

Choosing the correct statistical test made easy

N Gunawardana

Senior Lecturer in Community Medicine, Faculty of Medicine, University of Colombo

Gone are the days where researchers had to perform statistical calculations manually. Nowadays, many researchers have access to software which will perform whatever statistical test the researcher wants to perform. However such software does not have the ability to make decisions on selecting the statistical test to be used in a given situation. In other words, they do not have the ability to match the correct statistical test to the correct situation. If the researcher blindly orders the software to perform all possible statistical tests the software will present him/her with a whole array of tests, a mix of relevant and irrelevant. Therefore knowledge on choosing the correct test is a must for the researcher. The aim of this article is to present an easy method to choose the correct statistical test. This method is applicable for both descriptive and experimental study designs.

When should the statistical test to be used be chosen? Is it at the stage of analysis? Ideally the results that need to be presented and statistical tests to be performed to achieve each of the specific objectives should be decided at the planning stage. The method that is presented to you in this article requires the researcher to respond to a checklist of four questions and to follow a selected flow chart until the relevant statistical test is reached.

Following is the checklist of four questions.

- Q1. What scales of measurement has been used?
- Q2. Which hypothesis has been tested?
- Q3. If the hypothesis of difference has been tested, are the samples independent or dependent?
- Q4. How many sets of measures are involved?

Q1. What scales of measurements have been used?

Measurement is assigning numbers to observations. As an example, for weight observations/measurements in research, the researcher may assign a number using relevant

units such as kilogrammes (kg) and this number would reflect the actual magnitude of the weight of the study unit. When observing/inquiring the ethnicity or socio-economic status the researcher may assign a code number to the individual. The basis on which the numbers are assigned to observations determines the scale of measurement being used. Traditional classification of scales of measurement describes three types; nominal scale, ordinal scale and interval scale.

Nominal scale: If the researcher simply uses numbers to label categories which do not represent any order, then the scale of measurement used is the nominal scale. In a descriptive study the researcher may specify the sex of study units and assign number '1' for males and number '2' for females. By doing this, the researcher is not indicating that females are 'more' or 'less' in relation to sex, indicating that categories do not represent any order. In an experimental study, the researcher may use nominal scale to categorize study units in the experimental group and control group based on type of the side effect that they may experience following administering of a drug/placebo. Though the researcher may assign consecutive numbers for different side effects it does not indicate that one side effect is greater or lesser than the other. Nominal scale is considered the lowest form of scale of measurement as it does not provide any information on the relationship between the categories. Results of research data measured in nominal scale would be presented using frequency distributions.

Ordinal scale: If the researcher use numbers to label categories in which an order can be identified, then the scale of measurement used is the ordinal scale. In other words, the relationship between categories in terms of 'greater than' or 'lesser than' status can be identified. In descriptive studies the researcher may categorize people based on their level of satisfaction of a health service provided using categories of 'highly satisfied', 'somewhat satisfied' and 'dissatisfied'. This

categorization also provides information on the relationship between them. In experimental studies we may use ordinal scale to categorize study units in both experimental and control groups based on the type of response that they will experience following administering of a drug using the categories such as ‘very good response’, ‘good response’ and ‘poor response’. Again, this categorization provides information on the relationship between categories. It indicates that ‘very good response’ would be better than ‘good response’ and ‘good response’ would be better than ‘poor response’. However, it should be noted that ordinal scale does not numerically quantify how much greater ‘very good response’ is compared to ‘good response’. Results of research data measured in ordinal scale would be presented using frequency distributions.

Interval scale: When the researcher assigns numbers to observations, if the difference (interval) between two such numbers that are assigned is meaningful (numbers are assigned to weigh observations/measurements and the difference between two such observations/measurements will denote how much ‘greater than’ or ‘lesser than’ one measure is from the other), this scale of measurement is called interval scale. In interval scale the number we assign each observation/measurement represents the actual magnitude of it. The distances between successive points in an interval scale are equal.

In descriptive studies we may measure heights of individuals and assign them values in the interval scale using the unit centimetre (cm). A person who is assigned the number 150 cm we know is taller than the person assigned the number 140 cm and we also know that the difference of height between these two persons is 10 cm. In experimental studies we may use the interval scale to measure the response to a drug among study units in both experimental and control groups in terms of improvement of Haemoglobin level. We would then know that a person who has an improvement of 3 g/dl had responded to the drug better than a person who had an improvement of 1 g/dl and that the difference of improvement is 2 g/dl. We also know that the difference between two persons; one with 2 g/dl improvement and one with 3 g/dl improvement (i.e.: 1 g/dl) is equal to the difference between another two persons; one with 4 g/dl improvement and the other with 5 g/dl improvement.

Results of research data measured in interval scale would be presented using measures of central tendency (mean, mode, median) and dispersion (standard deviation, standard error).

At this stage, when we have determined the type of the scale that has been used to measure the outcome that has to be statistically tested, we can make some decisions regarding the collective group of statistical tests that we may use (Table 1).

Table 1: Groups of statistical tests to be used for data measured using different scales of measurement

Nominal	Chi-square or one of its variation
Ordinal	Ordinal tests / non-parametric tests (U,H,T, Spearmen r)
Interval	Is the measure normally distributed in the population? YES - parametric tests (t, F, Pearson r) NO – consider data as being measured in the ordinal scale. Use ordinal tests / non-parametric tests (U,H,T, Spearmen r)

As shown above, when the scales of measurements are either nominal or ordinal, the groups of statistical tests to be used can be decided without answering any further question. In case of interval scale, we need to answer a further question which inquires whether or not the measure is normally distributed in the population. The normally distributed measures will conform to a normal curve if we do this measurement on the whole population and plot a graph with values of the measurement presented in the X axis and frequency of occurrence presented in the Y axis. The normal curve as Gauss described it (also known as the Gaussian curve) is actually a theoretical distribution. The good news for researchers is that many human related measurements we perform such as weight and height come close to this ideal distribution. As shown in Table 1, if a search of the body of literature indicates that the measure is normally distributed in the population, the researcher should make the decision to use one of the parametric statistical tests to test for significance. If it is found that the measure that the researcher is dealing with

is NOT normally distributed in the population, the researcher has to treat these measurements as measures done using ordinal scale.

Upon answering the first question in the checklist, the researcher is able to pick up the relevant flow chart to be used (Figures 1-3).

Figure 1: Flow chart on statistical tests to be used for data measured in nominal scale

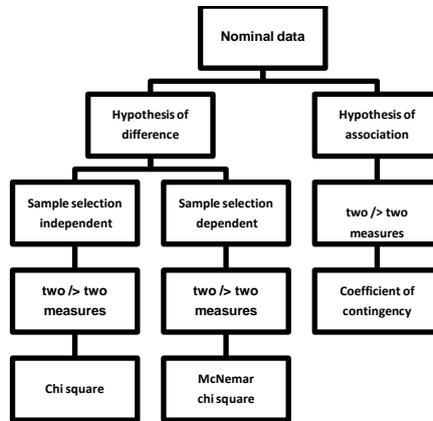


Figure 2: Flow chart on statistical tests to be used for data measured in ordinal scale

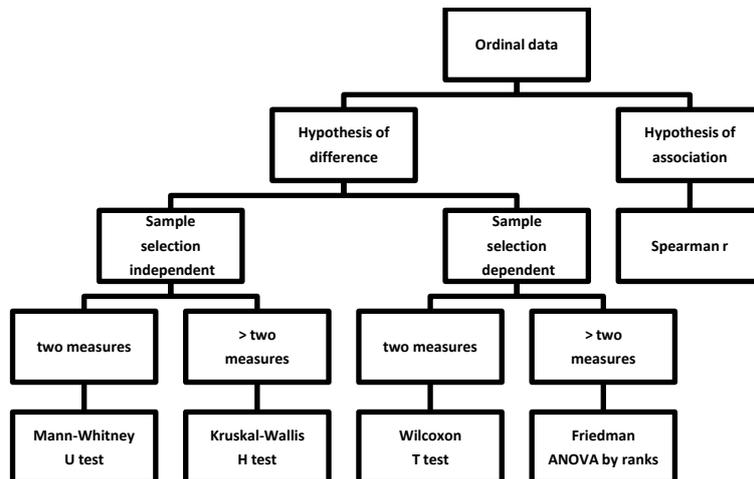
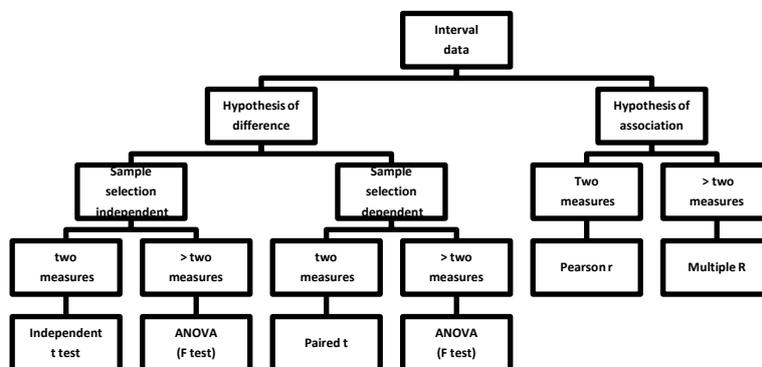


Figure 3: Flow chart on statistical tests to be used for data measured in interval scale



Q2. Which hypothesis has been tested?

The second question to answer is regarding the hypothesis that is to be tested. There are only two types of hypotheses that can be statistically tested in research; hypothesis of difference or hypothesis of association. Hypothesis of difference states that the difference that is shown in the results obtained from the samples are also different in the larger populations from which the samples came. In contrast hypothesis of association states that the relationship of the two (or more) sets of outcome that we see in the results obtained from the sample is also present in the larger populations from which the sample came.

In descriptive studies the researcher may test either the hypothesis of difference or the hypothesis of association. For example, if in a descriptive study, the research question to be answered by statistical testing is ‘whether there is a difference in the prevalence of childhood malnutrition in urban and rural sectors’ or ‘whether the mean heights of the three groups of basketball players are different’ the hypothesis to be tested is hypothesis of difference. In contrast to this, if the research question to be answered by statistical testing in a descriptive study is ‘whether there is a relationship between stunting and wasting of pre-school children in rural sector’ it indicates that the hypothesis of association is being tested. Testing of hypothesis of association involves measurements of two or more sets of outcome within a single sample whereas the testing of hypothesis of difference will always involve a measurement of a single outcome made on two or more samples.

Experimental research should always test only the hypotheses of difference. For example, the research question to be tested using statistical testing in an experimental study will be ‘whether the outcomes are different in the study and control groups’.

Once the researcher answers this second question in the checklist, he/she is now able to pick up the path to follow in the chosen flow chart (Figures 1-3).

Q3. If the hypothesis of difference has been tested, are the samples independent or dependent?

The third question is applicable only if the hypothesis of difference is being tested. As indicated earlier, testing for hypothesis of

association involves measurement of two or more sets of outcomes within a single sample; hence this checklist question becomes irrelevant in such studies.

In the instances where the hypothesis of difference is being tested, if the selection of one of the samples is in any way influenced by the selection of the other samples we call them dependently selected. An example of dependent sample selection will be when ‘matching’ criteria have been used in selecting the groups to be tested in either descriptive or experimental studies. Experimental studies in which the same subject acts as his/her own control (within group designs), are also another example of an instance in which samples are dependent on each other. Following checking on selection of sample, if the researcher is convinced that selection of one sample has in no way influenced the selection of the other sample, he/she should consider them as independent samples. In descriptive studies or in experimental studies if the samples are selected using random sampling techniques, this will indicate that the samples are independent.

Upon answering this third question in the checklist, the researcher is only one step away in the path to choose the correct statistical test (Figures 1-3).

Q4. How many sets of measures are involved?

The last question to answer is the easiest. It is about the number of sets of groups or outcome measures that are involved in the analysis. The question inquires whether the hypothesis of difference is being tested on only two groups or whether it is more than two. For example, in a descriptive study the research question to be answered by statistical testing is ‘whether there is a difference in the prevalence of childhood malnutrition in urban and rural sectors’, it indicates that two groups are being tested. If an experimental study involves three groups (one experimental group and two control groups) and if the research question is ‘whether the response to the drug is different among the three groups’, it indicates that more than two groups are being tested. Similarly in a descriptive study if hypothesis of association is being tested and the question to be answered by statistical testing is ‘whether there is a relationship between stunting and wasting of pre-school children in rural sector’ it indicates that two sets of outcomes are being tested. If the research question is to test ‘whether

there is a relationship between stunting, wasting and head circumference of pre-school children in rural sector', it indicates that more than two outcomes are being tested.

Answering this fourth question in the checklist, and following the flow chart appropriately will now allow the researcher to choose the correct statistical test to be used (Figures 1-3).

By following these simple principles a researcher will be able to apply the most appropriate statistical test to a given situation thus enabling the researcher to analyze and present data in the best possible way.