



The following resources are associated:
 Choosing the right test, Advanced choosing the right test

Terminology for choosing the right test

Choosing the most appropriate analysis to answer your research question is one of the most difficult aspects of applying statistics to your own research. There are many resources to help you decide but each assumes you understand the terminology which can vary depending on the author. This resource covers the terminology needed for choosing the right test by using the associated sheets or other resources.

In order to choose the correct test, clearly define a research question and decide which variables will be used to answer the research question. For large research projects, you may have several research questions so deal with each individually.

Example dataset 1: To help explain some of the concepts, this example dataset will be used which contains information on babies and their mothers. Each row is a different baby.

Research question: What affects the birthweight of a baby?

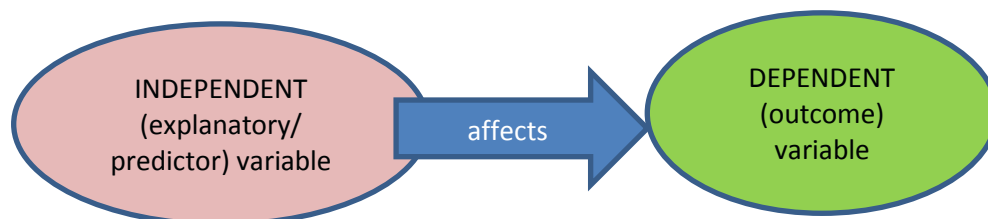
Baby ID	Birth weight (lbs)	Gestational age (weeks)	Mother smokes	Pre-pregnancy weight (lbs)	Previous pregnancies
1313	5.8	33	No	99	2
431	4.2	33	Yes	109	3
808	6.4	34	No	140	0

DATA: Measurements of individuals or answers to questionnaires.

VARIABLE: A characteristic which varies between individuals e.g. birth weight or whether or not the baby smokes.

DEPENDENT (OUTCOME) VARIABLE: The outcome of interest which may depend on or be influenced by other variables of interest e.g. birthweight.

INDEPENDENT (EXPLANATORY/PREDICTOR) VARIABLE: Variables we think are influencing the dependent variable e.g. gestational age or whether or not the mother smokes.

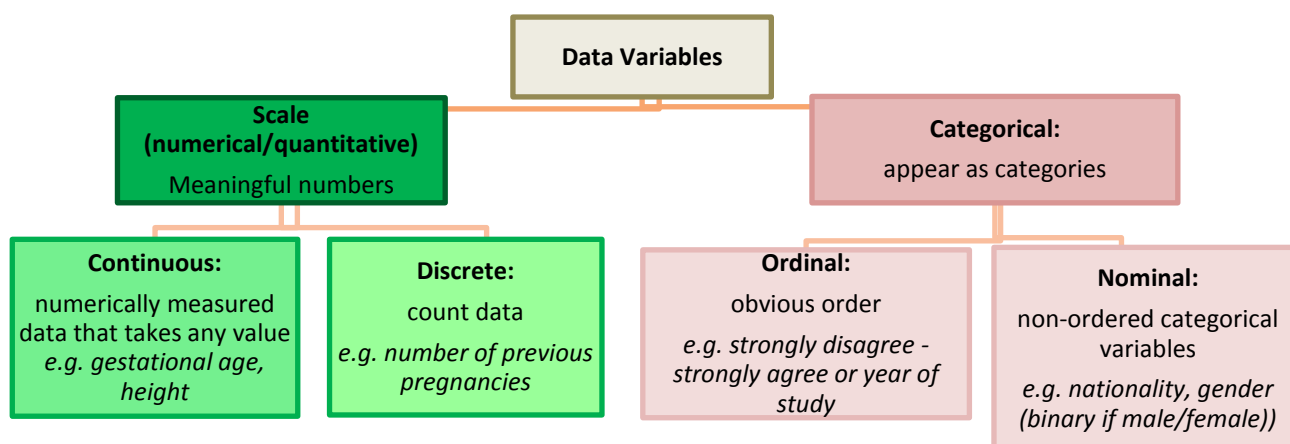


DATA TYPES

In order to choose the most appropriate test, you need to know the type of data you have for each variable. There are two main groups of variables; categorical and scale.

CATEGORICAL VARIABLES: Examples of categorical data are the answers to questions with tick box options or anything where individuals are grouped. If there is a natural order to the groups e.g. opinion questions with options ranging from strongly disagree to strongly agree, the data is ordinal. Nominal data are groups with no natural ordering. Binary is a special case of nominal when there are only two groups e.g. the mother smokes/ does not smoke.

SCALE/NUMERICAL VARIABLES: Scale data can also be referred to as numerical or quantitative depending on the author but they all used when describing meaningful measurements e.g. height. Scale data can be broken down into two further groups; discrete and continuous although the distinction is less important. Discrete data is generally count data e.g. number of previous pregnancies which can only take whole numbers. Although gestational age is measured in whole weeks, it could be potentially be measured using decimals or alternative units so it is classified as continuous.



Issues with ordinal data: Ordinal data is often represented as numbers e.g. 1 = strongly disagree → 5 strongly agree but these are NOT meaningful numbers. For example, when running a race, runners can be classified as 1st, 2nd, 3rd or gold, silver and bronze but this does not take into account the exact differences between runners in the same way that time would. The person who was first may be a long way ahead but 2nd and 3rd place may finish almost together so the gaps between the numbers are not the same.

There are different tests for ordinal variables but some disciplines use tests for scale data on ordinal variables. If there are a large range of options e.g. 7+ and the assumptions of the individual test are met, this could be considered acceptable by some although the interpretation of output can be tricky as the numbers are not meaningful.

In some questionnaires, the answers of sets of related ordinal questions can be combined to represent an underlying latent variable which cannot be measured directly e.g. a score to measure level of depression or self-confidence. Each ordinal question is called an item and the set of related items is called a scale. Usually, each item is numbered e.g. 1 → 5 and an overall scale score is calculated by summing or averaging the responses to the set of items. This score can then be treated as scale (numerical). There are a large number of these scales available which have

been validated and checked for reliability. It is not recommended that you create your own as the validation process is lengthy.

In order to choose the right test, think about how many dependent and independent variables there are, what data types they are, and whether you are wanting to:

- Test for differences e.g. do the babies of smokers weigh less?
- Investigate relationships e.g. is there a relationship between gestational age and birthweight?
- Test for agreement or reliability e.g. do two machines for weighing babies give the same result?

Testing for differences

Tests for differences commonly compare means but you also need to decide whether you have repeated measurements or independent groups and how many you are comparing.

INDEPENDENT GROUPS: This type of design occurs when individuals can only appear in one group and these groups are being compared (also known as between-groups design). Mother's either smoke or do not smoke so can only be in one of these groups. You can have more than one grouping variable e.g. comparisons of the race of the baby as well as whether or not the mother smokes. Smoker is a binary variable but race has more groups.

REPEATED MEASURES: Sometimes you will take multiple measurements of the same variable on each individual either at different time points or under different conditions e.g. measuring cholesterol before and after individuals change to a particular brand of margarine. This can also be known as a within-groups design. The independent variable is either time or condition.

Example: Cholesterol measurements were taken on participants prior to changing to a new margarine (Margarine A), after 4 weeks on the diet and after 8 weeks to test whether using Margarine A. This is a repeated measures design, the dependent variable (Cholesterol) is scale and the independent variable (Time point) is categorical.

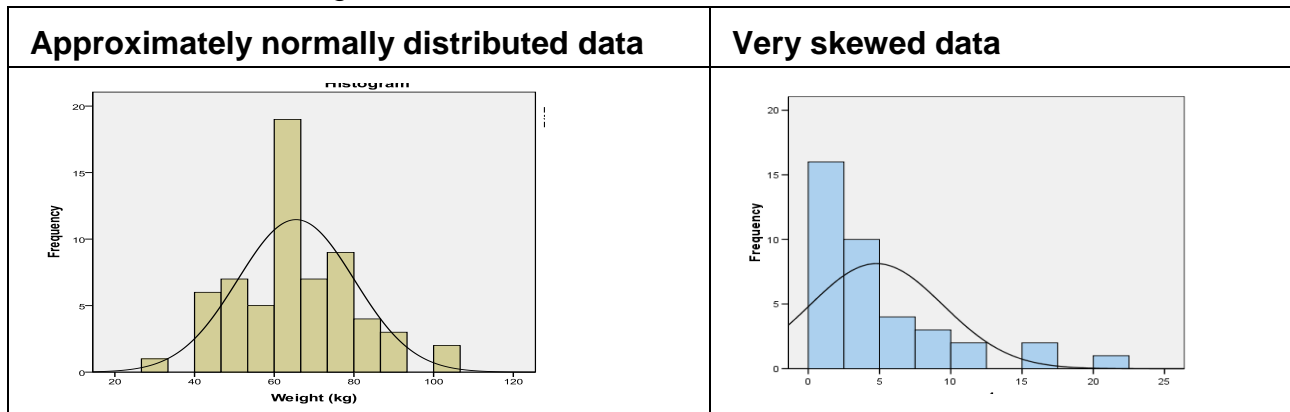
ID	Cholesterol Before	Cholesterol after 4 weeks	Cholesterol after 8 weeks
1	6.42	5.83	5.75
2	6.76	6.2	6.13
3	6.56	5.83	5.71
4	4.8	4.27	4.15

There are tests available for multiple independent variables which can each be either independent groups or repeated measurements. Imagine that the cholesterol study involves two Margarines A and B. If one group has A and the other B, the nominal variable 'Margarine' is an independent grouping variable. However, if each participant has margarine A during one time period and margarine B during another, then margarine is a repeated measurement.

PARAMETRIC TESTS: A parametric test requires the data to meet certain assumptions (see individual test resources for details on the exact assumptions) and if these assumptions are not met, for some tests, a non-parametric alternative can be used. For many common tests, one assumption is that something is normally distributed (see individual test resource and the 'Checking normality' resource for details). For the independent t-test, the dependent variable needs to be normally distributed by group but for other tests; output from the procedure should be checked. For an independent t-test, if the data are very skewed, a Mann-Whitney test can be used instead. When testing means of scale data, choose the test from the 'Parametric' column

and carry out the test. Check the assumptions and if the data are very skewed, choose the non-parametric equivalent from the 'Non-parametric' column instead.

Data only needs to be approximately normal so only worry of a histogram of the data is very skewed to the left or the right.



Investigating relationships

There are specific tests which look for general relationships (associations) between two variables of the same type and there are other tests which allow relationships between many to be summarised by grouping variables together or reducing the number of variables. There are also a group of tests which allow multiple independent variables to be tested simultaneously for one dependent variable e.g. regression. As well as testing relationships, these tests create mathematical equations (models) which allow prediction of the dependent for individuals.

When investigating relationships consider the following:

- Are you interested in the general relationship between two variables?
- Is there a clear dependent variable and one or more independent variables?
- Do you want to produce a mathematical equation (model) to predict values of the dependent?
- Do you want to look for patterns using multiple variables?
- Do you want to reduce or group multiple variables?

Techniques for assessing agreement or reliability

If you wish to compare measurements of the same thing for different observers or equipment, there are special techniques. You can't use a test for differences and assume that a non-significant result indicates that the results are the same. Equally, using correlation to show there is a relationship is not appropriate. The test varies depending on the data type of the measurement of interest.

If you have a set of ordinal questions (items) measuring an underlying variable (a scale), you should check that the scale is reliable using Cronbach's alpha. You should also check the validity of the scale if it is one you have created yourself. Generally, creating your own scale requires a lot of checks so using one created by someone else is advisable.