

LECTURE 8: CORRELATION

So far, we have looked at the statistical tests examining causal relationships and differences between variables. In this lecture, we cover statistical tests examining the relationship between variables. In statistics, correlation refers to statistical tests that investigate the degree to which two or more variables are linked to each other. For example, a correlational study in Applied Linguistics would look at the relationship between the number of years living abroad and ESL/ EFL learners' pronunciation accuracy. "The statistical procedure to achieve [this purpose] is called '*correlation analysis*' and it allows us to look at two variables and evaluate the strength and direction of their relationship." (Dörnyei, 2007, p. 223)¹. "Correlation research attempts to determine the relationship between or among variables; it does not determine causation" (Mackey & Gass, 2015, p. 284)². In other words, the fact that two variables are linearly linked to each other does not mean that one is causing the other. For example, just because British people spend more money in shops during the winter, it does not mean that the cold temperature is causing people in the UK to spend more money. It just happened that the winter season coincides with the British holiday season, a period that is known for increased spending.

I. Detecting a correlation

Generally, correlations can be detected or noticed when summarizing and describing data. For example, the table below reports the percentage of public library members in relation to their social class.

Percentage of public library members by their social class origin

PUBLIC LIBRARY MEMBERSHIP	SOCIAL CLASS STATUS	
	<i>Middle class</i>	<i>Working class</i>
Member	86%	37%
Non-member	14%	63%
Total	100%	100%

(Adapted from Cohen, Manion, & Morrison, 2007, p. 529)³

However, for a more accurate detection and measurement of correlation, two main statistical tests are used, namely: the 1) **Pearson product-moment correlation** (or the Pearson correlation) for interval and ratio data and 2) **Spearman rank order correlation** (or the Spearman's correlation) for ordinal data. Correlational tests are designed to answer the following questions:

- 1) *Is there a relationship between two (or more) variables?*
- 2) *What is the direction of the relationship?*
- 3) *What is the strength of the relationship?*

1. Pearson correlation

"Pearson product-moment correlation is the standard type, computed between two continuous variables. When we talk about 'correlation' in general, this is what we usually mean. The Pearson product-moment correlation coefficient is symbolized by the lower-case letter '*r*'". (Dörnyei, 2007, p.224)¹. This test is used when the data of the variables being correlated meet the conditions for parametric tests (i.e., 1) **numerical data**, 2) **large sample size**, and 3) **normal distribution of data**).

2. The Spearman rank order correlation

The Spearman correlation is the non-parametric equivalent for the Pearson correlation. "As the name suggests, the correlation is based on the ranks of the data (in a rank order) rather than the actual [numerical] values"

¹ Dörnyei, Z. (2007). Research methods in applied linguistics. Oxford university press.

² Mackey, A., & Gass, S.M. (2015). Second Language Research: Methodology and Design (2nd ed.). Routledge. <https://doi.org/10.4324/9781315750606>

³ Cohen, L., Manion, L., & Morrison, K. (2002). Research methods in education. Routledge.

(Dornyei, 2007, p. 230)¹. In other words, this test is used when the data of the variables being correlated do not meet the conditions for parametric tests. The Spearman correlation test is appropriate when the data are ordinal (rather than numerical), the sample size is very small, and the data are not normally distributed. The Spearman rank order correlation is represented with the Greek letter ' ρ '.

II. Interpreting correlation test results

When investigating correlation using the Pearson or the Spearman correlation tests, a coefficient is generated. This coefficient indicates the direction as well as the strength of a relationship between two variables.

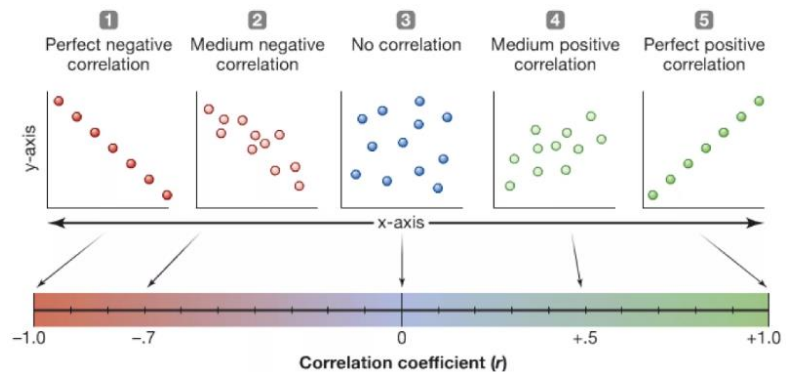
1. The direction of a correlation

A correlation can either be positive or negative. Where the two variables (or sets of data) fluctuate in the same direction, i.e., as one increases so does the other, or as one decreases so does the other, a **positive correlation** (i.e., a positive relationship) is said to exist. Correlations reflecting this pattern are prefaced with a plus sign (+) to indicate the positive nature of the relationship. Thus, **+1.0** would indicate a perfect positive correlation.

On the other hand, a **negative correlation** is to be found when an increase in one variable is accompanied by a decrease in the other variable. Negative correlations are prefaced with a minus sign (-). Thus, **-1.0** would represent perfect negative correlation.

2. The strength of a correlation

Researchers are often interested in the strength of a correlation rather than its direction. Theoretically speaking, correlation tests can give us *no correlation, low correlation, medium correlation, strong correlation, or a perfect correction*. The following figure provide a visualization of the possible correlation magnitudes.



However, “perfect correlations of +1.00 or -1.00 [or even 0.00] are rarely found and most coefficients of correlation in social research are around +0.50 or less”. (Cohen, Manion, & Morrison, 2007, p. 530)³.

PRACTICE

Calculate the Pearson correlation with the following dataset.

Participants	Attendance (x)	Marks (y)
Student 1	4	12
Student 2	7	16
Student 3	2	8
Student 4	5	12
Student 5	3	10
Student 6	6	14
	$\Sigma =$	$\Sigma =$
	$\bar{x} =$	$\bar{y} =$
	$r =$	