3

TESTING OF HYPOTHESIS

3.1 INTRODUCTION

Many problems in management require that we decide whether to accept or reject a statement about some parameter. The statement is called a hypothesis, and the decision-making procedure about the hypothesis is called hypothesis testing. This is one of the most useful aspects of statistical inference, since many types of decision-making problems, tests or experiments in the management world can be formulated as hypothesis-testing problems. Furthermore, as we will see, there is a very close connection between hypothesis testing and confidence intervals. Statistical hypothesis testing and confidence interval estimation of parameters are the fundamental methods used at the data analysis stage of a comparative experiment.

Before giving the notion of sampling, we will first define population.

3.1.1 Population

A population in statistics means a set of objects or mainly the set of numbers which are measurements or observations pertaining to the objects.

The population is finite or infinite according to the number of elements of the set is finite or infinite.

3.1.2 Sampling

A part selected from the population is called a sample. The process of selection of a sample is called sampling.

3.1.3 Random Sampling

A Random sampling is one in which each number of population has an equal chance of being included in it. There are NC_n different samples of size n that can be picked up from a population of size N.

3.1.4 Parameters and Statistics

The statistical constants of the population such as mean (μ), standard deviation (σ) are called parameters.

Parameters are denoted by Greek letters.

The mean $\bar{x}$, standard deviation S of a sample are known as statistics. Statistics are denoted by Roman letters.

3.1.5 Symbols for Population and Samples

Characteristic	Population	Sample
	Parameter	**Statistic**
Symbols	Population size = N Population mean = μ Population standard deviation = σ Population proportion = p	Sample size = n Sample mean = $\bar{x}$ Sample standard deviation = s Sample proportion = $\tilde{p}$

3.1.6 Sampling Distribution

From a population a number of samples are drawn of equal size n. Find out the mean of each sample. The means of sample are not equal. The means with their respective frequencies are grouped.

The frequency distribution so formed is known as sampling distribution of the mean, similarly, sampling distribution of standard deviation we can have.

3.1.7 Standard Error (S.E)

S.E is the standard deviation of the sampling distribution. For assessing the difference between the expected value and observed value, standard error is used.

Reciprocal of standard error is known as precision.

S.No	Statistic	Standard Error
1.	$\bar{x}$	$\frac{\sigma}{\sqrt{n}}$
2.	S	$\sqrt{\frac{\sigma^2}{2n}}$
3.	S^2	$\sigma^2\sqrt{\frac{2}{n}}$
4.	$\bar{x}_1 - \bar{x}_2$	$\sqrt{\frac{\sigma_1{}^2}{n_1} + \frac{\sigma_2{}^2}{n_2}}$
5.	$S_1 - S_2$	$\sqrt{\frac{\sigma_1{}^2}{2n_1} + \frac{\sigma_2{}^2}{2n_2}}$
6.	$p_1 - p_2$	$\sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}$
7.	Observed sample portion p	$\sqrt{\frac{PQ}{n}}$

If t is any statistic, for large samples $Z = \frac{t - E(t)}{S\,.\,E(t)}$ is normally distributed with mean zero and variance unity.

$$\text{i.e., } Z = \frac{t - E(t)}{S \cdot E(t)} \sim N(0,1)$$

3.2 TESTS OF SIGNIFICANCE

An important aspects of the sampling theory is to study the tests of significance, which will enable us to decide, on the basis of the results of the samples, whether

- ✓ the deviation between the observed sample statistic and the hypothetical parameter value or
- ✓ the deviation between two samples statistics is significant or might be attributed due to chance or the fluctuations of the sampling.

If n is large, all the distributions like, Binomial, Poisson, Chi-square, t distribution, F distribution can be approximated by a normal curve.

3.2.1 Hypothesis

A hypothesis is some statement about a population parameter. The hypothesis is tested on the basis of the outcome of a random sample.

Null Hypothesis

In any testing of hypothesis problem we are faced with a pair of hypothesis such that one and only one of them is always true. One of this pair is called null hypothesis, and the other one is called alternatiye hypothesis.

The Null hypothesis is represented as H_0 and the alternative hypothesis is represented by H_1.

If the population mean is represented by μ.

$$H_0 : \mu \le 50; \; H_1 : \mu > 50$$

End of the testing if we conclude that H_0 is to be rejected, then H_1 should be accepted.

3.2.2 Type - I and Type - II Errors

In testing the hypothesis if we wrongly reject H_0, when in reality H_0 is true, the error is called a Type I error. Similarly, if we wrongly accept H_0 when H_0 is false, the error is called a Type II error.

We should not commit both the errors and should be reduced to the minimum they can be completely eliminated when the full population is examined. The probability of Type I error would be kept down to lower limits.

3.2.3 The Significance Level

In testing of hypothesis, Type I error is assumed to be more serious than Type II error and so the probability of Type I error needs to be explicitly controlled. This is done through the significance level of which the test is conducted. The significance level sets a limit to the probability of Type I errors and test procedures are designed so as to get the lowest probability of Type II error subject to the significance level. The probability of Type I error is denoted by α and probability of Type II error is denoted by β. Most of the test are conducted at $\alpha = 0.1$, $\alpha = 0.05$, $\alpha = 0.01$, by convection as well as by convenience. Generally, we use to test at 5% level.

3.2.4 Hypothesis Testing Procedure

Step 1 : State the null and the alternative hypothesis.

Step 2 : Choose the test statistic.

Step 3 : Specify a level of significance of α.

Step 4 : Define the critical region in terms of the test statistic.

Step 5 : Compare the observed value of the test statistic with the cut-off value to accept or reject the null hypothesis.

Level of Significance and Critical Region

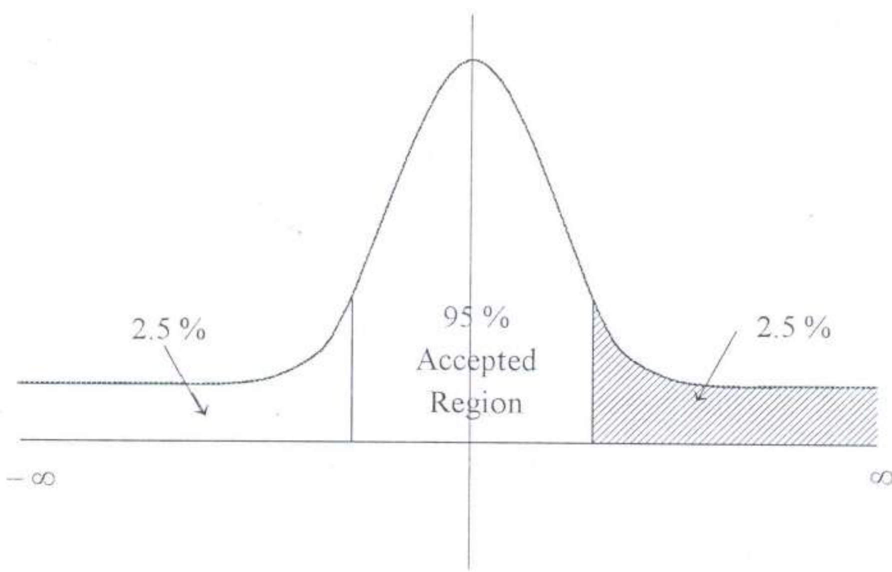


Fig.3.1

In testing a hypothesis, the maximum probability with which we are willing to risk a Type I error is called the level of significance of the test. Generally, we take either 5% or 1% level of significance, that is there are about 5 cases in 100 that we would reject the hypothesis when it should be accepted. That is we are about 95% confident that we have made that the right decision similarly for 1% level of significance.

From the fig. 3.1 the test statistic $Z = \frac{t - E(t)}{S.E \text{ of } t}$ of a sample statistic lies between −1.96 and 1.96, we are 95% confident that the hypothesis is true.

(i.e) $P[-1.96 \le Z \le 1.96] = 0.95$

If for a random sample the test statistic Z lies outside the range −1.96 to 1.96 i.e., if $|Z| > 1.96$, we say that event will happen with probability of only 0.05. If the given hypothesis were true. In this case we say that Z-score differed significantly from the value expected under the hypothesis and hence the hypothesis us to be rejected at 5% level of significance. The total shaded are 0.5 represents the probability of our H_0 being wrong in rejecting the hypothesis. 0.05 is the probability of making Type - I error. Thus if $|Z| > 1.96$ the hypothesis is rejected at a 5% level significance.

(i.,e.) $|Z| > 1.96$ constitutes critical region or region of rejection of the hypothesis or the region of significance. Thus the critical region is the area under the sampling distribution in which the test statistic value has to fall for the null hypothesis to be rejected. The set of Z scores inside the range −1.96 to 1.96 is called the region of the acceptance of the hypothesis.

Decision Rule

1. Reject the null hypothesis at 5% level of significance if the test statistic $|Z| > 1.96$.

   Accept the null hypothesis at 5% level of significance if $|Z| \le 1.96$.

2. Reject the null hypothesis at 1% level of significance if $|Z| > 2.58$.

   Accept the null hypothesis at 1% level of significance of $|Z| \le 2.58$.

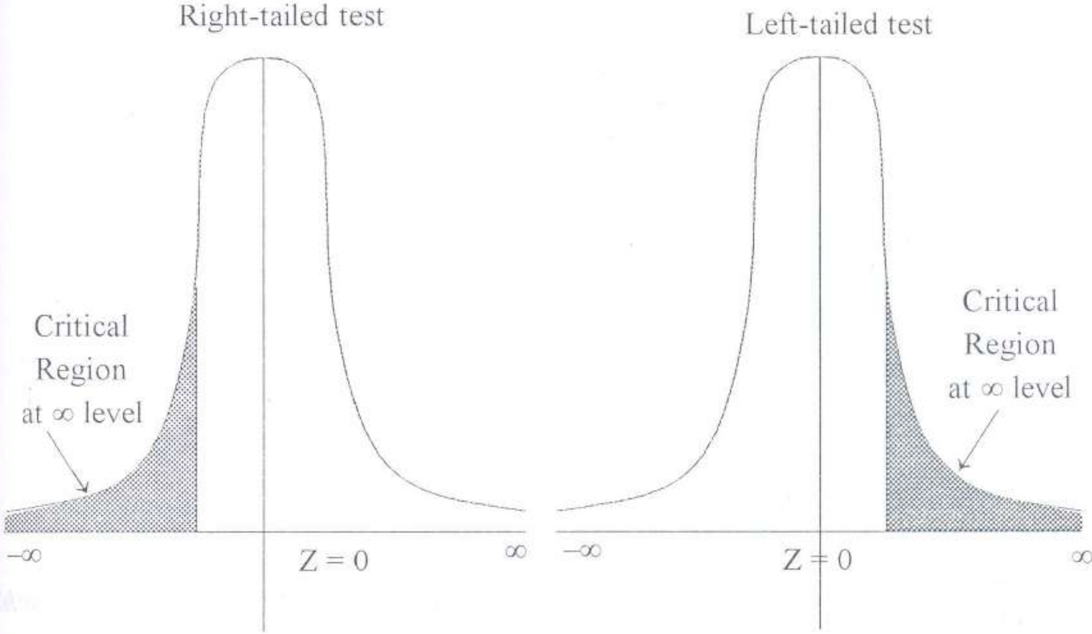
3.2.5 One – Tailed and Two – Tailed Test

In testing whether the population mean $\mu = \mu_0$, we have $H_0 : \mu = \mu_0$ against the alternative hypothesis H_1 given by

$H_1 : \mu > \mu_0$ (right-tailed) or $\mu < \mu_0$ (left-tailed).

In the right-tailed test $H_1 : \mu > \mu_0$, the critical region $Z > Z_\alpha$ lies entirely in the right tail of the sampling distribution of sample mean x with area equal to the level of significance α. Similarly, in the left-tailed $H_1 : \mu < \mu_0$, the critical region $[Z < -Z_\alpha]$ lies entirely in the left of the sampling distribution of x with area equal to the level of significance α.

If the alternative hypothesis H_1 in a test be two-tailed (both right and left tailed).



(i.e) $H_1 : \mu \neq \mu_0$ ($\mu > \mu_0$ or $\mu < \mu_0$) then the test is called two-tailed test and in such a case the critical region lies in both right and left tailed of the sampling distribution of the test statistic, with total area equal to the level of the significance. We apply two-tailed or one-tailed according as H_1 is two-tailed or one-tailed.

Critical Values of Z

Level of the significance (α)	**1%**	**5%**	**10%**
Critical values of two-tailed tests	$\lvert Z_\alpha \rvert = 2.58$	$\lvert Z_\alpha \rvert = 1.96$	$\lvert Z_\alpha \rvert = 1.645$
Critical values of right-tailed tests	$Z_\alpha = 2.33$	$Z_\alpha = 1.645$	$Z_\alpha = 1.28$
Critical values of left-tailed tests	$Z_\alpha = -2.33$	$Z_\alpha = -1.645$	$Z_\alpha = -1.28$

3.3 STUDENT'S 't' TEST FOR SINGLE MEAN

Suppose we want to test

- ✓ If a random sample x_i of size n has been drawn from a normal population with a specified mean μ_0.
- ✓ If the sample mean differs significantly from the hypothetical value μ_0 of the population mean.

$$t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}} \quad \text{(or)} \quad t = \frac{\bar{x} - \mu}{\frac{S.D}{\sqrt{n-1}}}$$

where

$\bar{x} \rightarrow$ Sample mean

$\mu \rightarrow$ Population mean

$S \rightarrow$ Standard deviation

$n \rightarrow$ Number of observation

$n - 1 \rightarrow$ is degree of freedom

Example : 1 ***A machinist is making engine parts with axle diameter of 0.700 inch. A random sample of 10 parts shows a mean diameter of 0.742 inch with S.D of 0.040 inch. Compute the statistic you would use to test whether the work is meeting the specification.***

Solution :

Null Hypothesis H_0 :

$\mu = 0.700$

Alternative Hypothesis H_1 :

$\mu \neq 0.700$

$n = 10$

$\bar{x} = 0.742$

$S = 0.040$

$$t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n-1}}}$$

$$= \frac{0.742 - 0.700}{\frac{0.040}{\sqrt{10-1}}}$$

$$= \frac{0.042}{\frac{0.04}{\sqrt{9}}}$$

$$= \frac{0.042 \times 3}{0.040}$$

$$= 3.15$$

Table value of t at 5% with degree of freedom 9.

$t_{0.05} = 2.26.$

Result :

Calculated value > Table value $\Rightarrow$ Reject H_0.

$3.15 > 2.26 \Rightarrow$ Reject H_0.

Example : 2 ***The mean weekly sales of soap bars in departmental stores was 146.3 bars per store. After an advertising campaign the mean weekly sales in 22 stores for a typical week increased to 153.7 and showed a S.D of 17.2 was the advertising campaign successful.***

Null Hypothesis H_0 :

$\mu = 146.3$

Alternative Hypothesis H_1 :

$\mu \neq 146.3$

$n = 22$

$\overline{x} = 153.7$

$S.D = 17$

$$t = \frac{\overline{x} - \mu}{\frac{S.D}{\sqrt{n-1}}}$$

$$= \frac{153.7 - 146.3}{\frac{17.2}{\sqrt{22-1}}}$$

$= 1.97$

$\therefore$ Calculated value of $t = 1.97$.

Table value of t at 5% level with 21 degree of freedom.

$t_{0.05} = 1.72$

Result :

Calculated value > Table value $\Rightarrow$ Reject H_0.

$1.97 > 1.72 \Rightarrow$ Reject H_0.

Example : 3 ***A sample of 26 bulbs gives a mean life of 990 hours with a S.D of 20 hours. The manufacturer claims that the mean life of bulbs is 1000 hours. Is the sample not upto the standard.***

Null Hypothesis H_0 :

$\mu = 1000$

Alternative Hypothesis H_1 :

$\mu \neq 1000$

$n = 226$

$\overline{x} = 990$

$S = 200$

$$t = \frac{\bar{x} - \mu}{\frac{S.D}{\sqrt{25}}}$$

$$= \frac{900 - 1000}{\frac{20}{\sqrt{25}}}$$

$$|t| = 2.5$$

$\therefore$ Calculated value of t = 2.5.

Tabulated 't' at 5% level with 25 degree of freedom.

$$t_{0.05} = 1.708$$

Result :

Calculated value > Table value $\Rightarrow$ Reject H_0.

$2.5 > 1.708 \Rightarrow$ Reject H_0.

Example : 4 ***A random sample of size 16 values from a normal population should have a mean of 53 and a sum of squares of deviation from the mean equals to 150. Can this sample be regarded as taken from the population having 56 as mean? Obtain 95% confidence limits of the mean of the population.***

Null Hypothesis H_0 :

$$\mu = 56$$

Alternative Hypothesis H_1 :

$$\mu \neq 56$$

$n = 16$

$\bar{x} = 53$

$\Sigma(x - x)^2 = 150$

$$\therefore\ S^2 = \frac{\Sigma(x-x)^2}{n-1}$$

$$= \frac{150}{15} = 10.$$

$$\boxed{S = \sqrt{10}}$$

$$t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}$$

$$= \frac{53 - 56}{\frac{\sqrt{10}}{\sqrt{16}}} = -3.79$$

$$|t| = 3.79$$

$\therefore$ Calculated value of = 3.79

The table value of t at 5% level of significance for 15 degree of freedom.

$t_{0.05} = 2.13.$

n is degree of freedom.

Result :

Calculated value > Table value $\Rightarrow$ Reject H_0.

$3.79 > 2.13 \Rightarrow$ Reject H_0.

(i.e) The sample cannot be regarded as taken from the population.

The 95% confidence limit of the mean of the population is given by

$$x \pm t_{0.05}\ \frac{S}{\sqrt{n}}$$

$$= 53 \pm 2.13 \times 0.79$$

$$= 53 \pm 1.6827$$

$$= 54.68 \text{ and } 51.31.$$

Hence 95% of confidence limit is

[54.68 and 51.31].

Example : 5 *A random sample of 10 boys had the following I.Q's 70, 120, 110, 101, 88, 83, 95, 98, 107, 100. Do these data support the assumption of a population mean I.Q of 100? Find a reasonable range in which most of the mean I.Q values of samples of 10 boys lie.*

Null Hypothesis H_0 :

$\mu = 100$

(i.e) The data support the assume of population mean I.Q of 100 in the population.

Alternative Hypothesis H_1 :

$\mu \neq 100$

x	d = x – A	d^2
70	–30	900
120	20	400
110	10	100
101	1	1
88	–12	144
83	–17	289
95	–5	25
98	–2	4
107	7	49
100	0	0
	$\Sigma d = -28$	$\Sigma d^2 = 1912$

$$\bar{x} = A + \frac{\Sigma d}{n}$$

$$= 100 + \left(\frac{-28}{10}\right)$$

$$= 100 - 2.8$$

$$\boxed{\bar{x} = 97.2}$$

$$S.D = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2}$$

$$= \sqrt{\frac{1912}{10} - \left(\frac{-28}{10}\right)^2}$$

$$= 13.54$$

$$t = \frac{\bar{x} - \mu}{\left(\frac{S}{\sqrt{n-1}}\right)}$$

$$= \frac{97.2 - 100}{\frac{13.5}{\sqrt{10-1}}}$$

$$= \frac{-2.8}{13.5} \times 3$$

$$= 0.6221$$

Table Value :

n – 1 Degrees of freedom

10 – 1 = 9

= 2.26

Result :

Calculated value > Table value ⇒ Reject H_0.

$0.622 > 2.26 \Rightarrow$ Accept H_0.

Conclusion :

The data supports the assumption of mean I.Q of 100 in the population.

Example : 6 ***The following table gives the length of 12 samples of Egyptian cotton taken from a large consignment. 48, 46, 49, 46, 52, 45, 43, 47, 47, 46, 47, 50. Test if mean length of the consignment be taken as 46.***

Null Hypothesis H_0 :

The mean length is 46

$\therefore \mu = 46$

Alternative Hypothesis H_1 :

$\mu \neq 46$.

x	d = x − A	d^2	A = 50
48	−2	4	
46	−4	16	
49	−1	1	
46	−4	16	
52	2	4	
45	−5	25	
43	−7	49	
47	−3	9	
47	−3	9	
46	−4	16	
47	−3	9	
50	0	0	
	$\Sigma d = -34$	$\Sigma d^2 = 158$	

$$\bar{x} = A + \frac{\Sigma d}{n}$$

$$= 100 + \frac{-34}{12}$$

$$= 47.166$$

$$S.D = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2}$$

$$= \sqrt{\frac{158}{12} - \left(\frac{-34}{12}\right)^2}$$

$$= \sqrt{13.16 - 8.02}$$

$$= 2.267$$

$$t = \frac{\bar{x} - \mu}{\left(\frac{S}{\sqrt{n-1}}\right)}$$

$$= \frac{47.16 - 46}{\frac{2.267}{\sqrt{11}}}$$

$$= 1.69$$

Table Value : $= n - 1 = 12 - 1 = 11$ at 5%

$$= 2.2$$

Result :

Calculated value > Table value $\Rightarrow$ Reject H_0.

$1.69 > 2.2 \Rightarrow$ Accept H_0.

Conclusion :

The mean length of consignment can be taken as 46.

3.3.1 STUDENT'S 't' TEST FOR DIFFERENCE OF MEANS

To test the significant difference between two means $\overline{x}_1$ and $\overline{x}_2$ of samples of size $\overline{n}_1$ and $\overline{n}_2$, use the statistic.

$$t = \frac{\overline{x}_1 - \overline{x}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where

$$S^2 = \frac{\Sigma(x_1 - \overline{x}_1)^2 + \Sigma(x_2 - \overline{x}_2)^2}{n_1 + n_1 - 2}$$

or

$$S^2 = \frac{1}{n_1 + n_1 - 2}[n_1 s_1^{\,2} + n_2 s_2^{\,2}]$$

S_1, S_2 sample standard deviation.

Degree of freedom d.f = $n_1 + n_2 - 2$.

Example : 1 ***Samples of two types of electric light bulbs were tested for length of life and following data were obtained.***

	Type I	*Type II*
Sample Number	$n_1 = 8$	$n_2 = 7$
Sample Means	$\overline{x}_1 = 124$ *hrs*	$\overline{x}_2 = 1036$ *hrs*
Sample S.D	$S_1 = 36$ *hours*	$S_2 = 40$*hrs.*

Is the difference in the means sufficient to warrant that type I superior to type II regarding length of life.

Null Hypothesis H_0 :

$\mu_1 = \mu_2$

There is no significant difference between two types.

$$S = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_1 - 2}}$$

$$= \sqrt{\frac{8 \times 36^2 + 7 \times 40^2}{8 + 7 - 2}}$$

$$= \sqrt{\frac{21568}{13}}$$

$$S = 40.73$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$= \frac{1234 - 1036}{40.73\sqrt{\frac{1}{8} + \frac{1}{7}}}$$

$$= \frac{198}{21.079}$$

$$t = 9.39$$

Degree of freedom $= n_1 + n_1 - 2$

$= 8 + 7 - 2$

$= 13$ at 5% level

$= 1.77$

Result :

Calculated value > Table value $\Rightarrow$ Reject H_0.

$9.39 > 1.77 \Rightarrow$ Accept H_0.

Example : 2 ***The average number of articles produced by two machines per day are 200 and 250 with standard deviations 20 and 25 respectively. On the basis of records of 25 days production. Can you regard both the machines equally efficient at 1% level of significance.***

Null Hypothesis H_0 :

$$\mu_1 = \mu_2$$

Alternative Hypothesis H_1 :

$$\mu_1 \neq \mu_2$$

$$n_1 = 25 \qquad n_2 = 25$$

$$\bar{x}_1 = 200 \qquad \bar{x}_2 = 250$$

$$S_1 = 20 \qquad S_2 = 25$$

$$S = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_1 - 2}}$$

$$= \sqrt{\frac{25 \times 20^2 + 25 \times 25^2}{25 + 25 - 2}}$$

$$= \sqrt{\frac{25625}{48}}$$

$$S = 23.10$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$= \frac{200-250}{23.10\sqrt{\frac{1}{25}+\frac{1}{25}}}$$

$$= \frac{200-250}{23.10\sqrt{\frac{1}{25}+\frac{1}{25}}}$$

$$= \frac{-50}{6.53}$$

$$|t| = 7.65$$

Tabulated value $= n_1 + n_1 - 2$

$= 25 + 25 - 2$

$= 48$ at 1% level

$= 2.58$

Result :

Calculated value > Table value $\Rightarrow$ Reject H_0.

$7.65 > 2.58 \Rightarrow$ Reject H_0.

Example : 3 ***The heights of size randomly chosen sailors are inches 63, 65, 68, 69, 71 and 72. Those of 10 randomly chosen soldiers are 61, 62, 65, 66, 69, 69, 70, 71, 72 and 73. Discuss the light that these data throw on the suggestion that sailors are on the average taller than soldiers.***

Null Hypothesis H_0 :

The sailors are not on the average taller than the soldiers. Both have the same average weight.

$\mu_1 = \mu_2$

Alternative Hypothesis H_1 :

$\mu_1 \neq \mu_2$

x_1	$d_1 = x_1 - 60$	d^2
63	3	9
65	5	25
68	8	64
69	9	81
71	11	121
72	12	144
	$\Sigma d_1 = 48$	$\Sigma d_1^2 = 444$

x_2	$d_2 = x_2 - 65$	d^2
61	–4	16
62	–3	9
65	0	0
66	1	1
69	4	16
69	4	16
70	5	25
71	6	36
72	7	49
73	8	64
	$\Sigma d_2 = 28$	$\Sigma d_2 = 232$

$$\bar{x}_1 = A + \frac{\Sigma d_1}{n}$$

$$= 60 + \frac{48}{6}$$

$$\boxed{\bar{x}_1 = 68}$$

$$\bar{x}_2 = A + \frac{\Sigma d_2}{n}$$

$$= 65 + \frac{28}{6}$$

$$\boxed{\bar{x}_2 = 69.6}$$

$$S_1 = \sqrt{\frac{\Sigma d_1^2}{n_1} - \left(\frac{\Sigma d_1}{n_1}\right)^2}$$

$$= \sqrt{\frac{444}{6} - \left(\frac{48}{6}\right)^2}$$

$$\boxed{S_1 = 3.16}$$

$$S_2 = \sqrt{\frac{\Sigma d_2^2}{n_2} - \left(\frac{\Sigma d_2}{n_2}\right)^2}$$

$$= \sqrt{\frac{232}{10} - \left(\frac{28}{10}\right)^2}$$

$$\boxed{S_2 = 3.91}$$

$$S = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_1 - 2}}$$

$$= \sqrt{\frac{6 \times 3.16^2 + 10 \times 3.91^2}{6 + 10 - 2}}$$

$$S = 3.89$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$= \frac{68 - 69.6}{3.89\sqrt{\frac{1}{0} + \frac{1}{0}}}$$

$$= 0.79$$

$$\boxed{|t| = 0.79}$$

$$\text{Table value} = n_1 + n_1 - 2$$

$$= 6 + 10 - 2$$

$$= 14 \text{ at } 5\% \text{ level}$$

$$= 1.76$$

Result :

Calculated value > Table value $\Rightarrow$ Accept H_0.

$0.79 > 1.76 \Rightarrow$ Accept H_0.

Conclusion :

The sailors are not on the average taller than the soldiers.

3.3.2 PAIRED t-TEST

Example : 4 ***Memory capacity of a student was tested before/after a course of meditation for a month state whether the course was effective or not from the following data?***

Before Training (x)	:	*10*	*15*	*9*	*3*	*7*	*12*	*16*	*17*	*4*
After Training (y)	:	*12*	*17*	*8*	*5*	*6*	*11*	*18*	*20*	*3*

Solution :

Null Hypothesis H_0 :

The course was effective

x	y	d = y − x	d^2
10	12	2	4
15	17	2	4
9	8	−1	1
3	5	2	4
7	6	−1	1
12	11	−1	1
16	18	2	4
17	20	3	9
4	3	−1	1
		$\Sigma d = 7$	$\Sigma d^2 = 29$

$$\bar{d} = \frac{\Sigma d}{n}$$

$$= \frac{7}{9}$$

$$= 0.77$$

$$S = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2}$$

$$= \sqrt{\frac{29}{9} - \left(\frac{7}{9}\right)^2}$$

$$= \sqrt{3.22 - 0.59}$$

$$= \sqrt{2.63}$$

$$S = 1.621$$

$$|t| = \frac{\bar{d}}{\frac{S}{\sqrt{n-1}}} \sim n-1 \text{ d.f}$$

$$= \frac{0.77}{\frac{1.621}{\sqrt{8}}}$$

$$|t| = 1.343$$

Table Value :

= n – 1 degree of freedom at 5%

= 9 – 1

= 8

= 2.31

Conclusion :

Calculated value > Table value

1.343 > 2.31

We accept the H_0

∴ The course is effective.

Example : 5 ***Poor students were given intensive coaching and test whether given before and after coaching if any improvement in the coaching class use parity test.***

Before Coaching	:	*50*	*42*	*51*	*26*	*35*	*42*	*60*	*41*	*70*	*55*	*62*	*38*
After Coaching	:	*62*	*40*	*61*	*35*	*30*	*52*	*68*	*51*	*84*	*63*	*72*	*50*

Solution :

Null Hypothesis H_0 :

The coaching classes is effective.

Alternate Hypothesis H_1 :

The coaching classes is not effective.

x	y	d = y – x	d^2
50	62	12	144
42	40	–2	4
51	61	10	100
26	35	9	81
35	30	–5	25
42	52	10	100
60	68	8	64
41	51	10	100
70	84	14	196
55	63	8	64
62	72	10	100
38	50	12	144
		$\Sigma d = 96$	$\Sigma d^2 = 1122$

$$\bar{d} = \frac{\Sigma d}{n} = \frac{97}{12}$$

$$= 8$$

$$S = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2}$$

$$= \sqrt{\frac{1122}{12} - (8)^2}$$

$$= 5.43$$

$$|t| = \frac{\bar{d}}{\frac{S}{\sqrt{n-1}}}$$

$$= \frac{8}{\frac{5.43}{\sqrt{12-1}}}$$

$$|t| = 4.87$$

Table Value :

= n – 1 degree of freedom at 5%

= 12 – 1 = 11

= 2.20

Conclusion :

Calculated value > Table value

1.343 > 2.31

We reject the H_0

∴ The coaching is not effective.

3.4 SMALL SAMPLES - F TEST

G.W. Snedecor has discovered a continuous probability distributing called Snedecor's F distributing. Here F is named after R.A Fishes who has contributed a lot to the development of Mathematical statistics.

$$F = \frac{S_1^2}{S_2^2}$$

Where

$$S_1^2 = \frac{n_1}{n_1 - 1} S_1^2$$

(or)

$$S_1^2 = \frac{\Sigma(x - \overline{x})^2}{n_1}$$

$$S_2^2 = \frac{n_2}{n_2 - 1} S_2^2$$

(or)

$$S_2^2 = \frac{\Sigma(y - \overline{y})^2}{n_2}$$

Example : 1 *From the following data test if the difference between the variances is significant at 5% level of significance.*

Sum of squaes of deviations from the mean	*84.4*	*102.6*
Size	*8*	*10*
Sample	*A*	*B*

Solution :

$H_0 :$ $\sigma_1^2 = \sigma_2^2$ (The samples are drawn from the populations with equal variance)

$H_1 :$ $\sigma_1^2 \neq \sigma_2^2$ (The samples are drawn from the populations with unequal variances)

Variance of the sample A is

$$S_1^2 = \frac{\Sigma(x-\bar{x})^2}{n_1}$$

$$S_1^2 = \frac{84.4}{8}$$

Variance of the sample B is

$$S_2^2 = \frac{\Sigma(y-\bar{y})^2}{n_2}$$

$$S_2^2 = \frac{102.6}{10}$$

The estimated variances of the populations from which the samples A and B are drawn are given by

$$S_1^2 = \frac{n_1 S_1^2}{n_1 - 1}$$

$$= \frac{8}{7} \times \frac{84.4}{8}$$

$$S_1^2 = 12.06$$

$$S_2^2 = \frac{n_2 S_2^2}{n_2 - 1}$$

$$= \frac{10}{9} \times \frac{102.6}{10}$$

$$S_2^2 = 11.4$$

Here $S_1^2 > S_2^2$

To carry out the test we use the F-statistic given by

$$F = \frac{S_1^2}{S_2^2}$$

$$= \frac{12.06}{11.4}$$

$$F = 1.058$$

$$d.f = (n_1 - 1, n_2 - 1) = (7, 9)$$

Table value of F(7, 9) at 5% level = 3.29.

Conclusion :

H_0 is accepted since the calculated value of F < the table value of F.

Hence the population variances are equal.

Example : 2 ***In a sample of 8 observations the sum of the squared deviations of items from the mean was 94.5. In another sample of 10 observations, the value was found to be 101.7. Test whether the difference in the variances is significant at 5% level.***

Solution :

$$n_1 = 8,\ \Sigma(x-\bar{x})^2 = 94.5$$

$$n_2 = 10,\ \Sigma(y-\bar{y})^2 = 101.7$$

$\therefore$ The sample variances are

$$S_1^2 = \frac{\Sigma(x-\bar{x})^2}{n_1} = \frac{94.5}{8}$$

$$S_2^2 = \frac{\Sigma(y-\bar{y})^2}{n_2} = \frac{101.7}{10}$$

$H_0 :$ $\sigma_1^2 = \sigma_2^2$ (The samples are taken from populations whose variance are equal)

$H_1 :$ $\sigma_1^2 \neq \sigma_2^2$ (The samples are taken from populations whose variance are not equal)

$$S_1^2 = \frac{n_1 S_1^2}{n_1 - 1}$$

$$= \frac{8}{7} \times \frac{94.5}{8}$$

$$S_1^2 = 13.5$$

$$S_2^2 = \frac{n_2 S_2^2}{n_2 - 1}$$

$$= \frac{10}{9} \times \frac{101.7}{10}$$

$$S_2^2 = 11.3$$

Here $S_1^2 > S_2^2$

The test statistic is given by

$$F = \frac{S_1^2}{S_2^2}$$

$$= \frac{13.5}{11.3}$$

$$F = 1.174$$

$$d.f = (n_1 - 1, n_2 - 1) = (7, 9)$$

Table value of F(7, 9) at 5% level = 3.29.

Conclusion :

H_0 is accepted at 5% level of significance since the calculated value of F < the table value of F.

∴ The samples belong to populations with equal variance.