16

Small Sampling Theory

INTRODUCTION

The techniques examined in earlier chapters (13, 14, 15) under the general headings of sampling distributions, estimation of parameters and tests of hypothesis were based on a knowledge of the underlying sampling distribution of the sample statistic for large samples. We have discussed earlier, that if the original population is normally distributed, all sampling distributions of the mean shall be normally distributed regardless of the sample size (central limit theorem). If the original population is normally distributed and the standard deviation of the population is unknown (and therefore, has to be estimated from a sample), the sampling distribution of the mean derived from large samples will also be normally distributed, but if the sample size is small (say 30, or less) then the sample statistic will follow a t-distribution. Problems of estimation and tests of hypothesis for large samples were developed in previous chapters and this chapter extends these concepts for small samples, when the underlying sampling distribution of the mean follows a Student's t-distribution.

The Student's t-distribution obtained by W.S. Gosset was published under the pen name of "Student" in the year 1908. It is reported that Gosset was a statistician for a brewery, and that the management did not want him to publish his scholarly theoretical work under his real name and bring shame to his employer. Consequently, he selected the pen name of Student.

As a matter of fact, procedures of statistical inference for small samples are the same as those presented in preceding two chapters. 'The study of statistical inference with the small samples is called small sampling theory or exact sampling theory'. In this chapter, we shall discuss in detail the "t" and "F" distributions. These two distributions are defined in terms of *number of degrees of freedom*. It is appropriate at this stage to clarify this concept.

Degrees of freedom. The number of degrees of freedom, usually denoted by the Greek symbol v (read as nu) can be interpreted as the number of useful items of information generated by a sample of given size with respect to the estimation of a given population parameter. Thus, a sample of size 1 generates one piece of useful information if one is estimating the population mean, but none, if one is estimating the population variance. In order to know about the variance, one need at least a sample of size $n \geq 2$. The number of degrees of freedom, in general, is the total number of observations minus the number of independent constraints imposed on the observations.

Suppose the expression $\Sigma X = X_1 + X_2 + X_3$ has four terms. We can arbitrarily assign values to any three of these four values (for example, $15 = X_1 + 2 + 8$) but the value of the fourth is automatically determined (for example, $X_1 = 5$).

In this example, there are 3 degrees of freedom. If n is the number of observations and k is the number of independent constants (the number of constants that have to be estimated from the original data) then $n - k$ is the number of degrees of freedom.

If we consider sample of size n drawn from a normal (or approximately normal) population with mean μ and if for each sample we compute t, using the sample mean $\bar{x}$ and sample standard deviation s, the distribution for t can be obtained. The probability density function of the t-distribution is given by

$$f(t) = \frac{Y_0}{\left(1 + \frac{t^2}{v}\right)^{(v+1)/2}} \qquad -\infty < t < \infty$$

Y_0 is a constant depending on n such that the total area under the curve is one.

$v = n - 1$ is called the number of degrees of freedom.

Properties of *t*-Distribution

(1) The t-distribution ranges from $-\infty$ to ∞ just as does a normal distribution.

(2) The t-distribution like the standard normal distribution is bell-shaped and symmetrical around mean *zero*.

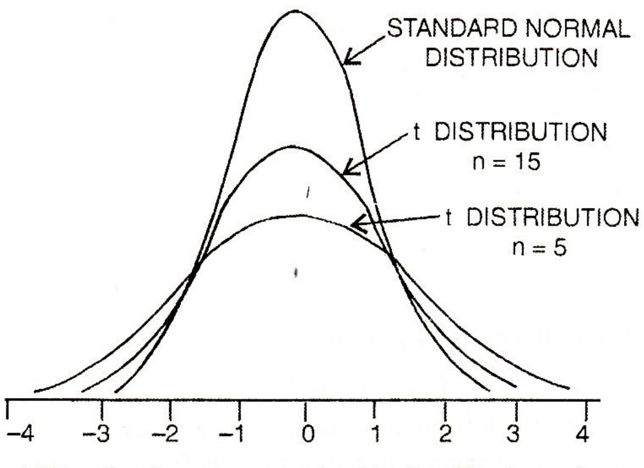
(3) The shapes of the t-distribution changes as the number of degrees of freedom changes. Therefore, for different degrees of freedom, the t-distribution has a family of t-distributions. Hence, the degrees of freedom v is a parameter of the t-distribution.

(4) The variance of the t-distribution is always greater than one and is defined only when $v \geq 3$ and is given as

$$\text{Var}(t) = \left(\frac{v}{v-2}\right)$$

(5) The t-distribution is more of platykurtic (less peaked at the centre and higher in tails) than the normal distribution.

(6) The t-distribution has a greater dispersion than the standard normal distribution. As n gets larger, the t-distribution approaches the normal form. When n is as large as 30, the difference is very small. Relation between the t-distribution and standard normal distribution is shown in the diagram.



Standard Normal Distribution compared with distribution when $n = 5$ and $n = 15$

The t-distribution has different shapes depending on the size of the sample. When the sample is quite small, for example, if n equals five, the height of the t-distribution is shorter than the normal distribution and the tails are wider. As n nears 30, however, the t-distribution approaches the normal distribution in shape.

The *t*-table. The t-table given at the end of the book is the probability integral of t-distribution. It gives over a range of values of v at different levels of significance. By selecting a particular degrees of freedom and level of significance, we determine the tabular value of t. We establish a null hypothesis, and if our computed t is greater than the tabular t, we reject the null hypothesis ; if our computed t is smaller than the tabular t, we accept the null hypothesis.

Applications of t-distribution. The following are some important applications of the t-distribution :

(1) Test of Hypothesis about the population mean.

(2) Test of Hypothesis about the difference between two means.

(3) Test of hypothesis about the difference between two means with dependent samples.

(4) Test of hypothesis about coefficient of correlation.

(1) **Test of Hypothesis about the Population Mean** (σ unknown and sample size is small). When the population distribution is normal and standard deviation σ is unknown then the "t" statistic is defined as :

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

follows the Student's t-distribution with $(n - 1)$ $d.f.$

where $\bar{x}$ = sample mean

μ = hypothesised population mean

n = sample size

and s is the standard deviation of the sample calculated by the formula :

$$s = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}}$$

The null hypothesis to be tested is whether there is a significant difference between $\bar{x}$ and μ.

If the calculated value of t exceeds the table value of t at a specified level of significance, the null hypothesis is rejected and the difference between $\bar{x}$ and μ is regarded significant. If the calculated value of t is less than the table value, the difference between $\bar{x}$ and μ is not considered to be significant. It may be noted that this test is based on $n - 1$ degrees of freedom.

Confidence Interval for the Population Mean. When sampling is from a normally distributed population with unknown σ, the 100 $(1 - \alpha)$ per cent confidence interval for the population is given by

$$\bar{x} \pm t_{\alpha/2, v} \; s / \sqrt{n}$$

Thus,

$$Pr. [-t_{\alpha/2, v} < t < t_{\alpha/2, v}] = 1 - \alpha$$

$$Pr.[-t_{\alpha/2, v} < \frac{\bar{x} - \mu}{s / \sqrt{n}} < t_{\alpha/2, v}] = 1 - \alpha$$

Hence, 100 $(1-\alpha)$% confidence interval is given by

$$\bar{x} - t_{\alpha/2, v} \; s / \sqrt{n} < \mu < \bar{x} + t_{\alpha/2, v} \; s / \sqrt{n}$$

Illustration 1. Ten oil tins are taken at random from an automatic filling machine. The mean weight of the tins is 15.8 kg and standard deviation is 0.50 kg. Does the sample mean differ significantly from the intended weight of 16 kg ?

(MBA, DU, 1999)

Solution. Let the null hypothesis be that the sample mean weight is not different from the intended weight.

Given that $n = 10$, $\bar{x} = 15.8$, $s = 0.50$, $\mu = 16$

Using the t-test, we have

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}} = \frac{15.8 - 16}{0.50 / \sqrt{10}} = -\frac{0.2}{0.158} = -1.266$$

The table value of t for 9 d.f. at 5% level of significance is 2.26. The computed value of t is smaller than the table value of t. Therefore, the difference is insignificant and the null hypothesis is accepted. Hence the difference between sample mean weight and the intended weight is insignificant.

Illustration 2. Prices of shares (in Rs.) of a company on the different days in a month were found to be :
66, 65, 69, 70, 69, 71, 70, 63, 64 and 68
Test whether the mean price of the shares in the month is 65. (*MBA, Delhi Univ., 2001*)

Solution. Null hypothesis $H_0 : \mu = 65$, Assuming the population to be normally distributed and the population standard deviation is unknown, the appropriate test statistic to be used is

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

$\bar{x}$ and s can be computed from the sample values from the following table :

x	$(x - 65)$ d	d^2
66	1	1
65	0	0
69	4	16
70	5	25
69	4	16
71	6	36
70	5	25
63	–2	4
64	–1	1
68	3	9
	$\Sigma d = 25$	$\Sigma d^2 = 133$

$$\bar{x} = A + \frac{\Sigma d}{N} = 65 + \frac{25}{10} = 67.5$$

$$s^* = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} = \sqrt{\frac{\Sigma x^2}{n-1} - \frac{(\Sigma x)^2}{n(n-1)}} = \sqrt{\frac{\Sigma d^2}{n-1} - \frac{(\Sigma d)^2}{n(n-1)}} = \sqrt{\frac{133}{9} - \frac{(25)^2}{10 \times 9}}$$

$$= \sqrt{14.78 - 6.94} = \sqrt{7.84} = 2.8$$

Therefore,

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{67.5 - 65}{2.8/\sqrt{10}} = \frac{2.5}{0.89} = 2.81.$$

The table value t for 9 degrees of freedom at 5% level of significance is 2.26. Since the computed value of $t = 2.81$ is greater than the table value, we reject the null hypothesis. Hence, the mean price of the shares in the month is not 65.

(2) Test of Hypothesis about the Difference between Two Means

In testing a hypothesis concerning the difference between the means of two normally distributed populations when the population variances are unknown, the t-test can be used in two types of cases : (a) the case in which variances are equal, *i.e.*, $\sigma_1^2 = \sigma_2^2$, (b) the case in which variances are not equal, *i.e.*, $\sigma_1^2 \neq \sigma_2^2$.

(a) **Case of equal variances.** Let the null hypothesis be that there is no significant difference between the means of the two populations, *i.e.*, $H_0 : \mu_1 = \mu_2$. When the population variances (though unknown) are equal then the appropriate test statistic to be used is

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

$$^*\frac{\Sigma(x-\bar{x})^2}{n-1} = \frac{\Sigma(x^2 + \bar{x}^2 - 2x\bar{x})}{n-1} = \frac{\Sigma x^2 - n\bar{x}^2}{n-1} = \frac{\Sigma x^2}{n-1} - \frac{(\Sigma x)^2}{n(n-1)}$$

will follow t-distribution with $(n_1 + n_2 - 2)$ d.f., where $\bar{x}_1$ and $\bar{x}_2$ are sample means of sample 1 of size n_1 and sample 2 of size n_2 respectively; μ_1 and μ_2 are the population means, and s is "pooled" estimate of the common population standard deviation obtained by pooling the data from both the samples as given below :

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

where $$s_1^2 = \frac{\Sigma(x_1 - \bar{x}_1)^2}{n_1 - 1}; \text{ and } s_2^2 = \frac{\Sigma(x_2 - \bar{x}_2)^2}{n_2 - 1}$$

Therefore, alternatively s can be computed from

$$s = \sqrt{\frac{\Sigma(x_1 - \bar{x}_1)^2 + \Sigma(x_2 - \bar{x}_2)^2}{n_1 + n_2 - 2}}.$$

If the computed value of t is less than the table value of t at a specified level of significance, the null hypothesis is accepted and the difference between the two means is regarded as insignificant. If the computed value of t is more than the table value of t, the null hypothesis is rejected and the difference between the sample means is regarded as significant.

(*b*) **Case of unequal variances.** When the population variances are not equal, *i.e.*, $\sigma_1^2 \neq \sigma_2^2$, we use the unbiased estimators s_1^2 and s_2^2 to replace σ_1^2 and σ_2^2. In this case, the difficulty arises because the sampling distribution has large variability than the population variability. The statistic :

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

may not strictly follow t-distribution but may be approximated by t-distribution with a modified value for the degrees of freedom given by

$$d.f. = \frac{\left[s_1^2/n_1 + s_2^2/n_2 \right]^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

Illustration 3. Two different types of drugs A and B were tried on certain patients for increasing weight, 5 persons were given drug A and 7 persons were given drug 'B'. The increase in weight (in pounds) is given below :

Drug A :	8	12	13	9	3		
Drug B :	10	8	12	15	6	8	11

Do the two drugs differ significantly with regard to their effect in increasing weight ?

Solution. Null hypothesis $H_0 : \mu_1 = \mu_2$, *i.e.*, there is no significant difference in the efficacy of the two drugs. Applying t-test:

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

CALCULATION FOR $\bar{x}_1$, $\bar{x}_2$ AND s

x_1	$(x_1 - \bar{x}_1)$	$(x_1 - \bar{x}_1)^2$	x_2	$(x_2 - \bar{x}_2)$	$(x_2 - \bar{x}_2)^2$
8	–1	1	10	0	0
12	+3	9	8	–2	4
13	+4	16	12	+2	4
9	0	0	15	+5	25
3	–6	36	6	–4	16
			8	–2	4
			11	+1	1
$\Sigma x_1 = 45$	$\Sigma(x_1 - \bar{x}_1) = 0$	$\Sigma(x_1 - \bar{x}_1)^2 = 62$	$\Sigma x_2 = 70$	$\Sigma(x_2 - \bar{x}_2) = 0$	$\Sigma(x_2 - \bar{x}_2)^2 = 54$

$$\bar{x}_1 = \frac{\Sigma x_1}{n_1} = \frac{45}{5} = 9, \quad \bar{x}_2 = \frac{\Sigma x_2}{n_2} = \frac{70}{7} = 10$$

$$s = \sqrt{\frac{\Sigma(x_1 - \bar{x}_1)^2 + \Sigma(x_2 - \bar{x}_2)^2}{n_1 + n_2 - 2}} = \sqrt{\frac{62+54}{5+7-2}} = \sqrt{\frac{116}{10}} = 3.406$$

Therefore,
$$t = \frac{\bar{x}_1 - \bar{x}_2}{s}\sqrt{\frac{n_1 n_2}{n_1 + n_2}} = \frac{9-10}{3.406}\sqrt{\frac{5\times 7}{5+7}} = -\frac{1}{3.406} \times 1.708 = -0.5$$

$$v = n_1 + n_2 - 2 = 5 + 7 - 2 = 10. \qquad \text{For } v = 10,\ t_{0.05} = 2.23.$$

The calculated value of t is less than the table value. Our null hypothesis is accepted. Hence, we conclude that there is no significant difference in the efficacy of the two drugs in the matter of increasing weight.

Illustration 4. Two salesmen A and B are working in a certain district. From a sample survey conducted by the Head Office, the following results were obtained. State whether there is any significant difference in the average sales between the two salesmen.

	A	B
No. of Sales	10	18
Average sales (in lakh Rs.)	190	205
Standard deviation (in lakh Rs.)	20	25

Solution. Null hypothesis $H_0 : \mu_1 = \mu_2$, *i.e.*, there is no significant difference in the average sales between the two salesmen. Applying t-test

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

where
$$s = \sqrt{\frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{n_1 + n_2 - 2}}$$

$$= \sqrt{\frac{9\,(20)^2 + 7\,(25)^2}{10+18-2}} = \sqrt{\frac{3600+10625}{26}} = 23.39$$

$$t = \frac{190-205}{23.39}\sqrt{\frac{10\times 18}{10+18}} = \frac{15}{23.39} \times 2.54 = 1.63$$

The table value of t at 5% level of significance for 26 *d.f.* is 2.056. The calculated value of t is less than the table value. The hypothesis holds true. Hence, we conclude that there is no significant difference in the average sales between the two salesmen.

Confidence Interval for the Difference between the Two Means

Two samples of sizes n_1 and n_2 are randomly and independently drawn from two normally distributed populations with unknowns but equal variances. The 100 (1–α) per cent confidence interval for $\mu_1 - \mu_2$ is given by

$$\left(\bar{x}_1 - \bar{x}_2\right) \pm t_{\alpha/2,\, v}\, s/\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

(3) Test of hypothesis about the difference between two means with dependent samples*. In the previous section, we assumed that the two random samples drawn from the two populations were independent. In many practical situations, this may not be true. The samples are *dependent* if they are paired so that each observation in one sample is associated with some particular observation in the second sample. Because of this property, the test we are going to use will be called *paired t-test.* In this test, it is necessary that the observations in the two samples be collected in the form called *matched*

* This is also known as the paired t-test.

pairs. If two samples are dependent, they must have the same number of units. Instead of obtaining two random samples, we can get one random sample of pairs, and the two measurements associated with a pair will be related to each other. This kind of a problem arises in cases such as before and after type experiment or when observations are matched by rise or some other criterion. Suppose that two training methods are to be compared on the basis of average scores by management trainees divided into two equal size classes, one taught by each method. When experimental results are available, we test the null hypothesis that the means associated with the two methods are equal, *i.e.*, $H_0 : \mu_1 = \mu_2$. The appropriate test statistic to be used here is

$$t = \frac{\bar{d}\sqrt{n}}{s}$$

follows *t*-distribution with $(n - 1)$ *d.f.* where $\bar{d}$ = mean of the differences is given by $\bar{d} = \Sigma d/n$, s is the standard deviation of the differences and is given by

$$s = \sqrt{\frac{\Sigma(d - \bar{d})^2}{n-1}} = \sqrt{\frac{\Sigma d^2 - n(\bar{d})^2}{n-1}} = \sqrt{\frac{\Sigma d^2}{n-1} - \frac{(\Sigma d)^2}{n(n-1)}}$$

and n is the number of paired observations in the samples. If the computed value of t (at a specified level of significance for a given number of degrees of freedom) is less than the table value of t, our null hypothesis is accepted, otherwise rejected.

Illustration 5. Ten persons were appointed in an officer cadre in an office. Their performance was noted by giving a test and the marks were recorded out of 100. They were given 4 months training and a test was held and marks were recorded out of 100.

Employee	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
Before training	80	76	92	60	70	56	74	56	70	56
After training	84	70	96	80	70	52	84	72	72	50

By applying the *t*-test, can it be concluded that the employees have benefited by the training?

Solution. Let us take the null hypothesis that the employees have not benefited by the training. Applying *t*-test :

Employees	*Before* 1*st*	*After* 2*nd*	(1*st* – 2*nd*) *d*	d^2
A	80	84	–4	16
B	76	70	+6	36
C	92	96	–4	16
D	60	80	–20	400
E	70	70	0	0
F	56	52	+4	16
G	74	84	–10	100
H	56	72	–16	256
I	70	72	–2	4
J	56	50	+6	36
$n = 10$			$\Sigma d = -40$	$\Sigma d^2 = 880$

$$t = \frac{\bar{d}\sqrt{n}}{s}, \text{ where } \bar{d} = \frac{\Sigma d}{n} = \frac{-40}{10} = -4$$

$$s = \sqrt{\frac{\Sigma d^2 - n\left(\bar{d}\right)^2}{n-1}}$$

$$= \sqrt{\frac{880 - 10(-4)^2}{9}} = \sqrt{\frac{880 - 160}{9}} = 8.944$$

$$t = \frac{-4\sqrt{10}}{8.944} = -\frac{4 \times 3.162}{8.944} = -1.414.$$

$v = 10 - 1 = 9$, For $v = 9$, $t_{0.05} = 2.62$.

The calculated value is less than the table value. The null hypothesis holds true. Hence, it can be concluded that the employees have not benefited by the training.

Illustration 6. Ten workers were given a training programme with a view to shorten their assembly time for a certain mechanism. The results of the time and motion studies before and after the training programme are given below :

Worker	1	2	3	4	5	6	7	8	9	10
First study (in mnts)	15	18	20	17	16	14	21	19	13	22
Second study (in mnts)	14	16	21	10	15	18	19	16	14	20

On the basis of this data, can it be concluded that the training programme has shortened the average assembly time?

Solution. Let us take the null hypothesis that the training programme has not helped in reducing the average assembly time. Applying t-test :

$$t = \frac{\bar{d}\sqrt{n}}{s}.$$

CALCULATIONS FOR $\bar{d}$ AND s

Worker	*1st study*	*2nd study*	*(1st–2nd)* d	d^2
1	15	14	+1	1
2	18	16	+2	4
3	20	21	–1	1
4	17	10	+7	49
5	16	15	+1	1
6	14	18	–4	16
7	21	19	+2	4
8	19	16	+3	9
9	13	14	–1	1
10	22	20	+2	4
			$\Sigma d = 12$	$\Sigma d^2 = 90$

$$\bar{d} = \frac{\Sigma d}{n} = \frac{12}{10} = 1.2.$$

$$s = \sqrt{\frac{\Sigma d^2}{n-1} - \frac{(\Sigma d)^2}{n(n-1)}} = \sqrt{\frac{90}{9} - \frac{(12)^2}{10 \times 9}} = \sqrt{10 - 1.6} = \sqrt{8.4} = 2.898$$

$$t = \frac{\bar{d}\sqrt{n}}{s} = \frac{1.2\sqrt{10}}{2.898} = \frac{1.2 \times 3.162}{2.898} = 1.309.$$

For $v = 9$, the table value of t at 5% level of significance is 2.262. Since the computed value of t is less than the table value, the null hypothesis is accepted. Hence, the training programme has not shortened the average assembly time.

Confidence Interval for the Mean of the Difference. When the population for the mean of differences is normally distributed with unknown variance for dependent samples, a 100 $(1 - \alpha)$ per cent confidence interval is given by

$$\bar{d} \pm t_{\alpha\ 2, v}\ s/\sqrt{n}\ .$$

(4) Test of Hypothesis about Coefficient of Correlation

Case 1 : Testing the hypothesis when the population coefficient of correlation equals zero, *i.e.*, $H_0 : \rho = 0$.

Here the null hypothesis is that there is no correlation in the population, *i.e.*, $H_0 : \rho = 0$. The population coefficient of correlation ρ measures the degree of relationship between the variables. When $\rho = 0$, there is no statistical relationship between the variables. In order to test this hypothesis, it is necessary to know the sample coefficient of correlation r (which is the best estimate of ρ). The appropriate test statistic to be used here is given by :

$$t = \frac{r}{\sqrt{1-r^2}} \times \sqrt{n-2}$$

which follows t-distribution with $n-2$ degrees of freedom.

If the computed value of t is greater than the table value of t, the null hypothesis is rejected which indicates that sample data provides sufficient evidence to indicate that $\rho \neq 0$. Hence, it can be concluded that there is a linear relationship between the variables.

Illustration 7. In a study of the relationship between expenditure (X) and annual sales volume (Y), a sample of 10 firms yielded the coefficient of correlation $r = 0.93$. Can we conclude on the basis of this data that X and Y are linearly related ?

Solution. The null hypothesis is $H_0 : \rho = 0$, *i.e.*, there is no relationship between two variables. Using the t-test

$$t = \frac{r}{\sqrt{1-r^2}} \times \sqrt{n-2} = \frac{0.93}{\sqrt{1-(0.93)^2}} \sqrt{10-2}$$

$$= \frac{0.93}{\sqrt{0.14}} \times \sqrt{8} = \frac{0.93 \times 2.828}{0.374} = 7.03.$$

The degrees of freedom or $v = n - 2 = 10 - 2 = 8$.

The table value of t at 5% level of significance for 8 *d.f.* is 2.306. Since the computed value is much greater than the table value, the null hypothesis is rejected. Hence, it may be concluded that X and Y are linearly related.

Case 2 : Testing the hypothesis when the population coefficient of correlation equals some other value than zero, *i.e.*, $H_0 : \rho = \rho_0$.

In this case when $\rho \neq 0$, the test based on t-distribution will not be appropriate. In testing the hypothesis, the use of Fisher's z-transformation will be applicable. Here, r is transformed into z by

$$z = \frac{1}{2} \log_e \frac{1+r}{1-r}$$

Here, $\log_e$ is a *natural* logarithm. Common logarithms may be shifted to natural logarithms by multiplying by the factor 2.3026, *i.e.*,

$$\log_e X = 2.3026 \log_{10} X$$

where X is a positive integer.

Since $\frac{1}{2}$ (2.306) = 1.1513, the transformation formula may be used as :

$$z = 1.1513 \log_{10} \frac{1+r}{1-r}$$

Now, it can be shown that Z is normally distributed with mean

$$z_\rho = \frac{1}{2} \log_e \frac{1+\rho}{1-\rho}$$

and standard deviation

$$\sigma_z = \frac{1}{\sqrt{n-3}}$$

Therefore, to test the null hypothesis that $\rho = \rho_0$, the test statistic would be :

$$z = \frac{z - z_\rho}{\sigma_z}$$

which follows approximately the standard normal distribution. This test is more appropriate if sample size is large. The approximation is reasonably good if the sample size is at least 10.

Case 3 : Testing the hypothesis for the difference between two independent correlation coefficients.

To test the hypothesis of two correlation coefficients derived from two separate samples, we have to compare the difference of the two corresponding values of z with the standard error of that difference. In this case, the formula used will be

$$z = \frac{z_1 - z_2}{\sigma_{z_1 - z_2}} = \frac{z_1 - z_2}{\sqrt{\dfrac{1}{n_1 - 3} + \dfrac{1}{n_2 - 3}}}$$

where
$$z_1 = \frac{1}{2} \log_e \frac{1 + r_1}{1 - r_2} = 1.1513 \log_{10} \frac{1 + r_1}{1 - r_1}$$

and
$$z_2 = \frac{1}{2} \log_e \frac{1 + r_2}{1 - r_2} = 1.1513 \log_{10} \frac{1 + r_2}{1 - r_2}.$$

If the absolute value of this statistic is greater than 1.96, the difference will be significant at 5% level.

Illustration 8. The following data give sample size :

Sample Size	Value of r
5	0.87
12	0.56

Test the significance of the difference between two values using Fisher's z-transformation.

Solution. Let the null hypothesis be that the experiment provides no evidence that the samples are drawn from the same population. Applying z-test :

$$z = \frac{z_1 - z_2}{\sqrt{\dfrac{1}{n_1 - 3} + \dfrac{1}{n_2 - 3}}}$$

$$z_1 = \frac{1}{2} \log_e \frac{1 + r_1}{1 - r_1} = 1.1513 \log_{10} \frac{1 + r_1}{1 - r_1}$$

$$= 1.1513 \log_{10} \frac{1 + 0.87}{1 - 0.87} = 1.1513 \log_{10} \frac{1.87}{0.13}$$

$$= 1.1513 \log_{10} 14.385 = 1.1513 \times 1.579 = 1.82$$

$$z_2 = \frac{1}{2} \log_e \frac{1 + r_2}{1 - r_2} = 1.1513 \log_{10} \frac{1 + r_2}{1 - r_2}$$

$$= 1.1513 \log_{10} \frac{1 + 0.56}{1 - 0.56} = 1.1513 \log_{10} = \frac{1.56}{0.44}$$

$$= 1.1513 \log 3.545 = 1.1513 \times 0.5495 = 0.63$$

$$z = \frac{z_1 - z_2}{\sqrt{\dfrac{1}{n_1 - 3} + \dfrac{1}{n_2 - 3}}} = \frac{1.82 - 0.63}{\sqrt{\dfrac{1}{5 - 3} + \dfrac{1}{12 - 3}}}$$

$$= \frac{1.19}{\sqrt{0.5 + 0.1}} = \frac{1.19}{\sqrt{0.6}} = \frac{1.19}{0.77} = 1.54.$$

Since the computed value of z is less than the table value of z at 5% level of significance, therefore, the null hypothesis is accepted. Hence, the experiment provides no evidence against the hypothesis that the samples are drawn from the same population.

Illustration 9. Suppose we want to test whether $r = 0.884$ for a pair of 20 observations is significantly different from a hypothesised value $\rho = 0.92$.

Solution. We transform r into z by :

$$z = 1.5113 \log_{10} \frac{1 + 0.884}{1 - 0.884} = 1.3938.$$

The distribution of z is approximately normal around the hypothesised value

$$\rho_0 = 0.92 = z\rho_0, \text{ where}$$

$$z\rho_0 = 1.1513 \log_{10} \frac{1 + 0.92}{1 - 0.92} = 1.5890.$$

The distribution of z has a standard deviation

$$\sigma_z = \frac{1}{\sqrt{20-3}} = \frac{1}{\sqrt{17}} = \frac{1}{4.123} = 0.2425.$$

Therefore, the statistic is

$$z = \frac{1.3938-1.5890}{0.2425} = -0.80.$$

From the table of areas for the normal curve, we find that in about 20 per cent times we may expect a difference as large or larger than this. This hypothesis that $r = 0.884$ can be rejected at a low level of confidence.

The *F*-Distribution

The F-distribution is named in honour of R.A. Fisher who *first studied it* in 1924. This distribution is usually defined in terms of the ratio of the variances of two normally distributed populations. The quantity

$$\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}$$

is distributed as F-distribution with $v_1 = n_1 - 1$ and $v_2 = n_2 - 1$ degrees of freedom,

where $s_1^2 = \dfrac{\Sigma(x_1-\bar{x}_1)^2}{n_1-1}$ is the unbiased estimator of σ_1^2 and $s_2^2 = \dfrac{\Sigma(x_2-\bar{x}_2)^2}{n_2-1}$ is the unbiased estimator of σ_2^2.

If $\sigma_1^2 = \sigma_2^2$, then the statistic

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

follows F-distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom.

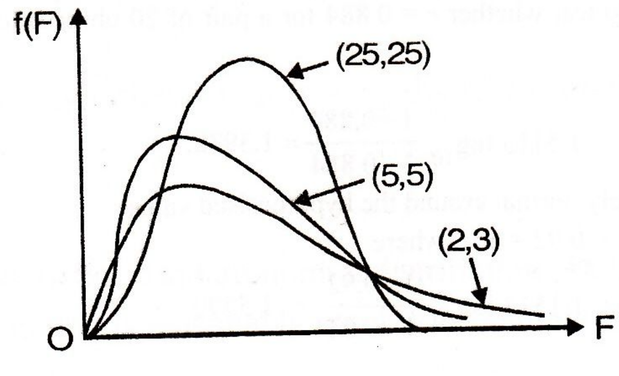
The F-distribution sometimes is also called *Variance Ratio* distribution which can be seen from the definition. The F-distribution depends on the degrees of freedom v_1 for the numerator and v_2 for the denominator. Therefore, the parameters for F-distribution are v_1 and v_2. For different values of v_1 and v_2 we shall have different distributions.

The probability density function of F-distribution is given by

$$f(F) = Y_0 \frac{F_{v_1/2-1}}{\left[1+\frac{v_1}{v_2}\right](v_1+v_2)/2} \qquad 0 \le F \le \infty$$

where Y_0 is a constant depending on the values v_1 and v_2 such that the area under the curve is unity. A typical F-distribution is given as below :

Some of the important properties of F-distribution are given below :



(1) The F-distribution is positively skewed and its skewness decreases with increase in v_1 and v_2.

(2) The value of F must always be positive or zero since variances are squares and can never assume negative values. Its value will always lie between 0 and ∞.

(3) The mean and variance of the F-distribution are

$$\text{Mean} = \frac{v_2}{-v_2 - 2}, \text{ for } v_2 > 2$$

$$\text{Variance} = \frac{2v_2^2(v_1 + v_2 - 2)}{v_1(v_2 - 2)^2(v_2 - 4)}, \text{ for } v_2 > 4.$$

(4) The shape of the F-distribution depends upon the number of degrees of freedom.

(5) The areas in the left-hand side of the distribution can be found by taking the reciprocal of F values corresponding to the right-hand side, when the number of degrees of freedom in the numerator and in the denominator are interchanged. This is also known as *reciprocal property* and *can be expressed as*

$$F_{1-\alpha, v_1, v_2} = \frac{1}{F_{\alpha, v_2, v_1}}$$

where the symbols have their usual meanings. This property is of great help when we want to know the lower tail F values from corresponding upper tail F values which are given in the Appendix.

Testing of Hypothesis for Equality of two Variances

The test of equality of two population variances is based on the variances in two independently selected random samples drawn from two normal populations. Under the null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$.

$$F = \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \text{ reduces to } F = \frac{s_1^2}{s_2^2}.$$

which follows F-distribution with v_1 and v_2 degrees of freedom. It is convenient to place larger sample variance in the numerator for computational purpose. If we do so, the ratio of the sample variance will be equal to or greater than one.

If the computed value of F exceeds the table value of F, we reject the null hypothesis, *i.e.*, the alternate hypothesis is accepted.

Illustration 10. Two sources of raw materials are under consideration by a company. Both sources seem to have similar characteristics but the company is not sure about their respective uniformity. A sample of 10 lots from source A yields a variance of 225 and a sample of 11 lots from source B yields a variance of 200. Is it likely that the variance of source A is significantly greater than the variance of source B at $\alpha = 0.01$?

Solution. Null hypothesis is $H_0 : \sigma_1^2 = \sigma_2^2$, *i.e.*, the variances of source A and that of source B are same. The F statistic to be used here is

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

where

$$s_1^2 = 225, \text{ and } s_2^2 = 200$$

$$F = \frac{225}{200} = 1.1$$

The table value of F for $v_1 = 9$ and $v_2 = 10$ at 1% level of significance is 4.49. Since the computed value of F is smaller than the table value of F, the null hypothesis is accepted. Hence, the population variances of the two populations are same.

Confidence Interval for the Ratio of Two Variances

A 100 (1–α) per cent confidence interval for the ratio of the variances of two normally distributed populations is given by

$$\frac{s_1^2 / s_2^2}{F_{(1-\alpha/2)}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2 / s_2^2}{F_{\alpha/2}}$$

where the symbols have their usual meanings.

MISCELLANEOUS ILLUSTRATIONS

Illustration 11. The nine items of a sample had the following values :

45, 47, 50, 52, 48, 47, 49, 53, 50

The mean is 49 and the sum of squares of deviations taken from mean is 52. Can this sample be regarded as taken from the population having 47 as mean ? Also obtain 95% and 99% confidence limits of the population mean. (*MBA, Delhi Univ., 2002*)

Solution. The null hypothesis is $H_0 : \mu = 47$, *i.e.*, the population mean is 47, we are given that

$\bar{x} = 49$, $\Sigma (x - \bar{x})^2 = 52$, and $\mu = 47$, $n=9$. Applying t-test

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

where $s = \sqrt{\frac{\Sigma (x - \bar{x})^2}{n - 1}} = \sqrt{\frac{52}{8}} = 2.55$

Substituting the values, we have

$$t = \frac{49 - 47}{2.55 / 3} = 2.35$$

The table value of t for 8 *d.f.* at 5% level of significance is 2.31. Since the computed value is slightly greater than the table value, the null hypothesis is rejected. Hence, the samples are not drawn from the population having 47 as mean.

95% confidence interval of the population mean is given by :

$$\bar{x} \pm t_{0.05}\ s/\sqrt{n}$$

$$= 49 \pm \frac{2.31 \times 2.55}{3} = 49 \pm 1.96 = 47.04 \text{ to } 50.96$$

99% confidence interval of the population mean is given by

$$\bar{x} \pm t_{0.01}\ s/\sqrt{n}$$

$$= 49 \pm \frac{3.36 \times 2.55}{3} = 49 \pm 2.86 = 46.14 \text{ to } 51.86.$$

Illustration 12. A company is interested in knowing if there is a significant difference in the average salary received by foremen in two divisions. Accordingly, samples of 12 foremen in the first division and 10 foremen in the second division are selected at random. Based upon experience, foremen's salaries are known to be approximately normally distributed, and standard deviations are about the same.

	First division	*Second division*
Sample size	12	10
Average weekly salary of foremen (Rs.)	1050	980
Standard deviation of salaries (Rs.)	68	74

Solution. Let us take the null hypothesis that the average salary received by foremen in the two divisions does not differ significantly. Applying t-test:

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

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$$= \sqrt{\frac{(12-1)68^2 + (10-1)74^2}{12+10-2}} = \sqrt{\frac{50864 + 49284}{20}} = 70.76$$

$$t = \frac{1050 - 980}{70.76} \sqrt{\frac{12 \times 10}{12 + 10}} = \frac{70}{70.76} \times 2.34 = 2.31$$

$$v = n_1 + n_2 - 2 = 12 + 10 - 2 = 20.$$

For $v = 20$, $t_{0.05} = 2.086$. The calculated value of t is greater than the table value. The null hypothesis does not hold good. Hence, there is significant difference in the salary received by foremen in the two divisions.

Illustration 3. Ten accountants were given intensive coaching and four tests were conducted in a month. The scores of tests 1 and 4 are given below:

S. No. of Accountants	:	1	2	3	4	5	6	7	8	9	10
Marks in 1st test	:	50	42	51	42	60	41	70	55	62	38
Marks in 4th test	:	62	40	61	52	68	51	64	63	72	50

Does the score from test 1 to test 4 show an improvement ? Test at 5% level of significance.

Solution. Let us take the null hypothesis that there is no improvement from test 1 to 4. Applying t-test:

$$t = \frac{\bar{d}\sqrt{n}}{s}.$$

CALCULATION FOR $\bar{d}$ AND s

S. No.	*1st test*	*4th test*	*(4th–1st)* d	d^2
1	50	62	+12	144
2	42	40	–2	4
3	51	61	+10	100
4	42	52	+10	100
5	60	68	+8	64
6	41	51	+10	100
7	70	64	–6	36
8	55	63	+8	64
9	72	62	+10	100
10	38	50	+12	144
			$\Sigma d = 72$	$\Sigma d^2 = 856$

$$\bar{d} = \frac{\Sigma d}{n} = \frac{72}{10} = 7.2$$

$$s = \sqrt{\frac{\Sigma d^2 - n(\bar{d})^2}{n-1}} = \sqrt{\frac{856 - 10(7.2)^2}{9}} = \sqrt{\frac{856 - 518.4}{9}} = 6.125$$

$$t = \frac{7.2\sqrt{10}}{6.125} = \frac{7.2 \times 3.162}{6.125} = 3.72$$

For $v = 9$, $t_{0.05} = 1.83$

The calculated value of t is greater than table value. The null hypothesis is rejected. Hence, the scores from test 1 to test 4 show an improvement.

Illustration 14. The means of two random samples of sizes 9 and 7 are 196.42 and 198.82 respectively. The sum of the squares of the deviations from the mean are 26.94 and 18.73 respectively. Can the sample be considered to have been drawn from the same normal population?

(*MBA, Delhi Univ., 2006*)

Solution. Let us take the null hypothesis that the samples are drawn from the same normal population. Applying t-test, *i.e.*,

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

$$s = \sqrt{\frac{\Sigma(x_1 - \bar{x}_1)^2 + \Sigma(x_2 - \bar{x}_2)^2}{n_1 + n_2 - 2}} = \sqrt{\frac{26.94 + 18.73}{9 + 7 - 2}} = \sqrt{\frac{45.67}{14}} = 1.81$$

$$t = \frac{196.42 - 198.82}{1.81}\sqrt{\frac{9 \times 7}{9 + 7}} = \frac{2.40 \times 1.984}{1.81} = \frac{4.76}{1.81} = 2.63$$

For $v = 14$, $t_{0.05} = 2.145$. Since the calculated value of t is greater than the table value, we reject the null hypothesis. Hence, the difference between the two means is significant. Therefore, the samples cannot be said to have been drawn from the same normal population.

Illustration 15. Strength tests carried out on samples of two yarns spun to the same count gave the following results :

	Sample size	*Sample mean*	*Sample variance*
Yarn *A*	4	52	42
Yarn *B*	9	42	56

The strengths are expressed in pounds. Is the difference in mean strengths significant of real difference in the mean strengths of the sources from which the samples are drawn ? *(MBA. Delhi Univ., 2004, 2007)*

Solution. Let us take the null hypothesis that there is no significant difference in the mean strengths of the two types of yarns. Applying *t*-test,

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

where,

$$s = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{3\times 42 + 8\times 56}{4+9-2}} = \sqrt{\frac{574}{11}} = 7.224$$

$$t = \frac{52-42}{7.224}\sqrt{\frac{4\times 9}{4+9}} = \frac{10\times 1.664}{7.224} = 2.303$$

$$v = n_1 + n_2 - 2 = 4 + 9 - 2 = 11. \text{ For } v = 11, t_{0.05} = 1.796$$

The calculated value of *t* is more than the table value of *t*. The null hypothesis is rejected. The difference in the mean strengths of the two types of yarn is significant.

Illustration 16. To verify whether a course in accounting improved performance, a similar test was given to 12 participants both before and after the course. The original marks recorded in alphabetical order of the participants were 44, 40, 61, 52, 32, 44, 70, 41, 67, 72, 53, and 72. After the course, the marks were in the same order : 53, 38, 69, 57, 46, 39, 73, 48, 73, 74, 60, and 78. Test whether the course was useful ?

Solution. Let us take the null hypothesis that the course has not improved the performance of the participants. Applying *t*-test,

$$t = \frac{\bar{d}\sqrt{n}}{s}$$

CALCULATION FOR $\bar{d}$ AND s

Before	*After*	*(2nd–1st)* d	d^2
44	53	+9	81
40	38	–2	4
61	69	+8	64
52	57	+5	25
32	46	+14	196
44	39	–5	25
70	73	+3	9
41	48	+7	49
67	73	+6	36
72	74	+2	4
53	60	+7	49
72	78	+6	36
		$\Sigma d = 60$	$\Sigma d^2 = 578$

$$\bar{d} = \frac{\Sigma d}{n} = \frac{60}{12} = 5$$

$$s = \sqrt{\frac{\Sigma d^2 - n(\bar{d})^2}{n-1}} = \sqrt{\frac{578 - 12(5)^2}{12-1}} = \sqrt{\frac{578-300}{11}} = 5.027$$

$$t = \frac{5\sqrt{12}}{5.027} = \frac{17.32}{5.027} = 3.45$$

For $v = 11$, $t_{0.05} = 2.201$. The calculated value of *t* is more than the table value of *t*. The null hypothesis is rejected. Hence, course has improved the performance of the participants.

Illustration 17. Samples of two different types of bulbs were tested for length of life, and the following data were obtained :

	Type I	*Type II*
Sample Size	8	7
Sample Mean	1234 hrs.	1136 hrs.
Sample S.D.	36 hrs.	40 hrs.

Is the difference in mean life of two types of bulbs significant ? (*MBA, Delhi Univ., 2003*)

Solution. Let us take the null hypothesis that there is no significant difference in the mean life of the two types of bulbs. Applying t-test,

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

$$s = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$$

$$= \sqrt{\frac{(8-1)\,36^2 + (7-1)\,40^2}{8+7-2}} = \sqrt{\frac{9072+9600}{13}} = 37.9$$

$$t = \frac{1234-1136}{37.9}\sqrt{\frac{8 \times 7}{8+7}} = \frac{98}{37.9} \times 1.932 = 5$$

For $v = 13$, $t_{0.05} = 2.16$. The calculated value of t is greater than the table value. The null hypothesis is rejected. Hence, there is a significant difference in the mean life of two types of bulbs.

Illustration 18. Two random samples drawn from normal populations are :

Sample I : 20 16 26 27 23 22 18 24 25 19

Sample II : 27 33 42 35 32 34 38 28 41 43 30 37

Obtain estimates of the variances of the population and test whether two populations have the same variance.

Solution. Let us take the null hypothesis that two population have the same variance. Applying F-test :

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

Sample I			*Sample II*		
x_1	$(x_1 - \bar{x}_1)$	$(x_1 - \bar{x}_1)^2$	x_2	$(x_2 - \bar{x}_2)$	$(x_2 - \bar{x}_2)^2$
20	−2	4	27	−8	64
16	−6	36	33	−2	4
26	+4	16	42	+7	49
27	+5	25	35	0	0
23	+1	1	32	−3	9
22	0	0	34	−1	1
18	−4	16	38	+3	9
24	+2	4	28	−7	49
25	+3	9	41	+6	36
19	−3	9	43	+8	64
			30	−5	25
			37	+2	4
$\Sigma x_1 = 220$		$\Sigma (x_1 - \bar{x}_1)^2 = 120$	$\Sigma x_2 = 420$		$\Sigma (x_2 - \bar{x}_2)^2 = 314$

$$s_1^2 = \frac{\Sigma(x_1-\bar{x}_1)^2}{n_1-1} = \frac{120}{9} = 13.333\ ;$$

$$s_2^2 = \frac{\Sigma(x_2-\bar{x}_2)^2}{n_2-1} = \frac{314}{11} = 28.545$$

$$F = \frac{s_1^2}{s_2^2} = \frac{13.333}{28.545} = 0.467$$

Since numerator is greater than denominator, therefore,

$$F = \frac{28.545}{13.33} = 2.14$$

The critical value of F for $v_1 = 9$ and $v_2 = 11$ at 5% level is 4.63. Since the calculated value of F is less than the table value, therefore, there is no reason to reject the null hypothesis. Hence, it may be concluded that the two populations have the same variance.

Illustration 19. A drug manufacturer has installed a machine which fills automatically 5 gms of drug in each phial. A random sample of phials was taken and it was found to contain 5.02 gms on an average in a phial. The S.D. of the sample was 0.002 gms. Test at 5% level of significance, if the adjustment in the machine is in order. *(MBA, DU, 1999)*

Solution. Let us take the null hypothesis that there is no significant difference between $\bar{x}$ and μ. Applying t-test,

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

$$\bar{x} = 5.02,\ \mu = 5,\ s = 0.002,\ n = 10$$

$$t = \frac{5.02 - 5}{0.002} \times \sqrt{10} = \frac{0.02}{0.002} \times 3.162 = 31.62$$

For $v = 9$, $t_{0.05} = 1.833$. The calculated value is much higher than the table value. The null hypothesis is rejected. Hence, the adjustment in the machine is not in order.

Illustration 20. A random sample of 12 families in one city showed an average weekly food expenditure of Rs. 1380 with a standard deviation of Rs. 100 and a random sample of 15 families in another city showed an average monthly food expenditure of Rs. 1320 with a standard deviation of Rs. 120. Test, whether the difference between the two means is significant at a level of significance of 0.01.

Solution. Let us take the null hypothesis that there is no significant difference in the mean expenditure of the families in the two cities. Applying t-test,

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

$$S = \sqrt{\frac{(n_1 - 1)\,s_1^2 + (n_2 - 1)\,s_2^2}{n_1 + n_2 - 2}}$$

$$= \sqrt{\frac{(12-1)\,100^2 + (15-1)\,120^2}{12+15-2}}$$

$$= \sqrt{\frac{110000 + 201600}{25}} = 111.64$$

$$t = \frac{1380 - 1320}{111.64}\sqrt{\frac{12 \times 15}{12+15}} = \frac{60 \times 2.582}{111.64} = 1.39$$

$$v = 12 + 15 - 2 = 25$$

For $v = 25$, $t_{0.01} = 2.485$.

The calculated value t is less than the table value. The null hypothesis is accepted. Hence, the difference between the two means is not significant.

Illustration 21. Eight students were given a test in statistics, and after one month's coaching, they were given another test of the similar nature. The following table gives the increase in their marks in the second test over the first :

Roll No.	:	1	2	3	4	5	6	7	8
Increase in marks	:	4	–2	6	–8	12	5	–7	2

Do the marks indicate that the students have gained from the coaching ?

Solution. Let us take the null hypothesis that the students have not gained from the coaching. Applying the t-test :

$$t = \frac{\bar{d}\sqrt{n}}{s}$$

d	d^2
+4	16
–2	4
+6	36
–8	64
+12	144
+5	25
–7	49
+2	4
$\Sigma d = 12$	$\Sigma d^2 = 342$

$$\bar{d} = \frac{\Sigma d}{n} = \frac{12}{8} = 1.5$$

$$s = \sqrt{\frac{\Sigma d^2 - n(\bar{d})^2}{n-1}} = \sqrt{\frac{342 - 8(1.5)^2}{8-1}} = \sqrt{\frac{342-18}{7}} = 6.8$$

$$t = \frac{1.5\sqrt{8}}{6.8} = 0.624$$

For $v = 7$, $t_{0.05} = 1.895$.

The calculated value of t is less than the table value. The null hypothesis is accepted. Hence, the students have not gained from the coaching.

Illustration 22. You are given the following data about the life of two brands of bulbs :

	Mean life	*Standard deviation*	*Sample size*
Brand A	2,000 hrs	250 hrs	12
Brand B	2,230 hrs	300 hrs	15

Do you think there is a significant difference in the quality of the two brands of bulbs ? (*MBA, DU, 2002*)

Solution :

Given : $\bar{x}_1 = 2000$, $s_1 = 250$, $n_1 = 12$

$\bar{x}_2 = 2230$, $s_2 = 300$, $n_2 = 15$

Let the null hypothesis be that there is no significant difference in the quality of the two brands of bulbs.

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

where,

$$s = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{11\,(62500) + 14\,(90000)}{12+15-2}}$$

$$= \sqrt{\frac{1947500}{25}} = \sqrt{77900} = 279.11$$

$$t = \frac{2000-2230}{279.11}\sqrt{\frac{12 \times 15}{12+15}} = \frac{-230}{279.11}\sqrt{\frac{180}{27}}$$

$$= \frac{-230(2.58)}{279.11} = -\frac{593.4}{279.11} = -2.126$$

The table value of t for 25 *d.f.* at 5% level of significance is 1.708. Since computed value is greater than the table value, therefore, we reject the null hypothesis. Hence, the quality of the two brands of bulbs differ significantly.

Illustration 23. Samples of final examination scores for two statistics classes with different instructors provided the following results :

	Sample Size	*Mean*	*Standard Deviation*
Instructor A	12	72	8
Instructor B	15	78	10

Test, whether these data are sufficient to conclude that the mean scores for the two classes differ. (*MBA, D.U., 2003*)

Solution. Let us take the hypothesis that there is no significant difference in the mean scores because of different instructors. Applying t-test of difference of means :

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

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{(12-1)\,8^2 + (15-1)10^2}{12 + 15 - 2}}$$

$$= \sqrt{\frac{704 + 1400}{25}} = \sqrt{\frac{2104}{25}} = 9.14$$

$$t = \frac{72 - 18}{9.17} \sqrt{\frac{12 \times 15}{12 + 15}} = \frac{-6}{9.17} \times 2.58 = -1.69$$

For $\qquad v = 25,\ t_{0.05} = 2.06.$

The calculated value of t is less than the table value. Hence, there is no significant difference in the mean scores of different instructors.

Illustration 24. The average middle class family spends Rs. 9,000 per month. A random sample of 25 families in a city showed a sample mean monthly expenditure of Rs. 8,450 with a standard deviation of Rs. 1,450. Test $H_0 : \mu =$ Rs. 9,000 and $H_a \neq$ Rs. 9,000 with $\alpha = 0.05$. Use Two Tailed test.

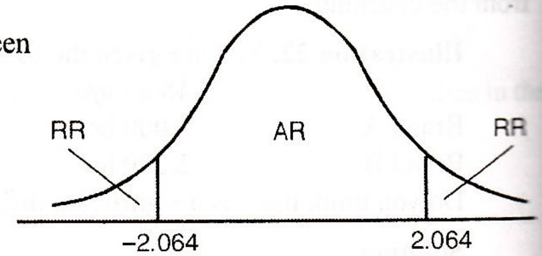
(*i*) What are the critical values of the test statistic, and what is the rejection region ?

(*ii*) Compute the value of the test statistic.

(*iii*) What is your conclusion ? (*MBA, Delhi Univ., 2007*)

Solution. Given $\mu = 9000$, $n = 25$, $\bar{x} = 8450$, $s = 1450$

Let us take the null hypothesis that there is no significant difference between sample mean monthly expenditure and population monthly expenditure, *i.e.*,



$H_0 : \mu =$ Rs. 9000, $H_a : \mu \neq$ Rs. 9000, Using t-test.

(*i*) Critical values of t for 24 *d.f.* at 5% level of significance are ± 2.064.

(*ii*) $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} = \dfrac{8450 - 9000}{1450/\sqrt{25}} = -\dfrac{550 \times 5}{1450} = -1.897$

(*iii*) Since the computed value of $t = -1.897$ is less than the table value of $t = -2.064$, therefore, it lies in the acceptance region. Hence, there is no significant difference between sample mean monthly expenditure and population mean monthly expenditure. Therefore, the samples have been drawn from the given population.

Illustration 25. A market research firm used a sample of individuals to rate the purchase potential of a particular product before and after the individuals saw a new television commercial about the product. The purchase potential ratings were based on 0 to 10 scale, with higher values indicating a higher purcahse potential. Test the hypothesis that the commercial improved the mean purchase potential rating. Use level of significance 5% and comment on the value of the commercial.

Individual :	1	2	3	4	5	6	7	8
Purchase rating (After) :	6	6	7	4	3	9	7	6
Purchase rating (Before) :	5	4	7	3	5	8	5	6

(*MBA, Delhi Univ., 2009*)

Solution. Let us take the hypothesis that the new television commercial has not improved the mean purchase potential rating. Applying the t-test :

Individual	*Purchase rating before 1st*	*Purchase rating after 2nd*	*(2nd – 1st)* d	d^2
1	5	6	1	1
2	4	6	2	4
3	7	7	0	0
4	3	4	1	1
5	5	3	–2	4
6	8	9	1	1
7	5	7	2	4
8	6	6	0	0
			$\Sigma d = 5$	$\Sigma d^2 = 15$

$$t = \frac{\bar{d}\sqrt{n}}{s}$$

$$\bar{d} = \frac{\Sigma d}{n} = \frac{5}{8} = 0.625$$

$$s = \sqrt{\frac{\Sigma d^2 - n(\bar{d})^2}{n-1}}$$

$$= \sqrt{\frac{15 - 8(0.625)^2}{8-1}} = \sqrt{\frac{15 - 3.125}{7}} = \sqrt{\frac{11.875}{7}} = 1.302$$

$$t = \frac{0.625\sqrt{8}}{1.302} = \frac{0.625 \times 2.828}{1.302} = \frac{1.7675}{1.302} = 1.358$$

$$v = 8 - 1 = 7, \quad \text{For } v = 7, t_{0.05} = 1.895$$

Since the calculated value of t (1.358) is less than the table value (1.895), the null hypothesis is accepted. Hence the new television commercial has not improved the mean purchase potential rating.

Illustration 26. The variance in production process is an important measure of the quality process. A large variance of ten signals an opportunity for improvement in the process by finding ways to reduce the process variance. Conduct a statistical test to determine whether there is a significant difference between the variances in the bag weights for the two machines. Use a 10% level of significance. What is your conclusion? Which machine, if either, provides the greater opportunity for quality improvements ?

	No. of observation	*Mean*	*Standard deviation*
Machine 1 :	25	5.9	2.0
Machine 2 :	22	6.3	1.9

(*MBA, Delhi Univ., 2009*)

Solution. Let us take the hypothesis that there is no significant difference in the two machines for providing opportunity in quality improvement. Applying the t-test of difference of means :

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s}\sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$$= \sqrt{\frac{(25-1)4 + (22-1)3.61}{25 + 22 - 2}} = \sqrt{\frac{96 + 75.81}{45}} = \sqrt{\frac{171.81}{45}} = 1.954$$

$$\bar{X}_1 = 5.9, \bar{X}_2 = 6.3, n_1 = 25, n_2 = 22, s = 1.954$$

$$t = \frac{5.9 - 6.3}{1.954}\sqrt{\frac{25 \times 22}{25 + 22}}$$

$$= \frac{0.4}{1.954} \times 3.421 = 0.7$$

$$v = 25 + 22 - 2 = 45$$

$$\text{For } v = 45, t_{0.10} = 1.301$$

Since the calculated value of t is less than the table value, the hypothesis is accepted. Hence there is no significant difference in the two machines for providing opportunity in quality improvement.

PROBLEMS

I-A : Answer the following questions, each question carries **one** mark:

(*i*) What is a t-distribution ?

(*ii*) Write the formula for difference of two means in case of small sample tests.

(*iii*) Define t-test. (*M. Com., Madurai-Kamaraj Univ., 2002*)

(*iv*) Give two important properties of t-distribution.

(*v*) Give at least two important applications of t-distribution.

(*vi*) What do you understand by degrees of freedom ?

(*vii*) What is paired t-test? Give its formula.

(*viii*) What is F-distribution ?

(*ix*) Give two important properties of F-distribution.

(*x*) Give any important application of F-distribution.

1-B : Answer the following questions, each question carries **four** marks:

(*i*) Explain the difference between the means of two samples by using *t*-distribution.

(*ii*) Explain *t*-test distribution and note its properties.

(*iii*) In what ways small sampling theory differs from large sampling theory.

(*iv*) What is the procedure involved in testing hypothesis of a coefficient of correlation.

(*v*) What are the assumptions involved in using the *F*-test for testing the equality of two sample variances?

2. How does small sampling theory differ from large sampling theory ?
3. (*a*) What is *t*-distribution ? Give its important properties.
   (*b*) What is Students '*t*' distribution? Point out its usefulness.
4. (*a*) Give some important applications of the *t*-test and explain how it helps in arriving at business decisions.
   (*b*) How can "*t*" test be applied for testing the significance of the difference between two sample means ?
5. Discuss the *F*-test for testing the equality of two sample variances. State clearly assumptions involved.
6. 12 persons were appointed in clerical position in an office. Their performance was noted by giving a test and the marks recorded out of 10. They were given 3 month's training and again they were given a test and marks recorded out of 12.

Employees :	A	B	C	D	E	F	G	H	I	J	K	L
Before training :	4	5	3	7	8	6	5	9	10	6	4	3
After training :	5	4	6	8	7	5	9	9	10	6	5	4

Can it be concluded that the training has improved the performance of the employees ?

7. A random sample of 25 pairs of observations from a normal population gives a correlation coefficient of 0.46. Is it likely that the variables in the population uncorrelated ?
8. 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.
9. Two laboratories *A* and *B* carry out independent estimates of fat content in ice-cream made by a firm. A sample is taken from each batch, halved and the separate halves sent to the two laboratories. The fat content obtained by the laboratories is recorded below :

Batch No. :	1	2	3	4	5	6	7	8	9	10
Lab. *A* :	3	5	7	3	8	6	9	4	7	8
Lab. *B* :	9	8	8	4	7	7	9	6	6	6

Is there a significant difference between the mean fat content obtained by the two laboratories *A* and *B* ?

10. An automobile tyre manufacturer claims that the average life of certain grade of tyre is greater than 25,000 km when used under normal driving conditions on a car of a certain weight. A random sample of 15 tyres was tested, and a mean and standard deviation of 27,000 and 5,000 kms respectively, were computed. Can we conclude that the manufacturer's product is as good as claimed ? (*M. Com., Madurai-Kamaraj Univ., 2007*)
    [$t = 1.55$]
11. The quality control department of a food processing firm specifies that the mean net weight per package of a certain food must be 20 gms. Experience has shown that the weights are approximately normally distributed with a standard deviation of 15 gms. If a random sample of 15 packages yields a mean weight of 19.5 gms, is this sufficient evidence to indicate that the true mean weight of the package has decreased ?
    [$t = 1.29$]
12. Two working designs are under consideration for adoption in a plant. A time and motion study shows that 12 workers using design *A* have mean assembly time of 300 seconds with a standard deviation of 12 seconds and that 15 workers using design *B* have a mean assembly time of 335 seconds with a standard deviation of 15 seconds. Is the difference in the mean assembly time between the two working designs significant at 1% level of significance ?
    [$t = 23.52$]
13. The mean life of a sample of 10 electric light bulbs was found to be 1,456 hours with standard deviation of 423 hours. A second sample of 17 bulbs chosen from a different batch showed a mean life of 1,280 hours with standard deviation of 398 hours. Is there a significant difference between the means of the two batches ? (*MBA, Kumaun Univ., 2009*)
    [$t = 1.085$]
14. The variability in the tensile strength of two types of steel wire is to be compared. Given a sample of 14 observations of type *A* wire yielding a variance of 31.5, and a sample of 15 observations of type *B* wire yielding a variance of 29.3. Test the hypothesis that the two populations have equal variances.
15. In an *F* ratio with $v_1 = 4$ and $v_2 = 15$ is found to be 3.64. Is this value of *F*, significantly different from zero at 5% level of significance ?

16. A sample of the monthly earnings records of 15 employees of company A has a variance of Rs. 15.90, while a similar sample of 87 employees for company B has a variance of Rs. 17.50. Is it safe to assume that there is less variance in company A than in company B ?

17. The nicotine contents in milligrams of two samples of tobacco were found to be as follows :

Sample A :	20	16	26	27	23	22	
Sample B :	27	33	42	35	32	34	38

Can it be said that two samples come from normal population having the same mean ?

18. Two laboratories carry out independent estimates of a particular chemical in a medicine produced by a certain firm. A sample is taken from each batch, halved and the separate halves sent to the two laboratories. The following data are obtained :

Number of samples	10
Mean value of the difference of estimates	0.6
Sum of the squares of the differences (from their means)	20

Is the difference significant ?

19. A fertiliser mixing machine is set to give 12 kg of nitrate for every quintal bag of fertiliser. Ten 100 kg bags are examined. The percentage of nitrate are as follows :

11, 14, 13, 12, 13, 12, 13, 14, 11, 12.

Is there reason to believe that the machine is defective ?

20. The marks obtained by two groups of students in a Statistics test are given below :

	Group A	*Group B*
No. of students	12	11
Mean Marks	42	38
Standard Deviation of Marks	10	15

On the basis of this data, can it be concluded that there is a significant difference in the mean marks obtained by the two groups ?

21. Two types of batteries are tested for their length of life and the following data are obtained :

	Size of sample	*Mean life in hours*	*Variance*
Type A	9	600	121
Type B	8	640	144

Is there a significant difference in the two means ?

$[t = 7.17]$ (*MBA, Delhi Univ., 2008*)

22. A correlation coefficient of 0.2 is found in a sample of 28 pairs. Use Z-test to find out if this is significantly different from zero.

23. Two different types of drugs 'A' and 'B' were tried on certain patients for increasing weight, 6 persons were given drug A and 8 persons were given drug B. The increase in pounds is given below :

Drug A :	7	10	13	12	4	8		
Drug B :	12	8	3	18	16	9	8	3

Do the two drugs differ significantly with regard to their effect in increase in weight ?

$[t = 0.42]$ (*MBA, Hyderabad Univ., 2005*)

24. The mean weekly sales of the chocolate bar in general stores was 146.3 bars per store. After an advertising campaign, the mean weekly sales in 22 stores for a typical week increased to 157.7 bars and showed a standard deviation of 17.2. Was the advertising campaign successful ?

25. A company desires to compare the effects on cavities of its brand A with a competitor's brand B. To eliminate some of the variation in test population pairs of identical twins are used. A brand is randomly assigned to each twin and is used for two years. The number of cavities developed during the period are reported below :

Pair :	1	2	3	4	5
Brand A :	4	5	7	6	3
Brand B :	3	5	4	2	1

Test at 5% level of significance, whether the data indicate a difference in cavities developed between the two brands.

26. Calculate the value of t and test the hypothesis of the difference between the average proteins for the two States as given below :

Protein results

State I :	12.6	13.4	11.9	12.8	13.0		
State II :	13.1	13.4	12.8	13.5	13.3	12.7	12.4

27. As a part of an industrial training programme, some trainees are instructed by method A, which is straight teaching machine instruction, and some are instructed by method B, which involves the personal attention of the instructor. The trainees instructed by each method, and the scores they obtained in an appropriate achievement test are :

Method A :	71	75	65	69	73	66	68	71	74	68
Method B :	72	77	84	78	69	70	77	73	65	75

Test the claim that method B is more effective. Use 5% level of significance.

[$t = 1.977$]

28. Two independent samples of 8 and 7 items respectively gave the following values :

Sample A :	9	11	13	11	15	9	12	14
Sample B :	10	12	10	14	9	8	10	

Examine, whether the difference between the means of the two samples is significant ?

29. To test the effect of a fertiliser on rice production, 24 plots of land having equal areas were chosen. Half of these plots were treated with fertiliser and the other half were untreated. Other conditions were the same. The mean yield of rice on the untreated plots was 4.8 quintals with a standard deviation of 0.4 quintal, while the mean yield on the treated plots was 5.1 quintals with a standard deviation of 0.26 quintal. Can we conclude that there is significant improvement in rice production because of the fertiliser at 5% level of significance ?

[t=2.18]

30. To compare the efficiency of standard and electric typewriters, ten typists are chosen at random and trained in the use of both kinds of typewriters. They are then asked to type on each kind of typewriter for half an hour and their speeds measured average number of words typed per minute, are observed and given in the table below :

Typist :	A	B	C	D	E	F	G	H	I	J
Standard :	60	64	72	76	75	75	79	74	84	82
Electric :	55	62	70	90	70	72	78	70	90	100

Are you of the opinion that there is a vast difference in the efficiency of the two types of typewriters ?

Suppose your company decides to buy the standard typewriter for use only when the electric typewriter gives 20 words per minute greater than that of a standard typewriter. Based on the result obtained above, how should the company act ?

31. In an assignment, subjects were assigned at random between two conditions, five to each. Their scores are given below. Can one say that there is a significant difference between these two conditions ? What must be assumed in carrying out this test ?

Condition A :	128	115	120	110	103
Condition B :	123	115	130	135	113

32. A company selects 9 salesmen at random and their sales figures (in thousand Rs.) for the previous month are recorded. They then undergo a course devised by a business consultant and their sales figures for the following month are compared as shown in the table. Has the training course caused an improvement in the salesmen's ability ? Use 5% level.

Previous month :	75	90	94	85	100	90	69	70	64
Following month :	77	101	93	92	105	88	73	76	68

[$t = 3$]

33. Two random samples were drawn from two normal populations and their values are :

A :	66	67	75	76	82	84	88	90	92		
B :	64	66	74	78	82	85	87	92	93	95	97

Test, whether the two populations have the same variance at 5% level of significance.

[$F = 1.415$]

34. The lifetime of electric bulbs for a random sample of 10 from a large consignment gave the following data :

Sample :	1	2	3	4	5	6	7	8	9	10
Life in 1000 hours :	4.2	4.6	3.9	4.1	5.2	3.8	3.9	4.3	4.4	5.6

Can we accept the hypothesis that the average lifetime of bulbs is 4000 hours ?

35. A sample of 25 college students is given an aptitude test. The mean test score for the sample is 475 and the standard deviation is 30. It is believed that the students of the college are above the normal average test score of 470. Conduct an appropriate test at 5% level of significance.

36. A Drug is given to 10 patients, and the increments in their blood pressure were recorded to be 3, 6, –2, 4, –3, 4, 6, 0, 0, 2. Is it reasonable to believe that the drug has no effect on change of blood pressure ?

37. A machine is producing ball bearings with diameters of 0.5 inches. It is known that the standard deviation of the bearings is 0.005 inches. A sample of 25 ball bearings is selected and their average diameter is found to be 0.498 inches. Determine the 99 per cent confidence interval.

38. Ten students are selected at random from a college and their height (in cms.) are found to be 100, 104, 108, 110, 118, 120, 122, 124, 126 and 128 cms. In the light of these data, discuss the suggestion that the mean height of the students of the college is 110 cms. (Use 5% level of significance).
[t = 1.936]

39. For a random sample of 10 persons, fed on diet A, the increases in weight in pounds in a certain period were :
10 6 16 17 13 12 8 14 15 9
For another random sample of 12 persons, fed on diet B, the increases in weight in the same period were :
9 13 22 15 12 14 18 8 21 13 10 17
Test, whether the diets A and B differ significantly as regards their effect on increase in weight.

40. The wages of 10 workers taken at random from a factory are given below :
Wages (Rs.) : 1578, 1572, 1570, 1568, 1572, 1578, 1570, 1572, 1596, 1584.
Is it possible that the mean wage of all workers of this factory is Rs. 1580 ?
[t=1.481, yes]

41. Eight students were given a test in mathematics and after one month's coaching, they were given another test of the similar nature. The following table gives the increases in their marks in the second test over the first :

Roll No. :	1	2	3	4	5	6	7	8
Increase in marks :	4	–2	6	–8	12	5	–7	2

Do the marks indicate that the students have gained from coaching ?

42. 10 workers are selected at random from a large number of workers in a factory. The number or items produced by them on a certain day are found to be :
51, 52, 53, 54, 55, 56, 57, 58, 59, 60.
In the light of these data, would it be appropriate to suggest that the mean of the number of items produced in the population is 58 ?
[t = 2.6, yes]

43. An automatic device is set to fill 170 pills in each bottle of a certain medicine. A sample of ten bottles was taken. They were found to contain : 168, 164, 166, 167, 168, 169, 170, 170, 170 and 171 pills, with a standard deviation of 2.16 pills. Discuss whether the device is properly adjusted.

44. A random sample of 9 items is taken of a certain measurement. From the data, it is found that ΣX = 108 and ΣX^2 = 1584. Find the confidence limits for the population mean at 5% level of significance and test the hypothesis that the population mean is 8. (Table values of 't' at 5% level for 8 d.f. and 9 d.f. are respectively 2.306 and 2.262.)

45. The following data show weekly sales of a manufacturer before and after reorganisation of the sales function, 10 weeks from Sept. to Dec. in the successive years were selected for comparisons :

Week No.	1	2	3	4	5	6	7	8	9	10
Sales (before re-organisation) (in '000 Rs.)	15	17	12	18	16	13	15	17	19	18
Sales (after re-organisation) (in '000 Rs.)	20	19	18	22	20	19	21	23	24	24

Apply 't' test to determine whether reorganisation had any effect on sales.

46. Eleven sales executives trainees are assigned selling jobs right after their recruitment. After a fortnight, they are withdrawn from their field duties and given a month's training for executive sales. Sales executed by them in thousands of rupees before and after the training in the same period, are listed below :

Sales ('000 Rs.) (Before training) :	23	20	19	21	18	20	18	17	23	16	19
Sales ('000 Rs.) (After training) :	24	19	21	18	20	22	20	20	23	20	27

Do these data indicate that the training has contributed to their performance ?
(M.Com., DU, 1999; M. Com., MD Univ., 2001)

47. (*a*) Road testing of a random sample of cars was carried out to determine if mean mileage is greater for model I cars than for model II. The sample data for model I: $n_1 = 8$, $\bar{x}_1 = 26.0$ and $s_1 = 1.4$, model II: $n_2 = 10$, $\bar{x}_2 = 23.6$ and $s_2 = 1.2$. Specify the null hypothesis. Perform a hypothesis test at 5% level and interpret the results. *(MBA, M.D. Univ., 2006)*

(*b*) Two types of drugs were used on 5 and 7 patients for reducing their weights.

Drug A :	10	12	13	11	14		
Drug B :	8	9	12	14	15	10	9

Is there a significant difference in the efficacy of the two drugs ? If not which drug should you buy ?

(*M. Com., Madurai Kamaraj Univ., 2007*)

48. Two random samples gave the following results :

Sample	*Size*	*Sample mean*	*Sum of squares of deviations from mean*
1	10	15	90
2	12	14	108

Assuming normal population, test for the equality of population at 5% level of significance. (*MBA, IGNOU, 2002*)

49. Twelve children, each one selected from 12 sets of identical twins were trained by a certain method A and the remaining 12 children were trained by method B. At the end of the year, the following I. Q scores were obtained :

Pair :	1	2	3	4	5	6	7	8	9	10	11	12
MethodA :	124	118	127	120	135	130	140	128	140	126	130	126
MethodB :	131	127	135	128	137	131	132	125	141	118	132	129

Is this a sufficient evidence to indicate a difference in the average I. Q scores of the two groups ? (*MBA, Anna Univ., 2003*)

50. In a certain experiment to compare two types of food A and B, the following results of increase in weights are observed in subjects :

Subject→		*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*Total*
Increase in weight	Food A	49	53	51	52	47	50	52	53	407
	Food B	52	55	52	53	50	54	54	53	423

Assuming that the two samples of subjects are independent, can we conclude that Food B is better than Food A in promoting weight gain ? (*MBA, IGNOU, 2006*)

51. A vending machine is supposed to discharge 8 ounces of coffee if the correct coins are inserted. To test whether the machine is operating properly, 16 cups of coffee are taken from the machine and measured. It is found that the mean and standard deviation of 16 measurements are 7.5 and 0.8 ounces respectively. Is the machine operating properly ?

(*M. Com., Allahabad Univ., 2004*)

52. Two designs A and B gave the following output in 9 trails of each, which is a better design. Why ?

	Output								
A :	16	16	53	15	31	17	14	30	20
B :	18	27	23	21	22	26	39	17	28

(*MBA, Bharathidasan Univ., 2007*)

53. Each day the major stock markets have a group of leading gainers in price (stocks that go up the most). On one day the standard deviation in the per cent change for a sample of 12 NASDAQ leading gainers was 16.8. On the same day, the standard deviation in the per cent change for a sample of 12 NYSE leading gainers was 8.9. Conduct a significance test for equal population variances to see whether it can be concluded that there is a difference in the volatility of the leading gainers on the two exchanges. What is your conclusion at 5% leve of significance?

(*MBA, Delhi Univ., 2009*)

*****