9

Testing of Hypotheses I

(Parametric or Standard Tests of Hypotheses)

Hypothesis is usually considered as the principal instrument in research. Its main function is to suggest new experiments and observations. In fact, many experiments are carried out with the deliberate object of testing hypotheses. Decision-makers often face situations wherein they are interested in testing hypotheses on the basis of available information and then take decisions on the basis of such testing. In social science, where direct knowledge of population parameter(s) is rare, hypothesis testing is the often used strategy for deciding whether a sample data offer such support for a hypothesis that generalisation can be made. Thus hypothesis testing enables us to make probability statements about population parameter(s). The hypothesis may not be proved absolutely, but in practice it is accepted if it has withstood a critical testing. Before we explain how hypotheses are tested through different tests meant for the purpose, it will be appropriate to explain clearly the meaning of a hypothesis and the related concepts for better understanding of the hypothesis testing techniques.

WHAT IS A HYPOTHESIS?

Ordinarily, when one talks about hypothesis, one simply means a mere assumption or some supposition to be proved or disproved. But for a researcher hypothesis is a formal question that he intends to resolve. Thus a hypothesis may be defined as a proposition or a set of proposition set forth as an explanation for the occurrence of some specified group of phenomena either asserted merely as a provisional conjecture to guide some investigation or accepted as highly probable in the light of established facts. Quite often a research hypothesis is a predictive statement, capable of being tested by scientific methods, that relates an independent variable to some dependent variable. For example, consider statements like the following ones:

"Students who receive counselling will show a greater increase in creativity than students not receiving counselling" Or

"the automobile *A* is performing as well as automobile *B*."

These are hypotheses capable of being objectively verified and tested. Thus, we may conclude that a hypothesis states what we are looking for and it is a proposition which can be put to a test to determine its validity.

Characteristics of hypothesis: Hypothesis must possess the following characteristics:

(i) Hypothesis should be clear and precise. If the hypothesis is not clear and precise, the inferences drawn on its basis cannot be taken as reliable.

(ii) Hypothesis should be capable of being tested. In a swamp of untestable hypotheses, many a time the research programmes have bogged down. Some prior study may be done by researcher in order to make hypothesis a testable one. A hypothesis "is testable if other deductions can be made from it which, in turn, can be confirmed or disproved by observation."[1]

(iii) Hypothesis should state relationship between variables, if it happens to be a relational hypothesis.

(iv) Hypothesis should be limited in scope and must be specific. A researcher must remember that narrower hypotheses are generally more testable and he should develop such hypotheses.

(v) Hypothesis should be stated as far as possible in most simple terms so that the same is easily understandable by all concerned. But one must remember that simplicity of hypothesis has nothing to do with its significance.

(vi) Hypothesis should be consistent with most known facts i.e., it must be consistent with a substantial body of established facts. In other words, it should be one which judges accept as being the most likely.

(vii) Hypothesis should be amenable to testing within a reasonable time. One should not use even an excellent hypothesis, if the same cannot be tested in reasonable time for one cannot spend a life-time collecting data to test it.

(viii) Hypothesis must explain the facts that gave rise to the need for explanation. This means that by using the hypothesis plus other known and accepted generalizations, one should be able to deduce the original problem condition. Thus hypothesis must actually explain what it claims to explain; it should have empirical reference.

BASIC CONCEPTS CONCERNING TESTING OF HYPOTHESES

Basic concepts in the context of testing of hypotheses need to be explained.

(a) *Null hypothesis and alternative hypothesis:* In the context of statistical analysis, we often talk about null hypothesis and alternative hypothesis. If we are to compare method *A* with method *B* about its superiority and if we proceed on the assumption that both methods are equally good, then this assumption is termed as the null hypothesis. As against this, we may think that the method *A* is superior or the method *B* is inferior, we are then stating what is termed as alternative hypothesis. The null hypothesis is generally symbolized as H_0 and the alternative hypothesis as H_a. Suppose we want to test the hypothesis that the population mean (μ) is equal to the hypothesised mean $(\mu_{H_0}) = 100$.

Then we would say that the null hypothesis is that the population mean is equal to the hypothesised mean 100 and symbolically we can express as:

$$H_0 : \mu = \mu_{H_0} = 100$$

[1] C. William Emory, *Business Research Methods*, p. 33.

If our sample results do not support this null hypothesis, we should conclude that something else is true. What we conclude rejecting the null hypothesis is known as alternative hypothesis. In other words, the set of alternatives to the null hypothesis is referred to as the alternative hypothesis. If we accept H_0, then we are rejecting H_a and if we reject H_0, then we are accepting H_a. For $H_0 : \mu = \mu_{H_0} = 100$, we may consider three possible alternative hypotheses as follows*:

Table 9.1

Alternative hypothesis	*To be read as follows*
$H_a : \mu \neq \mu_{H_0}$	(The alternative hypothesis is that the population mean is not equal to 100 i.e., it may be more or less than 100)
$H_a : \mu > \mu_{H_0}$	(The alternative hypothesis is that the population mean is greater than 100)
$H_a : \mu < \mu_{H_0}$	(The alternative hypothesis is that the population mean is less than 100)

The null hypothesis and the alternative hypothesis are chosen before the sample is drawn (the researcher must avoid the error of deriving hypotheses from the data that he collects and then testing the hypotheses from the same data). In the choice of null hypothesis, the following considerations are usually kept in view:

(a) Alternative hypothesis is usually the one which one wishes to prove and the null hypothesis is the one which one wishes to disprove. Thus, a null hypothesis represents the hypothesis we are trying to reject, and alternative hypothesis represents all other possibilities.

(b) If the rejection of a certain hypothesis when it is actually true involves great risk, it is taken as null hypothesis because then the probability of rejecting it when it is true is α (the level of significance) which is chosen very small.

(c) Null hypothesis should always be specific hypothesis i.e., it should not state about or approximately a certain value.

Generally, in hypothesis testing we proceed on the basis of null hypothesis, keeping the alternative hypothesis in view. Why so? The answer is that on the assumption that null hypothesis is true, one can assign the probabilities to different possible sample results, but this cannot be done if we proceed with the alternative hypothesis. Hence the use of null hypothesis (at times also known as statistical hypothesis) is quite frequent.

(b) *The level of significance:* This is a very important concept in the context of hypothesis testing. It is always some percentage (usually 5%) which should be chosen wit great care, thought and reason. In case we take the significance level at 5 per cent, then this implies that H_0 will be rejected

*If a hypothesis is of the type $\mu = \mu_{H_0}$, then we call such a hypothesis as simple (or specific) hypothesis but if it is of the type $\mu \neq \mu_{H_0}$ or $\mu > \mu_{H_0}$ or $\mu < \mu_{H_0}$, then we call it a composite (or nonspecific) hypothesis.

when the sampling result (i.e., observed evidence) has a less than 0.05 probability of occurring if H_0 is true. In other words, the 5 per cent level of significance means that researcher is willing to take as much as a 5 per cent risk of rejecting the null hypothesis when it (H_0) happens to be true. Thus the significance level is the maximum value of the probability of rejecting H_0 when it is true and is usually determined in advance before testing the hypothesis.

(c) *Decision rule or test of hypothesis:* Given a hypothesis H_0 and an alternative hypothesis H_a, we make a rule which is known as decision rule according to which we accept H_0 (i.e., reject H_a) or reject H_0 (i.e., accept H_a). For instance, if (H_0 is that a certain lot is good (there are very few defective items in it) against H_a) that the lot is not good (there are too many defective items in it), then we must decide the number of items to be tested and the criterion for accepting or rejecting the hypothesis. We might test 10 items in the lot and plan our decision saying that if there are none or only 1 defective item among the 10, we will accept H_0 otherwise we will reject H_0 (or accept H_a). This sort of basis is known as decision rule.

(d) *Type I and Type II errors:* In the context of testing of hypotheses, there are basically two types of errors we can make. We may reject H_0 when H_0 is true and we may accept H_0 when in fact H_0 is not true. The former is known as Type I error and the latter as Type II error. In other words, Type I error means rejection of hypothesis which should have been accepted and Type II error means accepting the hypothesis which should have been rejected. Type I error is denoted by α (alpha) known as α error, also called the level of significance of test; and Type II error is denoted by β (beta) known as β error. In a tabular form the said two errors can be presented as follows:

Table 9.2

	Decision	
	Accept H_0	Reject H_0
H_0 (true)	Correct decision	Type I error (α error)
H_0 (false)	Type II error (β error)	Correct decision

The probability of Type I error is usually determined in advance and is understood as the level of significance of testing the hypothesis. If type I error is fixed at 5 per cent, it means that there are about 5 chances in 100 that we will reject H_0 when H_0 is true. We can control Type I error just by fixing it at a lower level. For instance, if we fix it at 1 per cent, we will say that the maximum probability of committing Type I error would only be 0.01.

But with a fixed sample size, *n*, when we try to reduce Type I error, the probability of committing Type II error increases. Both types of errors cannot be reduced simultaneously. There is a trade-off between two types of errors which means that the probability of making one type of error can only be reduced if we are willing to increase the probability of making the other type of error. To deal with this trade-off in business situations, decision-makers decide the appropriate level of Type I error by examining the costs or penalties attached to both types of errors. If Type I error involves the time and trouble of reworking a batch of chemicals that should have been accepted, whereas Type II error means taking a chance that an entire group of users of this chemical compound will be poisoned, then in such a situation one should prefer a Type I error to a Type II error. As a result one must set very high level for Type I error in one's testing technique of a given hypothesis.[2] Hence, in the testing of hypothesis, one must make all possible effort to strike an adequate balance between Type I and Type II errors.

(e) *Two-tailed and One-tailed tests:* In the context of hypothesis testing, these two terms are quite important and must be clearly understood. A two-tailed test rejects the null hypothesis if, say, the sample mean is significantly higher or lower than the hypothesised value of the mean of the population. Such a test is appropriate when the null hypothesis is some specified value and the alternative hypothesis is a value not equal to the specified value of the null hypothesis. Symbolically, the two-tailed test is appropriate when we have $H_0: \mu = \mu_{H_0}$ and $H_a: \mu \neq \mu_{H_0}$ which may mean $\mu > \mu_{H_0}$ or $\mu < \mu_{H_0}$. Thus, in a two-tailed test, there are two rejection regions*, one on each tail of the curve which can be illustrated as under:

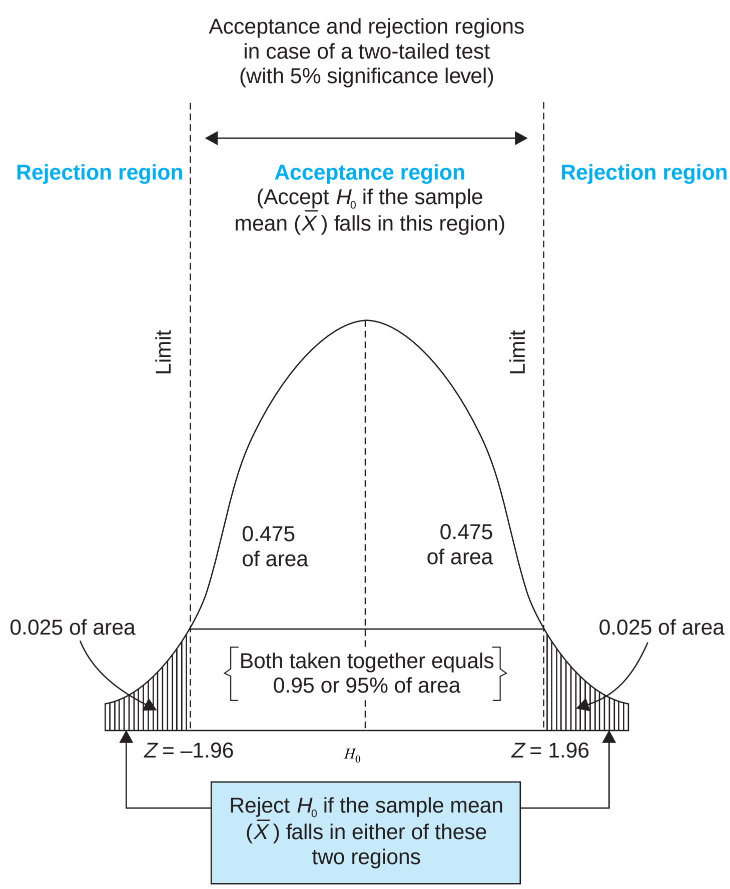

Fig. 9.1

[2] Richard I. Levin, *Statistics for Management*, p. 247–248.

*Also known as critical regions.

Mathematically we can state:

$$\textit{Acceptance Region} \quad A : |Z| \,\overline{<}\, 1.96$$

$$\textit{Rejection Region} \quad R : |Z| \,\overline{>}\, 1.96$$

If the significance level is 5 per cent and the two-tailed test is to be applied, the probability of the rejection area will be 0.05 (equally splitted on both tails of the curve as 0.025) and that of the acceptance region will be 0.95 as shown in the above curve. If we take $\mu = 100$ and if our sample mean deviates significantly from 100 in either direction, then we shall reject the null hypothesis; but if the sample mean does not deviate significantly from μ, in that case we shall accept the null hypothesis.

But there are situations when only one-tailed test is considered appropriate. A *one-tailed test* would be used when we are to test, say, whether the population mean is either lower than or higher than some hypothesised value. For instance, if our $H_0: \mu = \mu_{H_0}$ and $H_a: \mu < \mu_{H_0}$, then we are interested in what is known as left-tailed test (wherein there is one rejection region only on the left tail) which can be illustrated as below:

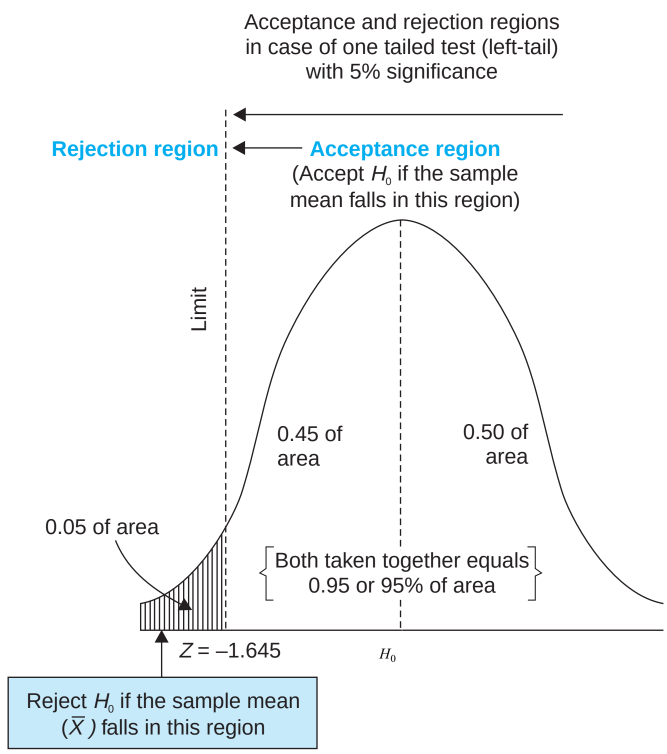

Fig. 9.2

Mathematically we can state:

$$\textit{Acceptance Region} \quad A : Z > -1.645$$

$$\textit{Rejection Region} \quad R : Z \,\overline{<}\, -1.645$$

If our $\mu = 100$ and if our sample mean deviates significantly from100 in the lower direction, we shall reject H_0, otherwise we shall accept H_0 at a certain level of significance. If the significance level in the given case is kept at 5%, then the rejection region will be equal to 0.05 of area in the left tail as has been shown in the above curve.

In case our $H_0 : \mu = \mu_{H_0}$ and $H_a : \mu > \mu_{H_0}$, we are then interested in what is known as one-tailed test (right tail) and the rejection region will be on the right tail of the curve as shown below:

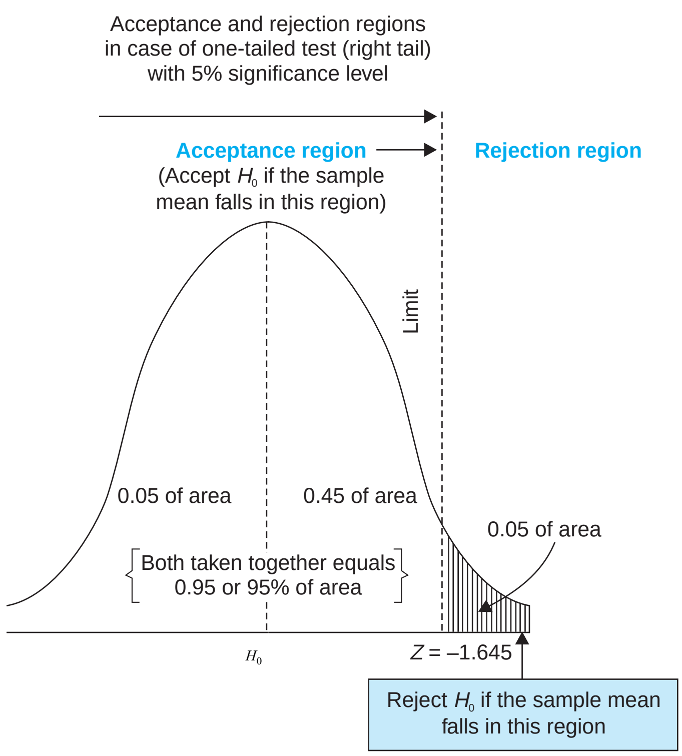

Fig. 9.3

Mathematically we can state:

$$\textit{Acceptance Region} \quad A : Z \leqq 1.645$$

$$\textit{Rejection Region} \quad A : Z > 1.645$$

If our $\mu = 100$ and if our sample mean deviates significantly from 100 in the upward direction, we shall reject H_0, otherwise we shall accept the same. If in the given case the significance level is kept at 5%, then the rejection region will be equal to 0.05 of area in the right-tail as has been shown in the above curve.

It should always be remembered that accepting H_0 on the basis of sample information does not constitute the proof that H_0 is true. We only mean that there is no statistical evidence to reject it, but we are certainly not saying that H_0 is true (although we behave as if H_0 is true).

PROCEDURE FOR HYPOTHESIS TESTING

To test a hypothesis means to tell (on the basis of the data the researcher has collected) whether or not the hypothesis seems to be valid. In hypothesis testing the main question is: whether to accept the null hypothesis or not to accept the null hypothesis? Procedure for hypothesis testing refers to all those steps that we undertake for making a choice between the two actions i.e., rejection and acceptance of a null hypothesis. The various steps involved in hypothesis testing are stated below:

(i) *Making a formal statement:* The step consists in making a formal statement of the null hypothesis (H_0) and also of the alternative hypothesis (H_a). This means that hypotheses should be clearly stated, considering the nature of the research problem. For instance, Mr. Mohan of the Civil Engineering Department wants to test the load bearing capacity of an old bridge which must be more than 10 tons, in that case he can state his hypotheses as under:

Null hypothesis $H_0 : \mu = 10$ tons

Alternative Hypothesis $H_a : \mu > 10$ tons

Take another example. The average score in an aptitude test administered at the national level is 80. To evaluate a state's education system, the average score of 100 of the state's students selected on random basis was 75. The state wants to know if there is a significant difference between the local scores and the national scores. In such a situation the hypotheses may be stated as under:

Null hypothesis $H_0 : \mu = 80$

Alternative Hypothesis $H_a : \mu \neq 80$

The formulation of hypotheses is an important step which must be accomplished with due care in accordance with the object and nature of the problem under consideration. It also indicates whether we should use a one-tailed test or a two-tailed test. If H_a is of the type greater than (or of the type lesser than), we use a one-tailed test, but when H_a is of the type "whether greater or smaller" then we use a two-tailed test.

(ii) *Selecting a significance level:* The hypotheses are tested on a pre-determined level of significance and as such the same should be specified. Generally, in practice, either 5% level or 1% level is adopted for the purpose. The factors that affect the level of significance are: (a) the magnitude of the difference between sample means; (b) the size of the samples; (c) the variability of measurements within samples; and (d) whether the hypothesis is directional or non-directional (A directional hypothesis is one which predicts the direction of the difference between, say, means). In brief, the level of significance must be adequate in the context of the purpose and nature of enquiry.

(iii) *Deciding the distribution to use:* After deciding the level of significance, the next step in hypothesis testing is to determine the appropriate sampling distribution. The choice generally remains between normal distribution and the *t*-distribution. The rules for selecting the correct distribution are similar to those which we have stated earlier in the context of estimation.

(iv) *Selecting a random sample and computing an appropriate value:* Another step is to select a random sample(s) and compute an appropriate value from the sample data concerning the test statistic utilizing the relevant distribution. In other words, draw a sample to furnish empirical data.

(v) *Calculation of the probability:* One has then to calculate the probability that the sample result would diverge as widely as it has from expectations, if the null hypothesis were in fact true.

(vi) *Comparing the probability:* Yet another step consists in comparing the probability thus calculated with the specified value for α, the significance level. If the calculated probability is equal to or smaller than the α value in case of one-tailed test (and $\alpha/2$ in case of two-tailed test), then reject the null hypothesis (i.e., accept the alternative hypothesis), but if the calculated probability is greater, then accept the null hypothesis. In case we reject H_0, we run a risk of (at most the level of significance) committing an error of Type I, but if we accept H_0, then we run some risk (the size of which cannot be specified as long as the H_0 happens to be vague rather than specific) of committing an error of Type II.

FLOW DIAGRAM FOR HYPOTHESIS TESTING

The above stated general procedure for hypothesis testing can also be depicted in the from of a flow-chart for better understanding as shown in Fig. 9.4:[3]

FLOW DIAGRAM FOR HYPOTHESIS TESTING

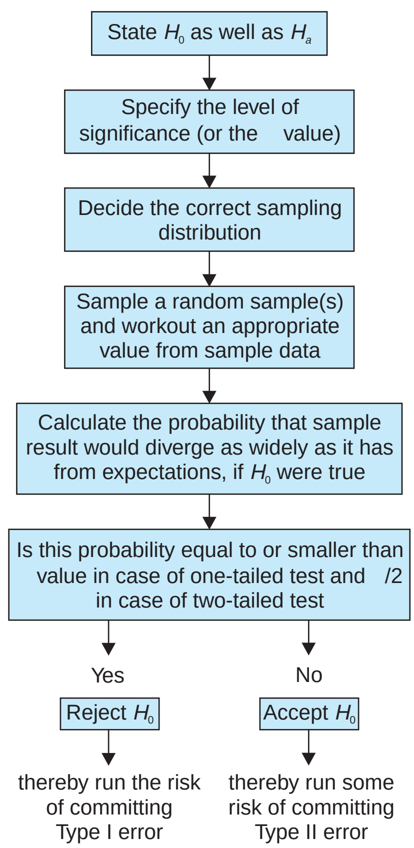

Fig. 9.4

[3] Based on the flow diagram in William A. Chance's *Statistical Methods for Decision Making*, Richard D. Irwin INC., Illinois, 1969, p.48.

MEASURING THE POWER OF A HYPOTHESIS TEST

As stated above we may commit Type I and Type II errors while testing a hypothesis. The probability of Type I error is denoted as α (the significance level of the test) and the probability of Type II error is referred to as β. Usually the significance level of a test is assigned in advance and once we decide it, there is nothing else we can do about α. But what can we say about β? We all know that hypothesis test cannot be foolproof; sometimes the test does not reject H_0 when it happens to be a false one and this way a Type II error is made. But we would certainly like that β (the probability of accepting H_0 when H_0 is not true) to be as small as possible. Alternatively, we would like that $1 - \beta$ (the probability of rejecting H_0 when H_0 is not true) to be as large as possible. If $1 - \beta$ is very much nearer to unity (i.e., nearer to 1.0), we can infer that the test is working quite well, meaning thereby that the test is rejecting H_0 when it is not true and if $1 - \beta$ is very much nearer to 0.0, then we infer that the test is poorly working, meaning thereby that it is not rejecting H_0 when H_0 is not true. Accordingly $1 - \beta$ value is the measure of how well the test is working or what is technically described as the *power of the test*. In case we plot the values of $1 - \beta$ for each possible value of the population parameter (say μ, the true population mean) for which the H_0 is not true (alternatively the H_a is true), the resulting curve is known as the power curve associated with the given test. Thus power curve of a hypothesis test is the curve that shows the conditional probability of rejecting H_0 as a function of the population parameter and size of the sample.

The function defining this curve is known as the *power function*. In other words, the power function of a test is that function defined for all values of the parameter(s) which yields the probability that H_0 is rejected and the value of the power function at a specific parameter point is called the power of the test at that point. As the population parameter gets closer and closer to hypothesised value of the population parameter, the power of the test (i.e., $1 - \beta$) must get closer and closer to the probability of rejecting H_0 when the population parameter is exactly equal to hypothesised value of the parameter. We know that this probability is simply the significance level of the test, and as such the power curve of a test terminates at a point that lies at a height of α (the significance level) directly over the population parameter.

Closely related to the power function, there is another function which is known as the *operating characteristic function* which shows the conditional probability of accepting H_0 for all values of population parameter(s) for a given sample size, whether or not the decision happens to be a correct one. If power function is represented as H and operating characteristic function as L, then we have $L = 1 - H$. However, one needs only one of these two functions for any decision rule in the context of testing hypotheses. How to compute the power of a test (i.e., $1 - \beta$) can be explained through examples.

Illustration 1

A certain chemical process is said to have produced 15 or less pounds of waste material for every 60 lbs. batch with a corresponding standard deviation of 5 lbs. A random sample of 100 batches gives an average of 16 lbs. of waste per batch. Test at 10 per cent level whether the average quantity of waste per batch has increased. Compute the power of the test for μ = 16 lbs. If we raise the level of significance to 20 per cent, then how the power of the test for μ = 16 lbs. would be affected?

Solution: As we want to test the hypothesis that the average quantity of waste per batch of 60 lbs. is 15 or less pounds against the hypothesis that the waste quantity is more than 15 lbs., we can write as under:

$$H_0 : \mu \le 15 \text{ lbs.}$$
$$H_a : \mu > 15 \text{ lbs.}$$

As H_a is one-sided, we shall use the one-tailed test (in the right tail because H_a is of more than type) at 10% level for finding the value of standard deviate (z), corresponding to .4000 area of normal curve which comes to 1.28 as per normal curve area table.* From this we can find the limit of μ for accepting H_0 as under:

Accept $\quad H_0$ if $\quad \overline{X} \le 15 + 1.28\ (\alpha_p/\sqrt{n})$

or $\quad \overline{X} \le 15 + 1.28\left(5/\sqrt{100}\right)$

or $\quad \overline{X} \le 15.64$

at 10% level of significance otherwise accept H_a.

But the sample average is 16 lbs. which does not come in the acceptance region as above. We, therefore, reject H_0 and conclude that average quantity of waste per batch has increased. For finding the power of the test, we first calculate β and then subtract it from one. Since β is a conditional probability which depends on the value of μ, we take it as 16 as given in the question. We can now write $\beta = p$ (Accept $H_0 : \mu \le 15 \mid \mu = 16$). Since we have already worked out that H_0 is accepted if $\overline{X} \le 15.64$ (at 10% level of significance), therefore $\beta = p\ (\overline{X} \le 15.64 \mid \mu = 16)$ which can be depicted as follows:

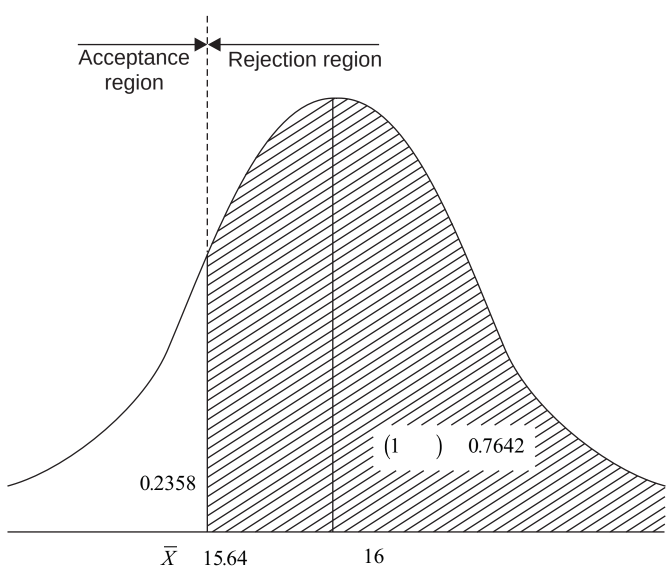

Fig. 9.5

* Table No. 1. given in appendix at the end of the book.

We can find out the probability of the area that lies between 15.64 and 16 in the above curve first by finding z and then using the area table for the purpose. In the given case $z = (\overline{X} - \mu)/(\sigma/\sqrt{n})$ $= (15.64 - 16)/(5/\sqrt{100}) = -0.72$ corresponding to which the area is 0.2642. Hence, β = 0.5000 – 0.2642 = 0.2358 and the power of the test = (1 – β) = (1 – .2358) = 0.7642 for μ = 16.

In case the significance level is raised to 20%, then we shall have the following criteria:

Accept H_0 if $\overline{X} \,\overline{<}\, 15 + (.84)\left(5/\sqrt{100}\right)$

or $\overline{X} \,\overline{<}\, 15.42$, otherwise accept H_a

$\therefore\ \beta = p\left(\overline{X} \,\overline{<}\, 15.42 \mid \mu = 16\right)$

or $\beta = .1230,$, using normal curve area table as explained above.

Hence, $(1 - \beta) = (1 - .1230) = .8770$

TESTS OF HYPOTHESES

As has been stated above that hypothesis testing determines the validity of the assumption (technically described as null hypothesis) with a view to choose between two conflicting hypotheses about the value of a population parameter. Hypothesis testing helps to decide on the basis of a sample data, whether a hypothesis about the population is likely to be true or false. Statisticians have developed several tests of hypotheses (also known as the tests of significance) for the purpose of testing of hypotheses which can be classified as: (a) Parametric tests or standard tests of hypotheses; and (b) Non-parametric tests or distribution-free test of hypotheses.

Parametric tests usually assume certain properties of the parent population from which we draw samples. Assumptions like observations come from a normal population, sample size is large, assumptions about the population parameters like mean, variance, etc., must hold good before parametric tests can be used. But there are situations when the researcher cannot or does not want to make such assumptions. In such situations we use statistical methods for testing hypotheses which are called non-parametric tests because such tests do not depend on any assumption about the parameters of the parent population. Besides, most non-parametric tests assume only nominal or ordinal data, whereas parametric tests require measurement equivalent to at least an interval scale. As a result, non-parametric tests need more observations than parametric tests to achieve the same size of Type I and Type II errors.[4] We take up in the present chapter some of the important parametric tests, whereas non-parametric tests will be dealt with in a separate chapter later in the book.

IMPORTANT PARAMETRIC TESTS

The important parametric tests are: (1) z-test; (2) t-test; (*3) χ^2-test, and (4) F-test. All these tests are based on the assumption of normality i.e., the source of data is considered to be normally distributed.

[4] Donald L. Harnett and James L. Murphy, *Introductory Statistical Analysis*, p. 368.

* χ^2 - test is also used as a test of goodness of fit and also as a test of independence in which case it is a non-parametric test. This has been made clear in Chapter 10 entitled χ^2 -test.

In some cases the population may not be normally distributed, yet the tests will be applicable on account of the fact that we mostly deal with samples and the sampling distributions closely approach normal distributions.

z-test is based on the normal probability distribution and is used for judging the significance of several statistical measures, particularly the mean. The relevant test statistic*, *z*, is worked out and compared with its probable value (to be read from table showing area under normal curve) at a specified level of significance for judging the significance of the measure concerned. This is a most frequently used test in research studies. This test is used even when binomial distribution or *t*-distribution is applicable on the presumption that such a distribution tends to approximate normal distribution as '*n*' becomes larger. *z*-test is generally used for comparing the mean of a sample to some hypothesised mean for the population in case of large sample, or when population variance is known. *z*-test is also used for judging he significance of difference between means of two independent samples in case of large samples, or when population variance is known. *z*-test is also used for comparing the sample proportion to a theoretical value of population proportion or for judging the difference in proportions of two independent samples when *n* happens to be large. Besides, this test may be used for judging the significance of median, mode, coefficient of correlation and several other measures.

t-test is based on *t*-distribution and is considered an appropriate test for judging the significance of a sample mean or for judging the significance of difference between the means of two samples in case of small sample(s) when population variance is not known (in which case we use variance of the sample as an estimate of the population variance). In case two samples are related, we use *paired t-test* (or what is known as difference test) for judging the significance of the mean of difference between the two related samples. It can also be used for judging the significance of the coefficients of simple and partial correlations. The relevant test statistic, *t*, is calculated from the sample data and then compared with its probable value based on *t*-distribution (to be read from the table that gives probable values of *t* for different levels of significance for different degrees of freedom) at a specified level of significance for concerning degrees of freedom for accepting or rejecting the null hypothesis. It may be noted that *t*-test applies only in case of small sample(s) when population variance is unknown.

χ^2-*test* is based on chi-square distribution and as a parametric test is used for comparing a sample variance to a theoretical population variance.

F-test is based on *F*-distribution and is used to compare the variance of the two-independent samples. This test is also used in the context of analysis of variance (ANOVA) for judging the significance of more than two sample means at one and the same time. It is also used for judging the significance of multiple correlation coefficients. Test statistic, *F*, is calculated and compared with its probable value (to be seen in the *F*-ratio tables for different degrees of freedom for greater and smaller variances at specified level of significance) for accepting or rejecting the null hypothesis.

The table on pages 198–201 summarises the important parametric tests along with test statistics and test situations for testing hypotheses relating to important parameters (often used in research studies) in the context of one sample and also in the context of two samples.

We can now explain and illustrate the use of the above stated test statistics in testing of hypotheses.

* The test statistic is the value obtained from the sample data that corresponds to the parameter under investigation.

HYPOTHESIS TESTING OF MEANS

Mean of the population can be tested presuming different situations such as the population may be normal or other than normal, it may be finite or infinite, sample size may be large or small, variance of the population may be known or unknown and the alternative hypothesis may be two-sided or one-sided. Our testing technique will differ in different situations. We may consider some of the important situations.

1. *Population normal, population infinite, sample size may be large or small but variance of the population is known, H_a may be one-sided or two-sided*:

   In such a situation z-test is used for testing hypothesis of mean and the test statistic z is worked our as under:

$$z = \frac{\overline{X} - \mu_{H_0}}{\sigma_p / \sqrt{n}}$$

2. *Population normal, population finite, sample size may be large or small but variance of the population is known, H_a may be one-sided or two-sided*:

   In such a situation z-test is used and the test statistic z is worked out as under (using finite population multiplier):

$$z = \frac{\overline{X} - \mu_{H_0}}{\left(\sigma_p / \sqrt{n}\right) \times \left[\sqrt{(N-n)/(N-1)}\right]}$$

3. *Population normal, population infinite, sample size small and variance of the population unknown, H_a may be one-sided or two-sided:*

   In such a situation t-test is used and the test statistic t is worked out as under:

$$t = \frac{\overline{X} - \mu_{H_0}}{\sigma_s / \sqrt{n}} \quad \text{with d.f.} = (n-1)$$

and

$$\sigma_s = \sqrt{\frac{\Sigma \left(X_i - \overline{X}\right)^2}{(n-1)}}$$

4. *Population normal, population finite, sample size small and variance of the population unknown, and H_a may be one-sided or two-sided:*

   In such a situation t-test is used and the test statistic 't' is worked out as under (using finite population multiplier):

$$t = \frac{\overline{X} - \mu_{H_0}}{\left(\sigma_s / \sqrt{n}\right) \times \sqrt{(N-n)/(N-1)}} \quad \text{with d.f.} = (n-1)$$

Table 9.3: Names of Some Parametric Tests along with Test Situations and Test Statistics used in Context of Hypothesis Testing

Unknown parameter	*Test situation (Population characteristics and other conditions. Random sampling is assumed in all situations along with infinite population*	*Name of the test and the test statistic to be used*		
		One sample	*Two samples*	
			Independent	*Related*
1	*2*	*3*	*4*	*5*
Mean (μ)	Population(s) normal *or* Sample size large (i.e., $n > 30$) *or* population variance(s) known	z-test and the test statistic $z = \dfrac{X - \mu_{H_0}}{\sigma_p / \sqrt{n}}$ In case σ_p is not known, we use σ_s in its place calculating $\sigma_s = \sqrt{\dfrac{\Sigma\left(X_i - \overline{X}\right)^2}{n-1}}$	z-test for difference in means and the test statistic $z = \dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{\sigma_p^2 \left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$ is used when two samples are drawn from the same population. In case σ_p is not known, we use σ_{s12} in its place calculating $\sigma_{s12} = \sqrt{\dfrac{n_1\left(\sigma_{s1}^2 + D_1^2\right) + n_2\left(\sigma_{s2}^2 + D_2^2\right)}{n_1 + n_2}}$ where $D_1 = \left(\overline{X}_1 - \overline{X}_{12}\right)$ $D_2 = \left(\overline{X}_2 - \overline{X}_{12}\right)$ $\overline{X}_{12} = \dfrac{n_1 \overline{X}_1 + n_2 \overline{X}_2}{n_1 + n_2}$	

Contd.

1	*2*	*3*	*4*	*5*
			OR $z\ \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{\sigma_{p1}^2}{n_1} + \dfrac{\sigma_{p2}^2}{n_2}}}$ is used when two samples are drawn from different populations. In case σ_{p_1} and σ_{p_2} are not known. We use σ_{s_1} and σ_{s_2} respectively in their places calculating $\sigma_{s1} = \sqrt{\Sigma\left(X_{1i} - \bar{X}_1\right)^2 \Big/ n_1 - 1}$ and $\sigma_{s2} = \sqrt{\Sigma\left(X_{2i} - \bar{X}_2\right)^2 \Big/ n_2 - 1}$	
Mean (μ)	Populations(s) normal *and* sample size small (i.e., $n < 30$) *and* population variance(s) unknown (but the population variances assumed equal in case of test on difference between means)	*t*-test and the test statistic $t = \dfrac{\bar{X} - \mu_{H_0}}{\sigma_s \big/ \sqrt{n}}$ with d.f. $= (n - 1)$ where	*t*-test for difference in means and the test statistic $t = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{\Sigma\left(X_{1i} - \bar{X}_1\right)^2 + \Sigma\left(X_{2i} - \bar{X}_2\right)^2}{n_1 + n_2 - 2}}} \times \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}$ with d.f. $= (n_1 + n_2 - 2)$	Paired *t*-test or difference test and the test statistic $t = \dfrac{\bar{D} - 0}{\sqrt{\dfrac{\Sigma D_i^2 - \bar{D}^2 \cdot n}{n - 1}} \Big/ \sqrt{n}}$ with d.f $= (n - 1)$ where n = number of

Contd.

1	2	3	4	5
		$\sigma_s = \sqrt{\dfrac{\Sigma\left(X_i - \bar{X}\right)^2}{n-1}}$	*Alternatively*, t can be worked out as under: $\left\{\dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{(n_1-1)\sigma_{s1}^2 + (n_2-1)\sigma_{s2}^2}{n_1+n_2-2}} \times \sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}\right.$ with d.f. $= (n_1 + n_2 - 2)$ $\Big\}$	pairs in two samples. D_i = differences (i.e., $D_i = X_i - Y_i$)
Proportion (p)	Repeated independent trials, sample size large (presuming normal approximation of binomial distribution)	z-test and the test statistic $z = \dfrac{\hat{p} - p}{\sqrt{p \cdot q/n}}$ If p and q are not known, then we use $\bar{p}$ and $\bar{q}$ in their places	z-test for difference in proportions of two samples and the test statistic $z = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\dfrac{\hat{p}_1\hat{q}_1}{n_1} + \dfrac{\hat{p}_2\hat{q}_2}{n_2}}}$ is used in case of heterogenous populations. But when populations are similar with respect to a given attribute, we work out the best estimate of the population proportion as under: $p_0 = \dfrac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$	

Contd.

1	*2*	*3*	*4*	*5*
		$z = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{p_0\, q_0 \left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$	and $q_0 = 1 - p_0$ in which case we calculate test statistic	
variance $\left(\sigma_p^2\right)$	Population(s) normal, observations are independent	χ^2-test and the test statistic $\chi^2 = \dfrac{\sigma_s^2}{\sigma_p^2}(n-1)$ with d.f. = $(n-1)$	*F*-test and the test statistic $F = \dfrac{\sigma_{s1}^2}{\sigma_{s2}^2} = \dfrac{\sum\left(X_{1i} - \overline{X}_1\right)^2 / n - 1}{\sum\left(X_{2i} - \overline{X}_2\right)^2 / n - 1}$ where σ_{s1}^2 is treated $> \sigma_{s2}^2$ with d.f. = $v_1 = (n_1 - 1)$ for greater variance and d.f. = $v_2 = (n_2 - 1)$ for smaller variance	

In the table the various symbols stand as under:

$\overline{X}$ = mean of the sample, $\overline{X}_1$ = mean of sample one, $\overline{X}_2$ = mean of sample two, n = No. of items in a sample, n_1 = No. of items in sample one, n_2 = No. of items in sample two, μ_{H_0} = Hypothesised mean for population, σ_p = standard deviation of population, σ_s = standard deviation of sample, p = population proportion, $q = 1 - p$, $\hat{p}$ = sample proportion, $\hat{q} = 1 - \hat{p}$.

and $$\sigma_s = \sqrt{\frac{\Sigma\left(X_i - \overline{X}\right)^2}{(n-1)}}$$

5. *Population may not be normal but sample size is large, variance of the population may be known or unknown, and H_a may be one-sided or two-sided:*

   In such a situation we use *z*-test and work out the test statistic *z* as under:

$$z = \frac{\overline{X} - \mu_{H_0}}{\sigma_p / \sqrt{n}}$$

(This applies in case of infinite population when variance of the population is known but when variance is not known, we use σ_s in place of σ_p in this formula.)

OR

$$z = \frac{\overline{X} - \mu_{H_0}}{\left(\sigma_p/\sqrt{n}\right) \times \sqrt{(N-n)/(N-1)}}$$

(This applies in case of finite population when variance of the population is known but when variance is not known, we use σ_s in place of σ_p in this formula.)

Illustration 2

A sample of 400 male students is found to have a mean height 67.47 inches. Can it be reasonably regarded as a sample from a large population with mean height 67.39 inches and standard deviation 1.30 inches? Test at 5% level of significance.

Solution: Taking the null hypothesis that the mean height of the population is equal to 67.39 inches, we can write:

$$H_0 : \mu_{H_0} = 67.39''$$
$$H_a : \mu_{H_0} \neq 67.39''$$

and the given information as $\overline{X} = 67.47''$, $\sigma_p = 1.30''$, $n = 400$. Assuming the population to be normal, we can work out the test statistic *z* as under:

$$z = \frac{\overline{X} - \mu_{H_0}}{\sigma_p/\sqrt{n}} = \frac{67.47 - 67.39}{1.30/\sqrt{400}} = \frac{0.08}{0.065} = 1.231$$

As H_a is two-sided in the given question, we shall be applying a two-tailed test for determining the rejection regions at 5% level of significance which comes to as under, using normal curve area table:

$$R : |\, z \,| > 1.96$$

The observed value of *z* is 1.231 which is in the acceptance region since $R : |\, z \,| > 1.96$ and thus H_0 is accepted. We may conclude that the given sample (with mean height = 67.47″) can be regarded to have been taken from a population with mean height 67.39″ and standard deviation 1.30″ at 5% level of significance.

Illustration 3

Suppose we are interested in a population of 20 industrial units of the same size, all of which are experiencing excessive labour turnover problems. The past records show that the mean of the distribution of annual turnover is 320 employees, with a standard deviation of 75 employees. A sample of 5 of these industrial units is taken at random which gives a mean of annual turnover as 300 employees. Is the sample mean consistent with the population mean? Test at 5% level.

Solution: Taking the null hypothesis that the population mean is 320 employees, we can write:

$$H_0 : \mu_{H_0} = 320 \text{ employees}$$

$$H_a : \mu_{H_0} \neq 320 \text{ employees}$$

and the given information as under:

$$\bar{X} = 300 \text{ employees}, \; \sigma_p = 75 \text{ employees}$$

$$n = 5;\ N = 20$$

Assuming the population to be normal, we can work out the test statistic z as under:

$$z^* = \frac{\bar{X} - \mu_{H_0}}{\sigma_p/\sqrt{n} \times \sqrt{(N-n)/(N-1)}}$$

$$= \frac{300 - 320}{75/\sqrt{5} \times \sqrt{(20-5)/(20-1)}} = -\frac{20}{(33.54)\,(.888)}$$

$$= -0.67$$

As H_a is two-sided in the given question, we shall apply a two-tailed test for determining the rejection regions at 5% level of significance which comes to as under, using normal curve area table:

$$R : |z| > 1.96$$

The observed value of z is –0.67 which is in the acceptance region since $R : |z| > 1.96$ and thus, H_0 is accepted and we may conclude that the sample mean is consistent with population mean i.e., the population mean 320 is supported by sample results.

Illustration 4

The mean of a certain production process is known to be 50 with a standard deviation of 2.5. The production manager may welcome any change is mean value towards higher side but would like to safeguard against decreasing values of mean. He takes a sample of 12 items that gives a mean value of 48.5. What inference should the manager take for the production process on the basis of sample results? Use 5 per cent level of significance for the purpose.

Solution: Taking the mean value of the population to be 50, we may write:

$$H_0 : \mu_{H_0} = 50$$

* Being a case of finite population.

$$H_a : \mu_{H_0} < 50 \text{ (Since the manager wants to safeguard against decreasing values of mean.)}$$

and the given information as $\overline{X} = 48.5$, $\sigma_p = 2.5$ and $n = 12$. Assuming the population to be normal, we can work out the test statistic z as under:

$$z = \frac{\overline{X} - \mu_{H_0}}{\sigma_p/\sqrt{n}} = \frac{48.5 - 50}{2.5/\sqrt{12}} = -\frac{1.5}{(2.5)/(3.464)} = -2.0784$$

As H_a is one-sided in the given question, we shall determine the rejection region applying one-tailed test (in the left tail because H_a is of less than type) at 5 per cent level of significance and it comes to as under, using normal curve area table:

$$R : z < -1.645$$

The observed value of z is – 2.0784 which is in the rejection region and thus, H_0 is rejected at 5 per cent level of significance. We can conclude that the production process is showing mean which is significantly less than the population mean and this calls for some corrective action concerning the said process.

Illustration 5

The specimen of copper wires drawn form a large lot have the following breaking strength (in kg. weight):

578, 572, 570, 568, 572, 578, 570, 572, 596, 544

Test (using Student's *t*-statistic)whether the mean breaking strength of the lot may be taken to be 578 kg. weight (Test at 5 per cent level of significance). Verify the inference so drawn by using Sandler's *A*-statistic as well.

Solution: Taking the null hypothesis that the population mean is equal to hypothesised mean of 578 kg., we can write:

$$H_0 : \mu = \mu_{H_0} = 578 \text{ kg.}$$
$$H_a : \mu \neq \mu_{H_0}$$

As the sample size is mall (since $n = 10$) and the population standard deviation is not known, we shall use *t*-test assuming normal population and shall work out the test statistic *t* as under:

$$t = \frac{\overline{X} - \mu_{H_0}}{\sigma_s/\sqrt{n}}$$

To find $\overline{X}$ and σ_s we make the following computations:

S. No.	X_i	$(X_i - \overline{X})$	$(X_i - \overline{X})^2$
1	578	6	36
2	572	0	0
3	570	– 2	4

Contd.

S. No.	X_i	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$
4	568	– 4	16
5	572	0	0
6	578	6	36
7	570	– 2	4
8	572	0	0
9	596	24	576
10	544	– 28	784
$n = 10$	$\sum X_i = 5720$		$\sum (X_i - \bar{X})^2 = 1456$

$$\therefore \qquad \bar{X} = \frac{\sum X_i}{n} = \frac{5720}{10} = 572 \text{ kg.}$$

and
$$\sigma_s = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n-1}} = \sqrt{\frac{1456}{10-1}} = 12.72 \text{ kg.}$$

Hence,
$$t = \frac{572 - 578}{12.72/\sqrt{10}} = -1.488$$

Degree of freedom = $(n - 1) = (10 - 1) = 9$

As H_a is two-sided, we shall determine the rejection region applying two-tailed test at 5 per cent level of significance, and it comes to as under, using table of *t*-distribution* for 9 d.f.:

$$R : | t | > 2.262$$

As the observed value of t (i.e., – 1.488) is in the acceptance region, we accept H_0 at 5 per cent level and conclude that the mean breaking strength of copper wires lot may be taken as 578 kg. weight.

The same inference can be drawn using Sandler's *A*-statistic as shown below:

Table 9.3: Computations for *A*-Statistic

S. No.	X_i	Hypothesised mean $m_{H_0} = 578$ kg.	$D_i = (X_i - \mu_{H_0})$	D_i^2
1	578	578	0	0
2	572	578	–6	36
3	570	578	–8	64
4	568	578	–10	100

contd.

* Table No. 2 given in appendix at the end of the book.

S. No.	X_i	*Hypothesised mean* $m_{H_0} = 578$ *kg.*	$D_i = \left(X_i - \mu_{H_0}\right)$	D_i^2
5	572	578	–6	36
6	578	578	0	0
7	570	578	–8	64
8	572	578	–6	36
9	596	578	18	324
10	544	578	–34	1156
$n = 10$			$\sum D_i = -60$	$\sum D_i^2 = 1816$

$$\therefore \qquad A = \sum D_i^2 / \left(\sum D_i\right)^2 = 1816/(-60)^2 = 0.5044$$

Null hypothesis $\quad H_0 : \mu_{H_0} = 578$ kg.

Alternate hypothesis $H_a : \mu_{H_0} \neq 578$ kg.

As H_a is two-sided, the critical value of *A*-statistic from the *A*-statistic table (Table No. 10 given in appendix at the end of the book) for $(n - 1)$ i.e., $10 - 1 = 9$ d.f. at 5% level is 0.276. Computed value of *A* (0.5044), being greater than 0.276 shows that *A*-statistic is insignificant in the given case and accordingly we accept H_0 and conclude that the mean breaking strength of copper wire' lot maybe taken as578 kg. weight. Thus, the inference on the basis of *t*-statistic stands verified by *A*-statistic.

Illustration 6

Raju Restaurant near the railway station at Falna has been having average sales of 500 tea cups per day. Because of the development of bus stand nearby, it expects to increase its sales. During the first 12 days after the start of the bus stand, the daily sales were as under:

550, 570, 490, 615, 505, 580, 570, 460, 600, 580, 530, 526

On the basis of this sample information, can one conclude that Raju Restaurant's sales have increased? Use 5 per cent level of significance.

Solution: Taking the null hypothesis that sales average 500 tea cups per day and they have not increased unless proved, we can write:

$H_0 : \mu = 500$ cups per day

$H_a : \mu > 500$ (as we want to conclude that sales have increased).

As the sample size is small and the population standard deviation is not known, we shall use *t*-test assuming normal population and shall work out the test statistic *t* as:

$$t = \frac{\overline{X} - \mu}{\sigma_s / \sqrt{n}}$$

(To find $\overline{X}$ and σ_s we make the following computations:)

Table 9.4

S. No.	X_i	$\left(X_i - \overline{X}\right)$	$\left(X_i - \overline{X}\right)^2$
1	550	2	4
2	570	22	484
3	490	–58	3364
4	615	67	4489
5	505	–43	1849
6	580	32	1024
7	570	22	484
8	460	–88	7744
9	600	52	2704
10	580	32	1024
11	530	–18	324
12	526	–22	484
$n = 10$	$\sum X_i = 6576$		$\sum\left(X_i - \overline{X}\right)^2 = 23978$

$\therefore$ $$\overline{X} = \frac{\sum X_i}{n} = \frac{6576}{12} = 548$$

and $$\sigma_s = \sqrt{\frac{\sum\left(X_i - \overline{X}\right)^2}{n-1}} = \sqrt{\frac{23978}{12-1}} = 46.68$$

Hence, $$t = \frac{548 - 500}{46.68/\sqrt{12}} = \frac{48}{13.49} = 3.558$$

Degree of freedom = $n - 1 = 12 - 1 = 11$

As H_a is one-sided, we shall determine the rejection region applying one-tailed test (in the right tail because H_a is of more than type) at 5 per cent level of significance and it comes to as under, using table of t-distribution for 11 degrees of freedom:

$$R : t > 1.796$$

The observed value of t is 3.558 which is in the rejection region and thus H_0 is rejected at 5 per cent level of significance and we can conclude that the sample data indicate that Raju restaurant's sales have increased.

HYPOTHESIS TESTING FOR DIFFERENCES BETWEEN MEANS

In many decision-situations, we may be interested in knowing whether the parameters of two populations are alike or different. For instance, we may be interested in testing whether female workers earn less than male workers for the same job. We shall explain now the technique of hypothesis testing for differences between means. The null hypothesis for testing of difference between means is generally stated as $H_0 : \mu_1 = \mu_2$, where μ_1 is population mean of one population and μ_2 is population mean of the second population, assuming both the populations to be normal populations. Alternative hypothesis may be of not equal to or less than or greater than type as stated earlier and accordingly we shall determine the acceptance or rejection regions for testing the hypotheses. There may be different situations when we are examining the significance of difference between two means, but the following may be taken as the usual situations:

1. *Population variances are known or the samples happen to be large samples:*

   In this situation we use *z*-test for difference in means and work out the test statistic *z* as under:

$$z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_{p1}^2}{n_1} + \frac{\sigma_{p2}^2}{n_2}}}$$

   In case σ_{p_1} and σ_{p_2} are not known, we use σ_{s_1} and σ_{s_2} respectively in their places calculating

$$\sigma_{s_1} = \sqrt{\frac{\Sigma\left(X_{1i} - \bar{X}_1\right)^2}{n_1 - 1}} \text{ and } \sigma_{s_2} = \sqrt{\frac{\Sigma\left(X_{2i} - \bar{X}_2\right)^2}{n_2 - 1}}$$

2. *Samples happen to be large but presumed to have been drawn from the same population whose variance is known:*

   In this situation we use *z* test for difference in means and work out the test statistic *z* as under:

$$z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\sigma_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

   In case σ_p is not known, we use $\sigma_{s_{1.2}}$ (combined standard deviation of the two samples) in its place calculating

$$\sigma_{s_{1.2}} = \sqrt{\frac{n_1\left(\sigma_{s_1}^2 + D_1^2\right) + n_2\left(\sigma_{s_2}^2 + D_2^2\right)}{n_1 + n_2}}$$

   where $D_1 = \left(\bar{X}_1 - \bar{X}_{1.2}\right)$

$$D_2 = \left(\bar{X}_2 - \bar{X}_{1.2}\right)$$

$$\bar{X}_{1.2} = \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2}{n_1 + n_2}$$

3. *Samples happen to be small samples and population variances not known but assumed to be equal:*

   In this situation we use t-test for difference in means and work out the test statistic t as under:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{\sum\left(X_{1i} - \bar{X}_1\right)^2 + \sum\left(X_{2i} - \bar{X}_2\right)^2}{n_1 + n_2 - 2}} \times \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

with d.f. = $(n_1 + n_2 - 2)$

Alternatively, we can also state

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{(n_1 - 1)\sigma_{s_1}^2 + (n_2 - 1)\sigma_{s_2}^2}{n_1 + n_2 - 2}} \times \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

with d.f. = $(n_1 + n_2 - 2)$

Illustration 7

The mean produce of wheat of a sample of 100 fields in 200 lbs. per acre with a standard deviation of 10 lbs. Another samples of 150 fields gives the mean of 220 lbs. with a standard deviation of 12 lbs. Can the two samples be considered to have been taken from the same population whose standard deviation is 11 lbs? Use 5 per cent level of significance.

Solution: Taking the null hypothesis that the means of two populations do not differ, we can write

$$H_0 : \mu = \mu_2$$

$$H_a : \mu_1 \neq \mu_2$$

and the given information as $n_1 = 100$; $n_2 = 150$;

$$\bar{X}_1 = 200 \text{ lbs.}; \qquad \bar{X}_2 = 220 \text{ lbs.};$$

$$\sigma_{s_1} = 10 \text{ lbs.}; \qquad \sigma_{s_2} = 12 \text{ lbs.};$$

and $$\sigma_p = 11 \text{ lbs.}$$

Assuming the population to be normal, we can work out the test statistic z as under:

$$z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\sigma_p^2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} = \frac{200 - 220}{\sqrt{(11)^2\left(\dfrac{1}{100} + \dfrac{1}{150}\right)}}$$

$$= -\frac{20}{1.42} = -14.08$$

As H_a is two-sided, we shall apply a two-tailed test for determining the rejection regions at 5 per cent level of significance which come to as under, using normal curve area table:

$$R : | z | > 1.96$$

The observed value of z is – 14.08 which falls in the rejection region and thus we reject H_0 and conclude that the two samples cannot be considered to have been taken at 5 per cent level of significance from the same population whose standard deviation is 11 lbs. This means that the difference between means of two samples is statistically significant and not due to sampling fluctuations.

Illustration 8

A simple random sampling survey in respect of monthly earnings of semi-skilled workers in two cities gives the following statistical information:

Table 9.5

City	*Mean monthly earnings (Rs)*	*Standard deviation of sample data of monthly earnings (Rs)*	*Size of sample*
A	695	40	200
B	710	60	175

Test the hypothesis at 5 per cent level that there is no difference between monthly earnings of workers in the two cities.

Solution: Taking the null hypothesis that there is no difference in earnings of workers in the two cities, we can write:

$$H_0 : \mu_1 = \mu_2$$
$$H_a : \mu_1 \neq \mu_2$$

and the given information as

Sample 1 (City A)	*Sample 2 (City B)*
$\overline{X}_1 = 695$ Rs	$\overline{X}_2 = 710$ Rs
$\sigma_{s_1} = 40$ Rs	$\sigma_{s_2} = 60$ Rs
$n_1 = 200$	$n_2 = 175$

As the sample size is large, we shall use z-test for difference in means assuming the populations to be normal and shall work out the test statistic z as under:

$$z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\dfrac{\sigma_{s_1}^2}{n_1} + \dfrac{\sigma_{s_2}^2}{n_2}}}$$

(Since the population variances are not known, we have used the sample variances, considering the sample variances as the estimates of population variances.)

$$\text{Hence } z = \frac{695 - 710}{\sqrt{\frac{(40)^2}{200} + \frac{(60)^2}{175}}} = -\frac{15}{\sqrt{8 + 20.57}} = -2.809$$

As H_a is two-sided, we shall apply a two-tailed test for determining the rejection regions at 5 per cent level of significance which come to as under, using normal curve area table:

$$R : |z| > 1.96$$

The observed value of z is – 2.809 which falls in the rejection region and thus we reject H_0 at 5 per cent level and conclude that earning of workers in the two cities differ significantly.

Illustration 9

Sample of sales in similar shops in two towns are taken for a new product with the following results:

Town	*Mean sales*	*Variance*	*Size of sample*
A	57	5.3	5
B	61	4.8	7

Is there any evidence of difference in sales in the two towns? Use 5 per cent level of significance for testing this difference between the means of two samples.

Solution: Taking the null hypothesis that the means of two populations do not differ we can write:

$$H_0 : \mu_1 = \mu_2$$
$$H_a : \mu_1 \neq \mu_2$$

and the given information as follows:

Table 9.6

Sample from town A as sample one	$\bar{X}_1 = 57$	$\sigma^2_{s_1} = 5.3$	$n_1 = 5$
Sample from town B As sample two	$\bar{X}_2 = 61$	$\sigma^2_{s_2} = 4.8$	$n_2 = 7$

Since in the given question variances of the population are not known and the size of samples is small, we shall use t-test for difference in means, assuming the populations to be normal and can work out the test statistic t as under:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1 - 1)\sigma^2_{s_1} + (n_2 - 1)\sigma^2_{s_2}}{n_1 + n_2 - 2}} \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

with d.f. = $(n_1 + n_2 - 2)$

$$= \frac{57 - 61}{\sqrt{\frac{4(5.3) + 6(4.8)}{5 + 7 - 2}} \times \sqrt{\frac{1}{5} + \frac{1}{7}}} = -3.053$$

Degrees of freedom = $(n_1 + n_2 - 2) = 5 + 7 - 2 = 10$

As H_a is two-sided, we shall apply a two-tailed test for determining the rejection regions at 5 per cent level which come to as under, using table of t-distribution for 10 degrees of freedom:

$$R : |t| > 2.228$$

The observed value of t is – 3.053 which falls in the rejection region and thus, we reject H_0 and conclude that the difference in sales in the two towns is significant at 5 per cent level.

Illustration 10

A group of seven-week old chickens reared on a high protein diet weigh 12, 15, 11, 16, 14, 14, and 16 ounces; a second group of five chickens, similarly treated except that they receive a low protein diet, weigh 8, 10, 14, 10 and 13 ounces. Test at 5 per cent level whether there is significant evidence that additional protein has increased the weight of the chickens. Use assumed mean (or A_1) = 10 for the sample of 7 and assumed mean (or A_2) = 8 for the sample of 5 chickens in your calculations.

Solution: Taking the null hypothesis that additional protein has not increased the weight of the chickens we can write:

$H_0 : \mu_1 = \mu_2$

$H_a : \mu_1 > \mu_2$ (as we want to conclude that additional protein has increased the weight of chickens)

Since in the given question variances of the populations are not known and the size of samples is small, we shall use t-test for difference in means, assuming the populations to be normal and thus work out the test statistic t as under:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1 - 1)\sigma_{s_1}^2 + (n_2 - 1)\sigma_{s_2}^2}{n_1 + n_2 - 2}} \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

with d.f. = $(n_1 + n_2 - 2)$

From the sample data we work out $\bar{X}_1, \bar{X}_2, \sigma_{s_1}^2$ and $\sigma_{s_2}^2$ (taking high protein diet sample as sample one and low protein diet sample as sample two) as shown below:

Table 9.7

Sample one				*Sample two*			
S.No.	X_{1i}	$X_{1i} - A_1$ ($A_1 = 10$)	$(X_{1i} - A_1)^2$	S.No.	X_{2i}	$X_{2i} - A_2$ ($A_2 = 8$)	$(X_{2i} - A_2)^2$
1.	12	2	4	1.	8	0	0
2.	15	5	25	2.	10	2	4
3.	11	1	1	3.	14	6	36
4.	16	6	36	4.	10	2	4
5.	14	4	16	5.	13	5	25
6.	14	4	16				
7.	16	6	36				
$n_1 = 7$;	$\Sigma(X_{1i} - A_1) = 28$;		$\Sigma(X_{1i} - A_1)^2 = 134$	$n_2 = 5$;	$\Sigma(X_{2i} - A_2) = 15$;		$\Sigma(X_{2i} - A_2)^2 = 69$

$\therefore$

$$\bar{X}_1 = A_1 + \frac{\Sigma(X_{1i} - A_1)}{n_1} = 10 + \frac{28}{7} = 14 \text{ ounces}$$

$$\bar{X}_2 = A_2 + \frac{\Sigma(X_{2i} - A_2)}{n_2} = 8 + \frac{15}{5} = 11 \text{ ounces}$$

$$\sigma^2_{s_1} = \frac{\Sigma(X_{1i} - A_1)^2 - \left[\Sigma(X_{1i} - A_1)\right]^2 / n_1}{(n_1 - 1)}$$

$$= \frac{134 - (28)^2/7}{7 - 1} = 3.667 \text{ ounces}$$

$$\sigma^2_{s_2} = \frac{\Sigma(X_{2i} - A_2)^2 - \left[\Sigma(X_{2i} - A_2)\right]^2 / n_2}{(n_2 - 1)}$$

$$= \frac{69 - (15)^2/5}{5 - 1} = 6 \text{ ounces}$$

Hence,

$$t = \frac{14 - 11}{\sqrt{\dfrac{(7-1)(3.667) + (5-1)(6)}{7 + 5 - 2}} \times \sqrt{\dfrac{1}{7} + \dfrac{1}{5}}}$$

$$= \frac{3}{\sqrt{4.6} \times \sqrt{.345}} = \frac{3}{1.26} = 2.381$$

Degrees of freedom $= (n_1 + n_2 - 2) = 10$

As H_a is one-sided, we shall apply a one-tailed test (in the right tail because H_a is of more than type) for determining the rejection region at 5 per cent level which comes to as under, using table of t-distribution for 10 degrees of freedom:

$$R : t > 1.812$$

The observed value of t is 2.381 which falls in the rejection region and thus, we reject H_0 and conclude that additional protein has increased the weight of chickens, at 5 per cent level of significance.

HYPOTHESIS TESTING FOR COMPARING TWO RELATED SAMPLES

Paired t-test is a way to test for comparing two related samples, involving small values of n that does not require the variances of the two populations to be equal, but the assumption that the two populations are normal must continue to apply. For a paired t-test, it is necessary that the observations in the two samples be collected in the form of what is called matched pairs i.e., "each observation in the one sample must be paired with an observation in the other sample in such a manner that these observations are somehow "matched" or related, in an attempt to eliminate extraneous factors which are not of interest in test."[5] Such a test is generally considered appropriate in a before-and-after-treatment study. For instance, we may test a group of certain students before and after training in order to know whether the training is effective, in which situation we may use paired t-test. To apply this test, we first work out the difference score for each matched pair, and then find out the average of such differences, $\overline{D}$, along with the sample variance of the difference score. If the values from the two matched samples are denoted as X_i and Y_i and the differences by D_i ($D_i = X_i - Y_i$), then the mean of the differences i.e.,

$$\overline{D} = \frac{\sum D_i}{n}$$

and the variance of the differences or

$$\left(\sigma_{diff.}\right)^2 = \frac{\sum D_i^2 - \left(\overline{D}\right)^2 \cdot n}{n - 1}$$

Assuming the said differences to be normally distributed and independent, we can apply the paired t-test for judging the significance of mean of differences and work out the test statistic t as under:

$$t = \frac{\overline{D} - 0}{\sigma_{diff} / \sqrt{n}} \text{ with } (n - 1) \text{ degrees of freedom}$$

where $\overline{D}$ = Mean of differences

[5] Donald L. Harnett and James L. Murphy, "*Introductory Statistical Analysis*", p. 364.

$\sigma_{diff.}$ = Standard deviation of differences

n = Number of matched pairs

This calculated value of t is compared with its table value at a given level of significance as usual for testing purposes. We can also use Sandler's A-test for this very purpose as stated earlier in Chapter 8.

Illustration 11

Memory capacity of 9 students was tested before and after training. State at 5 per cent level of significance whether the training was effective from the following scores:

Student	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
Before	10	15	9	3	7	12	16	17	4
After	12	17	8	5	6	11	18	20	3

Use paired t-test as well as A-test for your answer.

Solution: Take the score before training as X and the score after training as Y and then taking the null hypothesis that the mean of difference is zero, we can write:

$H_0 : \mu_1 = \mu_2$ which is equivalent to test $H_0 : \overline{D} = 0$

$H_a : \mu_1 < \mu_2$ (as we want to conclude that training has been effective)

As we are having matched pairs, we use paired t-test and work out the test statistic t as under:

$$t = \frac{\overline{D} - 0}{\sigma_{diff.}/\sqrt{n}}$$

To find the value of t, we shall first have to work out the mean and standard deviation of differences as shown below:

Table 9.8

Student	*Score before training* X_i	*Score after training* Y_i	*Difference* $(D_i = X_i - Y_i)$	*Difference Squared* D_i^2
1	10	12	– 2	4
2	15	17	– 2	4
3	9	8	1	1
4	3	5	– 2	4
5	7	6	1	1
6	12	11	1	1
7	16	18	– 2	4
8	17	20	– 3	9
9	4	3	1	1
$n = 9$			$\sum D_i = -7$	$\sum D_i^2 = 29$

$\therefore$ Mean of Differences or $\bar{D} = \dfrac{\sum D_i}{n} = \dfrac{-7}{9} = -0.778$

and Standard deviation of differences or

$$\sigma_{diff.} = \sqrt{\frac{\sum D_i^2 - \left(\bar{D}\right)^2 \cdot n}{n-1}}$$

$$= \sqrt{\frac{29 - (-.778)^2 \times 9}{9-1}}$$

$$= \sqrt{2.944} = 1.715$$

Hence, $$t = \frac{-0.778 - 0}{1.715/\sqrt{9}} = \frac{-.778}{0.572} = -1.361$$

Degrees of freedom = n – 1 = 9 – 1 = 8.

As H_a is one-sided, we shall apply a one-tailed test (in the left tail because H_a is of less than type) for determining the rejection region at 5 per cent level which comes to as under, using the table of t-distribution for 8 degrees of freedom:

$$R : t < -1.860$$

The observed value of t is –1.361 which is in the acceptance region and thus, we accept H_0 and conclude that the difference in score before and after training is insignificant i.e., it is only due to sampling fluctuations. Hence we can infer that the training was not effective.

Solution using A-test: Using A-test, we workout the test statistic for the given problem thus:

$$A = \frac{\sum D_i^2}{\left(\sum D_i\right)^2} = \frac{29}{(-7)^2} = 0.592$$

Since H_a in the given problem is one-sided, we shall apply one-tailed test. Accordingly, at 5% level of significance the table value of A-statistic for (n – 1) or (9 – 1) = 8 d.f. in the given case is 0.368 (as per table of A-statistic given in appendix). The computed value of A i.e., 0.592 is higher than this table value and as such A-statistic is insignificant and accordingly H_0 should be accepted. In other words, we should conclude that the training was not effective. (This inference is just the same as drawn earlier using paired t-test.)

Illustration 12

The sales data of an item in six shops before and after a special promotional campaign are:

Shops	*A*	*B*	*C*	*D*	*E*	*F*
Before the promotional campaign	53	28	31	48	50	42
After the campaign	58	29	30	55	56	45

Can the campaign be judged to be a success? Test at 5 per cent level of significance. Use paired t-test as well as A-test.

Solution: Let the sales before campaign be represented as X and the sales after campaign as Y and then taking the null hypothesis that campaign does not bring any improvement in sales, we can write:

$H_0: \mu_1 = \mu_2$ which is equivalent to test $H_0: \overline{D} = 0$

$H_a: \mu_1 < \mu_2$ (as we want to conclude that campaign has been a success).

Because of the matched pairs we use paired t-test and work out the test statistic 't' as under:

$$t = \frac{\overline{D} - 0}{\sigma_{diff.}/\sqrt{n}}$$

To find the value of t, we first work out the mean and standard deviation of differences as under:

Table 9.9

Shops	*Sales before campaign* X_i	*Sales after campaign* Y_i	*Difference* $(D_i = X_i - Y_i)$	*Difference squared* D_i^2
A	53	58	–5	25
B	28	29	–1	1
C	31	30	1	1
D	48	55	–7	49
E	50	56	–6	36
F	42	45	–3	9
$n = 6$			$\sum D_i = -21$	$\sum D_i^2 = 121$

$$\therefore \qquad \overline{D} = \frac{\sum D_i}{n} = -\frac{21}{6} = -3.5$$

$$\sigma_{diff.} = \sqrt{\frac{\sum D_i^2 - \left(\overline{D}\right)^2 \cdot n}{n-1}} = \sqrt{\frac{121 - (-3.5)^2 \times 6}{6-1}} = 3.08$$

Hence,
$$t = \frac{-3.5 - 0}{3.08/\sqrt{6}} = \frac{-3.5}{1.257} = -2.784$$

Degrees of freedom = $(n - 1) = 6 - 1 = 5$

As H_a is one-sided, we shall apply a one-tailed test (in the left tail because H_a is of less than type) for determining the rejection region at 5 per cent level of significance which come to as under, using table of t-distribution for 5 degrees of freedom:

$$R : t < -2.015$$

The observed value of t is -2.784 which falls in the rejection region and thus, we reject H_0 at 5 per cent level and conclude that sales promotional campaign has been a success.

Solution: *Using A-test:* Using A-test, we work out the test statistic for the given problem as under:

$$A = \frac{\sum D_i^2}{\left(\sum D_i\right)^2} = \frac{121}{(-21)^2} = 0.2744$$

Since H_a in the given problem is one-sided, we shall apply one-tailed test. Accordingly, at 5% level of significance the table value of A-statistic for $(n-1)$ or $(6-1) = 5$ d.f. in the given case is 0.372 (as per table of A-statistic given in appendix). The computed value of A, being 0.2744, is less than this table value and as such A-statistic is significant. This means we should reject H_0 (alternately we should accept H_a) and should infer that the sales promotional campaign has been a success.

HYPOTHESIS TESTING OF PROPORTIONS

In case of qualitative phenomena, we have data on the basis of presence or absence of an attribute(s). With such data the sampling distribution may take the form of binomial probability distribution whose mean would be equal to $n \cdot p$ and standard deviation equal to $\sqrt{n \cdot p \cdot q}$, where p represents the probability of success, q represents the probability of failure such that $p + q = 1$ and n, the size of the sample. Instead of taking mean number of successes and standard deviation of the number of successes, we may record the proportion of successes in each sample in which case the mean and standard deviation (or the standard error) of the sampling distribution may be obtained as follows:

Mean proportion of successes $= (n \cdot p)/n = p$

and standard deviation of the proportion of successes $= \sqrt{\dfrac{p \cdot q}{n}}$.

In n is large, the binomial distribution tends to become normal distribution, and as such for proportion testing purposes we make use of the test statistic z as under:

$$z = \frac{\hat{p} - p}{\sqrt{\dfrac{p \cdot q}{n}}}$$

where $\hat{p}$ is the sample proportion.

For testing of proportion, we formulate H_0 and H_a and construct rejection region, presuming normal approximation of the binomial distribution, for a predetermined level of significance and then may judge the significance of the observed sample result. The following examples make all this quite clear.

Illustration 13

A sample survey indicates that out of 3232 births, 1705 were boys and the rest were girls. Do these figures confirm the hypothesis that the sex ratio is 50 : 50? Test at 5 per cent level of significance.

Solution: Starting from the null hypothesis that the sex ratio is 50 : 50 we may write:

$$H_0 : p = p_{H_0} = \frac{1}{2}$$

$$H_a : p \neq p_{H_0}$$

Hence the probability of boy birth or $p = \frac{1}{2}$ and the probability of girl birth is also $\frac{1}{2}$.

Considering boy birth as success and the girl birth as failure, we can write as under:

$$\text{the proportion success or } p = \frac{1}{2}$$

$$\text{the proportion of failure or } q = \frac{1}{2}$$

and $n = 3232$ (given).

The standard error of proportion of success.

$$= \sqrt{\frac{p \cdot q}{n}} = \sqrt{\frac{\frac{1}{2} \times \frac{1}{2}}{3232}} = 0.0088$$

Observed sample proportion of success, or

$$\hat{p} = 1705/3232 = 0.5275$$

and the test statistic

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{0.5275 - .5000}{.0088} = 3.125$$

As H_a is two-sided in the given question, we shall be applying the two-tailed test for determining the rejection regions at 5 per cent level which come to as under, using normal curve area table:

$$R : | z | > 1.96$$

The observed value of z is 3.125 which comes in the rejection region since $R : | z | > 1.96$ and thus, H_0 is rejected in favour of H_a. Accordingly, we conclude that the given figures do not conform the hypothesis of sex ratio being 50 : 50.

Illustration 14

The null hypothesis is that 20 per cent of the passengers go in first class, but management recognizes the possibility that this percentage could be more or less. A random sample of 400 passengers includes 70 passengers holding first class tickets. Can the null hypothesis be rejected at 10 per cent level of significance?

Solution: The null hypothesis is

$H_0 : p = 20\%$ or 0.20

and $H_a : p \neq 20\%$

Hence, $p = 0.20$ and
$q = 0.80$

Observed sample proportion $(\hat{p}) = 70/400 = 0.175$

and the test statistic $z = \dfrac{\hat{p} - p}{\sqrt{\dfrac{p \cdot q}{n}}} = \dfrac{0.175 - .20}{\sqrt{\dfrac{.20 \times .80}{400}}} = -1.25$

As H_a is two-sided we shall determine the rejection regions applying two-tailed test at 10 per cent level which come to as under, using normal curve area table:

$$R : |z| > 1.645$$

The observed value of z is –1.25 which is in the acceptance region and as such H_0 is accepted. Thus the null hypothesis cannot be rejected at 10 per cent level of significance.

Illustration 15

A certain process produces 10 per cent defective articles. A supplier of new raw material claims that the use of his material would reduce the proportion of defectives. A random sample of 400 units using this new material was taken out of which 34 were defective units. Can the supplier's claim be accepted? Test at 1 per cent level of significance.

Solution: The null hypothesis can be written as $H_0 : p = 10\%$ or 0.10 and the alternative hypothesis $H_a : p < 0.10$ (because the supplier claims that new material will reduce proportion of defectives). Hence,

$$p = 0.10 \text{ and } q = 0.90$$

Observed sample proportion $\hat{p} = 34/400 = 0.085$ and test statistic

$$z = \frac{\hat{p} - p}{\sqrt{\dfrac{p \cdot q}{n}}} = \frac{.085 - .10}{\sqrt{\dfrac{.10 \times .90}{400}}} = \frac{.015}{.015} = -1.00$$

As H_a is one-sided, we shall determine the rejection region applying one-tailed test (in the left tail because H_a is of less than type) at 1% level of significance and it comes to as under, using normal curve area table:

$$R : z < -2.32$$

As the computed value of z does not fall in the rejection region, H_0 is accepted at 1% level of significance and we can conclude that on the basis of sample information, the supplier's claim cannot be accepted at 1% level.

HYPOTHESIS TESTING FOR DIFFERENCE BETWEEN PROPORTIONS

If two samples are drawn from different populations, one may be interested in knowing whether the difference between the proportion of successes is significant or not. In such a case, we start with the hypothesis that the difference between the proportion of success in sample one $(\hat{p}_1)$ and the proportion of success in sample two $(\hat{p}_2)$ is due to fluctuations of random sampling. In other words, we take the null hypothesis as $H_0 : \hat{p}_1 = \hat{p}_2$ and for testing the significance of difference, we work out the test statistic as under:

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1 \cdot \hat{q}_1}{n_1} + \frac{\hat{p}_2 \cdot \hat{q}_2}{n_2}}}$$

where $\hat{p}_1$ = proportion of success in sample one

$\hat{p}_2$ = proportion of success in sample two

$\hat{q}_1 = 1 - \hat{p}_1$

$\hat{q}_2 = 1 - \hat{p}_2$

n_1 = size of sample one

n_2 = size of sample two

and

$\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$ = the standard error of difference between two sample proportions.*

Then, we construct the rejection region(s) depending upon the H_a for a given level of significance and on its basis we judge the significance of the sample result for accepting or rejecting H_0. We can now illustrate all this by examples.

Illustration 6

A drug research experimental unit is testing two drugs newly developed to reduce blood pressure levels. The drugs are administered to two different sets of animals. In group one, 350 of 600 animals tested respond to drug one and in group two, 260 of 500 animals tested respond to drug two. The research unit wants to test whether there is a difference between the efficacy of the said two drugs at 5 per cent level of significance. How will you deal with this problem?

* This formula is used when samples are drawn from two heterogeneous populations where we cannot have the best estimate of the common value of the proportion of the attribute in the population from the given sample information. But on the assumption that the populations are similar as regards the given attribute, we make use of the following formula for working out the standard error of difference between proportions of the two samples:

$$\text{S.E.}_{Diff.\, p_1 - p_2} = \sqrt{\frac{p_0 \cdot q_0}{n_1} + \frac{p_0 \cdot q_0}{n_2}}$$

where $p_0 = \frac{n_1 \cdot \hat{p}_1 + n_2 \cdot \hat{p}_2}{n_1 + n_2}$ = best estimate of proportion in the population

$q_0 = 1 - p_0$

Solution: We take the null hypothesis that there is no difference between the two drugs i.e., $H_0 : \hat{p}_1 = \hat{p}_2$

The alternative hypothesis can be taken as that there is a difference between the drugs i.e., $H_a : \hat{p}_1 \neq \hat{p}_2$ and the given information can be stated as:

$$\hat{p}_1 = 350/600 = 0.583$$
$$\hat{q}_1 = 1 - \hat{p}_1 = 0.417$$
$$n_1 = 600$$
$$\hat{p}_2 = 260/500 = 0.520$$
$$\hat{q}_2 = 1 - \hat{p}_2 = 0.480$$
$$n_2 = 500$$

We can work out the test statistic z thus:

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\dfrac{\hat{p}_1\hat{q}_1}{n_1} + \dfrac{\hat{p}_2\hat{q}_2}{n_2}}} = \frac{0.583 - 0.520}{\sqrt{\dfrac{(.583)(.417)}{600} + \dfrac{(.520)(.480)}{500}}}$$

$$= 2.093$$

As H_a is two-sided, we shall determine the rejection regions applying two-tailed test at 5% level which comes as under using normal curve area table:

$$R : |z| > 1.96$$

The observed value of z is 2.093 which is in the rejection region and thus, H_0 is rejected in favour of H_a and as such we conclude that the difference between the efficacy of the two drugs is significant.

Illustration 17

At a certain date in a large city 400 out of a random sample of 500 men were found to be smokers. After the tax on tobacco had been heavily increased, another random sample of 600 men in the same city included 400 smokers. Was the observed decrease in the proportion of smokers significant? Test at 5 per cent level of significance.

Solution: We start with the null hypothesis that the proportion of smokers even after the heavy tax on tobacco remains unchanged i.e. $H_0 : \hat{p}_1 = \hat{p}_2$ and the alternative hypothesis that proportion of smokers after tax has decreased i.e.,

$$H_a : \hat{p}_1 > \hat{p}_2$$

On the presumption that the given populations are similar as regards the given attribute, we work out the best estimate of proportion of smokers (p_0) in the population as under, using the given information:

$$p_0 = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{500\left(\dfrac{400}{500}\right) + 600\left(\dfrac{400}{600}\right)}{500 + 600} = \frac{800}{1100} = \frac{8}{11} = .7273$$

Thus, $q_0 = 1 - p_0 = .2727$

The test statistic z can be worked out as under:

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{p_0 q_0}{n_1} + \frac{p_0 q_0}{n_2}}} = \frac{\frac{400}{500} - \frac{400}{600}}{\sqrt{\frac{(.7273)(.2727)}{500} + \frac{(.7273)(.2727)}{600}}}$$

$$= \frac{0.133}{0.027} = 4.926$$

As the H_a is one-sided we shall determine the rejection region applying one-tailed test (in the right tail because H_a is of greater than type) at 5 per cent level and the same works out to as under, using normal curve area table:

$$R : z > 1.645$$

The observed value of z is 4.926 which is in the rejection region and so we reject H_0 in favour of H_a and conclude that the proportion of smokers after tax has decreased significantly.

Testing the difference between proportion based on the sample and the proportion given for the whole population: In such a situation we work out the standard error of difference between proportion of persons possessing an attribute in a sample and the proportion given for the population as under:

Standard error of difference between sample proportion and

population proportion or $\text{S.E.}_{diff.\hat{p}-p} = \sqrt{p \cdot q \frac{N-n}{nN}}$

where p = population proportion

$q = 1 - p$

n = number of items in the sample

N = number of items in population

and the test statistic z can be worked out as under:

$$z = \frac{\hat{p} - p}{\sqrt{p \cdot q \frac{N-n}{nN}}}$$

All other steps remain the same as explained above in the context of testing of proportions. We take an example to illustrate the same.

Illustration 18

There are 100 students in a university college and in the whole university, inclusive of this college, the number of students is 2000. In a random sample study 20 were found smokers in the college and the proportion of smokers in the university is 0.05. Is there a significant difference between the proportion of smokers in the college and university? Test at 5 per cent level.

Solution: Let $H_0: \hat{p} = p$ (there is no difference between sample proportion and population proportion)

and $H_a: \hat{p} \neq p$ (there is difference between the two proportions)

and on the basis of the given information, the test statistic z can be worked out as under:

$$z = \frac{\hat{p} - p}{\sqrt{p \cdot q \dfrac{N-n}{nN}}} = \frac{\dfrac{20}{100} - .05}{\sqrt{(.05)(.95)\dfrac{2000-100}{(100)(2000)}}}$$

$$= \frac{0.150}{0.021} = 7.143$$

As the H_a is two-sided, we shall determine the rejection regions applying two-tailed test at 5 per cent level and the same works out to as under, using normal curve area table:

$$R: |z| > 1.96$$

The observed value of z is 7.143 which is in the rejection region and as such we reject H_0 and conclude that there is a significant difference between the proportion of smokers in the college and university.

HYPOTHESIS TESTING FOR COMPARING A VARIANCE TO SOME HYPOTHESISED POPULATION VARIANCE

The test we use for comparing a sample variance to some theoretical or hypothesised variance of population is different than z-test or the t-test. The test we use for this purpose is known as chi-square test and the test statistic symbolised as χ^2, known as the chi-square value, is worked out. The chi-square value to test the null hypothesis viz, $H_0: \sigma_s^2 = \sigma_p^2$ worked out 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

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

Then by comparing the calculated value of χ^2 with its table value for $(n – 1)$ degrees of freedom at a given level of significance, we may either accept H_0 or reject it. If the calculated value of χ^2 is equal to or less than the table value, the null hypothesis is accepted; otherwise the null hypothesis is rejected. This test is based on chi-square distribution which is not symmetrical and all the values happen to be positive; one must simply know the degrees of freedom for using such a distribution.*

TESTING THE EQUALITY OF VARIANCES OF TWO NORMAL POPULATIONS

When we want to test the equality of variances of two normal populations, we make use of *F*-test based on *F*-distribution. In such a situation, the null hypothesis happens to be $H_0 : \sigma^2_{p_1} = \sigma^2_{p_2}$, $\sigma^2_{p_1}$ and $\sigma^2_{p_2}$ representing the variances of two normal populations. This hypothesis is tested on the basis of sample data and the test statistic *F* is found, using $\sigma^2_{s_1}$ and $\sigma^2_{s_2}$ the sample estimates for $\sigma^2_{p_1}$ and $\sigma^2_{p_2}$ respectively, as stated below:

$$F = \frac{\sigma^2_{s_1}}{\sigma^2_{s_2}}$$

where $\sigma^2_{s_1} = \dfrac{\Sigma\left(X_{1i} - \overline{X}_1\right)^2}{\left(n_1 - 1\right)}$ and $\sigma^2_{s_2} = \dfrac{\Sigma\left(X_{2i} - \overline{X}_2\right)^2}{\left(n_2 - 1\right)}$

While calculating *F*, $\sigma^2_{s_1}$ is treated $> \sigma^2_{s_2}$ which means that the numerator is always the greater variance. Tables for *F*-distribution** have been prepared by statisticians for different values of *F* at different levels of significance for different degrees of freedom for the greater and the smaller variances. By comparing the observed value of *F* with the corresponding table value, we can infer whether the difference between the variances of samples could have arisen due to sampling fluctuations. If the calculated value of *F* is greater than table value of *F* at a certain level of significance for $(n_1 - 1)$ and $(n_2 - 2)$ degrees of freedom, we regard the *F*-ratio as significant. Degrees of freedom for greater variance is represented as v_1 and for smaller variance as v_2. On the other hand, if the calculated value of *F* is smaller than its table value, we conclude that *F*-ratio is not significant. If *F*-ratio is considered non-significant, we accept the null hypothesis, but if *F*-ratio is considered significant, we then reject H_0 (i.e., we accept H_a).

When we use the *F*-test, we presume that

(i) the populations are normal;
(ii) samples have been drawn randomly;
(iii) observations are independent; and
(iv) there is no measurement error.

The object of *F*-test is to test the hypothesis whether the two samples are from the same normal population with equal variance or from two normal populations with equal variances. *F*-test was initially used to verify the hypothesis of equality between two variances, but is now mostly used in the

*See Chapter 10 entitled Chi-square test for details.

** *F*-distribution tables [Table 4(a) and Table 4(b)] have been given in appendix at the end of the book.

context of analysis of variance. The following examples illustrate the use of *F*-test for testing the equality of variances of two normal populations.

Illustration 19

Two random samples drawn from two normal populations are:

Sample 1	20	16	26	27	23	22	18	24	25	19		
Sample 2	27	33	42	35	32	34	38	28	41	43	30	37

Test using variance ratio at 5 per cent and 1 per cent level of significance whether the two populations have the same variances.

Solution: We take the null hypothesis that the two populations from where the samples have been drawn have the same variances i.e., $H_0: \sigma^2_{p_1} = \sigma^2_{p_2}$. From the sample data we work out $\sigma^2_{s_1}$ and $\sigma^2_{s_2}$ as under:

Table 9.10

Sample 1			*Sample 2*		
X_{1i}	$(X_{1i} - \bar{X}_1)$	$(X_{1i} - \bar{X}_1)^2$	X_{2i}	$(X_{2i} - \bar{X}_2)$	$(X_{2i} - \bar{X}_2)^2$
20	–2	4	27	–8	64
16	–6	36	33	–2	4
26	4	16	42	7	49
27	5	25	35	0	0
23	1	1	32	–3	9
22	0	0	34	–1	1
18	–4	16	38	3	9
24	2	4	28	–7	49
25	3	9	41	6	36
19	–3	9	43	8	64
			30	–5	25
			37	2	4
$\sum X_{1i} = 220$		$\sum(X_{1i} - \bar{X}_1)^2 = 120$	$\sum X_{2i} = 420$		$\sum(X_{2i} - \bar{X}_2)^2 = 314$
$n_1 = 10$			$n_2 = 12$		

$$\bar{X}_1 = \frac{\sum X_{1i}}{n_1} = \frac{220}{10} = 22; \qquad \bar{X}_2 = \frac{\sum X_{2i}}{n_2} = \frac{420}{12} = 35$$

$$\therefore \qquad \sigma^2_{s_1} = \frac{\sum(X_{1i} - \bar{X}_1)^2}{n_1 - 1} = \frac{120}{10 - 1} = 13.33$$

and $$\sigma^2_{s_2} = \frac{\Sigma\left(X_{2i} - \bar{X}_2\right)^2}{n_2 - 1} = \frac{314}{12 - 1} = 28.55$$

Hence, $$F = \frac{\sigma^2_{s_2}}{\sigma^2_{s_1}} \qquad \left(\because \sigma^2_{s_2} > \sigma^2_{s_1}\right)$$

$$= \frac{28.55}{13.33} = 2.14$$

Degrees of freedom in sample 1 = $(n_1 - 1) = 10 - 1 = 9$

Degrees of freedom in sample 2 = $(n_2 - 1) = 12 - 1 = 11$

As the variance of sample 2 is greater variance, hence

$$v_1 = 11;\ v_2 = 9$$

The table value of F at 5 per cent level of significance for $v_1 = 11$ and $v_2 = 9$ is 3.11 and the table value of F at 1 per cent level of significance for $v_1 = 11$ and $v_2 = 9$ is 5.20.

Since the calculated value of $F = 2.14$ which is less than 3.11 and also less than 5.20, the F ratio is insignificant at 5 per cent as well as at 1 per cent level of significance and as such we accept the null hypothesis and conclude that samples have been drawn from two populations having the same variances.

Illustration 20

Given $n_1 = 9$; $n_2 = 8$

$$\Sigma\left(X_{1i} - \bar{X}_1\right)^2 = 184$$

$$\Sigma\left(X_{2i} - \bar{X}_2\right)^2 = 38$$

Apply F-test to judge whether this difference is significant at 5 per cent level.

Solution: We start with the hypothesis that the difference is not significant and hence, $H_0 : \sigma^2_{p_1} = \sigma^2_{p_2}$.

To test this, we work out the F-ratio as under:

$$F = \frac{\sigma^2_{s_1}}{\sigma^2_{s_2}} = \frac{\Sigma\left(X_{1i} - \bar{X}_1\right)^2/(n_1 - 1)}{\Sigma\left(X_{2i} - \bar{X}_2\right)^2/(n_2 - 1)}$$

$$= \frac{184/8}{38/7} = \frac{23}{5.43} = 4.25$$

$v_1 = 8$ being the number of d.f. for greater variance

$v_2 = 7$ being the number of d.f. for smaller variance.

The table value of F at 5 per cent level for $v_1 = 8$ and $v_2 = 7$ is 3.73. Since the calculated value of F is greater than the table value, the F ratio is significant at 5 per cent level. Accordingly we reject H_0 and conclude that the difference is significant.

HYPOTHESIS TESTING OF CORRELATION COEFFICIENTS*

We may be interested in knowing whether the correlation coefficient that we calculate on the basis of sample data is indicative of significant correlation. For this purpose we may use (in the context of small samples) normally either the t-test or the F-test depending upon the type of correlation coefficient. We use the following tests for the purpose:

(a) *In case of simple correlation coefficient:* We use t-test and calculate the test statistic as under:

$$t = r_{yx}\sqrt{\frac{n-2}{1-r_{yx}^2}}$$

with $(n - 2)$ degrees of freedom r_{yx} being coefficient of simple correlation between x and y.

This calculated value of t is then compared with its table value and if the calculated value is less than the table value, we accept the null hypothesis at the given level of significance and may infer that there is no relationship of statistical significance between the two variables.

(b) *In case of partial correlation coefficient:* We use t-test and calculate the test statistic as under:

$$t = r_p\sqrt{\frac{(n-k)}{1-r_p^2}}$$

with $(n - k)$ degrees of freedom, n being the number of paired observations and k being the number of variables involved, r_p happens to be the coefficient of partial correlation.

If the value of t in the table is greater than the calculated value, we may accept null hypothesis and infer that there is no correlation.

(c) *In case of multiple correlation coefficient:* We use F-test and work out the test statistic as under:

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$

where R is any multiple coefficient of correlation, k being the number of variables involved and n being the number of paired observations. The test is performed by entering tables of the F-distribution with

$v_1 = k - 1$ = degrees of freedom for variance in numerator.

$v_2 = n - k$ = degrees of freedom for variance in denominator.

If the calculated value of F is less than the table value, then we may infer that there is no statistical evidence of significant correlation.

*Only the outline of testing procedure has been given here. Readers may look into standard tests for further details.

LIMITATIONS OF THE TESTS OF HYPOTHESES

We have described above some important test often used for testing hypotheses on the basis of which important decisions may be based. But there are several limitations of the said tests which should always be borne in mind by a researcher. Important limitations are as follows:

(i) The tests should not be used in a mechanical fashion. It should be kept in view that testing is not decision-making itself; the tests are only useful aids for decision-making. Hence "proper interpretation of statistical evidence is important to intelligent decisions."[6]

(ii) Test do not explain the reasons as to why does the difference exist, say between the means of the two samples. They simply indicate whether the difference is due to fluctuations of sampling or because of other reasons but the tests do not tell us as to which is/are the other reason(s) causing the difference.

(iii) Results of significance tests are based on probabilities and as such cannot be expressed with full certainty. When a test shows that a difference is statistically significant, then it simply suggests that the difference is probably not due to chance.

(iv) Statistical inferences based on the significance tests cannot be said to be entirely correct evidences concerning the truth of the hypotheses. This is specially so in case of small samples where the probability of drawing erring inferences happens to be generally higher. For greater reliability, the size of samples be sufficiently enlarged.

All these limitations suggest that in problems of statistical significance, the inference techniques (or the tests) must be combined with adequate knowledge of the subject-matter along with the ability of good judgement.

Questions

1. Distinguish between the following:
   (i) Simple hypothesis and composite hypothesis;
   (ii) Null hypothesis and alternative hypothesis;
   (iii) One-tailed test and two-tailed test;
   (iv) Type I error and Type II error;
   (v) Acceptance region and rejection region;
   (vi) Power function and operating characteristic function.
2. What is a hypothesis? What characteristics it must possess in order to be a good research hypothesis? A manufacturer considers his production process to be working properly if the mean length of the rods the manufactures is 8.5″. The standard deviation of the rods always runs about 0.26″. Suppose a sample of 64 rods is taken and this gives a mean length of rods equal to 8.6″. What are the null and alternative hypotheses for this problem? Can you infer at 5% level of significance that the process is working properly?
3. The procedure of testing hypothesis requires a researcher to adopt several steps. Describe in brief all such steps.

[6] Ya-Lun-Chou, "Applied Business and Economic Statistics".

4. What do you mean by the power of a hypothesis test? How can it be measured? Describe and illustrate by an example.
5. Briefly describe the important parametric tests used in context of testing hypotheses. How such tests differ from non-parametric tests? Explain.
6. Clearly explain how will you test the equality of variances of two normal populations.
7. (a) What is a *t*-test? When it is used and for what purpose(s)? Explain by means of examples.
   (b) Write a brief note on "Sandler's *A*-test" explaining its superiority over *t*-test.
8. Point out the important limitations of tests of hypotheses. What precaution the researcher must take while drawing inferences as per the results of the said tests?
9. A coin is tossed 10,000 times and head turns up 5,195 times. Is the coin unbiased?
10. In some dice throwing experiments, *A* threw dice 41952 times and of these 25145 yielded a 4 or 5 or 6. Is this consistent with the hypothesis that the dice were unbiased?
11. A machine puts out 16 imperfect articles in a sample of 500. After machine is overhauled, it puts out three imperfect articles in a batch of 100. Has the machine improved? Test at 5% level of significance.
12. In two large populations, there are 35% and 30% respectively fair haired people. Is this difference likely to be revealed by simple sample of 1500 and 1000 respectively from the two populations?
13. In a certain association table the following frequencies were obtained:

    $(AB) = 309, (Ab) = 214, (aB) = 132, (ab) = 119.$

    Can the association between *AB* as per the above data can be said to have arisen as a fluctuation of simple sampling?
14. A sample of 900 members is found to have a mean of 3.47 cm. Can it be reasonably regarded as a simple sample from a large population with mean 3.23 cm. and standard deviation 2.31 cm.?
15. The means of the two random samples of 1000 and 2000 are 67.5 and 68.0 inches respectively. Can the samples be regarded to have been drawn from the same population of standard deviation 9.5 inches? Test at 5% level of significance.
16. A large corporation uses thousands of light bulbs every year. The brand that has been used in the past has an average life of 1000 hours with a standard deviation of 100 hours. A new brand is offered to the corporation at a price far lower than one they are paying for the old brand. It is decided that they will switch to the new brand unless it is proved with a level of significance of 5% that the new brand has smaller average life than the old brand. A random sample of 100 new brand bulbs is tested yielding an observed sample mean of 985 hours. Assuming that the standard deviation of the new brand is the same as that of the old brand,
    (a) What conclusion should be drawn and what decision should be made?
    (b) What is the probability of accepting the new brand if it has the mean life of 950 hours?
17. Ten students are selected at random from a school and their heights are found to be, in inches, 50, 52, 52, 53, 55, 56, 57, 58, 58 and 59. In the light of these data, discuss the suggestion that the mean height of the students of the school is 54 inches. You may use 5% level of significance (Apply *t*-test as well as *A*-test).
18. In a test given to two groups of students, the marks obtained were as follows:

First Group	18	20	36	50	49	36	34	49	41
Second Group	29	28	26	35	30	44	46		

    Examine the significance of difference between mean marks obtained by students of the above two groups. Test at five per cent level of significance.
19. The heights of six randomly chosen sailors are, in inches, 63, 65, 58, 69, 71 and 72. The heights of 10 randomly chosen soldiers are, in inches, 61, 62, 65, 66, 69, 69, 70, 71, 72 and 73. Do these figures indicate that soldiers are on an average shorter than sailors? Test at 5% level of significance.

20. Ten young recruits were put through a strenuous physical training programme by the army. Their weights (in kg) were recorded before and after with the following results:

Recruit	1	2	3	4	5	6	7	8	9	10
Weight before	127	195	162	170	143	205	168	175	197	136
Weight after	135	200	160	182	147	200	172	186	194	141

Using 5% level of significance, should we conclude that the programme affects the average weight of young recruits (Answer using *t*-test as well as *A*-test)?

21. Suppose a test on the hypotheses $H_0: \mu = 200$ against $H_a: \mu > 200$ is done with 1% level of significance, $\sigma_p = 40$ and $n = 16$.

(a) What is the probability that the null hypothesis might be accepted when the true mean is really 210? What is the power of the test for $\mu = 210$? How these values of β and $1 - \beta$ change if the test had used 5% level of significance?

(b) Which is more serious, a Type I and Type II error?

22. The following nine observations were drawn from a normal population:

27 19 20 24 23 29 21 17 27

(i) Test the null hypothesis $H_0: \mu = 26$ against the alternative hypothesis $H_a: \mu \neq 26$. At what level of significance can H_0 be rejected?

(ii) At what level of significance can $H_0: \mu = 26$ be rejected when tested against $H_a: \mu < 26$?

23. Suppose that a public corporation has agreed to advertise through a local newspaper if it can be established that the newspaper circulation reaches more than 60% of the corporation's customers. What H_0 and H_a should be established for this problem while deciding on the basis of a sample of customers whether or not the corporation should advertise in the local newspaper? If a sample of size 100 is collected and 1% level of significance is taken, what is the critical value for making a decision whether or not to advertise? Would it make any difference if we take a sample of 25 in place of 100 for our purpose? If so, explain.

24. Answer using *F*-test whether the following two samples have come from the same population:

Sample 1 17 27 18 25 27 29 27 23 17

Sample 2 16 16 20 16 20 17 15 21

Use 5% level of significance.

25. The following table gives the number of units produced per day by two workers *A* and *B* for a number of days:

A 40 30 38 41 38 35

B 39 38 41 33 32 49 49 34

Should these results be accepted as evidence that *B* is the more stable worker? Use *F*-test at 5% level.

26. A sample of 600 persons selected at random from a large city gives the result that males are 53%. Is there reason to doubt the hypothesis that males and females are in equal numbers in the city? Use 1% level of significance.

27. 12 students were given intensive coaching and 5 tests were conducted in a month. The scores of tests 1 and 5 are given below. Does the score from Test 1 to Test 5 show an improvement? Use 5% level of significance.

No. of students	1	2	3	4	5	6	7	8	9	10	11	12
Marks in 1st Test	50	42	51	26	35	42	60	41	70	55	62	38
Marks in 5th test	62	40	61	35	30	52	68	51	84	63	72	50

28. (i) A random sample from 200 villages was taken from Kanpur district and the average population per village was found to be 420 with a standard deviation of 50. Another random sample of 200 villages from the same district gave an average population of 480 per village with a standard deviation of 60. Is the difference between the averages of the two samples statistically significant? Take 1% level of significance.

(ii) The means of the random samples of sizes 9 and 7 are 196.42 and 198.42 respectively. The sums of he squares of the deviations from the mean are 26.94 and 18.73 respectively. Can the samples be constituted to have been drawn from the same normal population? Use 5% level of significance.

29. A farmer grows crops on two fields *A* and *B*. On *A* he puts Rs. 10 worth of manure per acre and on *B* Rs 20 worth. The net returns per acre exclusive of the cost of manure on the two fields in the five years are:

Year	1	2	3	4	5
Field A, Rs per acre	34	28	42	37	44
Field B, Rs per acre	36	33	48	38	50

Other things being equal, discuss the question whether it is likely to pay the farmer to continue the more expensive dressing. Test at 5% level of significance.

30. *ABC* Company is considering a site for locating their another plant. The company insists that any location they choose must have an average auto traffic of more than 2000 trucks per day passing the site. They take a traffic sample of 20 days and find an average volume per day of 2140 with standard deviation equal to 100 trucks.

Answer the following:

(i) If $\alpha = .05$, should they purchase the site?

(ii) If we assume the population mean to be 2140, what is the β error?

10

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."[1] 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.

* See Chapter 12 Testing of Hypotheses-II for more details.

[1] Neil R. Ullman, *Elementary Statistics—An Applied Approach*, p. 234.

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:

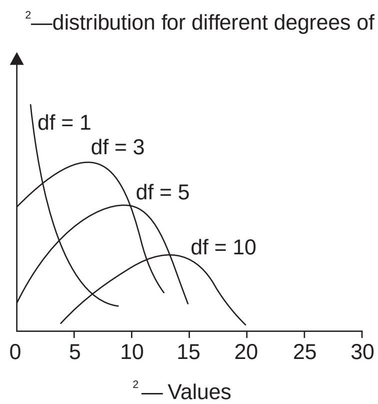

Fig. 10.1

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:

Table 10.1

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
$n = 10$	$\sum X_i = 470$	$\sum (X_i - \bar{X})^2 = 280$	

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

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

or $$\sigma_s^2 = 31.11.$$

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 $\chi^2 = 16.92$ 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: Given information is

$$n = 10$$

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

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

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 variance 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:

$$\chi^2 = \Sigma \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

where

O_{ij} = observed frequency of the cell in *i*th row and *j*th column.

E_{ij} = expected frequency of the cell in *i*th row and *j*th column.

If two distributions (observed and theoretical) are exactly alike, $\chi^2 = 0$; 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:

$$\text{d.f.} = (c - 1)\,(r - 1)$$

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

CONDITIONS FOR THE APPLICATION OF χ^2 TEST

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 itmes 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.

STEPS INVOLVED IN APPLYING CHI-SQUARE TEST

The various steps involved are as follows:

* For d.f. greater than 30, the distribution of $\sqrt{2\chi^2}$ approximates the normal distribution wherein the mean of $\sqrt{2\chi^2}$ distribution is $\sqrt{2\text{d.f.} - 1}$ and the standard deviation = 1. Accordingly, when d.f. exceeds 30, the quantity $\left[\sqrt{2\chi^2} - \sqrt{2\text{d.f.} - 1}\right]$ may be used as a normal variate with unit variance, i.e.,

$$z_\alpha = \sqrt{2\chi^2} - \sqrt{2\text{d.f.} - 1}$$

(i) First of all calculate the expected frequencies on the basis of given hypothesis or on the basis of null hypothesis. Usually in case of a 2 × 2 or any contingency table, the expected frequency for any given cell is worked out as under:

$$\text{Expected frequency of any cell} = \left[\frac{\text{(Row total for the row of that cell)} \times \text{(Column total for the column of that cell)}}{\text{(Grand total)}}\right]$$

(ii) Obtain the difference between observed and expected frequencies and find out the squares of such differences i.e., calculate $(O_{ij} - E_{ij})^2$.

(iii) Divide the quantity $(O_{ij} - E_{ij})^2$ obtained as stated above by the corresponding expected frequency to get $(O_{ij} - E_{ij})^2/E_{ij}$ and this should be done for all the cell frequencies or the group frequencies.

(iv) Find the summation of $(O_{ij} - E_{ij})^2/E_{ij}$ values or what we call $\sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$. This is the required χ^2 value.

The χ^2 value obtained as such should be compared with relevant table value of χ^2 and then inference be drawn as stated above.

We now give few examples to illustrate the use of χ^2 test.

Illustration 3

A die is thrown 132 times with following results:

Number turned up	1	2	3	4	5	6
Frequency	16	20	25	14	29	28

Is the die unbiased?

Solution: Let us take the hypothesis that the die is unbiased. If that is so, the probability of obtaining any one of the six numbers is 1/6 and as such the expected frequency of any one number coming upward is 132 ×1/6 = 22. Now we can write the observed frequencies along with expected frequencies and work out the value of χ^2 as follows:

Table 10.2

No. turned up	*Observed frequency* O_i	*Expected frequency* E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$(O_i - E_i)^2/E_i$
1	16	22	–6	36	36/22
2	20	22	–2	4	4/22
3	25	22	3	9	9/22
4	14	22	–8	64	64/22
5	29	22	7	49	49/22
6	28	22	6	36	36/22

$$\therefore \qquad \Sigma\,[(O_i - E_i)^2/E_i] = 9.$$

Hence, the calculated value of $\chi^2 = 9$.

$\because$ Degrees of freedom in the given problem is

$(n - 1) = (6 - 1) = 5.$

The table value* of χ^2 for 5 degrees of freedom at 5 per cent level of significance is 11.071. Comparing calculated and table values of χ^2, we find that calculated value is less than the table value and as such could have arisen due to fluctuations of sampling. The result, thus, supports the hypothesis and it can be concluded that the die is unbiased.

Illustration 4

Find the value of χ^2 for the following information:

Class	*A*	*B*	*C*	*D*	*E*
Observed frequency	8	29	44	15	4
Theoretical (or expected) frequency	7	24	38	24	7

Solution: Since some of the frequencies less than 10, we shall first re-group the given data as follows and then will work out the value of χ^2 :

Table 10.3

Class	*Observed frequency* O_i	*Expected frequency* E_i	$O_i - E_i$	$(O_i - E_i)^2/E_i$
A and *B*	$(8 + 29) = 37$	$(7 + 24) = 31$	6	36/31
C	44	38	6	36/38
D and *E*	$(15 + 4) = 19$	$(24 + 7) = 31$	–12	144/31

$$\therefore \qquad \chi^2 = \Sigma \frac{(O_i - E_i)^2}{E_i} = 6.76 \text{ app.}$$

Illustration 5

Genetic theory states that children having one parent of blood type *A* and the other of blood type *B* will always be of one of three types, *A*, *AB*, *B* and that the proportion of three types will on an average be as 1 : 2 : 1. A report states that out of 300 children having one *A* parent and *B* parent, 30 per cent were found to be types *A*, 45 per cent per cent type *AB* and remainder type *B*. Test the hypothesis by χ^2 test.

Solution: The observed frequencies of type *A*, *AB* and *B* is given in the question are 90, 135 and 75 respectively.

*Table No. 3 showing some critical values of χ^2 for specified degrees of freedom has been given in Appendix at the end of the book.

The expected frequencies of type *A*, *AB* and *B* (as per the genetic theory) should have been 75, 150 and 75 respectively.

We now calculate the value of χ^2 as follows:

Table 10.4

Type	*Observed frequency* O_i	*Expected frequency* E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$(O_i - E_i)^2/E_i$
A	90	75	15	225	225/75 = 3
AB	135	150	–15	225	225/150 = 1.5
B	75	75	0	0	0/75 = 0

$$\therefore \qquad \chi^2 = \Sigma \frac{(O_i - E_i)^2}{E_i} = 3 + 1.5 + 0 = 4.5$$

$$\because \qquad \text{d.f.} = (n - 1) = (3 - 1) = 2.$$

Table value of χ^2 for 2 d.f. at 5 per cent level of significance is 5.991.

The calculated value of χ^2 is 4.5 which is less than the table value and hence can be ascribed to have taken place because of chance. This supports the theoretical hypothesis of the genetic theory that on an average type *A*, *AB* and *B* stand in the proportion of 1 : 2 : 1.

Illustration 6

The table given below shows the data obtained during outbreak of smallpox:

	Attacked	*Not attacked*	*Total*
Vaccinated	31	469	500
Not vaccinated	185	1315	1500
Total	216	1784	2000

Test the effectiveness of vaccination in preventing the attack from smallpox. Test your result with the help of χ^2 at 5 per cent level of significance.

Solution: Let us take the hypothesis that vaccination is not effective in preventing the attack from smallpox i.e., vaccination and attack are independent. On the basis of this hypothesis, the expected frequency corresponding to the number of persons vaccinated and attacked would be:

$$\text{Expectation of } (AB) = \frac{(A) \times (B)}{N}$$

when *A* represents vaccination and *B* represents attack.

$$\therefore \qquad (A) = 500$$
$$(B) = 216$$
$$N = 2000$$

$$\text{Expectation of } (AB) = \frac{500 \times 216}{2000} = 54$$

Now using the expectation of (*AB*), we can write the table of expected values as follows:

	Attacked: B	*Not attacked: b*	*Total*
Vaccinated: *A*	(*AB*) = 54	(*Ab*) = 446	500
Not vaccinated: *a*	(*aB*) = 162	(*ab*) = 1338	1500
Total	216	1784	2000

Table 10.5: Calculation of Chi-Square

Group	*Observed frequency* O_{ij}	*Expected frequency* E_{ij}	$(O_{ij} - E_{ij})$	$(O_{ij} - E_{ij})^2$	$(O_{ij} - E_{ij})^2/E_{ij}$
AB	31	54	–23	529	529/54 = 9.796
Ab	469	446	+23	529	529/44 = 1.186
aB	158	162	+23	529	529/162 = 3.265
ab	1315	1338	–23	529	529/1338 = 0.395

$$\chi^2 = \Sigma \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = 14.642$$

$\because$ Degrees of freedom in this case = $(r - 1)(c - 1) = (2 - 1)(2 - 1) = 1$.

The table value of χ^2 for 1 degree of freedom at 5 per cent level of significance is 3.841. The calculated value of χ^2 is much higher than this table value and hence the result of the experiment does not support the hypothesis. We can, thus, conclude that vaccination is effective in preventing the attack from smallpox.

Illustration 7

Two research workers classified some people in income groups on the basis of sampling studies. Their results are as follows:

Investigators	*Income groups*			*Total*
	Poor	*Middle*	*Rich*	
A	160	30	10	200
B	140	120	40	300
Total	300	150	50	500

Show that the sampling technique of at least one research worker is defective.

Solution: Let us take the hypothesis that the sampling techniques adopted by research workers are similar (i.e., there is no difference between the techniques adopted by research workers). This being so, the expectation of *A* investigator classifying the people in

(i) Poor income group $= \dfrac{200 \times 300}{500} = 120$

(ii) Middle income group $= \dfrac{200 \times 150}{500} = 60$

(iii) Rich income group $= \dfrac{200 \times 50}{500} = 20$

Similarly the expectation of *B* investigator classifying the people in

(i) Poor income group $= \dfrac{300 \times 300}{500} = 180$

(ii) Middle income group $= \dfrac{300 \times 150}{500} = 90$

(iii) Rich income group $= \dfrac{300 \times 50}{500} = 30$

We can now calculate value of χ^2 as follows:

Table 10.6

Groups	*Observed frequency* O_{ij}	*Expected frequency* E_{ij}	$O_{ij} - E_{ij}$	$(O_{ij} - E_{ij})^2 E_{ij}$
Investigator A				
classifies people as poor	160	120	40	1600/120 = 13.33
classifies people as middle class people	30	60	–30	900/60 = 15.00
classifies people as rich	10	20	–10	100/20 = 5.00
Investigator B				
classifies people as poor	140	180	–40	1600/180 = 8.88
classifies people as middle class people	120	90	30	900/90 = 10.00
classifies people as rich	40	30	10	100/30 = 3.33

Hence, $$\chi^2 = \Sigma \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} = 55.54$$

$\because$ Degrees of freedom $= (c - 1)\ (r - 1)$
$= (3 - 1)\ (2 - 1) = 2.$

The table value of χ^2 for two degrees of freedom at 5 per cent level of significance is 5.991.

The calculated value of χ^2 is much higher than this table value which means that the calculated value cannot be said to have arisen just because of chance. It is significant. Hence, the hypothesis does not hold good. This means that the sampling techniques adopted by two investigators differ and are not similar. Naturally, then the technique of one must be superior than that of the other.

Illustration 8

Eight coins were tossed 256 times and the following results were obtained:

Numbers of heads	0	1	2	3	4	5	6	7	8
Frequency	2	6	30	52	67	56	32	10	1

Are the coins biased? Use χ^2 test.

Solution: Let us take the hypothesis that the coins are not biased. If that is so, the probability of any one coin falling with head upward is 1/2 and with tail upward is 1/2 and it remains the same whatever be the number of throws. In such a case the expected values of getting 0, 1, 2, … heads in a single throw in 256 throws of eight coins will be worked out as follows*.

Table 10.7

Events or No. of heads	*Expected frequencies*
0	${}^8C_0\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^8 \times 256 = 1$
1	${}^8C_1\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^7 \times 256 = 8$
2	${}^8C_2\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^6 \times 256 = 28$

contd.

*The probabilities of random variable i.e., various possible events have been worked out on the binomial principle viz., through the expansion of $(p + q)^n$ where $p = 1/2$ and $q = 1/2$ and $n = 8$ in the given case. The expansion of the term ${}^nC_r\, p^r\, q^{n-r}$ has given the required probabilities which have been multiplied by 256 to obtain the expected frequencies.

Events or No. of heads	*Expected frequencies*
3	${}^{8}C_{3}\left(\frac{1}{2}\right)^{3}\left(\frac{1}{2}\right)^{5} \times 256 = 56$
4	${}^{8}C_{4}\left(\frac{1}{2}\right)^{4}\left(\frac{1}{2}\right)^{4} \times 256 = 70$
5	${}^{8}C_{5}\left(\frac{1}{2}\right)^{5}\left(\frac{1}{2}\right)^{3} \times 256 = 56$
6	${}^{8}C_{6}\left(\frac{1}{2}\right)^{6}\left(\frac{1}{2}\right)^{2} \times 256 = 28$
7	${}^{8}C_{7}\left(\frac{1}{2}\right)^{7}\left(\frac{1}{2}\right)^{1} \times 256 = 8$
8	${}^{8}C_{8}\left(\frac{1}{2}\right)^{8}\left(\frac{1}{2}\right)^{0} \times 256 = 1$

The value of χ^2 can be worked out as follows:

Table 10.8

No. of heads	*Observed frequency* O_i	*Expected frequency* E_i	$O_i - E_i$	$(O_i - E_i)^2/E_i$
0	2	1	1	1/1 = 1.00
1	6	8	–2	4/8 = 0.50
2	30	28	2	4/28 = 0.14
3	52	56	–4	16/56 = 0.29
4	67	70	–3	9/70 = 0.13
5	56	56	0	0/56 = 0.00
6	32	28	4	16/28 = 0.57
7	10	8	2	4/8 = 0.50
8	1	1	0	0/1 = 0.00

$$\therefore \qquad \chi^2 = \Sigma \frac{(O_i - E_i)^2}{E_i} = 3.13$$

$\therefore$ Degrees of freedom = $(n - 1) = (9 - 1) = 8$

The table value of χ^2 for eight degrees of freedom at 5 per cent level of significance is 15.507. The calculated value of χ^2 is much less than this table and hence it is insignificant and can be ascribed due to fluctuations of sampling. The result, thus, supports the hypothesis and we may say that the coins are not biased.

ALTERNATIVE FORMULA

There is an alternative method of calculating the value of χ^2 in the case of a (2×2) table. If we write the cell frequencies and marginal totals in case of a (2×2) table thus,

a	*b*	(*a* + *b*)
c	*d*	(*c* + *d*)
(*a* + *c*)	(*b* + *d*)	*N*

then the formula for calculating the value of χ^2 will be stated as follows:

$$\chi^2 = \frac{(ad - bc)^2 \cdot N}{(a+c)(b+d)(a+b)(c+d)}$$

where *N* means the total frequency, *ad* means the larger cross product, *bc* means the smaller cross product and $(a + c)$, $(b + d)$, $(a + b)$, and $(c + d)$ are the marginal totals. The alternative formula is rarely used in finding out the value of chi-square as it is not applicable uniformly in all cases but can be used only in a (2×2) contingency table.

YATES' CORRECTION

F. Yates has suggested a correction for continuity in χ^2 value calculated in connection with a (2×2) table, particularly when cell frequencies are small (since no cell frequency should be less than 5 in any case, through 10 is better as stated earlier) and χ^2 is just on the significance level. The correction suggested by Yates is popularly known as Yates' correction. It involves the reduction of the deviation of observed from expected frequencies which of course reduces the value of χ^2. The rule for correction is to adjust the observed frequency in each cell of a (2×2) table in such a way as to reduce the deviation of the observed from the expected frequency for that cell by 0.5, but this adjustment is made in all the cells without disturbing the marginal totals. The formula for finding the value of χ^2 after applying Yates' correction can be stated thus:

$$\chi^2(\text{corrected}) = \frac{N \cdot \left(|ad - bc| - 0.5N\right)^2}{(a+b)(c+d)(a+c)(b+d)}$$

In case we use the usual formula for calculating the value of chi-square viz.,

$$\chi^2 = \Sigma \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}},$$

then Yates' correction can be applied as under:

$$\chi^2(\text{corrected}) = \frac{\left[|O_1 - E_1| - 0.5\right]^2}{E_1} + \frac{\left[|O_2 - E_2| - 0.5\right]^2}{E_2} + \ldots$$

It may again be emphasised that Yates' correction is made only in case of (2 × 2) table and that too when cell frequencies are small.

Illustration 9

The following information is obtained concerning an investigation of 50 ordinary shops of small size:

	Shops		*Total*
	In towns	*In villages*	
Run by men	17	18	35
Run by women	3	12	15
Total	20	30	50

Can it be inferred that shops run by women are relatively more in villages than in towns? Use χ^2 test.

Solution: Take the hypothesis that there is no difference so far as shops run by men and women in towns and villages. With this hypothesis the expectation of shops run by men in towns would be:

$$\text{Expectation of } (AB) = \frac{(A) \times (B)}{N}$$

where A = shops run by men

B = shops in towns

$(A) = 35$; $(B) = 20$ and $N = 50$

$$\text{Thus, expectation of } (AB) = \frac{35 \times 20}{50} = 14$$

Hence, table of expected frequencies would be

	Shops in towns	*Shops in villages*	*Total*
Run by men	14 (*AB*)	21 (*Ab*)	35
Run by women	6 (*aB*)	9 (*ab*)	15
Total	20	30	50

Calculation of χ^2 value:

Table 10.9

Groups	*Observed frequency* O_{ij}	*Expected frequency* E_{ij}	$(O_{ij} - E_{ij})$	$(O_{ij} - E_{ij})^2/E_{ij}$
(*AB*)	17	14	3	9/14 = 0.64
(*Ab*)	18	21	–3	9/21 = 0.43
(*aB*)	3	6	–3	9/6 = 1.50
(*ab*)	12	9	3	9/9 = 1.00

$$\therefore \qquad \chi^2 = \Sigma \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} = 3.57$$

As one cell frequency is only 3 in the given 2 × 2 table, we also work out χ^2 value applying Yates' correction and this is as under:

$$\chi^2 \text{(corrected)} = \frac{\left[\left|17 - 14\right| - 0.5\right]^2}{14} + \frac{\left[\left|18 - 21\right| - 0.5\right]^2}{21} + \frac{\left[\left|3 - 6\right| - 0.5\right]^2}{6} + \frac{\left[\left|12 - 9\right| - 0.5\right]^2}{9}$$

$$= \frac{(2.5)^2}{14} + \frac{(2.5)^2}{21} + \frac{(2.5)^2}{6} + \frac{(2.5)^2}{9}$$

$$= 0.446 + 0.298 + 1.040 + 0.694$$

$$= 2.478$$

$\because$ Degrees of freedom = $(c - 1)(r - 1) = (2 - 1)(2 - 1) = 1$

Table value of χ^2 for one degree of freedom at 5 per cent level of significance is 3.841. The calculated value of χ^2 by both methods (i.e., before correction and after Yates' correction) is less than its table value. Hence the hypothesis stands. We can conclude that there is no difference between shops run by men and women in villages and towns.

Additive property: An important property of χ^2 is its additive nature. This means that several values of χ^2 can be added together and if the degrees of freedom are also added, this number gives the degrees of freedom of the total value of χ^2. Thus, if a number of χ^2 values have been obtained from a number of samples of similar data, then because of the additive nature of χ^2 we can combine the various values of χ^2 by just simply adding them. Such addition of various values of χ^2 gives one value of χ^2 which helps in forming a better idea about the significance of the problem under consideration. The following example illustrates the additive property of χ^2.

Illustration 10

The following values of χ^2 from different investigations carried to examine the effectiveness of a recently invented medicine for checking malaria are obtained:

Investigation	χ^2	*d.f.*
1	2.5	1
2	3.2	1
3	4.1	1
4	3.7	1
5	4.5	1

What conclusion would you draw about the effectiveness of the new medicine on the basis of the five investigations taken together?

Solution: By adding all the values of χ^2, we obtain a value equal to 18.0. Also by adding the various d.f., as given in the question, we obtain the value 5. We can now state that the value of χ^2 for 5 degrees of freedom (when all the five investigations are taken together) is 18.0.

Let us take the hypothesis that the new medicine is not effective. The table value of χ^2 for 5 degrees of freedom at 5 per cent level of significance is 11.070. But our calculated value is higher than this table value which means that the difference is significant and is not due to chance. As such the hypothesis is rejected and it can be concluded that the new medicine is effective in checking malaria.

CONVERSION OF CHI-SQUARE INTO PHI COEFFICIENT (ϕ)

Since χ^2 does not by itself provide an estimate of the magnitude of association between two attributes, any obtained χ^2 value may be converted into Phi coefficient (symbolized as ϕ) for the purpose. In other words, chi-square tells us about the significance of a relation between variables; it provides no answer regarding the magnitude of the relation. This can be achieved by computing the Phi coefficient, which is a non-parametric measure of coefficient of correlation, as under:

$$\phi = \sqrt{\frac{\chi^2}{N}}$$

CONVERSION OF CHI-SQUARE INTO COEFFICIENT OF CONTINGENCY (*C*)

Chi-square value may also be converted into coefficient of contingency, especially in case of a contingency table of higher order than 2 × 2 table to study the magnitude of the relation or the degree of association between two attributes, as shown below:

$$C = \sqrt{\frac{\chi^2}{\chi^2 + N}}$$

While finding out the value of *C* we proceed on the assumption of null hypothesis that the two attributes are independent and exhibit no association. Coefficient of contingency is also known as coefficient of Mean Square contingency. This measure also comes under the category of non-parametric measure of relationship.

IMPORTANT CHARACTERISTICS OF χ^2 TEST

(i) This test (as a non-parametric test) is based on frequencies and not on the parameters like mean and standard deviation.

(ii) The test is used for testing the hypothesis and is not useful for estimation.

(iii) This test possesses the additive property as has already been explained.

(iv) This test can also be applied to a complex contingency table with several classes and as such is a very useful test in research work.

(v) This test is an important non-parametric test as no rigid assumptions are necessary in regard to the type of population, no need of parameter values and relatively less mathematical details are involved.

CAUTION IN USING χ^2 TEST

The chi-square test is no doubt a most frequently used test, but its correct application is equally an uphill task. It should be borne in mind that the test is to be applied only when the individual observations of sample are independent which means that the occurrence of one individual observation (event) has no effect upon the occurrence of any other observation (event) in the sample under consideration. Small theoretical frequencies, if these occur in certain groups, should be dealt with under special care. The other possible reasons concerning the improper application or misuse of this test can be (i) neglect of frequencies of non-occurrence; (ii) failure to equalise the sum of observed and the sum of the expected frequencies; (iii) wrong determination of the degrees of freedom; (iv) wrong computations, and the like. The researcher while applying this test must remain careful about all these things and must thoroughly understand the rationale of this important test before using it and drawing inferences in respect of his hypothesis.