r = **-.6** and r = **.6** are equal in strength despite one being a negative correlation and one being a positive correlation.

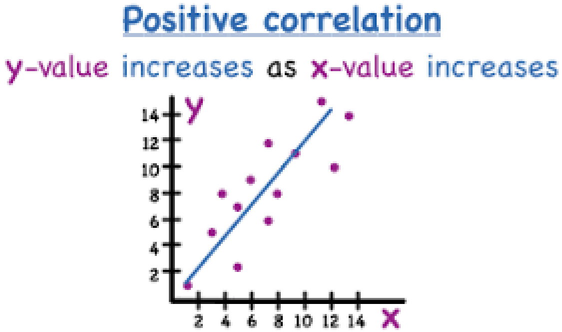
A Pearson r = **-.9** correlation is much, much stronger than an r = **.5** correlation.

The “trick” is to ignore the sign when you are making a decision about the strength, or degree, of the relationship between X and Y.

The DIRECTION of the Relationship Between X and Y

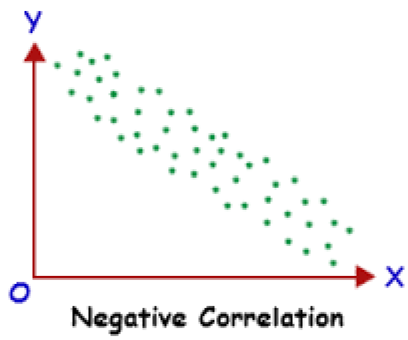
The DIRECTION of the relationship is provided by the SIGN of the correlation. For example, if the correlation is positive, it means that both variables, X and Y, in our case, SAT and GPA, are moving in the SAME direction: as SAT scores get higher, GPA’s get higher. Or, as SAT gets lower, GPA also gets lower. A positive correlation means that BOTH X and Y are moving in the SAME direction. As X goes up, Y goes up. As X comes down, Y comes down.

On a scatterplot, a positive correlation would look like this:



If the correlation is negative, it would mean that the variables X and Y are moving in OPPOSITE directions: as SAT scores increase, GPA's decrease or as SAT scores decrease, GPA's increase.

On a scatterplot, a negative correlation would look like this:



The **sign** of the correlation **only** indicates **direction**.

Much more detail about scatterplots (the graphs above) is provided in the LectureVideo in this unit; therefore, please make sure to review the LectureVideo more than once.

CORRELATION AND PREDICTION

A very useful aspect of correlational analysis is that it can be used to make PREDICTIONS. Consider, once again, the relationship between S.A.T. scores and college G.P.A., and you might understand why colleges request your SAT scores when you apply for admission. The reason is because researchers have found a relationship between SAT and final college GPA, which allows colleges to use your SAT scores to PREDICT your college GPA's (success in college). Again, the idea here is that if there is a relationship between X (SAT scores) and Y (college GPA), colleges can predict college success based on SAT scores.

More on PREDICTION

A correlation of zero (0) has zero predictive power. For example, if we had gotten an r = 0 for the SAT/GPA data, then there would be no relationship between SAT and GPA: these variables (X, SAT scores and Y, college GPA) do not move together in any systematic way. Because they do not move together systematically, knowing a student's SAT would allow us to predict nothing about future college GPA.

The closer the correlation value gets to 1, the higher the predictive ability. As a matter of fact, if we got a correlation of 1 (r = 1, also known as a perfect correlation), and we knew a student's SAT score, we would be able to predict, with 100% accuracy, that students final college GPA!! We would be psychics!! Clearly, we shouldn't expect correlations of 1 in psychological research.

Review the VideoLecture for further detail and explanation ☺