16.3.1 Definition of Chi-square (χ^2)

Chi-square (χ^2) is a descriptive measure of the magnitude of the discrepancies between the observed and expected frequencies. χ^2 may be defined as the sum of ratios of squared deviations of observed frequencies (f_o) from the corresponding expected frequencies (f_e), and the respective f_e values. Thus, the formula for the chi-square statistic (χ^2) is

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

where

f_o = observed frequencies of a category
f_e = expected frequencies of the corresponding category
Σ = directs one to sum over all the categories

As the above formula indicates, the value of χ^2 is computed by the following steps:

(i) Find the difference between f_o and f_e of each category.
(ii) Square the difference ($f_o - f_e$). This ensures that all values are positive.
(iii) Divide the squared difference of a particular category by its respective f_e.
(iv) Finally, sum the values from all the categories.

An alternative formula for chi-square reads as:

$$\chi^2 = \sum \frac{(f_o)^2}{f_e} - N$$

16.3.2 Characteristics of Chi-square (χ^2)

Inspection of the above formula for χ^2 leads to the following characteristics of the chi-square statistic:

(a) Since χ^2 is a descriptive measure of the magnitude of the discrepancies between the observed and expected frequencies, the greater the discrepancy between the f_o and f_e, relative to f_e, the greater is the amount of χ^2. If no discrepancies exist and the observed and expected frequencies are the same, χ^2 will be 0.

(b) χ^2 is always positive, since the difference between f_o and f_e is squared. The reader should note also that χ^2 is always 0 or a positive number. Negative values cannot occur.

(c) The magnitude of χ^2 varies from series to series, owing to the variations in the observed frequencies, the expected frequencies being constant over the categories. For example, one investigator tossed a coin 100 times and got 60 heads and 40 tails. Another investigator tossed a coin 100 times and got 45 heads and 55 tails. In both the cases, the expected frequencies are 50 in each category of responses, but the amount of χ^2 will be different in both the series. It is because of the different observed frequencies in the two categories of both the series.

(d) Different categories of responses are not independent. If the frequency of one category is known, the frequency of other category is determined. For example, in the tossing of 100 coins two frequencies are obtained, one for heads and one for tails. These frequencies are not independent. The frequency of tails is $100-45=55$, if the frequency of heads is 45. If the frequency of heads is 60, the frequency of tails is $100-60=40$. Thus, it is quite clear that, given either frequency, the other is determined.

(e) The typical chi-square distribution is positively skewed, and it is a continuous probability distribution at its right tail. The shape and form of the distribution depend on the df of the chi-square.

16.3.3 Assumptions of Chi-square (χ^2)

Since χ^2 is a non-parametric statistical test, most of the assumptions of non-parametric tests are also the assumptions of χ^2. The following are some of the important assumptions of χ^2 test:

(a) Observations must be independent. It implies that the selection of the samples from the population and the assignment of scores to the observations must be unbiased or impartial.

(b) The variable under study must have underlying continuity—nowhere in the sample or population, the trait is absent.

(c) The measurements or data should be in terms of nominal (categorical) scale of measurement. For example, the subjects are required to give their responses in different categories, such as 'Yes, Neutral or No', 'Favourable, Neutral or Unfavourable', 'Positive, Neutral or Negative', etc.

(d) The data are in terms of frequencies, percentage or proportions.

(e) Number of samples should be less so that maximum df would be 30 [$df=c-1$ or $df=(r-1)(c-1)$].

16.2.1 Parametric Statistical Tests (PST)

You will recall from our discussion in Chapter 1 that a parameter is a characteristic or a measure of a population. A parametric inference test is one that depends considerably on population characteristics, or parameters for its use. Thus, a parametric statistical test is defined as, 'A parametric statistical test is a test whose model specifies certain conditions about the parameters of the population from which the research sample was drawn.' Since these conditions are not ordinarily tested, they are assumed to hold. The meaningfulness of the results of a parametric test depends on the validity of these assumptions. The following are some of the assumptions of the parametric tests.

16.2.1.1 *Assumptions of Parametric Statistical Tests*

The assumptions of the parametric statistical tests are:

(i) The observations must be independent. That is, the selection of any one case from the population for inclusion in the sample must not bias the chances of any other case for inclusion, and the score which is assigned to any case must not bias the score which is assigned to any other case. In brief, independence of observation implies two things, that is, selection of samples from the population must be random, and assignment of scores to the observation must be unbiased.

(ii) The observations must be drawn from normally distributed populations. In other words, the trait or the dependent variable that the researcher is going to measure must be normally distributed in the population from which the research samples are drawn.

(iii) These populations from which the research samples are drawn must have the homogeneity of variances or same variance (or in special cases, they must have a known ratio of variances).

(iv) The variables involved must have been measured in an interval or ratio scales of measurement, so that it is possible to use the operations of arithmetic (adding, dividing, finding means, etc.) on the scores.

(v) The populations from which the research samples are drawn must have homoscedasticity of variances, which means the equality of variances between columns and rows. The means of these normal and homoscedastic populations must be linear combinations of effects due to columns and/or rows. That is, the effects must be additive.

All the above conditions [except Assumption (iv), which states the measurement requirement] are elements of the parametric statistical model. Assumption (v) regarding the homoscedastic population is found only in case of the analysis of variance model (the *F*-test). With the possible exception of the assumption of homoscedasticity (equal variances) these conditions are ordinarily not tested in the course of the performance of a statistical analysis. Rather, they are presumptions which are accepted, and their truth or falsity determines the meaningfulness of the probability statement arrived at by the parametric test.

When we have reason to believe that these conditions are met in the data under analysis, then we should certainly choose a parametric statistical test, such as '*t*' or *F*, for analysing those data. Such a choice is optimum because the parametric test will be most powerful for rejecting H_0 when it should be rejected.

But when the assumptions constituting the statistical model for a test are in fact not met, or when the measurement is not of the required strength, then it is difficult if not impossible to say what is really the power of the test. It is even difficult to estimate the extent to which a probability statement about the hypothesis in question is meaningful when that probability statement results from the unacceptable application of a test. Although some empirical evidence has been gathered to show that slight deviations in meeting the assumptions underlying parametric tests may not have radical effects on the obtained probability figure, there is as yet no general agreement as to what constitutes a 'slight' deviation.

16.2.2 Non-parametric Statistical Tests (NPSTs)

A non-parametric statistical test, on the contrary, is a test whose model does not specify conditions about the parameters of the population from which the research samples are drawn. For non-parametric statistical tests, the shape of the population is unimportant. For example, it is not necessary that the samples be random samples from normally distributed populations as with the parametric statistical tests. For non-parametric statistical tests all that is necessary is that the sample scores be random samples from population having the same distributions. For this reason, non-parametric inference tests are sometimes referred to as *distribution-free tests.*

16.2.2.1 *Assumptions of Non-parametric Statistical Tests*

The following are certain assumptions which are associated with most non-parametric tests.

(i) The observations are independent. The independence of observation implies two things—one, the selection of samples from the population must be random and, second, assignment of scores to the observation must be unbiased.

(ii) The trait or the dependent variable under study has underlying continuity. This means that the variable under study must be present in any amount in each and every element of the population from which the research samples are drawn; nowhere in the population, it is absent.

(iii) The variables involved must have been measured in an ordinal scale or in a nominal scale. In other words, the given data should be in terms of ranks or in categories. Non-parametric tests do not require measurement so strong as that required for the parametric tests; most non-parametric tests apply to data in an ordinal scale and some apply also to data in a nominal scale.

16.2.2.2 *Advantages of Non-parametric Statistical Tests*

The following are some of the advantages of the non-parametric statistical tests.

(i) Probability statements obtained from most non-parametric statistical tests are *exact* probabilities (except in the case of large samples, where excellent approximations are available), regardless of the shape of the population distribution from which the random sample was drawn. The accuracy of the probability statement does not depend on the shape of the population, although some non-parametric tests may assume identity of shape of two or more population distributions and some others assume symmetrical population distributions. In certain cases, the non-parametric tests do assume that the underlying distribution is continuous, an assumption which they share with parametric tests. In brief, exact probabilities can be determined by the help of non-parametric statistical tests.

(ii) If the sample size is small, as small as $n=6$ are used, there is no alternative to using a non-parametric statistical test unless the nature of the population distribution is *known exactly*.

(iii) There are suitable non-parametric statistical tests for treating samples made up of observations from several different populations. None of the parametric statistical tests can handle such data without requiring us to make seemingly unrealistic assumptions.

(iv) Non-parametric statistical tests are available to treat the data which are inherently in ranks as well as data whose seemingly numerical scores have the strength of ranks. That is, the researcher may only be able to say of his/her subjects that one has more or less of the characteristic than another, without being able to say how much more or less. For example, in studying such a variable as anxiety, we may be able to state that subject A is more anxious than subject B without knowing at all exactly how much more anxious A is. If data are inherently in ranks, or even if they can only be categorised as plus or minus (more or less, better or worse), they can be treated by non-parametric methods, whereas they cannot be treated by parametric methods unless precarious and perhaps unrealistic assumptions are made about the underlying distributions.

(v) Non-parametric statistical tests are available to treat data which are simply classificatory or categorical, that is, are measured in a nominal scale. No parametric statistical technique applies to such data.

(vi) Non-parametric statistical tests are typically much easier to learn and to apply than are parametric statistical tests.

16.2.2.3 *Disadvantages of Non-parametric Statistical Tests*

In spite of the so many advantages as cited above, the non-parametric statistical tests have the following disadvantages:

(i) If all the assumptions of the parametric statistical model are in fact met in the data, and if the measurement is of the required strength, then non-parametric statistical tests are wasteful of data, because the primary objectives of research are: (a) prediction about future outcomes and (b) drawing inference about population distribution. These two objectives are not met with non-parametric statistical tests.

(ii) The degree of wastefulness is expressed by the power efficiency of the non-parametric test. It will be remembered that if a non-parametric statistical test has power efficiency of, say, 90%, this means that where all the conditions of the parametric statistical test are satisfied the appropriate parametric test would be just as effective with a sample which is 10% smaller than that used in the non-parametric analysis.

(iii) Interaction effects among variables or factors cannot be measured by the help of non-parametric statistical tests. In other words, there are as yet no non-parametric methods for testing interactions in the analysis of variance model, unless special assumptions are made about additivity.

(iv) Another objection that has been raised against non-parametric methods is that the tests and their accompanying tables of significant values have been widely scattered in various publications and appear in different formats; many are highly specialised and they have, therefore, been comparatively inaccessible to the behavioural scientists.

16.2.3 Difference between Parametric and Non-parametric Statistical Tests

The following are some of the important points of difference between parametric and non-parametric statistical tests or simply statistical inference tests.

(i) A parametric statistical test depends considerably on population characteristics or parameters, for its use, whereas a non-parametric statistical test does not depend so much on the population characteristics or parameters for its use.

(ii) The model of the parametric statistical tests specifies certain conditions about the parameters of the population from which the research samples are drawn. However, the model of the non-parametric statistical tests does not specify conditions about the parameters of the population from which the samples are drawn.

(iii) For parametric statistical tests, the shape of the population is important but for non-parametric tests, the shape of population is unimportant.

(iv) Parametric statistical test can be applied to the data which are in terms of either interval or ratio scales of measurement, but when measurements are in terms of nominal or ordinal scales, the use of non-parametric tests will be more appropriate.

(v) Non-parametric statistical tests have fewer requirements or assumptions about population characteristics compared to the parametric statistical tests.

(vi) Many of the parametric statistical tests are robust with regard to violation of underlying assumptions. A test is said to be robust if violations in the assumptions do not greatly disturb the sampling distribution of its statistic. Thus, the '*t*'-test is robust regarding the violation of normality in the population. Even though, theoretically, normality in the population is required with small samples, it turns out empirically that unless the departures from normality are substantial, the sampling distribution of t remains essentially the same. Thus, the *t*-test can be used with data even though the data violate the assumptions of normality. This is one of the reasons of preferring parametric tests to non-parametric tests.

(vii) The main reasons for preferring parametric to non-parametric tests are that, in general, they are more powerful and versatile than non-parametric tests. We can see an example of the higher power of parametric tests when we will compare the *t*-test with the sign test for correlated groups (see Chapter 17). The factorial design discussed in Chapters 13–15 provides a good example of the versatility of parametric tests. With this design, we can test two, three, four or more variables and their interactions. No comparable statistical technique exists with non-parametric statistics.

(viii) As a general rule, investigators will use parametric statistical tests whenever possible. However, when there is an extreme violation of an assumption of the parametric test or if the investigator believes the scaling of the data makes the parametric test inappropriate, a non-parametric statistical test will be employed.

(ix) Prediction about future outcomes and drawing inferences about population distributions, which are the two basic objectives of research, are met with parametric statistical tests but not with non-parametric statistical tests.

(x) Non-parametric statistical tests are typically much easier to learn and to apply than are parametric statistical tests.

16.2.4 Precautions in Using Non-parametric Statistical Tests

The following precautions should be kept in mind while using non-parametric statistical tests:

(i) In situations where the assumptions underlying a parametric test are satisfied and both parametric and non-parametric tests can be applied, the preference should be for the parametric test because most parametric tests have greater power in such situations.

(ii) When measurements are in terms of interval or ratio scales, the transformation of the measurements on nominal or ordinal scales will lead to the loss of much information. Hence, as far as possible, parametric tests should be applied in such situations; no non-parametric test should be used here, unless it is asked for.

(iii) The non-parametric tests, no doubt, provide a means for avoiding the assumption of *normality* of distribution. But these tests do nothing to avoid the assumptions of independence of observation or homoscedasticity wherever applicable.

(iv) The *F*-test and *t*-test are generally considered to be robust test because the violation of the underlying assumptions does not invalidate the inferences about the population parameters. It is customary to justify the use of a normal theory test in a situation where normality cannot be guaranteed, by arguing that it is robust under non-normality.

(v) Behavioural scientists should specify the null hypothesis, alternative hypothesis, statistical test, sampling distribution and level of significance in advance of the collection of data. Hunting around for a statistical test after the data have been collected tends to maximise the effects of any chance differences which favour one test over another. As a result, the possibility of rejecting the null hypothesis (H_0) when it is true (Type I error) is greatly increased. However, this caution is equally applicable to both parametric and non-parametric statistical tests.

In the following sections of this chapter, we shall present some important and commonly used non-parametric statistical tests. Let us begin with chi-square (χ^2) test.