Psychology Factsheets

www.curriculum-press.co.uk Number 28

CORRELATION AND CAUSE

This Factsheet discusses what is meant by experimental and correlational studies and how correlation is measured.

Experiments

In an **experiment**, the psychologist manipulates the **independent variable** to observe the effect on the **dependent variable**. There are two ways to do this:

1. Actually **creating a difference** in subjects' experiences eg
- allocating children in different reading schemes
- playing/not playing music to subjects in a recall experiment
- giving/not giving subjects a drink of coffee in a reaction time experiment

2. **Grouping** subjects according to **pre-existing differences** eg
- grouping children according to spelling ability
- grouping people by sex

The experiment explicitly addresses **cause and effect** - the experimenter is testing whether changing the **independent variable** causes a change in the **dependent variable**. Examples include:
- whether a particular reading scheme improves children's spelling ability
- whether a particular type of music being played improves recall
- whether caffeine decreases reaction time
- whether children who spell well benefit more from a particular reading scheme than those who do not
- whether men have better spatial ability than women.

The data from these experiments are **differences** in the dependent variable between people allocated to different experimental conditions.

An experiment involves deciding between a **null hypothesis** (H_0) and an **alternative** or **experimental hypothesis** (H_1).
- H_0 is that the independent variable has **no effect** on the dependent variable (so any fluctuations are due to chance)
- H_1 is the relationship that the experimenter predicts

A suitable statistical test (see Factsheet 03) can be used to decide between the hypotheses.

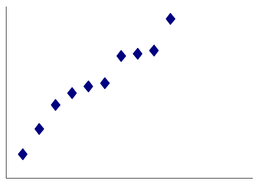
> Example - an experiment to determine whether men have better spatial ability than women
>
> *The **independent variable** is **sex***
> *The **dependent variable** is **spatial ability***
> *The data collected are measurements of spatial ability from groups of each sex*
> *The **null hypothesis** is that sex has no effect on spatial ability*
> *The **experimental hypothesis** is that men have better spatial ability than women.*

There are many types of experimental design; (will be dealt with in future Factsheets)

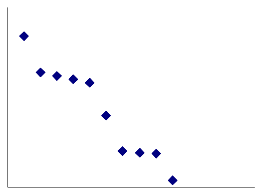
Correlational Studies

In a correlational study, the psychologist is looking to see whether two variables are linked (or **correlated**). This could mean that high values for one are consistently associated with high values for the other, while low values for one are associated with low values for the other (positive correlation), or that high values for one variable are associated with low values for the other (negative correlation).

A **scatter diagram** is one of the easiest ways to show correlation diagrammatically; the values of the two variables for each individual are plotted as coordinates, and an "upward sloping" line shows positive correlation, with a "downward sloping" line showing negative correlation



positive correlation



negative correlation

In a correlational study, the psychologist is not manipulating either of the variables. So we do not refer to an independent and a dependent variable - instead we can call the two variables **covariables.**

It is important to note that because neither variable is manipulated, we **cannot refer to cause and effect**. This is because if we do find a correlation, we can generally give at least three possible "causal" explanations:
- change in variable A causes change in variable B
- change in variable B causes change in variable A
- changes in both variables are both a consequence of some external factor

For example, if a study found there was a negative correlation between the number of hours of television watched by children and their scores on on school examinations, we might reasonably suggest:
- watching a lot of TV reduces children's study time, hence reducing their test scores (or similarly, it means they sleep less, or it reduces their concentration span...)
- children who do badly in school examinations are likely to be demotivated, and hence spend more time watching TV as opposed to doing school work.
- both variables are due to home circumstances - parents who place a high value on education are likely to permit less TV viewing and both support and promote achievement at school.

You can no doubt come up with many more explanations. But it is very important to remember that a correlational study cannot support one of these any more than any of the others!

Why use correlational analysis rather than an experiment?

In some cases, it would be **unethical** to manipulate a variable - for example, when studies were done on the link between cigarette smoking and lung cancer, it would not have been ethical to encourage people to smoke (although it would be possible to place people into groups according to their existing smoking habits).

There are a number of cases also where it is inherently not sensible to regard one variable as "causing" changes in the other. For example, if you wished to examine the relationship between participants' verbal and non-verbal IQ scores, it would not be reasonable to talk about a change in their verbal IQ score "causing" a change in their non-verbal IQ score, or vice versa.

Correlation is also useful because it looks at a relationship between the variables across the whole range of values. Suppose you were examining the relationship between a child's IQ score and their score on a mental arithmetic test. The experimental approach would be to group children according to IQ score, and then examine the differences in mental arithmetic score (we would consider IQ to be the independent variable, as it is meant to be a fixed quantity for a given individual, and a measure of their general intelligence). Although this might well give us a significant result, we have actually lost a lot of useful information from the data by grouping in this way - a correlational approach is likely to be more sensitive and informative.

Requirements for correlational analysis

The same requirements for sampling methods apply for a correlational study as for any other - your conclusions are only as good as the data you gather. In order to carry out any statistical test, at least **five** individuals must be considered - more is advisable. Further, your data must be **ordinal** - i.e. able to be put in rank order.

Measuring correlation and testing its significance

Correlation is measured on a scale of -1 to 1.

-1	0	1
perfect negative correlation	no correlation	perfect positive correlation

← *increasing negative correlation*   *increasing positive correlation* →

The hypotheses used in any test for correlation are:

- H_0: there is no correlation between the two variables
- H_1: there is a positive correlation between the two variables

OR there is a negative correlation between the two variables
OR there is some correlation between the two variables

The form of the alternative hypothesis should be chosen by the psychologist **before commencing the study**. One of the first two forms (**directional** hypotheses) should be chosen if there is a sound psychological reason for expecting a particular type of correlation (for example, we would normally expect positive correlation between verbal and non-verbal IQ scores). If it is unclear what type of correlation to expect, the third form - the **non-directional** hypothesis - should be used.

The correlation coefficient obtained is significant if it is greater (ignoring signs) than the tabulated value at the appropriate degree of significance.

Correlation coefficients

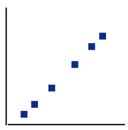
These are the measures of correlation that you are likely to encounter:-

- Pearson's product moment correlation coefficient (r)
- Spearman's rank correlation coefficient (ρ)
- Kendall's rank correlation coefficient (τ)

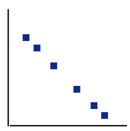
Of these, Spearman's rank correlation coefficient is the most commonly used, and is discussed in detail in Factsheet ??. However, the other two can also be useful.

Pearson's product moment correlation coefficient

Unlike the other two, this is a **parametric** test - it requires that you have **interval** data, that the values are approximately normally distributed and that the variability of each variable is roughly similar. It measures specifically how close to a **straight line** your data lie.



$r = 1$

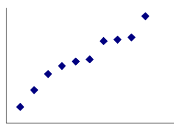


$r = -1$

If the requirements for this test are met, and a scatter graph indicates the data lie close to a straight line, it is worth using this measure, because it is actually taking account of the values of the data, rather than just their order, and hence is a more powerful test. If you have large quantities of data, it is also possible to calculate r automatically on an Excel spreadsheet; if your data were in cells A1 to A100 and B1 to B100, say, then in an empty cell, type in =CORREL(A1:A100,B1:B100).

Rank correlation

Rank correlation is a non-parametric test, and just requires that data be ordinal - in other words, the actual values are irrelevant, provided the data can be ordered. It just measures whether the highest value of one variable goes with the highest value of the other, and the second highest with the second highest etc; this does **not** require that the scatter diagram give a straight line.



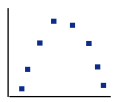
$\rho = 1$
$\tau = 1$
$r \neq 1$

Generally, you would use Spearman's rank correlation coefficient rather than Kendall's, as it is easier to calculate. However, there is one problem with using Spearman's - it is not valid if there are too many **ties -** i.e. if there are a lot of identical values for one or both variables (two or three ties for each variable is OK). In these circumstances, Kendall's is preferable.

Correlation when the scatter diagram gives a curve

Some correlational studies may produce a scatter graph that looks like a curve. In these cases, a rank correlation coefficient can be used provided the curve goes reasonably consistently upwards or downwards.

For other curves, such as the one below, the standard correlation coefficients will not give useful results, although there clearly is some link between the variables. However, suitable calculations for this type of data are beyond the scope of A-level.



Acknowledgements: *This PsychFactsheet was researched and written by Cath Brown.* ***Curriculum Press.*** *Bank House, 105 King Street, Wellington, TF1 1NU. Psychpress Factsheets may be copied free of charge by teaching staff or students, provided that their school is a registered subscriber. No part of these Factsheets may be reproduced, stored in a retrieval system, or transmitted, in any other form or by any other means, without the prior permission of the publisher.* ***ISSN 1351-5136***