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 $(\mu)$ 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 $\alpha$ (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 $\alpha$ (alpha) known as $\alpha$ error, also called the level of significance of test; and Type II error is denoted by $\beta$ (beta) known as $\beta$ 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 ($\alpha$ error) |
| $H_0$ (false) | Type II error ($\beta$ 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:

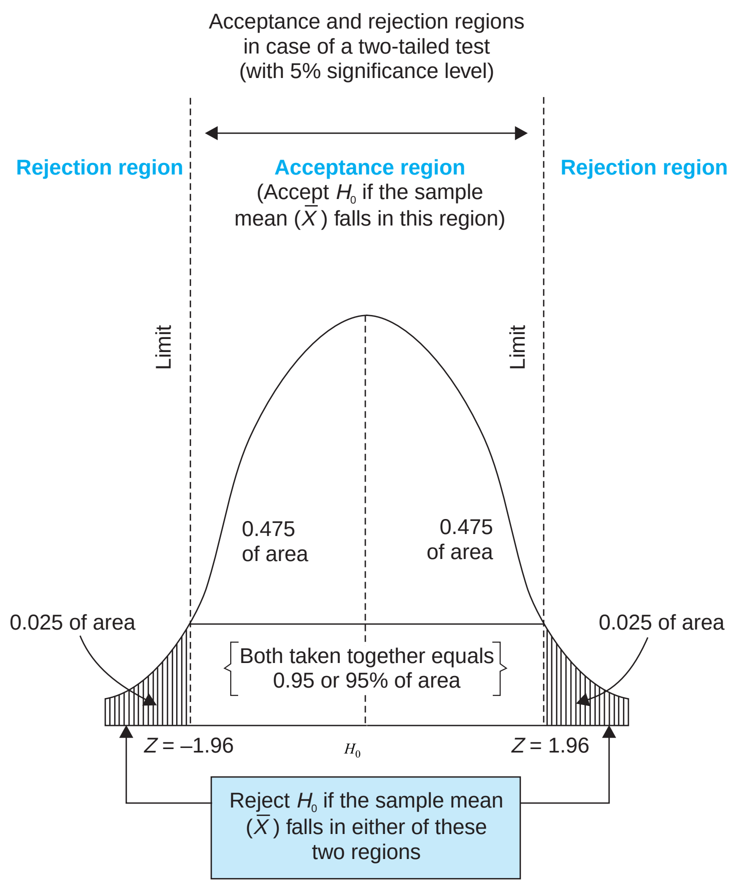


**Fig. 9.1**

[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 $\mu$, 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:

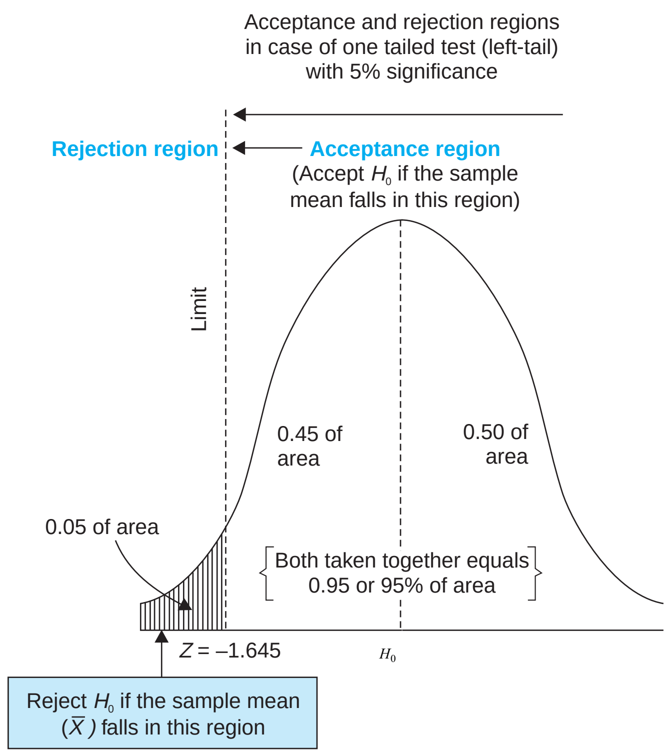


**Fig. 9.2**

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:

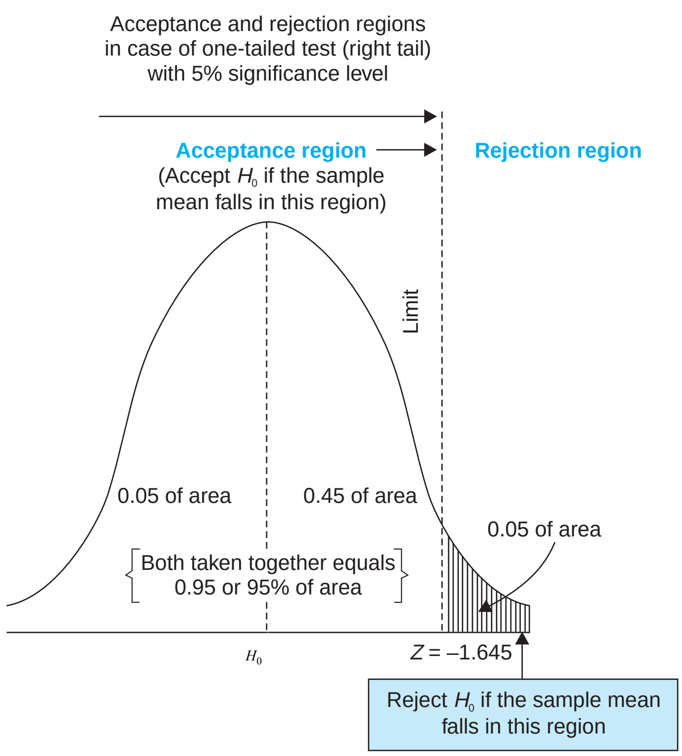


**Fig. 9.3**

Mathematically we can state:

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

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

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 $\alpha$, the significance level. If the calculated probability is equal to or smaller than the $\alpha$ 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

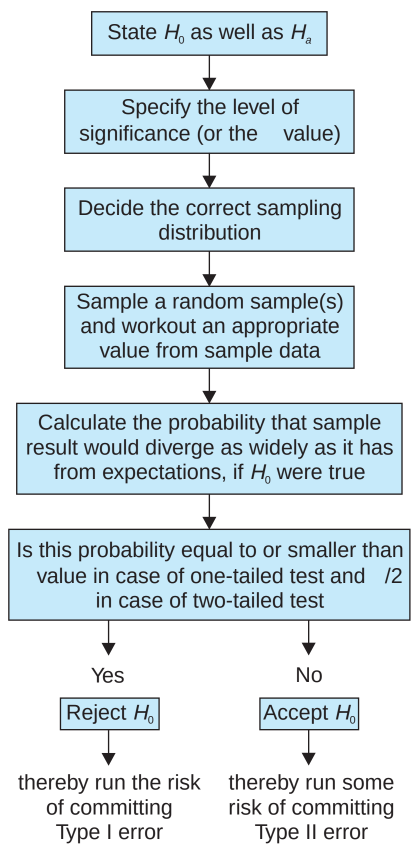


Fig. 9.4

[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 $\alpha$ (the significance level of the test) and the probability of Type II error is referred to as $\beta$. Usually the significance level of a test is assigned in advance and once we decide it, there is nothing else we can do about $\alpha$. But what can we say about $\beta$? 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 $\beta$ (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 $\mu$, 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 $\alpha$ (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 $\mu = 16$ lbs. If we raise the level of significance to 20 per cent, then how the power of the test for $\mu = 16$ 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:

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

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 $\mu$ for accepting $H_0$ as under:

Accept $H_0$ if $\bar{X} \leq 15 + 1.28\left(\alpha_p/\sqrt{n}\right)$

or $\bar{X} \leq 15 + 1.28\left(5/\sqrt{100}\right)$

or $\bar{X} \leq 15.64$

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 $\beta$ and then subtract it from one. Since $\beta$ is a conditional probability which depends on the value of $\mu$, we take it as 16 as given in the question. We can now write $\beta = p\,(\text{Accept } H_0 : \mu \leq 15 \mid \mu = 16)$. Since we have already worked out that $H_0$ is accepted if $\bar{X} \leq 15.64$ (at 10% level of significance), therefore $\beta = p\,(\bar{X} \leq 15.64 \mid \mu = 16)$ which can be depicted as follows:

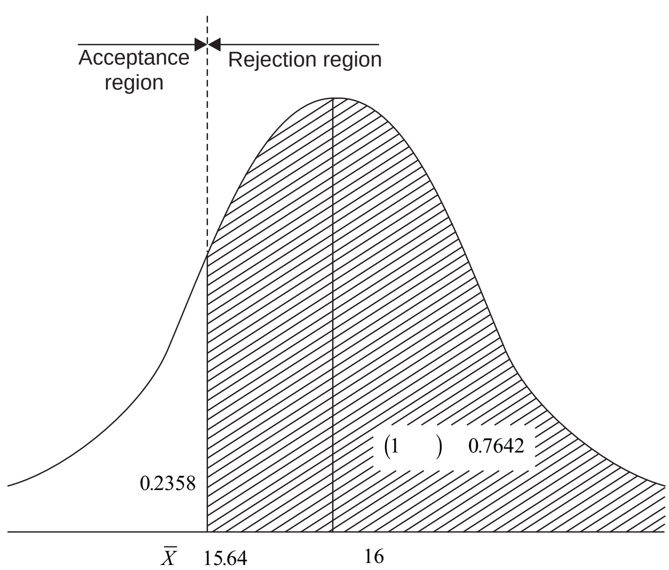


**Fig. 9.5**

* 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 $z = (\bar{X} - \mu)/(\sigma/\sqrt{n})$ $= (15.64 - 16)/(5/\sqrt{100}) = -0.72$ corresponding to which the area is 0.2642. Hence, $\beta$ = 0.5000 – 0.2642 = 0.2358 and the power of the test = (1 – $\beta$) = (1 – .2358) = 0.7642 for $\mu$ = 16.

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

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

or $\bar{X} \lessdot 15.42$, otherwise accept $H_a$

$\therefore\ \beta = p\left(\bar{X} \lessdot 15.42 \mid \mu = 16\right)$

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) $\chi^2$-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.

* $\chi^2$ - 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 $\chi^2$ -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.

$\chi^2$-*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:

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

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):

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

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:

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

and

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

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):

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

**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* | *One sample* | *Name of the test and the test statistic to be used* <br> *Two samples* <br> *Independent* | *Related* |
|---|---|---|---|---|
| *1* | *2* | *3* | *4* | *5* |
| Mean $(\mu)$ | Population(s) normal *or* Sample size large (i.e., $n > 30$) *or* population variance(s) known | $z$-test and the test statistic <br> $$z = \frac{X - \mu_{H_0}}{\sigma_p / \sqrt{n}}$$ <br> In case $\sigma_p$ is not known, we use $\sigma_s$ in its place calculating <br> $$\sigma_s = \sqrt{\frac{\Sigma\left(X_i - \overline{X}\right)^2}{n-1}}$$ | $z$-test for difference in means and the test statistic <br> $$z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\sigma_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$ <br> is used when two samples are drawn from the same population. In case $\sigma_p$ is not known, we use $\sigma_{s12}$ in its place calculating <br> $$\sigma_{s12} = \sqrt{\frac{n_1\left(\sigma_{s1}^2 + D_1^2\right) + n_2\left(\sigma_{s2}^2 + D_2^2\right)}{n_1 + n_2}}$$ <br> where $D_1 = \left(\overline{X}_1 - \overline{X}_{12}\right)$ <br> $D_2 = \left(\overline{X}_2 - \overline{X}_{12}\right)$ <br> $$\overline{X}_{12} = \frac{n_1 \overline{X}_1 + n_2 \overline{X}_2}{n_1 + n_2}$$ | |

*Contd.*

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

*Contd.*

| *1* | *2* | *3* | *4* | *5* |
|---|---|---|---|---|
| | | $\sigma_s = \sqrt{\dfrac{\Sigma\left(X_i - \overline{X}\right)^2}{n-1}}$ | *Alternatively*, $t$ can be worked out as under: $\left\{ \dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{\dfrac{(n_1 - 1)\sigma_{s1}^2 + (n_2 - 1)\sigma_{s2}^2}{n_1 + n_2 - 2}} \times \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \text{ with d.f.} = (n_1 + n_2 - 2) \right\}$ | 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 $\overline{p}$ and $\overline{q}$ 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 | $\chi^2$-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 $F = \dfrac{\sigma_{s1}^2}{\sigma_{s2}^2} = \dfrac{\Sigma\left(X_{1i} - \overline{X}_1\right)^2/n-1}{\Sigma\left(X_{2i} - \overline{X}_2\right)^2/n-1}$ where $\sigma_{s1}^2$ 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, $\mu_{H_0}$ = Hypothesised mean for population, $\sigma_p$ = standard deviation of population, $\sigma_s$ = 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 $\sigma_s$ in place of $\sigma_p$ 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 $\sigma_s$ in place of $\sigma_p$ 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.

$$H_a : \mu_{H_0} < 50$$ (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 $\sigma_s$ 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. | $X_i$ | $(X_i - \overline{X})$ | $(X_i - \overline{X})^2$ |
|---|---|---|---|
| 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 |
| $n = 10$ | $\Sigma X_i = 5720$ | | $\Sigma (X_i - \overline{X})^2 = 1456$ |

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

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

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 $H_0 : \mu_{H_0} = 578$ 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:

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

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

**Table 9.4**

| *S. No.* | $X_i$ | $(X_i - \overline{X})$ | $(X_i - \overline{X})^2$ |
|---|---|---|---|
| 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 |
| $n = 10$ | $\sum X_i = 6576$ | | $\sum(X_i - \overline{X})^2 = 23978$ |

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

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

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 $\mu_1$ is population mean of one population and $\mu_2$ 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:

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

   In case $\sigma_{p_1}$ and $\sigma_{p_2}$ are not known, we use $\sigma_{s_1}$ and $\sigma_{s_2}$ 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:

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

   In case $\sigma_p$ 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:

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

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

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

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:

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

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: