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 $\left(\mu_{H_0}\right) = 100$.

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:

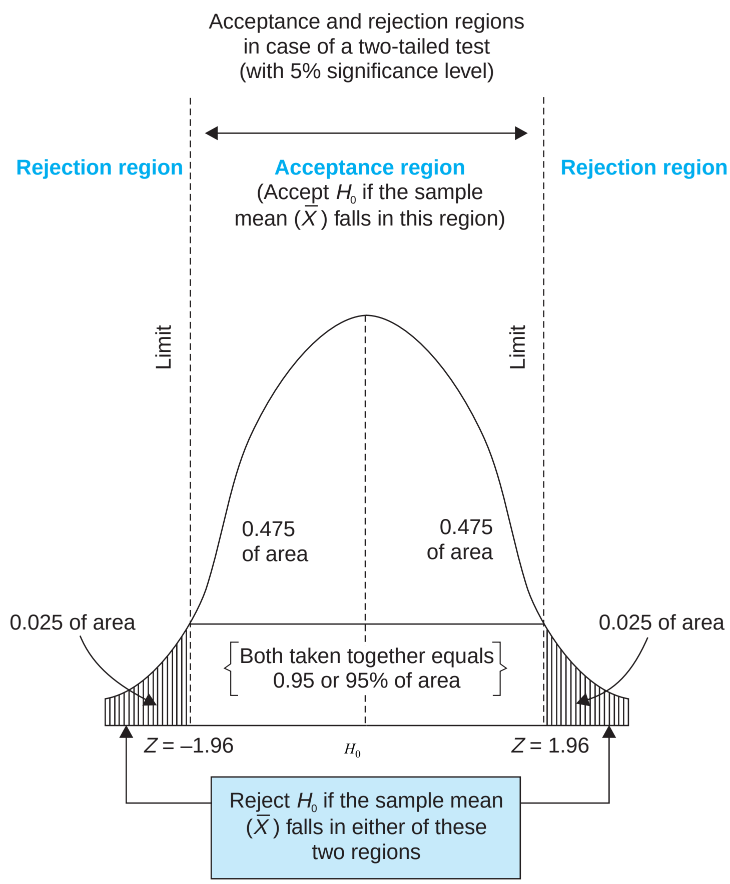

Fig. 9.1

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

*Also known as critical regions.

Mathematically we can state:

$$\textit{Acceptance Region} \quad A: |Z| \le 1.96$$

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

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:

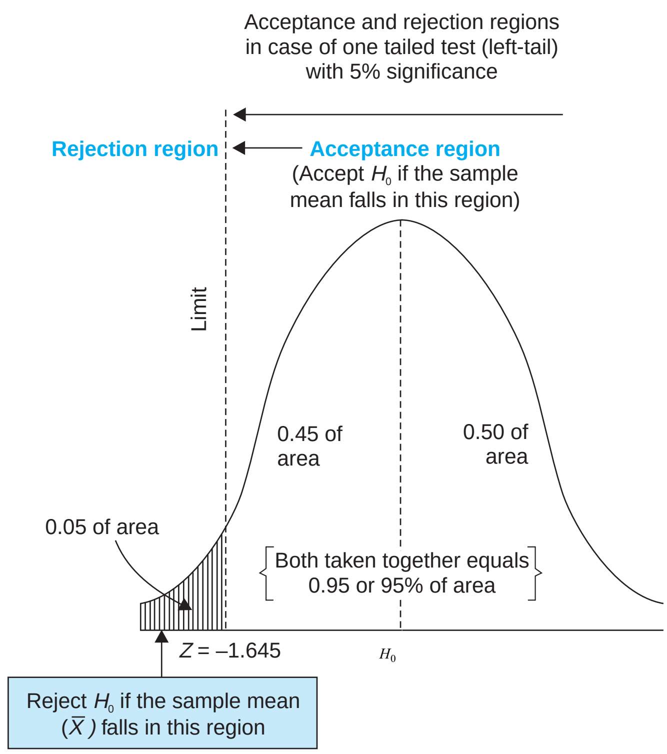

Fig. 9.2

Mathematically we can state:

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

$$\textit{Rejection Region} \quad R: Z \le -1.645$$

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:

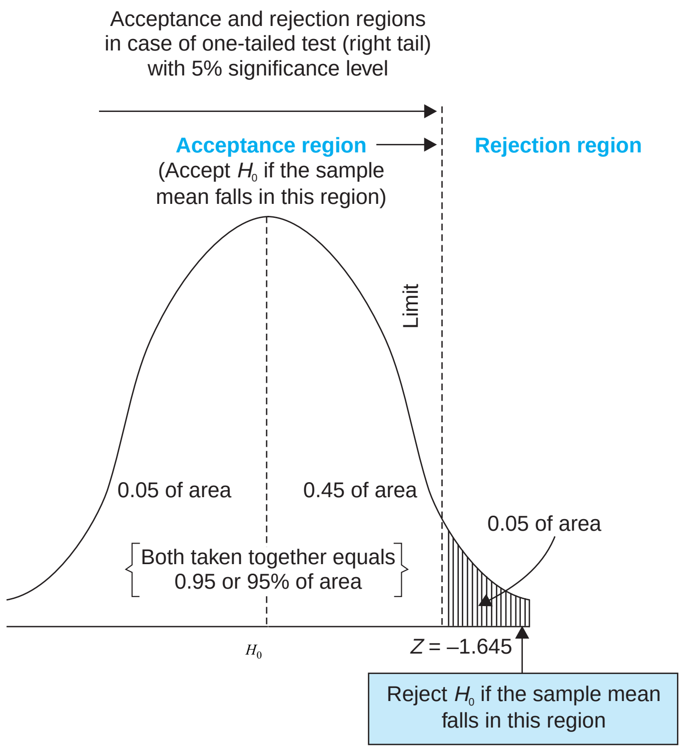

Fig. 9.3

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 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

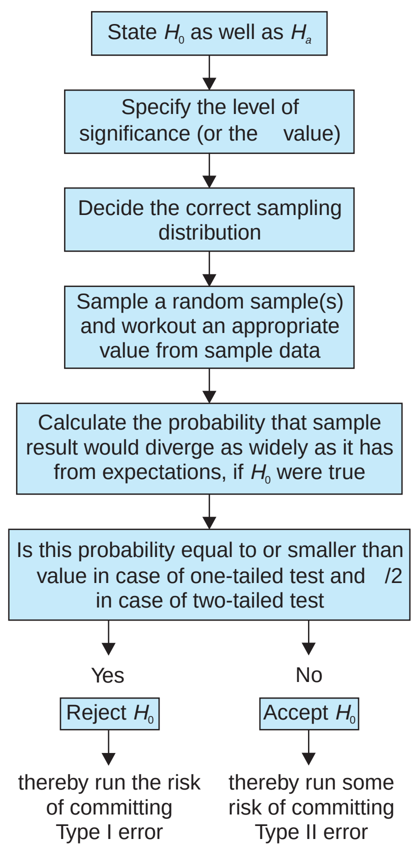

Fig. 9.4

[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:

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

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

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 μ for accepting H_0 as under:

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

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

or $\qquad \overline{X} \leq 15.64$

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 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 $H_0 : \mu \leq 15 \mid \mu = 16$). Since we have already worked out that H_0 is accepted if $\overline{X} \leq 15.64$ (at 10% level of significance), therefore $\beta = p\,(\overline{X} \leq 15.64 \mid \mu = 16)$ which can be depicted as follows:

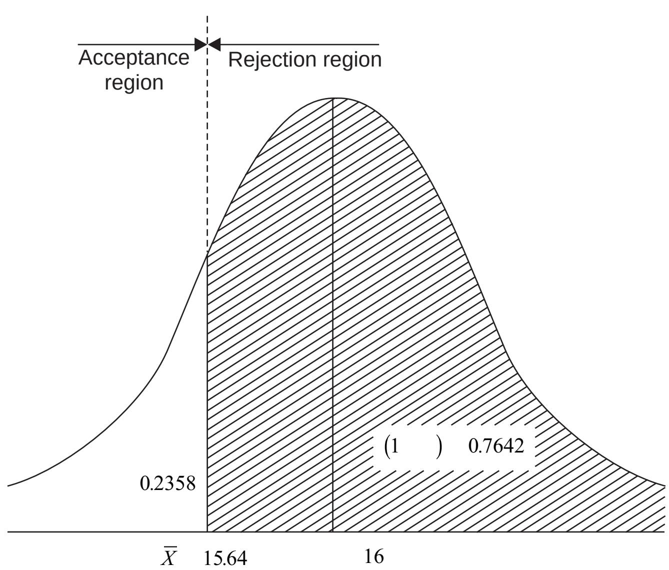

Fig. 9.5

* 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, 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 $\bar{X} \lesseqgtr 15 + (.84)\left(5/\sqrt{100}\right)$

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

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

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:

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

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*	*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 $$z = \frac{X - \mu_{H_0}}{\sigma_p / \sqrt{n}}$$ In case σ_p is not known, we use σ_s in its place calculating $$\sigma_s = \sqrt{\frac{\Sigma (X_i - \bar{X})^2}{n - 1}}$$	z-test for difference in means and the test statistic $$z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\sigma_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$ 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 $$\sigma_{s12} = \sqrt{\frac{n_1 \left(\sigma_{s1}^2 + D_1^2 \right) + n_2 \left(\sigma_{s2}^2 + D_2^2 \right)}{n_1 + n_2}}$$ where $D_1 = \left(\bar{X}_1 - \bar{X}_{12} \right)$ $D_2 = \left(\bar{X}_2 - \bar{X}_{12} \right)$ $$\bar{X}_{12} = \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2}{n_1 + n_2}$$	

Contd.

1	2	3	4	5
			OR $z \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_{p1}^2}{n_1} + \frac{\sigma_{p2}^2}{n_2}}}$ 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 $\sigma_{s1} = \sqrt{\Sigma\left(X_{1i} - \bar{X}_1\right)^2 / n_1 - 1}$ and $\sigma_{s2} = \sqrt{\Sigma\left(X_{2i} - \bar{X}_2\right)^2 / n_2 - 1}$	
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 $t = \frac{\bar{X} - \mu_{H_0}}{\sigma_s / \sqrt{n}}$ with d.f. $= (n - 1)$ where	*t*-test for difference in means and the test statistic $t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\Sigma\left(X_{1i} - \bar{X}_1\right)^2 + \Sigma\left(X_{2i} - \bar{X}_2\right)^2}{n_1 + n_2 - 2}}} \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ with d.f. $= (n_1 + n_2 - 2)$	Paired *t*-test or difference test and the test statistic $t = \frac{\bar{D} - 0}{\sqrt{\frac{\Sigma D_i^2 - \bar{D}^2 , n}{n - 1}} \Big/ \sqrt{n}}$ with d.f $= (n - 1)$ where n = number of

Contd.

1	2	3	4	5
		$\sigma_s = \sqrt{\dfrac{\Sigma\left(X_i - \overline{X}\right)^2}{n-1}}$	*Alternatively*, t can be worked out as under: $\left\{\begin{array}{c} \dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{\dfrac{(n_1-1)\sigma_{s1}^2 + (n_2-1)\sigma_{s2}^2}{n_1+n_2-2}} \times \sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} \\ \text{with d.f.} = (n_1+n_2-2) \end{array}\right\}$	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	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 $\bar{X} = 48.5$, $\sigma_p = 2.5$ and $n = 12$. Assuming the population to be normal, we can work out the test statistic z as under:

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

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:

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

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

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

Contd.

S. No.	X_i	$\left(X_i - \overline{X}\right)$	$\left(X_i - \overline{X}\right)^2$
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
$n = 10$	$\sum X_i = 5720$		$\sum\left(X_i - \overline{X}\right)^2 = 1456$

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

and

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

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.	$D_i = \left(X_i - \mu_{H_0}\right)$	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 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

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

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:

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

   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.)

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

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:

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

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 $\bar{D} = \frac{\sum D_i}{n} = \frac{-7}{9} = -0.778$

and Standard deviation of differences or

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

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

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

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:

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

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:

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

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 $\left(\hat{p}_2\right)$ 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:

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

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 $p_0 = \dfrac{n_1 \cdot \hat{p}_1 + n_2 \cdot \hat{p}_2}{n_1 + n_2}$ = 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:

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

$$= 2.093$$

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:

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

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

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}	$\left(X_{1i} - \bar{X}_1\right)$	$\left(X_{1i} - \bar{X}_1\right)^2$	X_{2i}	$\left(X_{2i} - \bar{X}_2\right)$	$\left(X_{2i} - \bar{X}_2\right)^2$
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
$\Sigma X_{1i} = 220$		$\Sigma\left(X_{1i} - \bar{X}_1\right)^2 = 120$	$\Sigma X_{2i} = 420$		$\Sigma\left(X_{2i} - \bar{X}_2\right)^2 = 314$
$n_1 = 10$			$n_2 = 12$		

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

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

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

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

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

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 $n_1 = 9$; $n_2 = 8$

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

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

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, $H_0 : \sigma_{p_1}^2 = \sigma_{p_2}^2$.

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

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

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

$v_1 = 8$ 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:

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

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".