LECTURE 25

HYPOTHESIS TESTING

BY

DR GAURAV SANKALP

INTRODUCTION

Hypothesis testing is a tool for making statistical inferences about the population data. It is an analysis tool that tests assumptions and determines how likely something is within a given standard of accuracy. Hypothesis testing provides a way to verify whether the results of an experiment are valid.

A null hypothesis and an alternative hypothesis are set up before performing the hypothesis testing. This helps to arrive at a conclusion regarding the sample obtained from the population. In this article, we will learn more about hypothesis testing, its types, steps to perform the testing, and associated examples.

What is Hypothesis Testing?

Hypothesis testing uses sample data from the population to draw useful conclusions regarding the population probability distribution. It tests an assumption made about the data using different types of hypothesis testing methodologies. The hypothesis testing results in either rejecting or not rejecting the null hypothesis.

Hypothesis Testing Definition

Hypothesis testing can be defined as a statistical tool that is used to identify if the results of an experiment are meaningful or not. It involves setting up a null hypothesis and an alternative hypothesis. These two hypotheses will always be mutually exclusive. This means that if the null hypothesis is true then the alternative hypothesis is false and vice versa. An example of hypothesis testing is setting up a test to check if a new medicine works on a disease in a more efficient manner.

Null Hypothesis

The null hypothesis is a concise mathematical statement that is used to indicate that there is no difference between two possibilities. In other words, there is no difference between certain characteristics of data. This hypothesis assumes that the outcomes of an experiment are based on chance alone. It is denoted as H_0. Hypothesis testing is used to conclude if the null hypothesis can be rejected or not. Suppose an experiment is conducted to check if girls are shorter than boys at the age of 5. The null hypothesis will say that they are the same height.

Alternative Hypothesis

The alternative hypothesis is an alternative to the null hypothesis. It is used to show that the observations of an experiment are due to some real effect. It indicates that there is a statistical significance between two possible outcomes and can be denoted as H_1or H_a. For the above-mentioned example, the alternative hypothesis would be that girls are shorter than boys at the age of 5.

Hypothesis Testing P Value

In hypothesis testing, the p value is used to indicate whether the results obtained after conducting a test are statistically significant or not. It also indicates the probability of making an error in rejecting or not rejecting the null hypothesis.This value is always a number between 0 and 1. The p value is compared to an alpha level, α or significance level. The alpha level can be defined as the acceptable risk of incorrectly rejecting the null hypothesis. The alpha level is usually chosen between 1% to 5%.

Hypothesis Testing Critical region

All sets of values that lead to rejecting the null hypothesis lie in the critical region. Furthermore, the value that separates the critical region from the non-critical region is known as the critical value.

Important Notes on Hypothesis Testing

Hypothesis testing is a technique that is used to verify whether the results of an experiment are statistically significant.

- It involves the setting up of a null hypothesis and an alternate hypothesis.
- There are three types of tests that can be conducted under hypothesis testing - z test, t test, and chi square test.
- Hypothesis testing can be classified as right tail, left tail, and two tail tests.

Introduction or Conceptual Framework of Hypothesis

A hypothesis (plural *hypotheses*) is a proposed explanation for a phenomenon. For a hypothesis to be a scientific hypothesis, the scientific method requires that one can test it. Scientists generally base scientific hypotheses on previous observations that cannot satisfactorily be explained with the available scientific theories. Even though the words "hypothesis" and "theory" are often used synonymously, a scientific hypothesis is not the same as a scientific theory. A working hypothesis is a provisionally accepted hypothesis proposed for further research.

A different meaning of the term hypothesis is used in formal logic, to denote the antecedent of a proposition; thus in the proposition "If P, then Q". P denotes the hypothesis (or antecedent); Q can be called a consequent. P is the assumption in a (possibly counterfactual) what if question.

The adjective hypothetical, meaning "having the nature of a hypothesis", or "being assumed to exist as an immediate consequence of a hypothesis", can refer to any of these meanings of the term "hypothesis".

Uses of Hypothesis

In its ancient usage, hypothesis referred to a summary of the plot of a classical drama the english word hypothesis comes from the ancient Greek ύπόθεσις (hupothesis), meaning "to put under" or "to suppose".

In Plato's Meno (86e-87b), Socrates dissects virtue with a method used by mathematicians, that of "investigating from a hypothesis". In this sense, 'hypothesis' refers to a clever idea or to a convenient mathematical approach that simplifies cumbersome calculations. Cardinal Bellarmine gave a famous example of this usage in the warning issued to Galileo in the early 17th century that he must not treat the motion of the earth as a reality, but merely as a hypothesis.

In common usage in the 21st century. a hypothesis refers to a provisional idea whose merit requires evaluation. For proper evaluation, the framer of a hypothesis needs to define specifics in operational terms. A hypothesis requires more work by the researcher in order to either confirm or disprove it in due course, a confirmed hypothesis may become part of a theory or occasionally may grew to become a theory itself. Normally, scientific hypotheses have the form of a mathematical model. Sometimes, but not always, one can also formulate them as existential statements, stating that some particular instance of the phenomenon under examination has some characteristic and causal explanations, which have the general form of universal statements, stating that every instance of the phenomenon has a particular characteristic.

Any useful hypothesis will enable predictions by reasoning (including deductive reasoning). It might predict the outcome of an experiment in a laboratory setting or the observation of a phenomenon in nature. The prediction may also invoke statistics and only talk about probabilities. Karl Popper, following others, has argued that a hypothesis must be falsifiable, and that one cannot regard a proposition or theory as scientific if it does not admit the possibility of being shown false. Other philosophers of science have rejected the criterion of falsifiability or supplemented it with other criteria, such as verifiability (e.g., verificationism) or coherence (e.g., confirmation holism). The scientific method involves experimentation, to test the ability of some hypothesis to adequately answer the question under investigation. In contrast, unfettered observation is not as likely to raise unexplained issues or open questions in science, as would the formulation of a crucial experiment to test the hypothesis. A thought experiment might also be used to test the hypothesis as well.

In framing a hypothesis, the investigator must not currently know the outcome of a test or that it remains reasonably under continuing investigation. Only in such cases does the experiment, test or study potentially increase the probability of showing the truth of a hypothesis. If the researcher already knows the outcome, it counts as a "consequence" and the researcher should have already considered this while formulating the hypothesis. If one cannot assess the predictions by observation or by experience, the hypothesis needs to be tested by others providing observations. For example, a new technology or theory might make the necessary experiments feasible.

Scientific Hypothesis

People refer to a trial solution to a problem as a hypothesis, often called an "educated guess" because it provides a suggested solution based on the evidence. However, some scientists reject the term "educated guess" as incorrect. Experimenters may test and reject several hypotheses before solving the problem.

Working Hypothesis

A working hypothesis is a hypothesis that is provisionally accepted as a basis for further research in the hope that a tenable theory will be produced, even if the hypothesis ultimately fails. Like all hypotheses, a working hypothesis is constructed as a statement of expectations, which can be linked to the exploratory research purpose in empirical investigation. Working hypotheses and are often used as a conceptual framework in qualitative research.

The provisional nature of working hypotheses make them useful as an organizing device in applied research. Here they act like a useful guide to address problems that are still in a formative phase.

In recent years, philosophers of science have tried to integrate the various approaches to evaluating hypotheses, and the scientific method in general, to form a more complete system that integrates the individual concerns of each approach. Notably, Imre Lakatos and Paul Feyerabend, Karl Popper's colleague and student, respectively, have produced novel attempts at such a synthesis.

Measurement of Hypothesis

Concepts in Hempel's deductive homological model play key role in the development and testing of hypothesis. Most formal hypothesis connect concepts by specifying the expected relationships between propositions. When a set of hypothesis are grouped together they become a type of conceptual framework. When a conceptual framework is complex and incorporates causality or explanation it is generally referred to as a theory. According to noted philosopher of science Carl Gustav Hempel " an adequate empirical interpretation turns a theoretical system into a testable theory. The hypothesis whose constituent terms have been interpreted become capable of test by reference to observable phenomena. Frequently the interpreted hypothesis will be derivative hypotheses of the theory; but their confirmation or disconfirmation by empirical data will then immediately strengthen or weaken also the primitive hypotheses from which they were derived."

Hempel provides a useful metaphor that describes the relationship between a conceptual framework and the framework as it is observed and perhaps tested (interpreted framework). "The whole system floats, as it were, above the plane of observation and is anchored to it by rules of interpretation. These might be viewed as strings which are not part of the network but link certain points of the latter with specific places in the plane of observation. By virtue of those interpretative connections, the network can function as a scientific theory". Hypotheses with concepts anchored in the plane of observation are ready to be tested. In "actual scientific practice the process of framing a theoretical structure and of interpreting it are not always sharply separated, since the intended interpretation usually guides the construction of the theoretician". It is, however, "possible and indeed desirable, for the purposes of logical clarification, to separate the two steps conceptually."

Statistical Hypothesis testing

When a possible correlation or similar relation between phenomena is investigated, such as whether a proposed remedy is effective in treating a disease, the hypothesis that a relation exists cannot be examined the same way one might examine a proposed new law of nature. In such an investigation, if the tested remedy shows no effect in a few cases, these do not necessarily falsify the hypothesis. Instead, statistical tests are used to determine how likely it is that the overall effect would be observed if the hypothesized relation does not exist. If that likelihood is sufficiently small (e.g., less than 1%), the existence of a relation may be assumed. Otherwise, any observed effect may be due to pure chance.

In statistical hypothesis testing, two hypotheses are compared. These are called the null hypothesis and the alternative hypothesis. The null hypothesis is the hypothesis that states that there is no relation between the phenomena whose relation is under investigation, or at least not of the form given by the alternative hypothesis. The alternative hypothesis, as the name suggests, is the alternative to the null hypothesis. it states that there is some kind of relation. The alternative hypothesis may take several forms, depending on the nature of the hypothesized relation; in particular it can be two-sided (for example: there is some effect, in a yet unknown direction) or one-sided (the direction of the hypothesized relation positive or negative, is fixed in advance).

Conventional significance levels for testing hypotheses (acceptable probabilities of wrongly rejecting a true null hypothesis) are .10, .05, and .01. Whether the null hypothesis is rejected and the alternative hypothesis is accepted, must be determined in advance, before the observations are collected or inspected. If these criteria are determined later, when the data to be tested are already known, the test is invalid.

The above procedure is actually dependent on the number of the participants (units or sample size) that is included in the study. For instance, the sample size may be too small to reject a null hypothesis and, therefore, it is recommended to specify the sample size from the beginning. It is advisable to define a small, medium and large effect size for each of a lumber of important statistical tests which are used to test the hypotheses.

Test of Goodness of fit

The goodness of fit of a statistical model describes how well it fits a set of observations. Measures of goodness of fit typically summarize the discrepancy between observed values and the values expected under the model in question. Such measures can be used in statistical hypothesis testing, ego to test for normality of residuals, to test whether two samples are drawn from identical distributions (see Kolmogorov-Smirnov test), or whether outcome frequencies follow a specified distribution (see Pearson's chi-squared test). In the analysis of variance, one of the components into which the variance is partitioned may be a lack-of-fit sum of squares.

Fit of distributions

In assessing whether a given distribution is suited to a data-set, the following tests and their underlying measures of fit can be used :

- Kolmogorov-Smirnov test;
- Cramér-von Mises criterion;
- Anderson-Darling test;
- Shapiro-Wilk test,
- Chi Square test;
- Akaike Information criterion;
- Hosmer-Lemeshow test;

Regression analysis

In regression analysis, the following relate to goodness of fit:

- Coefficient of determination (The R squared measure of goodness of fit);
- Lack-of-fit sum of squares.

Example

One way in which a measure of goodness of fit statistic can be constructed, in the case where the variance of the measurement error is known, is to construct a weighted sum of squared errors.

$$X^2 = \sum \frac{(O-E)^2}{\sigma^2}$$

where σ^2 is the known variance of the observation, O is the observed data and E is the theoretical data. This definition is only useful when one has estimates for the error on the measurements, but it leads to a situation where a chi-squared distribution can be used to test goodness of fit, provided that the errors can be assumed to have a normal distribution.

The reduced chi-squared statistic is simply the chi-squared divided by the number of degrees of freedom:

$$X^2_{red} = \frac{X^2}{\nu} = \frac{1}{\nu} \sum \frac{(O-E)^2}{\sigma^2}$$

where ν is the number of degrees of freedom, usually given by N–n–1, where N is the number of observations, and n is the number of fitted parameters, assuming that the mean value is an additional fitted parameter. The advantage of the reduced chi-squared is that it already normalizes for the number of data points and model complexity. This is also known as the mean square weighted deviation.

As a rule of thumb (again valid only when the variance of the measurement error is known a prior rather than estimated from the data), $a\ X^2_{red} > 1$ indicates a poor model fit $\mathrm{A}\ X^2_{red} > 1$ indicates that the fit has not fully captured the data (or that the error variance has been underestimated). In principle, a value of $X^2_{red} = 1$ indicates that the extent of the match between observation and estimates is in accord with the error variance $\mathrm{A}\ X^2_{red} < 1$ indicates that the model is 'over-fitting' the data either the model is improperly fitting noise, or the error variance has been overestimated.

Categorical Data

The following are examples that arise in the context of categorical data.

Pearson's Chi-Squared test

Pearson's chi-squared test uses a measure of goodness of fit which is the sum of differences between observed and expected outcome frequencies (that is, counts of observations), each squared and divided by the expectation

$$X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

Where :

O_i = an observed frequency (i.e. count) for bin i

E_i = an expected (theoretical) frequency for bin i, asserted by the null hypothesis.

The expected frequency is calculated by :

$$E_i = \left(F(Y_u) - F(Y_l)\right) N$$

where :

F = the cumulative distribution function for the distribution being tested.

Y_u = the upper limit for class *i*, and

Y_l = the lower limit for class i, and

N = the sample size

The resulting value can be compared to the chi-squared distribution to determine the goodness of fit. In order to determine the degrees of freedom of the chi-squared distribution, one takes the total number of observed frequencies and subtracts the number of estimated parameters. The test statistic follows, approximately, a chi-square distribution with (k – c) degrees of freedom where k is the number of non-empty cells and c is the number of estimated parameters (including location and scale parameters and shape parameters) for the distribution.

Example : Equal Frequencies of men and women

For example, to test the hypothesis that a random sample of 100 people has been drawn from a population in which men and women are equal in frequency, the observed number of men and women would be compared to the theoretical frequencies of 50 men and 50 women. If there were 44 men in the sample and 56 women, then

$$X^2 = \frac{(44-50)^2}{50} + \frac{(56-50)^2}{50} = 1.44$$

If the null hypothesis is true (i.e., men and women are chosen with equal probability in the sample), the test statistic will be drawn from a chi-squared distribution with one degree of freedom, though one might expect two degrees of freedom (one each for the men and women), we must take into account that the total number of men and women is constrained (100), and thus there is only one degree of freedom (2 – 1). Alternatively, if the male count is known the female count is determined, and vice-versa.

Consultation of the chi-squared distribution for 1 degree of freedom shows that the probability of observing this difference (or a more extreme difference than this) if men and women are equally numerous in the population is approximately 0.23. This probability is higher than conventional criteria for statistical significance (.001 – .05), so normally we would not reject the null hypothesis that the number of men in the population is the same as the number of women (i.e. we would consider our sample within the range of what we'd expect for a 50/50 male/female ratio.)

Binomial Case

A binomial experiment is a sequence of independent trials in which the trials can result in one of two outcomes, success or failure. There are n trials each with probability of success, denoted by p. Provided that pp_i » 1 for every i (where i = 1, 2, ... , k), then

$$X^2 = \sum_{i=1}^{k} \frac{(N_i - np_i)^2}{np_i} = \sum_{\text{all cells}} \frac{(O-E)^2}{E}$$

This has approximately a chi-squared distribution with k – 1 df. The fact that df = k – 1 is a consequence of the restriction $\sum N_i = n$. We know there are k observed cell counts, however, once any k – 1 are known, the remaining one is uniquely determined. Basically, one can say, there are only k – 1 freely determined cell counts, thus df = k – 1.

Other Measures of fit

The likelihood ratio test statistic is a measure of the goodness of fit of a model, judged by whether an expanded form of the model provides a substantially improved fit.

Chi-Square Test

The chi-square test is an important test amongst the several tests of significance developed by statisticians. Chi-square, symbolically written as χ^2 (Pronounced as Ki-square), is a statistical measure used in the context of sampling analysis for comparing a variance to a theoretical variance. As a non-parametric* test, it "can be used to determine if categorical data shows dependency or the two classifications are independent. It can also be used to make comparisons between theoretical populations and actual data when categories are used." Thus, the chi-square test is applicable in large number of problems. The test is, in fact, a technique through the use of which it is possible for all researchers to (i) test the goodness of fit; (ii) test the significance of association between two attributes, and (iii) test the homogeneity or the significance of population variance.

Chi-Square as a test for comparing variance

The chi-square value is often used to judge the significance of population variance i.e., we can use the test to judge if a random sample has been drawn from a normal population with mean (μ) and with a specified variance (σ_p^2). The test is based on χ^2 – distribution. Such a distribution we encounter when we deal with collections of values that involve adding up squares. Variances of samples require us to add a collection of squared quantities and, thus, have distributions that are related to χ^2 – distribution. If we take each one of a collection of sample variances, divided them by the known population variance and multiply these quotients by (n – 1), where n means the number of items in the sample, we shall obtain a χ^2 – distribution. Thus, $\frac{\sigma_s^2}{\sigma_p^2}(n-1)=\frac{\sigma_s^2}{\sigma_p^2}$ (d.f.) would have the same distribution as χ^2 –distribution with (n –1) degrees of freedom.

The χ^2 –distribution is not symmetrical and all the values are positive. For making use of this distribution, one is required to know the degrees of freedom since for different degrees of freedom we have different curves. The smaller the number of degrees of freedom, the more skewed is the distribution which is illustrated in Fig. 10.1:

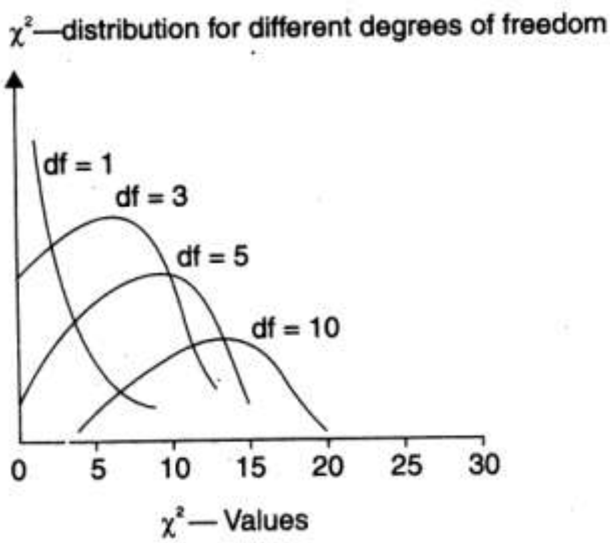

Fig. 10.1

Table given in the Appendix gives selected critical values of χ^2 for the different degrees of freedom. χ^2 -values are the quantities indicated on the x-axis of the above diagram and in the table are areas below that value.

In brief, when we have to use chi-square as a test of population variance, we have to work out the value of χ^2 to test the null hypothesis (viz., $H_0: \sigma_s^2 = \sigma_p^2$) as under :

$$\chi^2 = \frac{\sigma_s^2}{\sigma_p^2}(n-1)$$

where

σ_s^2 = variance of the sample;

σ_p^2 = variance of the population;

(n–1) = degrees of freedom, n being the number of items in the sample.

Then by comparing the calculated value with the table value of χ^2 for (n – 1) degrees of freedom at a given level of significance, we may either accept or reject the null hypothesis. If the calculated value of χ^2 is less than the table value, the null hypothesis is accepted, but if the calculated value is equal or greater than the table value, the hypothesis is rejected. All this can be made clear by an example.

Illustration 1

Weight of 10 students is as follows :

S.No.	1	2	3	4	5	6	7	8	9	10
Weight (kg.)	38	40	45	53	47	43	55	48	52	49

Can we say that the variance of the distribution of weight of all students from which the above sample of 10 students was drawn is equal to 20 kgs? Test this at 5 per cent and 1 per cent level of significance.

Solution

First of all we should work out the variance of the sample data or σ_s^2 and the same has been worked out as under:

S.No.	X_i (Weight in kgs.)	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$
1	38	–9	81
2	40	–7	49
3	45	–2	04

4	53	+6	36
5	47	+0	00
6	43	–4	16
7	55	+8	64
8	48	+1	01
9	52	+5	25
10	49	+2	04
n = 10	$\Sigma X_i = 470$		$\Sigma(X_i - \bar{X})^2 = 280$

$$\bar{X} = \frac{\Sigma X_i}{n} = \frac{470}{10} = 47 \text{ kgs.}$$

$$\therefore \quad \sigma_s = \sqrt{\frac{\Sigma(X_i - \bar{X})^2}{n-1}} = \sqrt{\frac{280}{10-1}} = \sqrt{31.11}$$

$$\text{or} \quad \sigma_s^2 = 31.11.$$

Let the null hypothesis be $H_0 : \sigma_p^2 = \sigma_s^2$. In order to test this hypothesis we work out the χ^2 value as under:

$$\chi^2 = \frac{\sigma_s^2}{\sigma_p^2}(n-1)$$

$$= \frac{31.11}{20}(10-1) = 13.999$$

Degrees of freedom in the given case is (n – 1) = (10 – 1) = 9. At 5 per cent level of significance the table value of χ^2 = 16.92 and at 1 per cent level of significance, it is 21.67 for 9 d.f. and both these values are greater than the calculated value of χ^2 which is 13.999. Hence we accept the null hypothesis and conclude that the variance of the given distribution can be taken as 20 kgs at 5 per cent as also at 1 per cent level of significance. In other words, the sample can be said to have been taken from a population with variance 20 kgs.

Illustration 2

A sample of 10 is drawn randomly from a certain population. The sum of the squared deviations from the mean of the given sample is 50. Test the hypothesis that the variance of the population is 5 at 5 per cent level of significance.

Solution.

n = 10

$$\Sigma(X_i - \overline{X})^2 = 50$$

$$\sigma_s^2 = \frac{\Sigma(X_i - \overline{X})^2}{n-1} = \frac{50}{9}$$

Take the null hypothesis as $H_0 : \sigma_p^2 = \sigma_s^2$. In order to test this hypothesis, we work out the χ^2 value as under:

$$\chi^2 = \frac{\sigma_s^2}{\sigma_p^2}(n-1) = \frac{\frac{50}{9}}{5}(10-1) = \frac{50}{9} \times \frac{1}{5} \times \frac{9}{1} = 10$$

Degrees of freedom = (10 – 1) = 9.

The table value of χ^2 at 5 per cent level for 9 d.f. is 16.92. The calculated value of χ^2 is less than this table value, so we accept the null hypothesis and conclude that the valiance of the population is 5 as given in the question.

Chi-Square as a Non-parametric test

Chi-square is an important non-parametric test and as such no rigid assumptions are necessary in respect of the type of population. We require only the degrees of freedom (implicitly of course the size of the sample) for using this test. As a non-parametric test, chi-square can be used (i) as a test of goodness of fit and (ii) as a test of independence.

As a test of goodness of fit, χ^2 test enables us to see how well does the assumed theoretical distribution (such as Binomial distribution, Poisson distribution or Normal distribution) fit to the observed data. When some theoretical distribution is fitted to the given data, we are always interested in knowing as to how well this distribution fits with the observed data. The chi-square test can give answer to this. If the calculated value of χ^2 is less than the table value at a certain level of significance, the fit is considered to be a good one which means that the divergence between the observed and expected frequencies is attributable to fluctuations of sampling. But if the calculated value of χ^2 is greater than its table value, the fit is not considered to be a good one.

As a test of independence, χ^2 test enables us to explain whether or not two attributes are associated. For instance, we may be interested in knowing whether a new medicine is effective in controlling fever or not, χ^2 test will helps us in deciding this issue. In such a situation, we proceed with the null hypothesis that the two attributes (viz., new medicine and control of fever) are independent which means that new medicine is not effective in controlling fever. On this basis we first calculate the expected frequencies and then work out the value of χ^2. If the calculated value of χ^2 is less than the table value at a certain level of significance for given degrees of freedom, we conclude that null hypothesis stands which means that the two attributes are independent or not associated (i.e., the new medicine is not effective in controlling the fever). But if the calculated value of χ^2 is greater than its table value, our inference then would be that null hypothesis does not hold good which means the two attributes are associated and the association is not because of some chance factor but it exists in reality (i.e., the new medicine is effective in controlling the fever and as such may be prescribed). It may, however, be stated here that χ^2 is not a measure of the degree of relationship or the form of relationship between two attributes, but is simply a technique of judging the significance of such association or relationship between two attributes.

In order that we may apply the chi-square test either as a test of goodness of fit or as a test to judge the significance of association between attributes, it is necessary that the observed as well as theoretical or expected frequencies must be grouped in the same way and the theoretical distribution must be adjusted to give the same total frequency as we find in case of observed distribution. χ^2 is then calculated as follows :

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

where,

O_{ij} = observed frequency of the cell in ith row and jth column.

Eij = expected frequency of the cell in ith row and jth column.

If two distributions (observed and theoretical) are exactly alike, χ^2 = 0; but generally due to sampling errors, χ^2 is not equal to zero and as such we must know the sampling distribution of χ^2 so that we may find the probability of an observed χ^2 being given by a random sample from the hypothetical universe. Instead of working out the probabilities, we can use ready table which gives probabilities for given values of χ^2. Whether or not a calculated value of χ^2 is significant can be ascertained by looking at the tabulated values of χ^2 for given degrees of freedom at a certain level of significance. If the calculated value of χ^2 is equal to or exceeds the table value, the difference between the observed and expected frequencies is taken as significant, but if the table value is more than the calculated value of χ^2, then the difference is considered as insignificant i.e., considered to have arisen as a result of chance and as such can be ignored.

As already stated, degrees of freedom play an important part in using the chi -square distribution and the test based on it, one must correctly determine the degrees of freedom. If there are 10 frequency classes and there is one independent constraint, then there are (10 – 1) = 9 degrees of freedom. Thus, if 'n' is the number of groups and one constraint is placed by making the totals of observed and expected frequencies equal, the d.f. would be equal to (n – 1). In the case of a contingency table (i.e., a table with 2 columns and 2 rows or a table with two columns and more than two rows or a table with two rows but more than two columns or a table with more than two rows and more than two columns), the d.f. is worked out as follows:

d.f. = (c – 1) (r – 1)

where 'c' means the number of columns and 'r' means the number of rows.

The following conditions should be satisfied before χ^2 test can be applied:

(i) Observations recorded and used are collected on a random basis.

(ii) All the items in the sample must be independent.

(iii) No group should contain very few items, say less than 10. In case where the frequencies are less than 10, regrouping is done by combining the frequencies of adjoining groups so that the new frequencies become greater than 10. Some statisticians take this number as 5, but 10 is regarded as better by most of the statisticians.

(iv) The overall number of items must also be reasonably large. It should normally be at least 50, howsoever small the number of groups may be.

(v) The constraints must be linear. Constraints which involve linear equations in the cell frequencies of a contingency table (i.e., equations containing no squares or higher powers of the frequencies) are known are know as linear constraints.