CHAPTER - 9

TEST OF SIGNIFICANCE
(SMALL SAMPLE TESTS)

9.1 INTRODUCTION

In the earlier chapter, we considered certian tests of significance based on the theory of the normal distribution. The assumptions made in deriving those tests will be valid only for large samples. When the sample is small ($n < 30$), we can use normal distribution to test for a specified population mean or difference of two population means as in large sample tests only when the sample is drawn from a normal population whose S.D., σ is known. If σ is not known, we cannot proceed as above. If a population is normally distributed, the sampling distribution of the sample mean for any sample size is also normally distributed whether σ is known or not.

9.2 TEST OF SIGNIFICANCE FOR SMALL SAMPLES

A very important aspect of the sampling theory is the study of tests of significance, which enable us to decide on the basis of the sample results, if

(*i*) The deviation between the observed sample statistic and the hypothetical parameter value is significant.

(*ii*) The deviation between two sample statistics is significant.

The following are some important tests for small samples :

(*i*) Student's '*t*' test (*ii*) *F*-test (*iii*) χ^2-test

9.3 STUDENT'S '*t*' TEST

Let $\bar{x}$ = Mean of a sample

n = Size of the sample

σ = Standard deviation of the sample

μ = Mean of the population supposed to be normal.

Then the student's t is defined by the statistic $t = \dfrac{\bar{x}-\mu}{s/\sqrt{n-1}}$

If s^2 be the sample variance, $s^2 = \dfrac{\sum_i (x_i - \bar{x})^2}{n}$

If $S^2 = \dfrac{\sum_i (x_i - \bar{x})^2}{n-1}$, then $\dfrac{S^2}{s^2} = \dfrac{n}{n-1}$ or $S = \sqrt{\dfrac{n}{n-1}} \cdot s$

$$\therefore \quad t = \frac{\bar{x}-\mu}{S/\sqrt{n}} = \frac{\bar{x}-\mu}{s/\sqrt{n-1}}$$

S^2 is called the unbiased estimate of population variance σ^2.

Note : 1. If the standard deviation of a sample is given directly, then the statistic is given by $t = \dfrac{\bar{x} - \mu}{S.D/\sqrt{n-1}}$

2. If the calculated value of t exceeds the table value of t at 5% level of significance, the null hypothesis H_0 is rejected. If the calculated value of t is *less than* the tabulated value of t at 5% level the null hypothesis is accepted.

Assumptions for Student's t-test :

The following are made in student's t test

(1)Sample size, $n < 30$

(2)The parent population from which sample is drawn is normal.

(3)The population standard deviation is unknown.

(4)The sample observations are independent, *i.e.*, sample is random.

Uses of t-test :

This test is used

(1) to test for a specified mean

(2) to test for equality of two means of two independent samples drawn from two normal populations, S.D. of the populations being unknown, and

(3) to test the significance of difference between the means of paired data.

9.4 CONFIDENCE OR FIDUCIAL LIMITS FOR μ

Suppose we want to find from the sample data the limits within which the population mean will lie with a probability of 0.95. The limits are called the 95% confidence limits of the population mean for the given sample.

If $t_{0.05}$ is the table value of t for $(n-1)$ degrees of freedom at 5% level of significance, then 95% confidence limits for μ are given by $\bar{x} \pm t_{0.05} \cdot \dfrac{S}{\sqrt{n}}$.

For $P(|t| > t_{0.05}) = 0.05$ *i.e.*, $P(|t| \le t_{0.05}) = 0.95$

95% confidence limits for μ are given by

$$|t| \le t_{0.05} \text{ i.e., } \left|\frac{\bar{x}-\mu}{S/\sqrt{n}}\right| \le t_{0.05} \text{ or } \frac{|\bar{x}-\mu|}{S/\sqrt{n}} \le t_{0.05}$$

$$\Rightarrow \quad -t_{0.05} \le \frac{\bar{x}-\mu}{S/\sqrt{n}} \le t_{0.05}$$

$$\Rightarrow \quad \bar{x} - t_{0.05} \cdot S/\sqrt{n} \le \mu \le \bar{x} + t_{0.05} \cdot S/\sqrt{n}$$

Similarly 99% confidence limits for μ are $\bar{x} \pm t_{0.01} \cdot S/\sqrt{n}$ where $t_{0.01}$ is the tabulated value of t for $(n-1)$ degrees of freedom at 1% level of significance.

9.5 STUDENT'S 't' TEST FOR SINGLE MEAN

Suppose we want to test

(*a*) if a random sample x_i of size n has been drawn from a normal population with a specified mean μ.

(*b*) if the sample mean differs significantly from the hypothetical value μ of the population mean.

In this case the statistic is given by $t = \dfrac{\bar{x}-\mu}{S/\sqrt{n}} \sim t_{(n-1)}$ where $\bar{x}$, μ, S, n have usual meanings.

Let a random sample of size $n\,(n<30)$ has a sample mean $\bar{x}$. To test the hypothesis that the population mean μ has a specified value μ_0 when population S. D. σ is not known.

Let the Null Hypothesis be $H_0 : \mu = \mu_0$

Then the Alternative Hypothesis is $H_1 : \mu \neq \mu_0$.

Assuming that H_0 is true, the test statistic given by $t = \dfrac{\bar{x}-\mu}{s/\sqrt{n-1}}$, where s is the sample S. D. follows t - distribution with $\nu = (n-1)$ d.f.

We calculate the value of $|t|$ and compare this value with the table value of t at α level of significance. If the calculated value of $t >$ the table value of t, we reject H_0 at α level. Otherwise we accept H_0.

In this case, 95% confidence limits for the population mean μ are $\bar{x} \pm t_\alpha \cdot \dfrac{s}{\sqrt{n-1}}$ where $\alpha = 0.025$ for two-tailed test and s = sample S. D. and 99% confidence limits for μ are $\bar{x} \pm t_\alpha \cdot \dfrac{s}{\sqrt{n-1}}$ where $\alpha = 0.05$.

For a two-tailed test at α level of significance, value of $\alpha/2$ is taken for α.

SOLVED EXAMPLES

Example 1 : A mechanist is making engine parts with axle diameters of 0.700 inch. A random sample of 10 parts shows a mean diameter of 0.742 inch with a S.D. of 0.040 inch. Compute the statistic you would use to test whether the work is meeting the specification at 0.05 level of significance.

[JNTU (A) Apr. 2012 (Set No. 2)]

Solution : Here the sample size $n = 10 < 30$

Hence the sample is small sample.

Also sample mean $\bar{x} = 0.742$ inches, the population mean $\mu = 0.700$ inches and $S.D. = 0.040$ inches are given.

$\therefore$ We use student's t-Test

1. **Null Hypothesis H_0 :** The product is confirming to specification.
2. **Alternative Hypothesis H_1 :** $\mu \neq 0.700$
3. **Level of significance,** $\alpha = 0.05$
4. **The test statistic is,** $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n-1}}$

Here $\bar{x} = 0.742$ inches, $\mu = 0.700$ inches, $S.D. = 0.040$ inches and $n = 10$.

Degrees of freedom (d.f) $= n - 1 = 10 - 1 = 9$

$$\therefore \quad t = \frac{0.742 - 0.700}{\dfrac{0.040}{\sqrt{10-1}}} = 3.15$$

Thus the calculated value of $t = 3.15$

The tabulated value of t at 5% level with 9 degrees of freedom is $t_{0.05} = 2.26$

Since calculated value of $t >$ tabulated value of t, therefore, H_0 is rejected.

$\therefore$ The product is not meeting the specification.

Example 2 : 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.

[JNTU (A) Nov. 2010 (Set No. 2,4)]

Solution : Here sample size, $n = 26 < 30$

$\therefore$ The sample is small sample.

Also given, sample mean, $\bar{x} = 990$

Population mean, $\mu = 1000$ and $S.D., s = 20$

Degrees of freedom $= n - 1 = 26 - 1 = 25$

Here we know $\bar{x}$, μ, $S.D.$ and n.

So we use students 't' test.

1. **Null Hypothesis H_0 :** The sample is upto the standard.
2. **Alternative Hypothesis H_1 :** $\mu' < 1000$

   (The sample is below standard) (left-tail test)

3. **Level of significance :** $\alpha = 0.05$
4. **The test statistic is** $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n-1}} = \dfrac{990 - 1000}{20/\sqrt{25}} = -2.5$

$\therefore \ |t| = 2.5$

i.e., Calculated value of $t = 2.5$

Tabulated value of 't' at 5% level with 25 degrees of freedom for left-tailed test is 1.708.

Since calculated $t >$ tabulated t, we reject the null hypothesis H_0 and conclude that the sample is not upto the standard.

Example 3 : A machine is designed to produce insulating washers for electrical devices of average thickness of 0.025 cm. A random sample of 10 washers was found to have a thickness of 0.024 cm with a S.D of 0.002 cm. Test the significance of the deviation. Value of t for 9 degrees of freedom at 5% level is 2.262.

Solution : Here the sample size is $10 < 30$

$\therefore$ The sample is small

Also given Sample mean, $\bar{x} = 0.024$ cm

Population mean , $\mu = 0.025$ cm

S.D. = 0.002 cm

Degrees of freedom (d.f) $= n - 1 = 10 - 1 = 9$

1. **Null Hypothesis H_0 :** The difference between $\bar{x}$ and μ is not significant.
2. **Alternative Hypothesis H_1 :** $\mu_1 \neq 0.025$
3. **Level of significance :** $\alpha = 0.05$
4. **The test statistic is** 't' $= \dfrac{\bar{x} - \mu}{s/\sqrt{n-1}} = \dfrac{0.024 - 0.025}{\dfrac{0.002}{\sqrt{10-1}}} = -1.5$

$\Rightarrow |t| = 1.5$

$\therefore$ Calculated value of $t = 1.5$ for two tailed test.

Tabulated value of t for 9 degrees of freedom at 5% level = 2.262

Since calculated $t <$ tabulated t, we accept the null hypothesis and conclude that the difference between $\bar{x}$ and μ is not significant.

Example 4 : The mean life time of a sample of 25 fluorescent light bulbs produced by a company is computed to be 157 hours with a S.D of 120 hours. The company claims that the average life of the bulbs produced by the company is 1600 hours using the level of significance of 0.05. Is the claim acceptable ?

Solution : Given sample size, $n = 25$

Sample mean, $\bar{x} = 1570$

Population mean, $\mu = 1600$

$S.D.(s) = 120$

Degrees of freedom $= n - 1 = 24$

1. **Null Hypothesis H_0 :** The claim is acceptable. $\mu = 1600$ hrs
2. **Alternative Hypothesis H_1 :** $\mu \neq 1600$ hrs
3. **Level of significance :** $\alpha = 0.05$
4. **The test statistic is** $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n-1}} = \dfrac{1570 - 1600}{120/\sqrt{24}} = \dfrac{-30}{24.49} = -1.22$

$\therefore \quad |t| = 1.22$

i.e., Calculated $t = 1.22$

The tabulated value of t at 5% level with 24 degrees of freedom for two tailed test is 2.06.

Since the calculated value of t < tabulated value of t, we accept the null hypothesis H_0 *i.e.*, the claim that the average life of the bulbs produced by the company is 1600 hrs is acceptable.

Example 5 : The average breaking strength of the steel rods is specified to be 18.5 thousand pounds. To test this sample of 14 rods were tested. The mean and standard deviations obtained were 17.85 and 1.955 respectively. Is the result of experiment significant ?

[JNTU (H) May 2011 (Set No. 3), (K) Nov. 2011 (Set No. 1)]

Solution : We are given

Sample size, $n = 14$

Sample mean, $\bar{x} = 17.85$

S.D. $(s) = 1.955$

Population mean, $\mu = 18.5$

Degrees of freedom $= n - 1 = 13$

1. **Null Hypothesis H_0 :** The result of the experiment is not significant.
2. **Alternative Hypothesis H_1 :** $\mu \neq 18.5$
3. **Level of significance :** $\alpha = 0.05$
4. **The test statistic** is, $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n-1}} = \dfrac{17.85 - 18.5}{1.955/\sqrt{13}} = \dfrac{0.65}{0.542} = -1.199$

$\therefore |t| = 1.199$

i.e., Calculated $t = 1.199$

Tabulated t at 5% level of significance for 13 d.f. for two tailed test $= 2.16$

Since calculated t < tabulated t, we accept the Null Hypothesis H_0 at 5% level and conclude that the result of the experiment is not significant.

Example 6 : A random sample of size 16 values from a normal population showed a mean of 53 and a sum of squares of deviations 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.

[JNTU (K) May 2012 (Set No.2)]

Solution : Given sample size, $n = 16$

Sample mean, $\bar{x} = 53$

Now $\sum (x_i - \bar{x})^2 = 150$

$$\therefore S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{150}{15} = 10 \Rightarrow S = \sqrt{10}$$

Degrees of freedom, $\nu = n - 1 = 16 - 1 = 15$

(*a*) 1. **Null Hypothesis H_0 :** The sample is taken from the population having 56 as mean *i.e.*, $\mu = 56$

2. **Alternative hypothesis H_1 :** $\mu \neq 56$

3. **Level of significance :** $\alpha = 0.05$

4. **The test statistic,** $t = \dfrac{\bar{x} - \mu}{S/\sqrt{n}}$ [Note that *S.D.* is not given directly]

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

$\therefore \ |t| = 3.79$

The tabulated value of t at 5% level of significance for 15 d.f for two tailed test is 2.13.

Since calculated t > tabulated t, the null hypothesis H_0 is rejected *i.e.*, the sample cannot be regarded as taken from the population.

(*b*) The 95% confidence limits of the mean of the population are given by

$$\bar{x} \pm t_{0.05} \frac{S}{\sqrt{n}} = 53 \pm 2.13 \times 0.79 = 53 \pm 1.6827$$

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

Hence 95% confidence limits are [51.31, 54.68]

Example 7 : A random sample of six steel beams has a mean compressive strength of 58,392 p.s.i (pounds per square inch) with a standard deviation of 648 p.s.i. Use this information and the level of significance $\alpha = 0.05$ to test whether the true average compressive strength of the steel from which this sample came is 58,000 p.s.i. Assume normality.

[JNTU 2005S, 2006S, (H) Nov. 2010, (H) May & Dec. 2011 (Set No. 2), (K) May 2013 (Set No. 1)]

Solution : We have

n = Sample size (number of steel beams) = 6 < 30. $\therefore$ The sample is small.

$\bar{x}$ = Sample mean (average compressive strength) = 58392 psi.

S = Standard deviation of six beams = 648 psi

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

In this problem σ is known and $n < 30$. Hence we use t - distribution.

1. **Null Hypothesis is H_0 :** $\mu = 58000$

2. **Alternative Hypothesis H_1 :** $\mu \neq 58000$

3. **Level of Significance :** $\alpha = 0.05$

4. **Critical Region :** Since Alternative Hypothesis is of the type ' $\neq$ ', the test is two-tailed and the critical region is $-3.365 < t < 3.365$

5. **The Test Statistic** is $t = \dfrac{\bar{x}-\mu}{s/\sqrt{n-1}} = \dfrac{58392-58000}{648/\sqrt{5}} = 1.353$

Since $t = 1.353 < 3.365 = t_{\alpha/2}$, we accept the null hypothesis H_0.

Hence the average compressive strength of the steel beam is not equal to 58000 psi.

Example 8 : A sample of 100 iron bars is said to be drawn from a large number of bars whose lengths are normally distributed with mean 4 feet and S.D. 6 ft. If the sample mean is 4.2 feet, Can the sample be regarded as a truly random sample?

[JNTU (H) Nov. 2010 (Set No. 2)]

Solution : Let the Null Hypothesis be H_0 : The sample can be regarded as a truly random sample.

Then the Alternative Hypothesis is H_1 : The sample cannot be regarded as a random sample.

Here $n =$ Sample size $= 100$

$\bar{x} =$ Sample mean $= 4.2$

$\mu =$ Population mean $= 4$

$\sigma =$ S.D. of population $= 0.6$

Let us assume that H_0 is true. The test statistic is

$$t = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{4.2-4}{0.6/10} = \frac{2}{0.6} = \frac{10}{3} = 3.33$$

$\therefore\ t = 3.33 > 3$

Hence we reject the Null Hypothesis and conclude that the given sample cannot be regarded as a truly random sample from the given population.

Note : Since $|z| > 3$ and level of significance is not given, we may reject H_0 at 5% or 1% level of significance. In the above problem, the Null Hypothesis is rejected without mentioning any level of significance.

Example 9 : A sample of 155 members has a mean 67 and S.D.5.2. Is this sample has been taken from a large population of mean 70? **[JNTU (H) Nov. 2010 (Set No. 3)]**

Solution : Let the Null Hypothesis be H_0 : The sample has been taken from population

Then the Alternative Hypothesis is H_1 : The sample has not been taken from population.

Here $n =$ Sample size $= 155$

$\bar{x}$ = Mean of the sample = 67

σ = S.D. of sample = 5.2

μ = Mean of the population = 70

$\therefore$ The test statistic is $t = \dfrac{\bar{x}-\mu}{S.D/\sqrt{n-1}} = \dfrac{67-70}{5.2/\sqrt{155-1}}$

$$= \frac{-3}{5.2/\sqrt{154}} = -7.16$$

$\therefore |t| = 7.16 > 3$

We reject the Null Hypothesis at 5% level of significance and conclude that the sample has not been taken from the given population.

Example 10 : In the past a machine has produced washers having a mean thickness of 0.050 inch. To determine whether the machine is in proper working order a sample of 10 washers is chosen for which the mean thickness is 0.053 inch and the standard deviation is 0.003 inch. Test the hypothesis that the machine is in proper working order using a level of significance of 0.05?

Solution : We have $\mu = 0.050, n = 10, \bar{x} = 0.053, s = 0.003$, where s is the sample S.D.

Since the sample size is small and population S.D σ is not known, we use t - test.

Let the Null Hypothesis be $H_0 : \mu = 0.050$

Then the Alternative Hypothesis is $H_1 : \mu \neq 0.050$

The test statistic is $t = \dfrac{\bar{x}-\mu}{s/\sqrt{n-1}} = \dfrac{0.053-0.050}{0.003/\sqrt{10-1}} = \dfrac{0.003}{0.001} = 3$

Degrees of freedom $= n-1 = 10-1 = 9$.

The tabulated value of t for 9 d.f. at 5% level = 2.262

Since the calculated value of $t >$ the tabulated value, we reject the Null Hypothesis at 5% level and conclude that the machine is not in proper working order.

Example 11 : The following are the times between six calls for an ambulance in a city and the patient's arrival at the hospital : 27,15,20,32,18 and 26 minutes. Use these figures to judge the reasonableness of the ambulance services claim that it takes on the average 20 minutes between the call for an ambulance and patient's arrival at the hospital?

[JNTU (A) Nov. 2011 (Set No. 1)]

Solution : We have n = Sample size = 6

$\bar{x}$ = mean of the sample $= \dfrac{27+15+20+32+18+26}{6} = \dfrac{138}{6} = 23$

μ = mean of the population = 20

and

$$S^2 = \text{sample variance} = \frac{\Sigma(x_i - \bar{x})^2}{n-1}$$

$$= \frac{1}{5}[(27-23)^2 + (15-23)^2 + (20-23)^2 + (32-23)^2 + (18-23)^2 + (26-23)^2]$$

$$= \frac{1}{5}(16+64+9+81+25+9) = \frac{204}{5} = 40.8$$

$\therefore$ S = Sample S. D. $= \sqrt{40.8} = 6.4$

Since the sample size is small and population S. D. σ is not known, we use t - test.

Let the Null Hypothesis be $H_0 : \mu = 20$

Then the Alternative Hypothesis is $H_1 : \mu \neq 20$

The test statistic is

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}} = \frac{23-20}{6.4/\sqrt{6}} = \frac{3\sqrt{6}}{6.4} = \frac{7.348}{6.4} = 1.148$$

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

The tabulated value of t for 5 d.f. at 5% level = 2.571

[$\because$ for two -tail test, $\alpha/2$ is taken for α]

$\therefore |t| = 1.148 < 2.571$

Since the calculated value of $|t| <$ the tabulated value, we accept the Null Hypothesis at 5% level and conclude that the average time between the call for an ambulance and patient's arrival at the hospital is signifiant.

Problems related to Student's t Test (When S.D of the sample is not given directly)

SOLVED EXAMPLES

Example 1 : A random sample of 10 boys had the following I.Q's : 70, 120, 110, 101, 88, 83, 95, 98, 107 and 100.

(*a*) Do these data support the assumption of a population mean I.Q of 100 ?

(*b*) Find a reasonable range in which most of the mean I.Q. values of samples of 10 boys lie. **[JNTU 2008, (A) Apr. 2012, (H) May 2011 (Set No. 4), (K) May 2013 (Set No. 2)]**

Solution : (*a*) Here *S.D.* and mean of the sample is not given directly.

We have to determine these *S.D.* and mean as follows.

$$\text{Mean, } \bar{x} = \frac{\Sigma x}{n} = \frac{972}{10} = 97.2$$

x	$x - \bar{x}$	$(x - \bar{x})^2$
70	−27.2	739.84
120	22.8	519.84
110	12.8	163.84
101	3.8	14.44
88	−9.2	84.64
83	−14.2	201.64
95	−2.2	4.84
98	0.8	0.64
107	9.8	96.04
100	2.8	7.84
972		1833.60

We know that $S^2 = \frac{1}{n-1}\sum_i (x_i - \bar{x})^2 = \frac{1833.60}{9}$

$\therefore$ Standard deviation, $S = \sqrt{203.73} = 14.27$

1. **Null Hypothesis H_0 :** The data support the assumption of a population mean I.Q of 100 in the population.
2. **Alternative Hypothesis H_1 :** $\mu \neq 100$
3. **Level of significance,** $\alpha = 0.05$
4. **The test statistic** is $t = \frac{\bar{x} - \mu}{S/\sqrt{n}} = \frac{97.2 - 100}{14.27/\sqrt{10}} = -0.62$

$\therefore |t| = 0.62$ *i.e.*, Calculated value of $t = 0.62$

Tabulated value of t for (10 – 1) d.f. *i.e.*, 9 d.f. at 5% level of significance is 2.26 (two-tailed list).

Since calculated value of $t <$ tabulated value of t, we accept the null hypothesis H_0. *i.e.*, the data support the assumption of mean I.Q of 100 in the population.

(*b*) The 95% confidence limits are given by $\bar{x} \pm t_{0.05} \cdot S/\sqrt{n}$

$= 97.2 \pm 2.26 \times 4.512 = 97.2 \pm 10.198 = 107.4$ and 87

$\therefore$ The 95% confidence limits within which the mean I.Q values of samples of 10 boys will lie is (87, 107.40).

Example 2 : The heights of 10 males of a given locality are found to be 70, 67, 62, 68, 61, 68, 70,64,64,66 inches. Is it reasonable to believe that the average height is greater than 64 inches? Test at 5% significance level assuming that for 9 degrees of freedom ($t = 1.833$ at $\alpha = 0.05$).

[JNTU (H) May 2011 (Set No. 2)]

Solution : Mean, $\bar{x} = \frac{\sum x}{n} = \frac{660}{10} = 66$

Calculation for sample mean and S.D.

x	$x - \bar{x}$	$(x - \bar{x})^2$
70	4	16
67	1	1
62	−4	16
68	2	4
61	−5	25
68	2	4
70	4	16
64	−2	4
64	−2	4
66	0	0
660		90

We know that $S^2 = \frac{1}{n-1} \sum_i (x_i - \bar{x})^2 = \frac{1}{9} \times 90 = 10$

$\therefore$ Sample $S.D.(S) = \sqrt{10} = 3.16$

1. **Null Hypothesis H_0 :** The average height is not greater than 64 inches *i.e.*, $\mu = 64$ inches.
2. **Alternative Hypothesis H_1 :** $\mu > 64$ inches
3. **Level of significance :** $\alpha = 0.05$
4. **Test statistic** is $t = \frac{\bar{x} - \mu}{S/\sqrt{n-1}} = \frac{66 - 64}{3.16/\sqrt{9}} = \frac{2}{1.05} = 1.9$

i.e., Calculated value $t = 1.9$

Tabulated t for $10 - 1 = 9$ d.f at 5% level of significance for single tail test is 1.833 (given).

Since calculated value of $t >$ tabulated value of t, we reject the null hypothesis H_0. *i.e.*, The average height is greater than 64 inches.

Example 3 : A random sample from a company's very extensive files shows that the orders for a certain kind of machinery were filled, respectively in 10, 12, 19, 14, 15, 18, 11 and 13 days. Use the level of significance $\alpha = 0.01$ to test the claim that on the average such orders are filled in 10.5 days. Choose the Alternative Hypothesis so that rejection of Null Hypothesis $\mu = 10.5$ days implies that it takes longer than indicated.

[JNTU 2004S, (K) May 2010S (Set No. 3)]

Solution : We have

$$n = 8, \ \bar{x} = \frac{1}{8}(10 + 12 + 19 + 14 + 15 + 18 + 11 + 13) = \frac{112}{8} = 14$$

and sample S.D, S is given by

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

$$= \frac{1}{7}[(10-14)^2 + (12-14)^2 + (19-14)^2 + \ldots + (13-14)^2]$$

$$= \frac{1}{7}(16 + 4 + 25 + 0 + 1 + 16 + 9 + 1)$$

$$= \frac{1}{7}(72) = 10.286$$

$$\therefore S = \sqrt{10.286} = 3.207$$

1. **Null Hypothesis H_0 :** $\mu = 10.5$ days
2. **Alternative Hypothesis H_1 :** $\mu > 10.5$ days
3. **Level of significance :** $\alpha = 0.01$
4. **Critical region :** $t > t_{0.01}$

   Reject H_0 if $t > t_{0.01} = 2.998$

5. **The Test Statistic** is $t = \dfrac{\bar{x} - \mu}{S/\sqrt{n-1}}$

$$= \frac{14 - 10.5}{3.207/\sqrt{7}} = 3.087$$

Tabulated value of t for $(8 - 1)$ d.f. at 1% level of significance is 2.998 (two tailed test).

Since calculated value of $t >$ tabulated value of t, we reject the null hypothesis H_0. *i.e.*, The orders on average are filled in more than 10.5 days.

Example 4 : The life time of electric bulbs for a random sample of 10 from a large consignment gave the following data.

Item :	1	2	3	4	5	6	7	8	9	10
Life in 1000 hrs :	1.2	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 life time of bulbs is 4000 hrs. Use a 0.05 level of significance.

[JNTU Nov. 2008 (Set No. 2)]

Solution : Here $n = 10$.

1. **Null Hypothesis** $H_0 : \mu = 4000$
2. **Alterntive Hypothesis** $H_1 : \mu \neq 4000$
3. **Level of significance :** $\alpha = 0.05$
4. **The test statistic** is $t = \dfrac{\bar{x} - \mu}{S/\sqrt{n}}$

From given data, $\bar{x}$ = average life time of bulbs

$$= \frac{1.2+4.6+3.9+4.1+5.2+3.8+3.9+4.3+4.4+5.6}{10}$$

$$= \frac{1}{10}(41) = 4.1 \simeq 4100$$

$$\text{Now } S^2 = \sum_{i=1}^{10} \frac{1}{n-1}(x_i - \bar{x})^2$$

$$= \frac{1}{9}\left[\begin{array}{l}(1.2-4.1)^2 + (4.6-4.1)^2 + (3.9-4.1)^2 + (4.1-4.1)^2 + (5.2-4.1)^2 \\ + (3.8-4.1)^2 + (3.9-4.1)^2 + (4.3-4.1)^2 + (4.4-4.1)^2 + (5.6-4.1)^2\end{array}\right]$$

$$= \frac{1}{9}(8.41 + 0.25 + 0.04 + 1.21 + 0.09 + 0.04 + 0.03 + 0.09 + 2.25)$$

$$= \frac{12.42}{9} = 1.38 \simeq 1380 \Rightarrow S = \sqrt{1380}$$

$$\text{Hence } t = \frac{\bar{x}-\mu}{S/\sqrt{n}} = \frac{4100-4000}{\sqrt{1380}/\sqrt{10}} = 8.512.$$

$\therefore |t| = 8.512$

No. of degrees of freedom = 9

$\therefore t_{\alpha/2} = 2.262$ (From tables)

Thus we see that $|t| > t_{\alpha/2}$.

Hence, we reject the Null Hypothesis H_0 at 5% level and conclude that the average life time bulbs is not equal to 4000 hrs.

Example 5 : The manufacturer of a certain make of electric bulbs claims that his bulbs have a mean life of 25 months with a S. D. of 5 months. A random sample of 6 such bulbs gave the following values. Life of months : 24, 26, 30, 20, 20, 18. Can you regard the producers, claims to be valid at 1% level of significance ? **[JNTU (A) 2009 (Set No. 2)]**

Solution : We have $n = 6$,

$$\text{Sample mean, } \bar{x} = \frac{24+26+30+20+20+18}{6} = \frac{138}{6} = 23$$

and sample S.D, S is given by $S^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1}$

$$i.e., S^2 = \frac{1}{5}[(24-23)^2 + (26-23)^2 + (30-23)^2 + (20-23)^2 + (20-23)^2 + (18-23)^2]$$

$$= \frac{1}{5}[1+9+49+9+9+25] = \frac{102}{5} = 20.4$$

$\therefore S = \sqrt{20.4} = 4.52$

1. Null Hypothesis $H_0 : \mu = 25$

2. Alternative Hypothesis $H_1 : \mu \neq 25$

3. Level of significance : $\alpha = 0.01$

4. The test statistic is $t = \dfrac{\bar{x} - \mu}{S/\sqrt{n}}$

$$i.e., \; t = \frac{23-25}{4.52/\sqrt{6}} = \frac{-2\sqrt{6}}{4.52} = -1.08$$

$\therefore \; |t| = 1.08$

i.e., calculated value of $t = 1.08$

Tabulated value of t for $(6-1)$ degrees of freedom at 1% level of significance is 3.365.

Since calculated value of $t <$ tabulated value of t, we accept the null hypothesis H_0 and conclude that the claim is valid.

Example 6 : Producer of 'gutkha', claims that the nicotine content in his 'gutkha' on the average is 1.83mg. Can this claim accepted if a random sample of 8 'gutkha' of this type have the nicotine contents of 2.0, 1.7, 2.1, 1.9, 2.2, 2.1, 2.0, 1.6mg? Use a 0.05 level of significance. **[JNTU (K) 2009, (A) Apr. 2012 (Set No. 1)]**

Solution : Given $n = 8$ and $\mu = 1.83$ mg

1. Null Hypothesis $H_0 : \mu = 1.83$

2. Alternative Hypothesis $H_1 : \mu \neq 1.83$

3. Level of significance : $\alpha = 0.05$

4. The test statistic is $t = \dfrac{\bar{x} - \mu}{S/\sqrt{n}} \sim t_{(n-1)}$

where $\bar{x}$ and S^2 are to be computed from the sample values gutkha.

	x	$x - \bar{x}$	$(x-\bar{x})^2$
	2.0	0.05	0.0025
	1.7	-0.25	0.0625
	2.1	0.15	0.0225
	1.9	-0.05	0.0025
	2.2	0.25	0.0625
	2.1	0.15	0.0225
	2.0	0.05	0.0025
	1.6	-0.35	0.1225
Total	15.6	0.3	

Now $\bar{x} = \frac{15.6}{8} = 1.95$ and $S^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} = \frac{0.3}{7}$ or $S = 0.21$

$$\therefore \quad t = \frac{1.95 - 1.83}{0.21/\sqrt{8}} = \frac{(0.12)(\sqrt{8})}{0.21} = 1.62$$

Tabulated $t_{0.05}$ for $(8-1)$ *i.e.*, 7d.f is 1.895.

Since calculated $t <$ tabulated $t_{0.05}$, the null hypothesis H_0 may be accepted at 5% level of significance and we may conclude that the data are consistent with the assumption of gutkha on the average is 1.83mg.

Example 7 : Eight students were given a test in STATISTICS and after one month coaching they were given another test of the similar nature. The following table gives the increase I their marks in the second test over the first.

Student No.	1	2	3	4	5	6	7	8
Increase of Marks	4	−2	6	−8	12	5	−7	2

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

[JNTU 2008, (A) Nov. 2011 (Set No. 2)]

Solution : We shall compute the mean and S.D. of the increase of marks as follows.

$$\bar{x} = \frac{\sum x_i}{8} = \frac{4-2+6-8+12+5-7+2}{8} = \frac{12}{8} = 1.5$$

We have $S^2 = \frac{1}{n-1}\sum_i (x_i - \bar{x})^2$

$$= \frac{1}{7}[(4-1.5)^2 + (-2-1.5)^2 + (6-1.5)^2 + (-8-1.5)^2 + (12-1.5)^2 + (5-1.5)^2 + (-7-1.5)^2 + (2-1.5)^2]$$

$$= \frac{1}{7}[6.25 + 12.25 + 20.25 + 90.25 + 110.25 + 12.25 + 72.25 + 0.25]$$

$$= \frac{324}{7} = 46.2857$$

$$\therefore \quad \text{S.D.} = S = \sqrt{46.2857} = 6.8$$

Assuming that the students have not been benefitted by coaching, it implies that the mean of the difference of the two tests is zero i.e., $\mu = 0$.

$$\text{Then } t = \frac{\bar{x} - \mu}{S/\sqrt{n-1}} = \frac{1.5 - 0}{(6.8)/\sqrt{7}} = 0.5836$$

No. of degrees of freedom $= 8 - 1 = 7$

$\therefore$ Tabulated $t_{0.05} = 2.36$

As the calculated value of $t < t_{0.05}$, the value of t is not significant at 5% level significance i.e., the test provides no evidence tht the students have been benefitted by coaching.

Example 8 : A random sample of 10 bags of pesticide are taken whose weights are 50, 49, 52, 44, 45, 48, 46, 45, 49, 45 (in kgs). Test whether the average packing can be taken to be 50 kgs. **[JNTU (H) Apr. 2012 (Set No. 1)]**

Solution : We have $\mu = 50, n = 10$

Let the Null Hypothesis be $H_0 : \mu = 50$

Then the Alternative Hypothesis is $H_1 : \mu \neq 50$.

The test statistic is $t = \dfrac{\bar{x} - \mu}{s / \sqrt{n-1}}$

Now $\bar{x} = \dfrac{1}{10}(50 + 49 + 51 + 44 + 45 + 48 + 46 + 45 + 49 + 45) = \dfrac{472}{10} = 47.2$

Variance, $s^2 = \dfrac{\Sigma(x_i - \bar{x})^2}{n-1}$

$$= \frac{1}{9}[(50-47.2)^2 + (49-47.2)^2 + (51-47.2)^2 + (44-47.2)^2 +$$

$$(45-47.2)^2 + (48-47.2)^2 + (46-47.2)^2$$

$$+ (45-47.2)^2 + (49-47.2)^2 + (45-47.2)^2]$$

$$= \frac{1}{9}[7.84 + 3.24 + 14.44 + 10.24 + 4.84$$

$$+ 0.64 + 1.44 + 4.84 + 3.24 + 4.84] = \frac{55.6}{9} = 6.1777$$

$\therefore$ Standard deviation, $s = \sqrt{6.18} = 2.48$

Hence $t = \dfrac{47.2 - 50}{2.48/3} = \dfrac{-8.4}{2.48} = -3.39$

Degrees of freedom $= n - 1 = 10 - 1 = 9$.

The tabulated value of t for 9 d.f. at 5% level $= 2.262$

$\therefore |t| = 3.39 > 2.262$ [$\because$ for two-tailed test, $\alpha/2$ is taken for α]

Example 9 : A stimulus administered to 12 patients resulted in the following changes in the blood pressure : $5, 2, 8, -1, 3, 0, -2, 1, 5, 0, 4, 6$. Can it be concluded that the stimulus will in general be accompanied by an increase in blood pressure? **[JNTU(K) May.2012 (Set No.1)]**

Solution : Let x denote the increment in blood pressure. We shall compute the mean and standard deviation of blood pressure as follows:

$$\bar{x} = \frac{\sum x_i}{12} = \frac{1}{12}(5 + 2 + 8 - 1 + 3 + 0 - 2 + 1 + 5 + 0 + 4 + 6) = \frac{31}{12} = 2.58$$

Also $S^2 = \frac{1}{n^2}\sum_i (x_i - \bar{x})^2$

$$= \frac{1}{144}[(5-2.58)^2 + (2-2.58)^2 + (8-2.58)^2 + (-1-2.58)^2 + (3-2.58)^2 + (0-2.58)^2 + (-2-2.58)^2 + (1-2.58)^2 + (5-2.58)^2 + (0-2.58)^2 + (4-2.58)^2 + (6-2.58)^2]$$

$$= 9.53$$

$\therefore S = 3.08$.

1. **Null Hypothesis:** $H_0 : \mu_1 = \mu_2$.
2. **Alternative Hypothesis:** $\mu_1 \neq \mu_2$.
3. **Level of significance:** $\alpha = 0.05$.
4. The test statistic is $t = \frac{\bar{x} - \mu}{S/\sqrt{n}} = \frac{2.58 - 0}{3.08/\sqrt{12}} = \frac{2.58}{0.89} = 2.89$

Tabulated $t_{0.05}$ for $(12-1)$ degrees of freedom $= 2.2$

Since calculated t > tabulated t, the Null hypothesis is rejected at 5% level of significance and we may conclude that the stimulus will in general be accompanied by an increase in blood pressure.

Example 10 : A new process of producing synthetic diamonds can be operated at a profitable level only if the average weight of the diamonds is greater than 0.5 carat. To test the profitability of the process, 6 diamonds are produced with weights 0.46,0.60,0.52,0.49,0.58 and 0.54 carat respectively. Do the 6 measurements present sufficient evidence to indicate that the average weight of the diamonds produced by the process is in excess of 0.5 carat ?

[JNTU(A) Dec. 2014]

Solution : Here $\bar{x} = \frac{1}{6}(0.45 + 0.60 + 0.52 + 0.49 + 0.58 + 0.54) = \frac{3.18}{6} = 0.53$

Calculation for Sample Mean and S. D.

x	$x - \bar{x}$	$(x - \bar{x})^2$
0.45	-0.08	0.0064
0.60	0.07	0.0049
0.52	-0.01	0.0001
0.49	-0.04	0.0016
0.58	0.05	0.0025
0.54	0.01	0.0001
3.18		0.0156

$$\therefore S^2 = \frac{1}{n-1}\Sigma(x_i - \bar{x})^2 = \frac{1}{5} \times 0.0156 = 0.00312$$

$$\Rightarrow S = \sqrt{0.00312} = 0.056$$

1. **Null Hypothesis** H_0 : The average weight of the diamond is not greater than 0.5 carat *i.e.* $\mu = 0.5$
2. **Alternative Hypothesis** $H_1 : \mu > 0.5$
3. **Level of significance :** $\alpha = 0.05$
4. **Test statistic** is $t = \dfrac{\bar{x} - \mu}{S/\sqrt{n}}$

   i.e. $t = \dfrac{0.53 - 0.5}{0.056/\sqrt{6}} = \dfrac{(0.03)\sqrt{6}}{0.056} = 1.3122$
5. **Critical Region :**

   Tabulated value of t for $(6-1)$ d.f. at 5% level of significance for one tail test is 2.015
6. **Decision :** Since calculated value of t < tabulated value of t, we accept the null hypothesis H_0. *i.e.*, the average weight of the diamond is not greater than 0.5 carat.

The calculated t does not fall within the rejection region. Therefore, we cannot reject H_0. Hence the data do not give sufficient evidence to indicate that the mean diamond weight exceeds 0.5 carat.

Example 11 : A random sample of 8 envelopes is taken from the letter box of a post office and their weights in grams are found to be: 12.1, 11.9, 12.4, 12.3, 11.5, 11.6, 12.1 and 12.4.

(*i*) Does this sample indicate at 1% level that the average weight of envelopes received at their post office is 12.35 grms.

(*ii*) Find 95% confidence limits for the mean weight of the envelopes received at that post office.

Solution :

(*i*)
1. **Null Hypothesis** is $H_0 : \mu = 12.35$ grms *i.e.*, the samples have been drawn from the same population of mean 12.35 grms.
2. **Alternative Hypothesis** is $H_1 : \mu \neq 12.35$ grms
3. **Level of significance :** $\alpha = 0.01$
4. **The test statistic** is $t = \dfrac{\bar{x} - \mu}{S/\sqrt{n}}$

Here $n = 8$.

Now, sample mean,

$$\bar{x} = \frac{1}{8}(12.1+11.9+12.4+12.3+11.5+11.6+12.1+12.4) = \frac{96.3}{8} = 12.0375.$$

Sample variance, $S^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}$

$$= \frac{1}{7}[(12.1-12.0375)^2 + (11.9-12.0375)^2 + (12.4-12.0375)^2 + (12.3-12.0375)^2 +$$

$$(11.5-12.0375)^2 + (11.6-12.0375)^2 + (12.1-12.0375)^2 + (12.4-12.0375)^2]$$

$$= \frac{1}{7}[0.0039+0.0189+0.1314+0.0689+0.2889+0.1914+0.0039+0.1314]$$

$$= \frac{0.8387}{7} = 0.1198$$

$$\therefore S = \sqrt{0.1198} = 0.3461$$

Thus $t = \frac{12.0375-12.35}{0.3461/\sqrt{8}} = \frac{2.8284(-0.3125)}{0.3461} = \frac{-0.883875}{0.3461} = -2.5538$

$\therefore |t| = 2.5538$ *i.e.*, calculated value of $t = 2.5538$

Tabulated value of t for $(8-1)$ degrees of freedom at $\frac{\alpha}{2} = \frac{1}{2}\%$ level of significance is

i.e., $t_{\alpha/2} = t_{0.005} = 3.499$

Since calculated value of $t <$ tabulated value of t, we accept the Null Hypothesis H_0 and conclude that the sample is taken from the population whose mean is 12.35.

The 95% confidence limits are given by $\bar{x} \pm t_{\alpha/2} \cdot \frac{S}{\sqrt{n}}$

i.e., $\left(\bar{x} - t_{\alpha/2} \cdot \frac{S}{\sqrt{n}},\ \bar{x} + t_{\alpha/2} \cdot \frac{S}{\sqrt{n}}\right)$

Now $t_{\alpha/2} = t_{0.025} = 2.365$ [From Tables]

$$\therefore t_{\alpha/2} \cdot \frac{S}{\sqrt{n}} = \frac{(2.365)(0.3461)}{\sqrt{8}} = \frac{0.8185}{2.8284} = 0.2894$$

Hence confidence limits $= (12.0375 - 0.2894, 12.0375 + 0.2894)$

$= (11.7481, 12.3269)$.

9.6 STUDENT'S t - TEST FOR DIFFERENCE OF MEANS

Let $\bar{x}$ and $\bar{y}$ be the means of two independent samples of sizes n_1 and n_2 $(n_1 < 30, n_2 < 30)$ drawn from two normal populations having means μ_1 and μ_2. To test whether the two population means are equal (*i.e.*, to test whether the difference $\mu_1 - \mu_2$ is significant).

Let the Null Hypothesis be $H_0 : \mu_1 = \mu_2$

Then the Alternative Hypothesis is $H_1 : \mu_1 \neq \mu_2$

If $\sigma_1 = \sigma_2 = \sigma$, then an unbiased estimate S^2 of the common variance σ^2 is given by $S^2 = \dfrac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$ where s_1^2, s_2^2 are the two sample variances.

Also S. E. of $(\bar{x} - \bar{y}) = S\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}$ where $S = \sqrt{\dfrac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}}$

Assuming that H_0 is true, the test statistic t given by $t = \dfrac{\bar{x} - \bar{y}}{S.\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$, follows t-distribution with $(n_1 + n_2 - 2)$ d.f.

Here $\bar{x} = \dfrac{1}{n_1}\sum_{i=1}^{n_1} x_i$, $\bar{y} = \dfrac{1}{n_2}\sum_{i=1}^{n_2} y_i$ and

$$S^2 = \frac{1}{n_1 + n_2 - 2}\left[\Sigma(x_i - \bar{x})^2 + \Sigma(y_i - \bar{y})^2\right]$$

or $S^2 = \dfrac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$

We calculate the value of $|t|$ and compare this value with the table value at α level of significance. If the calculated value of $|t| >$ the table value, we reject H_0 at α level. Otherwise we accept H_0.

95% confidence limits for the difference of two population means are

$$(\bar{x} - \bar{y}) \pm t_\alpha . S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \text{ where } \alpha = 0.025.$$

and 99% confidence limits for the difference of two population means are

$$(\bar{x} - \bar{y}) \pm t_\alpha . S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \text{ where } \alpha = 0.0005$$

SOLVED EXAMPLES

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 mean, $\bar{x} = 1234$ hours	$\bar{y} = 1036$ hrs
Sample S.D., $s_1 = 36$ hrs	$s_2 = 40$ hrs

Is the difference in the means sufficient to warrant that type I is superior to type II regarding length of life. **[JNTU (A) June 2016]**

Solution : Since the sample sizes are small and σ_1, σ_2 are not known, we use t - test. Let μ_1 and μ_2 be the two population means.

Null Hypothesis H_0 : The two types I and II of electric bulbs are identical

i.e., $\mu_1 = \mu_2$

Alternative Hypothesis H_1 : $\mu_1 \neq \mu_2$

Since two sample means $\bar{x}$ and $\bar{y}$ are given and also sample standard deviations S_1 and S_2 are given, we use the statistic

$$t = \frac{\bar{x} - \bar{y}}{S^2\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

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

$$= \frac{1}{8+7-2}[(8(36)^2 + 7(40)^2] = 1659.08$$

$$\therefore \; t = \frac{1234 - 1036}{\sqrt{1659.08\left(\frac{1}{8} + \frac{1}{7}\right)}} = 9.39$$

Degrees of freedom (d.f) = $n_1 + n_2 - 2 = 8 + 7 - 2 = 13$

Tabulated value of t for 13 d.f at 5% level is 2.160 (two-tail test)

Since calculated t > tabulated t, we reject the null hypothesis H_0 and conclude that the two types I and II of electric bulbs are not identical.

Example 2 : 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.

Solution : Given $n_1 = 9$, $n_2 = 7$, $\bar{x} = 196.42$, $\bar{y} = 198.82$

and $\Sigma(x_i - \bar{x})^2 = 26.94$, $\Sigma(y_i - \bar{y})^2 = 18.73$

$$\therefore S^2 = \frac{\Sigma(x_i - \bar{x})^2 + \Sigma(y_i - \bar{y})^2}{n_1 + n_2 - 2}$$

$$= \frac{26.94 + 18.73}{9 + 7 - 2} = 3.26$$

$$\Rightarrow S = 1.81$$

1. Null Hypothesis H_0 : The two samples are drawn from the same population *i.e.* $\mu_1 = \mu_2$
2. **Alternative Hypothesis** $H_1 : \mu_1 \neq \mu_2$
3. **Level of significance :** $\alpha = 0.05$
4. **The test statistic** is $t = \dfrac{\bar{x} - \bar{y}}{S\sqrt{\dfrac{1}{n_2} + \dfrac{1}{n_2}}} = \dfrac{196.42 - 198.82}{(1.81)\sqrt{\dfrac{1}{9} + \dfrac{1}{7}}}$

$$= \frac{-2.4}{0.912} = -2.63$$

i.e., calculated value of $|t| = 2.63$

Tabulated value of t for $9 + 7 - 2 = 14$ d.f at 5% level of significance is 2.15. Since calculated value of $t >$ tabulated value of t, we reject the null hypothesis H_0 and conclude that the two samples are not drawn from the same population.

Example 3 : Below are given the gain in weights (in lbs) of pigs fed on two diets A and B.

Diet A	25	32	30	34	24	14	32	24	30	31	35	25			
Diet B	44	34	22	10	47	31	40	30	32	35	18	21	35	29	22

Test, if the two diets differ significantly as regards their effect on increase in weight.

Solution :

1. **Null Hypothesis** H_0 : There is no significant difference between the mean increase in weight due to diets A and B *i.e.*, $\mu_1 = \mu_2$.
2. **Alternative Hypothesis** $H_1 : \mu_1 \neq \mu_2$
3. **Level of significance,** $\alpha = 0.05$
4. **Calculation for sample means and S.D's.**

Here $n_1 = 12$, $n_2 = 15$, $\Sigma x = 336$, $\Sigma y = 450$

$$\therefore \bar{x} = \frac{336}{12} = 28, \ \bar{y} = \frac{450}{15} = 30$$

x	$x-\bar{x}$	$(x-\bar{x})^2$	y	$(y-\bar{y})$	$(y-\bar{y})^2$
25	−3	9	44	14	196
32	4	16	34	4	16
30	2	4	22	−8	64
34	6	36	10	−20	400
24	−4	16	47	17	289
14	−14	196	31	1	1
32	4	16	40	10	100
24	−4	16	30	0	0
30	2	4	32	2	4
31	3	9	35	5	25
35	7	49	18	−12	144
25	−3	9	21	−9	81
			35	5	25
			29	−1	1
			22	−8	64
336		380	450		1410

$$\sum_{i=1}^{12}(x_i-\bar{x})^2 = 380, \quad \sum_{i=1}^{15}(y_i-\bar{y})^2 = 1410$$

$$\therefore S^2 = \frac{1}{n_1+n_2-2}\left[\sum_i (x_i-\bar{x})^2 + \sum_i (y_i-\bar{y})^2\right]$$

$$= \frac{1}{12+15-2}[380 + 1410] = 71.6$$

5. **The test statistic** is

$$t = \frac{\bar{x}-\bar{y}}{\sqrt{S^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{30-28}{\sqrt{71.6\left(\frac{1}{12}+\frac{1}{15}\right)}} = \frac{2}{\sqrt{10.74}} = 0.609$$

Degrees of freedom $= n_1 + n_2 - 2 = 12 + 15 - 2 = 25$

Tabulated t for 25 d.f at 5% level = 2.06

Since calculated t < tabulated t, therefore, we accept the null hypothesis H_0.

That is, the two diets do not differ significantly as regards their effect on increase in weight.

Example 4 : A group of 5 patients treated with medicine A weigh 42, 39, 48, 60 and 41 kgs. Second group of 7 patients from the same hospital treated with medicine B weigh 38, 42, 56, 64, 68, 69 and 62 kgs. Do you agree with the claim that medicine B increases the weight significantly. **[JNTU (K) Nov. 2011 (Set No. 3)]**

Solution : Calculation for sample means and S.D's

x	$x-\bar{x}$	$(x-\bar{x})^2$	y	$(y-\bar{y})$	$(y-\bar{y})^2$
42	–4	16	38	–19	361
39	–7	49	42	–15	225
48	2	4	56	–1	1
60	14	196	64	7	49
41	–5	25	68	11	21
			69	12	44
			62	5	25
230	0	290	399	0	926

Now $\bar{x} = \dfrac{230}{5} = 46, \quad \bar{y} = \dfrac{399}{7} = 57,$

and $\sum_{i=1}^{5}(x_i-\bar{x})^2 = 290, \quad \sum_{j=1}^{7}(y_j-\bar{y})^2 = 926$

$$\therefore \; S^2 = \frac{1}{n_1+n_2-2}[\Sigma(x_i-\bar{x})^2 + \Sigma(y_j-\bar{y})^2] = \frac{1}{5+7-2}[290+926] = 121.6$$

$\Rightarrow S = 11.03$

1. **Null Hypothesis H_0 :** There is no significant difference between the medicines A and B as regards their effect on increase in weight i.e., $H_0 : \mu_1 = \mu_2$
2. **Alternative Hypothesis H_1 :** $\mu_1 > \mu_2$
3. **Level of significance,** $\alpha = 0.05$
4. **The test statistic** is $t = \dfrac{\bar{x}-\bar{y}}{S\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} = \dfrac{46-57}{(11.03)\sqrt{\dfrac{1}{5}+\dfrac{1}{7}}} = \dfrac{-11}{6.46} = -1.7$

$\therefore$ Calculated value of $|t| = 1.7$

Tabulated value of t for $5 + 7 - 2 = 10$ d.f at 5% level of significant is 1.81 (right-tail test).

Since calculated $t <$ tabulated t, we accept H_0, i.e., The medicines A and B do not differ significantly as regards their effect on increase in weight.

Example 5 : Two horses A and B were tested according to the time (in seconds) to run a particular track with the following results.

Horse A	28	30	32	33	33	29	34
Horse B	29	30	30	24	27	29	

Test whether the two horses have the same running capacity.

[JNTU 2006S, (A) Nov. 2010 (Set No. 1), (H) Dec. 2019 (R18)]

Solution : Given $n_1 = 7$, $n_2 = 6$

We first compute the sample means and standard deviations.

$\bar{x}$ = Mean of first sample $= \frac{1}{7}(28 + 30 + 32 + 33 + 33 + 29 + 34)$

$= \frac{1}{7}(219) = 31.286$

$\bar{y}$ = Mean of second sample $= \frac{1}{6}(29 + 30 + 30 + 24 + 27 + 29)$

$= \frac{1}{6}(169) = 28.16$

x	$x - \bar{x}$	$(x - \bar{x})^2$	y	$(y - \bar{y})$	$(y - \bar{y})^2$
28	−3.286	10.8	29	0.84	0.7056
30	−1.286	1.6538	30	1.84	3.3856
32	0.714	0.51	30	1.84	3.3856
33	1.714	2.94	24	−4.16	17.3056
33	1.714	2.94	27	−1.16	1.3456
29	−2.286	5.226	29	0.84	0.7056
34	2.714	7.366			
219		31.4358	169		26.8336

Now $S^2 = \frac{1}{n_1 + n_2 - 2}\left[\Sigma(x_i - \bar{x})^2 + \Sigma(y_i - \bar{y})^2\right]$

$= \frac{1}{11}[31.4358 + 26.8336] = \frac{1}{11}(58.2694) = 5.23$

$\therefore S = \sqrt{5.23} = 2.3$

1. **Null Hypothesis H_0 :** $\mu_1 = \mu_2$
2. **Alternative Hypothesis H_1 :** $\mu_1 \neq \mu_2$
3. **Level of significance :** $\alpha = 0.05$
4. The test statistic is $t = \frac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{31.286 - 28.16}{(2.3)\sqrt{\frac{1}{7} + \frac{1}{6}}} = 2.443$

Tabulated value of t for $7 + 6 - 2 = 11$ d.f. at 5% level of significance is 2.2.

Since Calculated $t >$ Tabulated t, we reject the null hypothesis H_0 and conclude that both horses A and B do not have the same running capacity.

Example 6 : To examine the hypothesis that the husbands are more intelligent than the wives, an investigator took a sample of 10 couples and administered them a test which measures the I.Q. The results are as follows :

Husbands	117	105	97	105	123	109	86	78	103	107
Wives	106	98	87	104	116	95	90	69	108	85

Test the hypothesis with a reasonable test at the level of significance of 0.05.

[JNTU 2004, 2005, 2007S, (A) Nov. 2010, (K) May 2013 (Set No. 3)]

Solution : We have $n_1 = 10$, $n_2 = 10$ and

$$\bar{x} = \frac{1}{10}(117 + 105 + 97 + 105 + 123 + 109 + 86 + 78 + 103 + 107)$$

$$= \frac{1}{10}(1030) = 103$$

$$\bar{y} = \frac{1}{10}(106 + 98 + 87 + 104 + 116 + 95 + 90 + 69 + 108 + 85)$$

$$= \frac{1}{10}(958) = 95.8$$

Now we compute the standard deviations of both samples

x	$x - \bar{x}$	$(x - \bar{x})^2$	y	$(y - \bar{y})$	$(y - \bar{y})^2$
117	14	196	106	10.2	104.04
105	2	4	98	2.2	4.84
97	−6	36	87	−8.8	77.44
105	2	4	104	8.2	67.24
123	20	400	116	20.2	408.04
109	6	36	95	−0.8	0.64
86	−17	289	90	−5.8	33.64
78	−25	625	69	−26.8	718.24
103	0	0	108	12.2	148.84
107	4	16	85	−10.8	116.64
1030		1606	958		1679.6

Now $S^2 = \dfrac{1}{n_1 + n_2 - 2}[\Sigma(x_i - \bar{x})^2 + \Sigma(y_i - \bar{y})^2]$

$$= \frac{1}{18}[1606 + 1679.6] = \frac{1}{18}(3285.6) = 182.53$$

$\therefore S = 13.51$

1. **Null Hypothesis H_0 :** $\mu_1 = \mu_2$ (*i.e.*, no difference in I.Q.)

2. **Alternative Hypothesis H_1 :** $\mu_1 > \mu_2$

(*i.e.*, husbands are more intelligent than wives) (one tailed test, right)

3. **Level of significance :** $\alpha = 0.05$

4. **The test statistic is** $t = \dfrac{\bar{x}-\bar{y}}{S.\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} = \dfrac{103-95.8}{(13.51)\sqrt{\dfrac{1}{10}+\dfrac{1}{10}}} = 1.19168$

Since $t_{cal} = 1.19168 < t_{tab} = 1.734$, we accept the null hypothesis H_0 i.e., There is no difference in I.Q's.

Example 7 : Find the maximum difference that we can expect with probability 0.95 between the means of samples of sizes 10 and 12 from a normal population if their standard deviations are found to be 2 and 3 respectively.

[JNTU 2004S, (A) Nov. 2010, Apr. 2012 (Set No. 2)]

Solution : We have $n_1 = 10$, $n_2 = 12$, $s_1 = 2$ and $s_2 = 3$

$$\therefore S^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} = \frac{1}{10+12-2}[10(2)^2 + 12(3)^2] = 7.4$$

$$\Rightarrow S = \sqrt{7.4} = 2.72$$

1. **Null Hypothesis H_0 :** $\mu_1 = \mu_2$
2. **Alternative Hypothesis H_1 :** $\mu_1 \neq \mu_2$
3. **Level of significance :** $\alpha = 0.05$
4. **The test statistic is** $t = \dfrac{\bar{x}-\bar{y}}{S.\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}$

$$\Rightarrow |\bar{x}-\bar{y}| = |t| \cdot S \cdot \sqrt{\frac{1}{n_1}+\frac{1}{n_2}} = (2.086)\,(2.72)\sqrt{\frac{1}{10}+\frac{1}{12}} = 2.43$$

[$\because$ Tabulated value of t for 10+12−2=20 d.f. at 5% level of significance is 2.086 (two-tailed)]

Hence the maximum difference between the means is 2.43.

Example 8 : Two independent samples of 8 and 7 items respectively had the following values.

Sample I	11	11	13	11	15	9	12	14
Sample II	9	11	10	13	9	8	10	–

Is the difference between the means of samples significant ?

[JNTU 2005S (Set No. 1), (A) Dec. 2017]

Solution : Given $n_1 = 8$, $n_2 = 7$ and

$$\bar{x} = \frac{1}{8}(11 + 11 + 13 + 11 + 15 + 9 + 12 + 4) = \frac{96}{8} = 12$$

$$\bar{y} = \frac{1}{7}(9 + 11 + 10 + 13 + 9 + 8 + 10) = \frac{70}{7} = 10$$

x	$x - \bar{x}$	$(x - \bar{x})^2$	y	$(y - \bar{y})$	$(y - \bar{y})^2$
11	−1	1	9	−1	1
11	−1	1	11	1	1
13	1	1	10	0	0
11	−1	1	13	3	9
15	3	9	9	−1	1
9	−3	9	8	−2	4
12	0	0	10	0	0
14	2	4			
96		26	70		16

$$\text{Now } S^2 = \frac{1}{n_1 + n_2 - 2}\left[\Sigma(x_i - \bar{x})^2 + \Sigma(x_i - \bar{y})^2\right]$$

$$= \frac{1}{8+7-2}(26 + 16) = \frac{42}{13} = 3.23$$

$\therefore S = 1.8$

1. **Null Hypothesis** $H_0 : \mu_1 = \mu_2$
2. **Alternative Hypothesis** $H_1 : \mu_1 \neq \mu_2$ (two tailed test)
3. **Level of significance :** $\alpha = 0.05$
4. **The test statistic** is $t = \dfrac{\bar{x} - \bar{y}}{S.\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} = \dfrac{12 - 10}{(1.8)\sqrt{\dfrac{1}{8} + \dfrac{1}{7}}} = \dfrac{2}{0.9316} = 2.15$

Tabulated value of t for $8 + 7 - 2 = 13$ d.f. at 5% level of significance is 2.16.

Since calculated t < tabulated t, we accept H_0 *i.e.*, the difference between the means of samples is not significant.

Example 9 : Measuring specimens of nylon yarn, taken from two machines, it was found that 8 specimens from first machine had a mean denier of 9.67 with a standard deviation of 1.81 while 10 specimens from second machine had a mean denier of 7.43 with a standard deviation of 1.48. Assuming that the proportions are normal, test the hypothesis $H_0 : \mu_1 - \mu_2 = 1.5$ against $H_1 : \mu_1 - \mu_2 > 1.5$ at 0.05 level of significance.

[JNTU 2005S, 2006S (Set No. 1, 4)]

Solution : We have $n_1 = 8$, $n_2 = 10$, $\bar{x} = 9.67$, $\bar{y} = 7.43$, $s_1 = 1.81$ and $s_2 = 1.48$

$$\text{Now } S^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} = \frac{8\,(1.81)^2 + 10\,(1.48)^2}{8+10-2}$$

$$= \frac{26.2088 + 21.901}{16} = \frac{48.1128}{16} = 3.007$$

$$\therefore S = \sqrt{3.007} = 1.734$$

1. **Null Hypothesis** H_0 : $\mu_1 - \mu_2 = 1.5\ (= \delta)$
2. **Alternative Hypothesis** H_1 : $\mu_1 - \mu_2 > 1.5$
3. **Level of significance** : $\alpha = 0.05$
4. **The test statistic** is $t = \dfrac{(\bar{x} - \bar{y}) - \delta}{S.\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$

$$= \frac{(9.67 - 7.43) - 1.5}{(1.734)\sqrt{\dfrac{1}{8} + \dfrac{1}{10}}} = \frac{0.74}{60.8225} = 0.8997$$

Tabulated value of t for $8 + 10 - 2 = 16$ d.f. at 5% level of significance is 1.746 (one-tail test).

Since calculated t < tabulated t, we accept H_0.

Example 10 : Ten soldiers participated in a shooting competition in the first week. After intensive training they participated in the competition in the second week. Their scores before and after training are given as follows :

Scores before	67	24	57	55	63	54	56	68	33	43
Scores after	70	38	58	58	56	67	68	75	42	38

Do the data indicate that the soldiers have been benefited by the training.

[JNTU 2005S (Set No. 2)]

Solution : We have $n_1 = 10,\ n_2 = 10$

Let $\bar{x}$ = Mean of the scores before training

$$= \frac{1}{10}(67 + 24 + 57 + 55 + 63 + 54 + 56 + 68 + 33 + 43)$$

$$= \frac{1}{10}(520) = 52$$

and $\bar{y}$ = Mean of the scores after training

$$= \frac{1}{10}(70 + 38 + 58 + 58 + 56 + 67 + 68 + 75 + 42 + 38)$$

$$= \frac{1}{10}(570) = 57$$

Now we compute the standard deviations of the two samples.

x	$x-\bar{x}$	$(x-\bar{x})^2$	y	$(y-\bar{y})$	$(y-\bar{y})^2$
67	15	225	70	13	169
24	–28	784	38	–19	361
57	5	25	58	1	1
55	3	9	58	1	1
63	11	121	56	–1	1
54	2	4	67	10	100
56	4	16	68	11	121
68	16	256	75	18	324
33	–19	361	42	–15	225
43	–9	81	38	–19	361
520		1882	570		1664

$$\text{Now } S^2 = \frac{1}{n_1+n_2-2}\left[\sum_i (x_i-\bar{x})^2 + \sum_i (y_i-\bar{y})^2\right]$$

$$= \frac{1}{10+10-2}[1882 + 1664) = \frac{1}{18}(3546) = 197$$

$\Rightarrow S = 14.0357$

1. **Null Hypothesis H_0 :** $\mu_1 = \mu_2$ *i..e*, there is no benefit of training
2. **Alternative Hypothesis H_1 :** $\mu_1 < \mu_2$ *i.e*, there is a benefit of training.
3. **Level of significance :** $\alpha = 0.05$
4. **The test statistic** is $t = \dfrac{(\bar{x}-\bar{y})}{S.\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} = \dfrac{52-57}{(14.0357).\sqrt{\dfrac{1}{10}+\dfrac{1}{10}}}$

$$= \frac{-5}{6.277} = -0.796$$

Tabulated value of t for $10 + 10 - 2 = 18$ d.f. at 5% level of significance (left tailed test) is –1.734

Since $|t|_{cal} = 0.796 < 1.734 = |t|_{tab}$, we accept the null Hypothesis H_0 i.e., the soldiers have been benefited by the training.

Example 11 : To compare two kinds of bumper guards, 6 of each kind were mounted on a car and then the car was run into a concrete wall. The following are the costs of repairs.

Bumper Guard 1	107	148	123	165	102	119
Bumper Guard 2	134	115	112	151	133	129

Use the 0.01 level of significance to test whether the difference between the means of these two samples is significant.

[JNTU 2005, (H) Nov. 2010, (A) Nov. 2011 (Set No. 3), June 2018]

Solution : We have

$$n_1 = 6,\ n_2 = 6,\ \bar{x} = \frac{1}{6}(107 + 148 + 123 + 165 + 102 + 119) = \frac{764}{6} = 127.33$$

and $\bar{y} = \frac{1}{6}(134 + 115 + 112 + 151 + 133 + 129) = \frac{774}{6} = 129$

Now we compute the standard deviations of two samples.

x	$x - \bar{x}$	$(x - \bar{x})^2$	y	$(y - \bar{y})$	$(y - \bar{y})^2$
107	−20.33	413.31	134	5	25
148	20.67	427.25	115	−14	196
123	−4.33	18.75	112	−17	289
165	37.67	1419.03	151	22	484
102	−25.33	641.61	133	4	16
119	−8.33	69.39	129	0	0
764		2989.34	774		1010

$$\therefore\ S^2 = \frac{1}{n_1 + n_2 - 2}[\Sigma(x_1 - \bar{x})^2 + \Sigma(y_1 - \bar{y})^2]$$

$$= \frac{1}{16 + 6 - 2}[2989.34 + 1010] = \frac{1}{10}(3999.34) = 399.934$$

$\Rightarrow\ S = 19.998$

1. **Null Hypothesis H_0 :** $\mu_1 = \mu_2$
2. **Alternative Hypothesis H_1 :** $\mu_1 \neq \mu_2$ (two tailed test)
3. **Level of significance :** $\alpha = 0.01$
4. **The test statistic** is

$$t = \frac{\bar{x} - \bar{y}}{S.\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{127.33 - 129}{(19.998)\sqrt{\frac{1}{6} + \frac{1}{6}}} = \frac{-1.67}{11.546} = -0.1446$$

$\therefore$ Calculated value of $|t| = 0.1446$

Tabulated value of t for $6 + 6 - 2 = 10$ d.f. at 1% level of significance is 3.169

Since calculated t < tabulated t, we accept the Null Hypothesis H_0, i.e., there is no significant difference between the sample means.

Example 12 : Random samples of specimens of coal from two mines A and B are drawn and their heat-producing capacity (in millions of calories/ton) were measured yielding the following results:

Mine A	8350	8070	8340	8130	8260	---
Mine B	7900	8140	7920	7840	7890	7950

Is there significant difference between the means of these two samples at 0.01 level of significance. **[JNTU 2005, (K) Nov. 2009 (Set No. 3)]**

Solution : Here $n_1 = 5$, $n_2 = 6$

Calculation for mean's and S.D's of the samples:

Mine A			Mine B		
x	$x-\bar{x}$ $(x-8230)$	$(x-\bar{x})^2$	y	$(y-\bar{y})$ $(y-7940)$	$(y-\bar{y})^2$
8350	120	14400	7900	−40	1600
8070	−160	25600	8140	200	40000
8340	110	12100	7920	−20	400
8130	−100	10000	7840	100	10000
8260	30	900	7890	−50	2500
			7950	10	100
41150		63000	47640		54600

$$\bar{x} = \frac{\Sigma x}{n} = \frac{41150}{5} = 8230, \quad \bar{y} = \frac{\Sigma y}{n} = \frac{47640}{6} = 7940$$

$$\text{Now } S^2 = \frac{1}{n_1+n_2-2}\left[\Sigma(x_i-\bar{x})^2 + \Sigma(y_i-\bar{y})^2\right]$$

$$= \frac{1}{5+6-2}[63000+54600] = \frac{1}{9}(117600) = 13066.67$$

$\therefore$ S = 114.31.

1. **Null Hypothesis** $H_0 : \mu_1 = \mu_2$
2. **Alternative Hypothesis** $H_1 : \mu_1 \neq \mu_2$
3. **Level of significance,** $\alpha = 0.01$
4. **The test statistic is**

$$t = \frac{\bar{x}-\bar{y}}{S\cdot\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{8230-7940}{(114.31)\sqrt{\frac{1}{5}+\frac{1}{6}}} = \frac{290}{(114.31)(0.6055)} = 4.19$$

i.e., calculated value of $|t| = 4.19$

Tabulated value of t for $5+6-2=9$ d.o.f at $\alpha = 0.01$ is 2.821

Since calculated $t > t_{0.005}$, we reject the null hypothesis H_0.

Thus we conclude that there is significant difference between the means of the two samples.

Example 13 : The IQ s (intelligence quotients) of 16 students from one area of a city showed a mean of 107 with a standard deviation of 10, while the IQs of 14 students from another area of the city showed a mean of 112 with a standard deviation of 8. Is there a significant difference between the IQs of the two groups at a 0.05 level of significance?

[JNTU (K) May 2010 (Set No. 2)]

Solution : Let μ_1 and μ_2 be the means of the two populations.

Let the Null Hypothesis be $H_0 : \mu_1 = \mu_2$.

Then the Alternative Hypothesis is $H_1 : \mu_1 \neq \mu_2$.

We have $n_1 = 16, \bar{x} = 107, s_1 = 10$ and $n_2 = 14, \bar{y} = 112, s_2 = 8$

Since the sample sizes are small and σ_1, σ_2 are not known, we use t - test.

An unbiased estimate S^2 of the common variance is given by

$$S^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} = \frac{16 \times 100 + 14 \times 64}{16 + 14 - 2} = \frac{2496}{28} = 89.14$$

$\therefore S = \sqrt{89.14} = 9.44$

Let us assume that H_0 is true, *i.e.*, there is no difference between μ_1 and μ_2.

The test statistic is

$$t = \frac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{107 - 112}{9.44\sqrt{\frac{1}{16} + \frac{1}{14}}} = \frac{-5}{9.44 \times 0.366} = -1.447$$

Also d.f. $= 16 + 14 - 2 = 28$

$\therefore \ |t| = 1.447 < 1.701$ for 28 d.f. *i.e.*, the calculated value of $|t| <$ the tabulated value at 5% level with 28 d.f.

Hence, we accept the Null Hypothesis and conclude that the difference between the IQs of the two groups is not significant.

Example 14 : The table gives the biological values of protein from 6 cow's milk and 6 buffalo's milk. Examine whether the di"erences are significant.

[JNTU (H) Dec. 2011 (Set No. 3)]

Cow's milk	1.8	2.0	1.9	1.6	1.8	1.5
Buffalo's milk	2	1.8	1.8	2.0	2.1	1.9

Solution : Let x and y represent the protein values of cow's milk and buffalo's milk respectively.

Let the Null Hypothesis be $H_0 : \mu_1 = \mu_2$

Then the Alternative Hypothesis is $H_1 : \mu_1 \neq \mu_2$

Calculations for Sample Means and S. D.'s

x	$x-\bar{x}$ $=x-1.77$	$(x-\bar{x})^2$	y	$y-\bar{y}$ $=y-1.93$	$(y-\bar{y})^2$
1.8	0.03	0.0009	2	0.07	0.0049
2.0	0.23	0.0529	1.8	−0.13	0.0169
1.9	0.13	0.0169	1.8	−0.13	0.0169
1.6	−0.17	0.0289	2.0	0.07	0.0049
1.8	0.03	0.0009	2.1	0.17	0.0289
1.5	−0.27	0.0729	1.9	−0.03	0.0009
$10.6=\Sigma x_i$		0.1734	$11.6=\Sigma y_i$		0.0734

$$\bar{x}=\frac{\Sigma x_i}{n_1}=\frac{10.6}{6}=1.77 \text{ and } \bar{y}=\frac{\Sigma y_i}{n_2}=\frac{11.6}{6}=1.93$$

Now $$S^2=\frac{1}{n_1+n_2-2}\left[\Sigma(x-\bar{x})^2+\Sigma(y-\bar{y})^2\right]$$

$$=\frac{1}{6+6-2}[0.1734+0.0734]=\frac{0.2468}{10}=0.0247$$

$\therefore$ $S=\sqrt{0.0247}=0.157$

$\therefore$ The test statistic is

$$t=\frac{\bar{x}-\bar{y}}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}=\frac{1.77-1.93}{(0.157)\sqrt{\frac{1}{6}+\frac{1}{6}}}$$

$$=\frac{(-0.16)\sqrt{3}}{0.157}=\frac{-0.2771}{0.157}=-1.7651$$

$\therefore |t|=1.7651$

d.f. $=n_1+n_2-2=6+6-2=10$

Tabulated value of t for 10 d.f. at 5% level of significance is 2.228 (two -tailed test).

Since the calculated value of $|t|<$ the table value of t at 5% level with 10 d.f. for two - tailed test, we accept Null Hypothesis H_0 and conclude that there is no significant difference in their means.

10.7 PAIRED - SAMPLE *t* - TEST

Paired observations arise in many practical situations where each homogeneous experimental unit receives both population conditions. As a result, each experimental unit has a pair of observations, one for each population.

For instance, to test the effectiveness of "drug" some 11 persons blood pressure is measured "before" and "after" the intake of certain drug. Here the individual person is the experimental unit and the two populations are blood pressure "before" and "after" the drug is given. Thus for each observation in one sample, there is a corresponding observation in the other sample pertaining to the same character. Hence the two samples are not independent.

Consider another example. Suppose a business concern is interested to know whether a particular media of promoting sales of a product is really effective or not. In this case we have to test whether the average sales before and after the sales promotion are equal.

If $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ be the pairs of sales data before and after the sales promotion in a business concern, we apply paired t - test to examine the significance of the difference of the two situations .

Let $d_i = x_i - y_i$ or $y_i - x_i$ for $i = 1, 2, 3, \ldots, n$

Let the Null Hypothesis be $H_0 : \mu_1 = \mu_2$ (*i.e.* $\mu = 0$), there is no significant difference between the means in two situations.

Then the Alternative Hypothesis is $H_1 : \mu_1 \neq \mu_2$

Assuming that H_0 is true, the test statistic for n paired observations (which are dependent) by taking the differences $d_1, d_2, \ldots, d_n$ of the paired data.

$$t = \frac{\bar{d} - \mu}{S/\sqrt{n}} = \frac{\bar{d}}{S/\sqrt{n}} \quad (\because \mu = 0)$$

where $\bar{d} = \frac{1}{n}\Sigma d_i$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2$

or $$S^2 = \frac{\Sigma d^2 - n(\bar{d})^2}{n-1} \text{ or } \frac{1}{n-1}\left[\Sigma d^2 - \frac{(\Sigma d)^2}{n}\right]$$

are the mean and variance of the differences $d_1, d_2, \ldots, d_n$ respectively and μ is the mean of the population of differences.

The above statistic follows student's t - distribution with $(n-1)$ degrees of freedom.

SOLVED EXAMPLES

Example 1 : Ten workers were given a training programme with a view to study their assembly time for a certain mechanism. The results of the time and motion studies before and after the training programme are given below.

Workers	1	2	3	4	5	6	7	8	9	10
X_1	15	18	20	17	16	14	21	19	13	22
Y_1	14	16	21	10	15	18	17	16	14	20

X_1 = Time taken for assembling before training.

Y_1 = Time taken for assembling after training.

Test whether there is significant difference in assembly times before and after training.

[JNTU (A) May 2010S]

Solution : From the given paired data, we see that we are to use paired t-test. Let μ be the mean of population of differences.

1. **Null Hypothesis** H_0 : $\mu_1 = \mu_2$ or $\mu = 0$ *i.e.*, training is not useful.
2. **Alternative Hypothesis** H_1: $\mu_1 \neq \mu_2$ *i.e.*, training is useful in assembly time.
3. **Level of significance** $\alpha = 0.05$
4. **Computation :** Differences d_i's (before and after training) are

$$1, 2, -1, 7, 1, -4, 4, 3, -1, 2$$

$$\bar{d} = \text{mean of differences of sample data} = \frac{\Sigma d}{n} = \frac{14}{10} = 1.4$$

$$\text{Now} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2$$

$$= \frac{1}{9}[(1-1.4)^2 + (2-1.4)^2 + (-1-1.4)^2 + (7-1.4)^2 + (1-1.4)^2 + (-4-1.4)^2$$

$$+ (4-1.4)^2 + (3-1.4)^2 + (-1-1.4)^2 + (2-1.4)^2]$$

$$= \frac{1}{9}[0.16 + 0.36 + 5.76 + 31.36 + 0.16 + 29.16 + 6.76 + 2.56 + 5.76 + 0.36]$$

$$= \frac{82.4}{9} = 9.1555$$

$$\therefore S = 3.026$$

5. **The test statistic is** $t = \dfrac{\bar{d} - \mu}{S/\sqrt{n}} = \dfrac{\bar{d}}{S/\sqrt{n}} = \dfrac{1.4}{3.026/\sqrt{10}} = \dfrac{(1.4)(3.1623)}{3.026} = 1.46$

$\therefore$ Calculated $|t| = 1.46$

Tabulated $t_{0.05}$ with $10 - 1 = 9$ degrees of freedom is 1.833

Since calculated $t < t_{0.05}$, we accept the Null Hypothesis H_0 and conclude that there is no significant difference in assembly times before and after training.

Example 2 : Scores obtained in a shooting competition by 10 soldiers before and after intensive training are given below :

Before	67	24	57	55	63	54	56	68	33	43
After	70	38	58	58	56	67	68	75	42	38

Test whether the intensive training is useful at 0.05 level of significance.

Solution : Let us apply paired test.

Let μ be the mean of population of differences.

1. **Null Hypothesis** $H_0 : \mu_1 = \mu_2$ *i.e.*, $\mu = 0$ there is no significant effect of the training.
2. **Alternative Hypothesis** $H_1 : \mu_1 \neq \mu_2$ intensive training is useful.
3. **Level of significance,** $\alpha = 0.05$
4. **Computation :**

Differences d_i's (before and after training) are $-3, -14, -1, -3, 7, -13, -12, -7, -9, 5$

$$\bar{d} = \frac{1}{10}\sum_{i=1}^{10} d_i = \frac{-50}{10} = -5$$

Now $$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2 = \frac{1}{9}\sum_{i=1}^{10}(d_i - \bar{d})^2$$

$$= \frac{1}{9}[(2)^2 + (-9)^2 + (4)^2 + (2)^2 + (12)^2 + (-8)^2 + (-7)^2 + (-2)^2 + (-4)^2 + (10)^2]$$

$$= \frac{1}{9}[4 + 81 + 16 + 4 + 144 + 64 + 49 + 4 + 16 + 100]$$

$$= \frac{482}{9} = 53.5555$$

$\therefore S = 7.32$

5. **The test statistic** is $t = \dfrac{\bar{d}}{S/\sqrt{n}} = \dfrac{-5-0}{7.32/\sqrt{10}} = -2.16$

$\therefore |t| = 2.16$

Tabulated $t_{0.05}$ with $10 - 1 = 9$ degrees of freedom is 1.83.

Since calculated $t > t$ tabulated t, we reject the Null Hypothesis and conclude that the intensive training is useful.

Example 3 : The Blood Pressure of 5 women before and after intake of a certain drug are given below :

Before	110	120	125	132	125
After	120	118	125	136	121

Test whether there is significant change in Blood Pressure at 1 % level of significance .

[JNTU (H) Nov. 2010 (Set No. 3), (K) May 2013 (Set No. 4)]

Solution : Let the Null Hypothesis be $H_0 : \mu_1 = \mu_2$ *i.e.*, there is no significant difference in blood pressure before and after intake of a drug.

Then the Alternative Hypothesis is $H_1 : \mu_1 < \mu_2$

Assuming that H_0 is true, the test statistic is

$$t = \frac{d}{S/\sqrt{n}} \text{ where } \bar{d} = \frac{\Sigma d}{n}, d = y - x$$

and $S^2 = \frac{\Sigma(d-\bar{d})^2}{n-1} = \frac{\Sigma d^2 - (\bar{d})^2 \times n}{n-1}$

Calculations For $\bar{d}$ and S

Womens	B.P. before intake of drug (x)	B.P. after intake of drug (y)	$d = y - x$	d^2
1	110	120	10	100
2	120	118	–2	4
3	123	125	2	4
4	132	136	4	16
5	125	121	–4	16
Total			$\Sigma d = 10$	$\Sigma d^2 = 140$

$$\therefore \bar{d} = \frac{\Sigma d}{n} = \frac{10}{5} = 2 \text{ and } S^2 = \frac{140-(2)^2 \times 5}{4} = \frac{140-20}{4} = \frac{120}{4} = 30$$

$$\Rightarrow S = \sqrt{30}$$

$$\therefore t = \frac{\bar{d}}{S/\sqrt{n}} = \frac{2}{\sqrt{30}/\sqrt{5}} = \frac{2}{\sqrt{6}} = 0.82$$

Degrees of freedom $= n - 1 = 5 - 1 = 4$

Thus $t = 0.82 < 4.6$ at 1% level with 4 d.f. .

Since the calculated value of $t <$ the tabulated value with 4 d.f. at 1% level, we accept H_0 at 1% level and conclude that there is no significant change in Blood Pressure after intake of a certain drug.

Example 4 : Memory capacity of 10 students were tested before and after training. State whether the training was effective or not from the following scores.

Before training	12	14	11	8	7	10	3	0	5	6
After training	15	16	10	7	5	12	10	2	3	8

[JNTU (H) Nov. 2010, Apr. 2012 (Set No. 2), (A) Dec. 2018]

Solution : Let the Null Hypothesis be $H_0 : \mu_1 = \mu_2$

i.e., there is no significant effect of the training

Then the Alternative Hypothesis is $H_1 : \mu_1 < \mu_2$

Assuming that H_0 is true, the test statistic is

$t = \dfrac{\bar{d}}{S/\sqrt{n}}$ where $\bar{d} = \dfrac{\Sigma d}{n}$, $d = x - y$ and $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2 = \dfrac{\Sigma d^2 - (\bar{d})^2 \times n}{n-1}$ ... (1)

Calculations for $\bar{d}$ and S

Memory capacity before training (x)	Memory capacity after training (y)	$d = x - y$	d^2
12	15	−3	9
14	16	−2	4
11	10	1	1
8	7	1	1
7	5	2	4
10	12	−2	4
3	10	−7	49
0	2	−2	4
5	3	2	4
6	8	−2	4
		$-12 = \Sigma d$	$84 = \Sigma d^2$

$\therefore \bar{d} = \dfrac{\Sigma d}{n} = \dfrac{-12}{10} = -1.2$

Now $S^2 = \dfrac{84 - (-1.2)^2 \times 10}{10 - 1}$ [using (1)]

$= \dfrac{84 - 14.4}{9} = \dfrac{69.6}{9} = 7.73$

$\Rightarrow S = \sqrt{7.73} = 2.78$

$\therefore t = \dfrac{\bar{d}}{S/\sqrt{n}} = \dfrac{-1.2}{2.78/\sqrt{10}} = -1.365$ and d.f. $= n - 1 = 10 - 1 = 9$

Thus $|t| = 1.365 < 1.833$ at 5% level with 9 d.f.

Since the calculated value of t < the tabulated value of t with 9 d.f. at 5% level, we accept H_0 at 5% level and conclude that there is no significant change in memory capacity after the training programme.

Example 5 : The average losses of workers, before and after certain program are given below. Use 0.05 level of significance to test whether the program is effective (paired sample t- test). 40 and 35, 70 and 65, 45, −42, 120 and 116, 35 and 55 and 50, 77 and 73.

[JNTU (H) Dec. 2011 (Set No. 3)]

Solution : Let the Null Hypothesis be $H_0 : \mu_1 = \mu_2$ *i.e.*, there is no significant effect of the program.

Then the Alternative Hypothesis is $H_1 : \mu_1 \neq \mu_2$

Assuming that H_0 is true, the test statistic is $t = \dfrac{\bar{d}}{S/\sqrt{n}}$

where $\bar{d} = \dfrac{\Sigma d}{n}, d = y - x$ and $S^2 = \dfrac{\Sigma(d-\bar{d})^2}{n-1}$

Calculations for $\bar{d}$ and S

Before program (x)	After program (y)	$d_i = y - x$	$d_i - \bar{d} = d_i - 4$	$(d_i - \bar{d})^2$
40	35	5	1	1
70	65	5	1	1
45	42	3	−1	1
120	116	4	0	0
35	33	2	−2	4
55	50	5	1	1
77	73	4	0	0
		28		8

$$\bar{d} = \frac{\Sigma d_i}{n} = \frac{28}{7} = 4$$

$$\text{Now } S^2 = \frac{\Sigma(d_i - \bar{d})^2}{n-1} = \frac{8}{6} = \frac{4}{3} \Rightarrow S = \frac{2}{\sqrt{3}}$$

$$\therefore t = \frac{4}{(2/\sqrt{3})/\sqrt{7}} = 2\sqrt{21} = 9.16$$

d.f. $= n - 1 = 7 - 1 = 6$

Thus $t = 9.16 > 2.447$ at 5% level with 6 d.f.

Since the calculated value of $t >$ the tabulated value with 6 d.f. at 5% level, we reject the Null Hypothesis H_0 and conclude that the program is not effective.

9.8 SNEDECOR'S *F*- TEST OF SIGNIFICANCE

When testing the significance of the difference of the means of two samples, we assumed that the two samples came from the same population or from populations with equal variances. If the variances of the populations are not equal, a significant difference in the means may arise. Hence, before we apply the t - test for the significance of the difference of two means, we have to test for the equality of population variances using F - test of significance.

If s_1^2 and s_2^2 are the variances of two samples of sizes n_1 and n_2 respectively, then the population variances are given by

$$n_1 s_1^2 = (n_1 - 1)S_1^2 \text{ and } n_2 s_2^2 = (n_2 - 1)S_2^2$$

The quantities $\nu_1 = n_1 - 1$ and $\nu_2 = n_2 - 1$ are called degrees of freedom of these estimates. We want to test if these estimates S_1^2 and S_2^2 are significantly different or if the samples may be regarded as drawn from the same population or from two populations with same variance σ^2.

Test for Equality of Two Population Variances

Let two independent random samples of sizes n_1 and n_2 be drawn from two normal populations.

To test the hypothesis that the two population variances σ_1^2 and σ_2^2 are equal.

Let the Null Hypothesis be $H_0 : \sigma_1^2 = \sigma_2^2$

Then the Alternative Hypothesis is $H_1 : \sigma_1^2 \neq \sigma_2^2$

The estimates of σ_1^2 and σ_2^2 are given by

$$S_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = \frac{\Sigma(x_i - \bar{x})^2}{n_1 - 1} \text{ and}$$

$$S_2^2 = \frac{n_2 s_2^2}{n_2 - 1} = \frac{\Sigma(y_i - \bar{y})^2}{n_2 - 1}$$, where s_1^2 and s_2^2 are the variances of the two samples.

Assuming that H_0 is true, the test statistic $\boxed{F = \frac{S_1^2}{S_2^2} \text{ or } \frac{S_2^2}{S_1^2}}$ according as $S_1^2 > S_2^2$ or $S_2^2 > S_1^2$ follows F - distribution with $(n_1 - 1, n_2 - 1)$ degrees of freedom.

Conclusion : If the calculated value of F > the tabulated value of F at 5% level, we reject the Null Hypothesis H_0 and conclude that the variances σ_1^2 and σ_2^2 are not equal. Otherwise, we accept the Null Hypothesis H_0 and conclude that σ_1^2 and σ_2^2 are equal.

Note : 1. In numerical problems, we take the greater of the two variances S_1^2, S_2^2 in the numerator and the other in the denominator.

$$i.e., \; F = \frac{\text{Greater variance}}{\text{Smaller variance}}$$

2. When F is close to 1, the two sample variances S_1 and S_2 are nearly same.

3. If sample variance S^2 is given, we can obtain population variance σ^2 by using the relation $n\,\sigma^2 = (n - 1)\,S^2$ and vice-versa.

4. $F_\alpha(\nu_1, \nu_2)$ is the value of F with ν_1 and ν_2 degrees of freedom such that the area under the F - distribution to the right of F_α is α. In tables F_α tabulated for $\alpha = 0.05$ and $\alpha = 0.01$ for various combinations of the d.o.f ν_1 and ν_2. Clearly value of F at 5% significance is lower than at 1%

SOLVED EXAMPLES

Example 1 : In one sample of 8 observations from a normal population, the sum of the squares of deviations of the sample values from the sample mean is 84.4 and in another sample of 10 observations it was 102.6. Test at 5% level whether the populations have the same variance.

(or) In two independent samples of sizes 8 and 10 the sum of squares of deviations of the sample values from the respective sample means were 84.4 and 102.6. Test whether the difference of variences of the population is significant or not. Use a 0.05 level of significance.

[JNTU (A) June 2016]

Solution : Let σ_1^2 and σ_2^2 be the variances of the two normal populations from which the samples are drawn.

Let the Null Hypothesis be $H_0 : \sigma_1^2 = \sigma_2^2$

Then the Alternative Hypothesis is $H_1 : \sigma_1^2 \neq \sigma_2^2$

Here $n_1 = 8,\ n_2 = 10$

Also $\Sigma(x_i - \bar{x})^2 = 84.4,\ \Sigma(y_i - \bar{y})^2 = 102.6$

If S_1^2 and S_2^2 be the estimates of σ_1^2 and σ_2^2 then

$$S_1^2 = \frac{1}{n_1 - 1}\Sigma(x_i - \bar{x})^2$$

$$= \frac{84.4}{7} = 12.057$$

and $$S_2^2 = \frac{1}{n_2 - 1}\Sigma(y_i - \bar{y})^2$$

$$= \frac{102.6}{9} = 11.4$$

Let H_0 be true. Since $S_1^2 > S_2^2$, the test statistic is

$$F = \frac{S_1^2}{S_2^2} = \frac{12.057}{11.4} = 1.057$$

i.e. calculated $F = 1.057$.

Degrees of freedom are given by $\nu_1 = n_1 - 1 = 8 - 1 = 7$

and $\nu_2 = n_2 - 1 = 10 - 1 = 9$.

Tabulated value of F at 5% level for (7, 9) degrees of freedom is 3.29

i.e., $F_{0.05}(7, 9) = 3.29$

Since calculated $F <$ tabulated F, we accept the Null Hypothesis H_0 and conclude that the populations have the same variance.

Example 2 : In one sample of 10 observations from a normal population, the sum of the squares of the deviations of the sample values from the sample mean is 102.4 and in another sample of 12 observations from another normal population, the sum of the squares of the deviations of the sample values from the sample mean is 120.5. Examine whether the two normal populations have the same variance.

Solution : Let σ_1^2 and σ_2^2 be the variances of the two normal populations from which the samples are drawn.

Let the Null Hypothesis be $H_0 : \sigma_1^2 = \sigma_2^2$

Then the Alternative Hypothesis is $H_1 : \sigma_1^2 \neq \sigma_2^2$

Given $n_1 = 10$, $n_2 = 12$, $\Sigma(x-\bar{x})^2 = 102.4$, $\Sigma(y-\bar{y})^2 = 120.5$

If S_1^2 and S_2^2 be the estimates of σ_1^2 and σ_2^2 then

$$S_1^2 = \frac{\Sigma(x-\bar{x})^2}{n_1 - 1} = \frac{102.4}{9} = 11.37$$

$$S_2^2 = \frac{\Sigma(y-\bar{y})^2}{n_2 - 1} = \frac{120.5}{11} = 10.95$$

The test statistic is $F = \dfrac{S_1^2}{S_2^2}$ $[\because S_1^2 > S_2^2]$

$$= \frac{11.37}{10.95} = 1.038$$

i.e, calculated $F = 1.132$

Tabulated value of F for (9, 11) d.f. at 5% level of significance is 2.91 (approx.)

Since calculated $F <$ tabulated F, we accept the null hypothesis H_0 i.e., the two normal populations have the same variance.

Example 3 : Two random samples reveal the following results :

Sample	Size	Sample Mean	Sum of Squares of Deviations from the Mean
1	10	15	90
2	12	14	108

Test whether the samples came from the same normal population.

Solution :

Null Hypothesis H_0 : The two samples have been drawn from the same normal population.

i.e., $H_0 : \mu_1 = \mu_2$

and $\sigma_1^2 = \sigma_2^2$.

Here we have to use two tests (*i*) To test equality of variances by *F*-test

(*ii*) To test equality of means by *t*-test.

(*i*) F-test (equality of variances)

Given $n_1 = 10,\ n_2 = 12,\ \bar{x}_1 = 15,\ \bar{x}_2 = 14,$

$\Sigma(x_i - \bar{x})^2 = 90,\ \Sigma(y_i - \bar{y})^2 = 108$

$$\therefore\ S_1^2 = \frac{\Sigma(x_i - \bar{x})^2}{n_1 - 1} = \frac{90}{9} = 10;$$

$$S_2^2 = \frac{\Sigma(y_i - \bar{y})^2}{n_1 - 1} = \frac{108}{11} = 9.82$$

$$\text{Hence } F = \frac{S_1^2}{S_2^2} = \frac{10}{9.82} = 1.018$$

i.e., calculated $F = 1.018$

Tabulated F at 5% level for (9, 11) degrees of freedom is 2.90

i.e., $F_{0.05}\ (9, 11) = 2.90$

Since calculated $F <$ tabulated F, we accept the hypothesis

$\therefore\ \sigma_1^2 = \sigma_2^2$

i.e., the samples came from the same normal populations with same variance.

(*ii*) *t*-test (To test equality of means)

Null Hypothesis H_0 : $\mu_1 = \mu_2$

Given $\bar{x}_1 = 15,\ \bar{x}_2 = 14,\ n_1 = 10,\ n_2 = 12,$

$\Sigma\,(x_i - \bar{x})^2 = 90,\ \Sigma\,(y_i - \bar{y})^2 = 108$

$$\text{Now } S^2 = \frac{1}{n_1 + n_2 - 2}\left[\Sigma\,(x - \bar{x})^2 + \Sigma\,(y - \bar{y})^2\right]$$

$$= \frac{1}{10 + 12 - 2}\,[90 + 108] = 9.9$$

$\therefore\ S = 3.15$

The test statistic is $t = \dfrac{\bar{x} - \bar{y}}{S\sqrt{\dfrac{1}{n_2} + \dfrac{1}{n_2}}} = \dfrac{15 - 14}{3.15\sqrt{\dfrac{1}{10} + \dfrac{1}{12}}} = 0.74$

$\therefore$ Calculated value of $|t| = 0.74$

Tabulated value of t for 20 d.f. $(n_1 + n_2 - 2)$ at 5% level of significance is 2.086.

Since calculated value of $t <$ tabulated value of t, we accept the null hypothesis H_0 : *i.e.*, $\mu_1 = \mu_2$

Hence from (*i*) and (*ii*), the given samples have been drawn from the same normal population. Hence we accept the null hypothesis that $\mu_1 = \mu_2$ and $\sigma_1^2 = \sigma_2^2$.

Example 4 : It is known that the mean diameters of rivets produced by two firms A and B are practically the same, but the standard deviations may differ. For 22 rivets produced by firm A, the S.D is 2.9 mm, while for 16 rivets manufactured by firm B, the S.D is 3.8 mm. Compute the statistic you would use to test whether the products of firm A have the same variability as those of firm B and test its significance.

Solution : Let the Null Hypothesis be $H_0 : \sigma_1^2 = \sigma_2^2$

i.e., the variances of sources are equal.

The Alternative Hypothesis is $H_1 : \sigma_1^2 \neq \sigma_2^2$

Given $n_1 = 22$, $n_2 = 16$, $s_1 = 2.9$ mm, $s_2 = 3.8$ mm

Since the S.D's of the samples s_1 and s_2 are given,

$\therefore$ The population variances S_1^2 and S_2^2 are obtained by using the relations

$$n_1 s_1^2 = (n_1 - 1)S_1^2 \text{ and } n_2 s_2^2 = (n_2 - 1) S_2^2$$

$$i.e.\ S_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = \frac{22 \times (2.9)^2}{21} = 8.805 \text{ and}$$

$$S_2^2 = \frac{n_2 s_2^2}{n_2 - 1} = \frac{16 \times (3.8)^2}{15} = 15.393$$

Since $S_2^2 > S_1^2$, the test statistic is $F = \dfrac{S_2^2}{S_1^2}$

$$i.e.\ F = \frac{15.393}{8.805} = 1.74 \text{ } i.e., \text{ calculated } F = 1.74$$

Tabulated value of F with (15, 21) d.f at 5% level of significance is 2.20.

Since calculated value of $F <$ tabulated value of F, we accept the null hypothesis H_0, i.e., the products of both the firms A and B have the same variability. So we may conclude that the products of firm A are not superior to those of firm B.

Example 5 : Pumpkins were grown under two experimental conditions. Two random samples of 11 and 9 pumpkins, show the sample standard deviations of their weights as 0.8 and 0.5 respectively. Assuming that the weight distributions are normal, test hypothesis that the true variances are equal. **[JNTU (A) Nov. 2011, (H) May 2011 (Set No. 4)]**

Solution : Let the Null Hypothesis be $H_0 : \sigma_1^2 = \sigma_2^2$ *i.e.*, the variances of sources are equal.

The Alternative Hypothesis is $H_1 : \sigma_1^2 \neq \sigma_2^2$

Given $n_1 = 11,\ n_2 = 9,\ S_1 = 0.8,\ S_2 = 0.5$

Here samples S.D's are given.

We have to determine the population variances S_1^2 and S_2^2 by using the relations

$$n_1 . s_1^2 = (n_1 - 1)\, S_1^2 \text{ and } n_2 . s_2^2 = (n_2 - 1)\, S_2^2$$

$$\Rightarrow S_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = \frac{11 \times (0.8)^2}{10} = 0.704$$

$$\text{and } S_2^2 = \frac{n_2 s_2^2}{n_2 - 1} = \frac{9 \times (0.5)^2}{8} = 0.281$$

$$\text{The test statistic is } F = \frac{S_1^2}{S_2^2} = \frac{0.704}{0.281}$$

$$= 2.5 \quad [\because S_1^2 > S_2^2]$$

$\therefore$ Calculated $F = 2.5$

Tabulated value of F for (10, 8) d.f at 5% level of significance is 3.35.

Since calculated $F <$ tabulated F, we accept the null hypothesis H_0 at 5% level with d.f. (10,8) and conclude that the variances of the two populations is the same, and therefore, the two samples have the same variance

Example 6 : Time taken by the workers in performing a job by method I and method II is given below :

Method I	20	16	26	27	23	22	–
Method II	27	33	42	35	32	34	38

Do the data show that the variances of time distribution from population from which these samples are drawn do not differ significantly ?

Solution : Let the Null Hypothesis be $H_0 : \sigma_1^2 = \sigma_1^2$ where σ_1^2 and σ_2^2 are the variances of the two populations from with the samples are drawn.

The Alternative Hypothesis is $H_1 : \sigma_1^2 \neq \sigma_2^2$.

Calculation of Sample Variances

x	$x-\bar{x}$	$(x-\bar{x})^2$	y	$y-\bar{y}$	$(y-\bar{y})^2$
20	−2.3	5.29	27	−7.4	54.76
16	−6.3	39.69	33	−1.4	1.96
26	3.7	13.69	42	7.6	57.76
27	4.7	22.09	35	0.6	0.36
23	0.7	0.49	32	−2.4	5.76
22	−0.3	0.09	34	−0.4	0.16
			38	3.6	12.96
134		81.34	241		133.72

Given $n_1 = 6, n_2 = 7$

$$\therefore \bar{x} = \frac{\Sigma x}{n_1} = \frac{134}{6} = 22.3,$$

$$\bar{y} = \frac{\Sigma y}{n_2} = \frac{241}{7} = 34.4$$

and $\Sigma(x_i - \bar{x})^2 = 81.34$, $\Sigma(y_i - \bar{y})^2 = 133.72$

If S_1^2 and S_2^2 be the estimates of σ_1^2 and σ_2^2, then

$$S_1^2 = \frac{\Sigma(x_i - \bar{x})^2}{n_1 - 1} = \frac{81.34}{5} = 16.26$$

and $$S_2^2 = \frac{\Sigma(y_i - \bar{y})^2}{n_2 - 1} = \frac{133.72}{6} = 22.29$$

Let H_0 be true.

Since $S_2^2 > S_1^2$, the statistic is

$$F = \frac{S_2^2}{S_1^2} = \frac{22.29}{16.268} = 1.3699 \simeq 1.37$$

Tabulated value of F for (n_1-1, n_2-1) d.f. *i.e.*, (5, 6) d.f. at 5% level significance is 4.39.

Since calculated $F <$ tabulated F, we accept the null hypothesis H_0 at 5% level of significance *i.e.*, there is no significant difference between the variances of the time distribution by the workers.

Example 7 : The nicotine contents in milligrams in two samples of tobacco were found to be as follows :

Sample A	24	27	26	21	25	–
Sample B	27	30	28	31	22	36

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

[JNTU (A) Nov. 2011 (Set No. 4)]

Solution :

To test whether the two samples came from same normal population, we have to test (i) the equality of variances by using F-test (ii) the equality of means by using student's t-test.

Given $n_1 = 5$, $n_2 = 6$

Calculation for mean's and S.D's of the samples.

x	$x - \bar{x}$	$(x - \bar{x})^2$	y	$y - \bar{y}$	$(y - \bar{y})^2$
24	0.6	0.36	27	–2	4
27	2.4	5.76	30	1	1
26	1.4	1.96	28	–1	1
21	3.6	12.96	31	2	4
25	0.4	0.16	22	–7	49
			36	7	49
123		21.2	174		108

$$\therefore \bar{x} = \frac{\Sigma x}{n_1} = \frac{123}{5} = 24.6, \quad \bar{y} = \frac{\Sigma y}{n_2} = \frac{174}{6} = 29$$

$$\Sigma(x_i - \bar{x})^2 = 21.2, \ \Sigma(y_i - \bar{y})^2 = 108$$

$$\therefore \quad S_1^2 = \frac{\Sigma(x_i - \bar{x})^2}{n_1 - 1} = \frac{21.2}{4} = 5.3$$

$$\text{and } S_2^2 = \frac{\Sigma(y_i - \bar{y})^2}{n_2 - 1} = \frac{108}{5} = 21.6$$

(i) *F-Test :*

Null Hypothesis H_0 : $\sigma_1^2 = \sigma_2^2$

The test statistic is $F = \dfrac{S_2^2}{S_1^2} = \dfrac{21.6}{5.3} = 4.075$

$\therefore$ Calculated $F = 4.075$

Tabulated value of F for $(n_2 - 1, n_1 - 1) = (5, 4)$ d.f at 5% level of significance is 6.26. Since calculated $F <$ tabulated F, we accept the null hypothesis H_0 *i.e.*, the variances are equal.

i.e. $\sigma_1^2 = \sigma_2^2$

(ii) ***Student's t-test :*** The two means are equal i.e., $\mu_1 = \mu_2$

Given $\bar{x} = 24.6$, $\bar{y} = 29$

$\Sigma(x_i - \bar{x})^2 = 21.2$, $\Sigma(y_i - \bar{y})^2 = 108$

$$\text{Now } S^2 = \frac{1}{n_1 + n_2 - 2}[\Sigma(x_i - \bar{x})^2 + \Sigma(y_i - \bar{y})^2]$$

$$= \frac{1}{5+6-2}[21.2 + 108] = 14.35$$

$\therefore S = 3.78$

$$\text{The test statistic is } t = \frac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{24.6 - 29}{3.78\sqrt{\frac{1}{5} + \frac{1}{6}}} = -1.92$$

Calculated value of $|t| = 1.92$

Tabulated value of t for 9 d.f at 5% level of significance is 2.26.

Since calculated value of t < tabulated value of t, we accept H_0 i.e., the two means are equal.

$\therefore$ From (*i*) and (*ii*) we conclude that the two samples came from same normal population.

Example 8 : The measurements of the output of two units have given the following results. Assuming that both samples have been obtained from the normal populations at 10% significant level, test whether the two populations have the same variance.

Unit - A	*14.1*	*10.1*	*14.7*	*13.7*	*14.0*
Unit - B	*14.0*	*14.5*	*13.7*	*12.7*	*14.1*

[JNTU 2007, 2008S (Set No. 4)]

Solution : Let the Null Hypothesis be $H_0 : \sigma_1^2 = \sigma_2^2$

i.e., the variances of sources are equal.

The Alternative Hypothesis is $H_1 : \sigma_1^2 \neq \sigma_2^2$

Given $n_1 = 5$, $n_2 = 5$

$$\text{Now } \bar{x} = \frac{\Sigma x}{5} = \frac{1}{5}(14.1 + 10.1 + 14.7 + 13.7 + 14.0)$$

$$= \frac{66.6}{5} = 13.32$$

$$\text{and } \bar{y} = \frac{\Sigma y}{5} = \frac{1}{5}(14.0 + 14.5 + 13.7 + 12.7 + 14.1)$$

$$= \frac{69}{5} = 13.8$$

Computing Standard Deviations of the samples.

x	$x-\bar{x}$	$(x-\bar{x})^2$	y	$y-\bar{y}$	$(y-\bar{y})^2$
14.1	0.78	0.6084	14.0	0.2	0.04
10.1	−3.22	10.3684	14.5	0.7	0.49
14.7	1.38	1.9044	13.7	−0.1	0.01
13.7	0.38	0.1444	12.7	−1.1	1.21
14.0	0.68	0.4624	14.1	0.3	0.09
66.6	·	13.488	69		1.84

If S_1^2 and S_2^2 be the estimates of σ_1^2 and σ_2^2 then

$$S_1^2 = \frac{\Sigma(x_i-\bar{x})^2}{n_1-1} = \frac{13.488}{4} = 3.372$$

$$S_2^2 = \frac{\Sigma(y_i-\bar{y})^2}{n_2-1} = \frac{1.84}{4} = 0.46$$

Let H_0 be true. Since $S_1^2 > S_2^2$, the test statistic is

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

$$i.e., \quad F = \frac{3.372}{0.46} = 7.33$$

$\therefore$ Calculated $F = 7.33$

Degrees of freedom are (n_1-1, n_2-1) i.e., $(5-1, 5-1)$ or $(4,4)$

Tabulated value of F for (4, 4) d.f. at 10% level of significance is 6.39.

Since calculated $F >$ tabulated F, we reject the null hypothesis H_0.

i.e. There is significant difference between the variances.

Example 9 : The following random samples are measurements of the heat-producing capacity (in millions of calories per ton) of speciments of coal from two mines :

Mine 1:	8,260	8,130	8,350	8,070	8,340	---
Mine 2:	7,950	7,890	7,900	8,140	7,920	7,840

Use the 0.02 level of significance to test whether it is reasonable to assume that the variances of the two populations samples are equal.

[JNTU (K) 2009 (Set No. 1)]

Solution : Here $n_1 = 5$ and $n_2 = 6$

Calculation for mean's and S.D's of the samples:

	Mine1			Mine 2		
	x	$(x-\bar{x}) = (x-8230)$	$(x-\bar{x})^2$	y	$(y-\bar{y}) = (y-7940)$	$(y-\bar{y})^2$
	8260	30	900	7950	10	100
	8130	-100	10000	7890	-50	2500
	8350	120	14400	7900	-40	1600
	8070	-160	25600	8140	200	40000
	8340	110	12100	7920	-20	400
Total	41150		63000	7840	-100	10000
				47640		54600

Now $\bar{x} = \frac{\Sigma x}{n_1} = \frac{41150}{5} = 8230$, $\bar{y} = \frac{\Sigma y}{n_2} = \frac{47640}{6} = 7940$, $\Sigma(x-\bar{x})^2 = 63000$

and $\Sigma(y-\bar{y})^2 = 54600$

$$\therefore S_1^2 = \frac{\Sigma(x-\bar{x})^2}{n_1-1} = \frac{63000}{4} = 15750;\ S_2^2 = \frac{\Sigma(y-\bar{y})^2}{n_2-1} = \frac{54600}{5} = 10920$$

1. Null Hypothesis $H_0 : S_1^2 = S_2^2$

2. Alternative Hypothesis $H_1 : S_1^2 \neq S_2^2$

3. Level of significance : $\alpha = 0.02$

4. The test statistic, $F = \frac{S_1^2}{S_2^2}$ $(\because S_1^2 > S_2^2)$

i.e., $F = \frac{15750}{10920} = 1.44$

$\therefore$ Calculated F = 1.44

5. Conclusion : Tabulated value of F for (4,5) degrees of freedom at 5% level of significance is 5.19. So at 2% level of significance F(4,5) > 5.19 (value of F at 5% significance is lower than at 2%). Since calculated F< tabulated F, we accept the null hypothesis H_0 *i.e.,* the variances of the two populations samples are equal.

Example 10 : In one sample of 10 observations, the sum of the squares of the deviations of the sample values from sample mean was 120 and in the other sample of 12 observations, it was 314. Test whether the difference is significant at 5% level ?

[JNTU (H) Nov. 2010, Apr. 2012 (Set No. 2)]

Solution : Let the Null Hypothesis be $H_0 : \sigma_1^2 = \sigma_2^2$, where σ_1^2, σ_2^2 are the variances of the two populations from which the samples are drawn.

The Alternative Hypothesis is $H_1 : \sigma_1^2 \neq \sigma_2^2$

We have $S_1^2 = \dfrac{\Sigma(x-\bar{x})^2}{n_1 - 1}$ and $S_2^2 = \dfrac{\Sigma(y-\bar{y})^2}{n_2 - 1}$

Here $n_1 = 10$, $\Sigma(x-\bar{x})^2 = 120$, $n_2 = 12$, $\Sigma(y-\bar{y})^2 = 314$

$\therefore S_1^2 = \dfrac{120}{10-1} = 13.33$ and $S_2^2 = \dfrac{314}{12-1} = 28.54$

Let H_0 be true. Since $S_2^2 > S_1^2$, the test statistic is

$$F = \frac{S_2^2}{S_1^2} = \frac{28.54}{13.33} = 2.14$$

Degrees of freedom are $v_1 = n_1 - 1 = 10 - 1 = 9$ and $v_2 = n_2 - 1 = 12 - 1 = 11$.

The table value of F at 5% level for $v_1 = 9$ and $v_2 = 11$ degrees of freedom is 2.90 and $2.14 < 2.9$. Since the calculated value of F < the table value, we accept the Null Hypothesis H_0 and conclude that the difference is not significant at 5% level.

Example 11 : An instructor has two classes A and B, in a particular subject. Class A has 16 students while class B has 25 students. On the same examination, although there was no significant difference in mean grades, class A has a standard deviation of 9 while class B had a standard deviation of 12. Can conclude at the 0.01 level of significance that the variability of class B is greater than that of A?

[JNTU (K) May 2010 (Set No. 2)]

Solution : Let the Null Hypothesis be $H_0 : \sigma_1^2 = \sigma_2^2$ *i.e.*, the variances of sources A and B are equal.

The Alternative Hypothesis is $H_1 : \sigma_1^2 > \sigma_2^2$

We have $S_1^2 = \left(\dfrac{n_1}{n_1 - 1}\right) s_1^2$ and $S_2^2 = \left(\dfrac{n_2}{n_2 - 1}\right) s_2^2$

Here $n_1 = 16, s_1 = 9$ and $n_2 = 25, s_2 = 12$

$$\therefore S_1^2 = \left(\frac{16}{16-1}\right) \times 81 = \frac{16}{15} \times 81 = 86.4$$

and $$S_2^2 = \left(\frac{25}{25-1}\right) \times 144 = \frac{25}{24} \times 144 = 150$$

Assuming that H_0 is true, the test statistic is

$$F = \frac{S_2^2}{S_1^2} \left[\because S_2^2 > S_1^2\right]$$

$$= \frac{150}{86.4} = 1.736$$

Degrees of freedom are $v_1 = n_1 - 1 = 16 - 1 = 15$ and $v_2 = n_2 - 1 = 25 - 1 = 24$

The table value of F with d.f. $v_1 = 15, v_2 = 24$ at 1% level is 2.89.

Since the calculated value of F < the tabulated value, we accept the Null Hypothesis H_0 at 1% level with $v_1 = 15, v_2 = 24$ and conclude that the two variances of A and B are the same.

Example 12 : Two independent samples of 8 and 7 items respectively had the following values of the variables.

[JNTU (A) April 2012 (Set No. 3)]

Sample I	9	11	13	11	16	10	12	14
Sample II	11	13	11	14	10	8	10	–

Do the estimates of the population variance differ significantly.

Solution : Let the Null Hypothesis be $H_0 : \sigma_1^2 = \sigma_2^2$ where σ_1^2 and σ_2^2 are the variances of the two populations from which the samples are drawn.

Then the Alternative Hypothesis is $H_1 : \sigma_1^2 \neq \sigma_2^2$

Calculation of Sample Variances

x	$x-\bar{x}$ $= x-12$	$(x-\bar{x})^2$	y	$y-\bar{y}$ $= y-11$	$(y-\bar{y})^2$
9	–3	9	11	0	0
11	–1	1	13	2	4
13	1	1	11	0	0
11	–1	1	14	3	9
16	4	16	10	–1	1
10	–2	4	8	–3	9
12	0	0	10	–1	1
14	2	4	–	–	–
96		36	77		24

Given $n_1 = 8$ and $n_2 = 7$

$$\therefore \bar{x} = \frac{\Sigma x}{n_1} = \frac{96}{8} = 12 \text{ and } \bar{y} = \frac{\Sigma y}{n_2} = \frac{77}{7} = 11$$

If S_1^2 and S_2^2 be the estimates of σ_1^2 and σ_2^2, then

$$S_1^2 = \frac{\Sigma(x-\bar{x})^2}{n_1 - 1} = \frac{36}{8-1} = \frac{36}{7} = 5.14 ; \quad S_2^2 = \frac{\Sigma(y-\bar{y})^2}{n_2 - 1} = \frac{24}{7-1} = \frac{24}{6} = 4$$

The test statistic is $F = \frac{S_1^2}{S_2^2}$ $[\because S_1^2 > S_2^2]$

$$= \frac{5.14}{4} = 1.285$$

The table value of F for $(n_1 - 1, n_2 - 1)$ d.f. *i.e.*, (7,6) d.f. at 5% level of significance is 4.21.

Since, the calculated value of F < the tabulated value of F at 5% level with (7,6) d.f., we accept the Null Hypothesis H_0 and conclude that there is no significant difference between the variances.

9.9 CHI-SQUARE (χ^2) TEST

When a fair coin is tossed 100 times, the theoretical considerations leads us toe expect 50 heads and 50 tails. But in practice, these results are rarely achieved i.e., the results obtained in an experiment do not agree exactly with the theoretical results. The magnitude of discrepancy between the theory and observation is given by the quantity χ^2 (a Greek letter, pronounced as "chi-square"). If $\chi = 0$, the observed and expected frequencies completely coinside. As the value of χ^2 increases, the discrepancy between the observed and theoretical frequencies increases. Thus, χ^2 affords a measure of the correspondence between theory and observation.

Definition : If a set of events $A_1, A_2, \ldots, A_n$ are observed to occur with frequencies $O_1, O_2, \ldots, O_n$ respectively and according to probability rules $A_1, A_2, \ldots, A_n$ are expected to occur with frequencies $E_1, E_2, \ldots, E_n$ respectively with $O_1, O_2, \ldots, O_n$ are called observed frequencies and $E_1, E_2, \ldots, E_n$ are called expected frequencies.

If O_i $(i = 1, 2, \ldots, n)$ is a set of observed (experimental) frequencies and E_i $(i = 1, 2, \ldots, n)$ is the corresponding set of expected (theoretical) frequencies, then χ^2 is defined as $\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$ with $(n-1)$ degrees of freedom.

χ^2 is used to test whether differences between observed and expected frequencies are significant.

Note : If the data is given in a series of 'n' numbers then degrees of freedom = $n - 1$.

In case of Binomial distribution, d.f. = $n - 1$.

In case of Poisson distribution, d.f. = $n - 2$

In case of Normal distribution, d.f. = $n - 3$

Chi-square Distribution is an important continuous probability distribution and it is used in both large and small tests. In chi-square tests, χ^2 - distribution is mainly used

(*i*) to test the goodness of fit,

(*ii*) to test the independence of attributes,

(*iii*) to test if the population has a specified value of the variance σ^2.

We shall discuss these tests one by one as applications of χ^2 - distribution.

§.10 χ^2 TEST AS A TEST OF GOODNESS OF FIT

We use this test to decide whether the discrepancy between theory and experiment is significant or not i.e., to test whether the difference between the theoretical and observed values can be attributed to chance or not.

In other words, if the two curves *i.e.*, observed frequency curve and the expected frequency curve, are drawn, then the Chi-square statistic may be used to determine whether the two curves so drawn are fitted good or not.

Thus, the term goodness of fit is used to test the concordence of the fitness of these two curves. In this case we have only one variable

$\therefore$ No. of d.f. *i.e.*, $\nu = n-1$.

Let the Null Hypothesis H_0 be that there is no significant difference between the observed values and the corresponding expected values.

The the Alternative Hypothesis H_1 is that the above difference is significant.

Let $O_1, O_2, \ldots, O_n$ be a set of observed frequencies and $E_1, E_2, \ldots, E_n$ the corresponding set of expected frequencies. Then the test statistic χ^2 is given by

$$\chi^2 = \sum_{i=1}^{n}\left[\frac{(O_i - E_i)^2}{E_i}\right]$$

$$= \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)}{E_2} + \ldots + \frac{(O_n - E_n)^2}{E_n}$$

Assuming that H_0 is true, the test statistic χ^2 follows Chi-squre distribution with $(n-1)$ d.f., where

$$\sum_{i=1}^{n} O_i = \sum_{i=1}^{n} E_i \quad \text{or} \quad \sum_{i=1}^{n}(O_i - E_i) = 0$$

Conclusion : If the calculated value of $\chi^2 >$ tabulated value of χ^2 at α level, the Null Hypothesis H_0 is rejected. Otherwise, H_0 is accepted.

χ^2 test enables us to assertain how well the theoretical distributions such as Binomial, Poisson, Normal, etc., fit the distributions obtained from sample data.

9.11 CONDITIONS OF VALIDITY

Following are the conditions which should be satisfied before χ^2 test can be applied.

1. The sample observations should be independent.
2. N, the total frequency is large, *i.e.*, > 50.
3. The constraints on the cell frequencies, if any, are linear.
4. No theoretical (or expected) frequency should be less than 10. If small theoretical frequencies occur, the difficulty is overcome by grouping 2 or more classes together before calculating $(O-E)$. Note that the degrees of freedom is determined with the number of classes after regrouping.
5. It is wholly dependent on the degrees of freedom.

$50\% \times \frac{4}{100}$

SOLVED EXAMPLES

Example 1 : The number of automobile accidents per week in a certain community are as follows : 12, 8, 20, 2, 14, 10, 15, 6, 9, 4. Are these frequencies in agreement with the belief that accident conditions were the same during this 10 week period.

[JNTU(H) Dec. 2011 (Set No. 4)]

Solution : Expected frequency of accidents each week $= \frac{100}{10} = 10.$

Null Hypothesis H_0 : The accident conditions were the same during the 10 week period.

Alternative Hypothesis : The accidents conditions are different during the 10 week period.

Observed Frequency (O_i)	Expected Frequency (E_i)	$(O_i - E_i)$	$\frac{(O_i - E_i)^2}{E_i}$
12	10	2	0.4
8	10	−2	0.4
20	10	10	10.0
2	10	−8	6.4
14	10	4	1.6
10	10	0	0.0
15	10	5	2.5
6	10	−4	1.6
9	10	−1	0.1
4	10	−6	3.6
100	100		26.6

Now $\chi^2 = \Sigma \frac{(O_i - E_i)^2}{E_i} = 26.6$ *i.e.*, Calculated $\chi^2 = 26.6$

Here $n = 10$ observations are given

$\therefore$ Degrees of freedom (d.f) $= n-1 = 10-1 = 9$

Tabulated $\chi^2 = 16.9$

Since Calculated $\chi^2 >$ Tabulated χ^2, therefore, the Null Hypothesis is rejected and conclude that the accident conditions were not the same during the 10 week period.

Example 2 : A sample analysis of examination results of 500 students was made. It was found that 220 students had failed, 170 had secured a third class, 90 were placed in second class and 20 got a first class. Do these figures commensurate with the general examination result which is in the ratio of 4 : 3 : 2 : 1 for the various categories respectively.

Solution :

Null Hypothesis H_0 : The observed results commensurate with the general examination results.

Expected frequencies are in the ratio of 4 : 3 : 2 : 1

Total frequency = 500

If we divide the total frequency 500 in the ratio 4 : 3 : 2 : 1, we get the expected frequencies as 200, 150, 100, 50.

Calculations for χ^2

Class	Observed Frequency (O_i)	Expected Frequency (E_i)	$(O_i - E_i)$	$\frac{(O_i - E_i)^2}{E_i}$
Failed	220	200	20	2.00
Third	170	150	20	2.667
Second	90	100	–10	1.000
First	20	50	–30	18.00
	500	500		23.667

$$\text{Calculated } \chi^2 = \Sigma \frac{(O_i - E_i)^2}{E_i} = 23.667$$

Degrees of freedom = 4 – 1 = 3 *i.e.* $\nu = 3$

For $\nu = 3$, $\chi^2_{0.05} = 7.81$

i.e. tabulated value of χ^2 at 5% level for 3 d.f = 7.81

Since calculated $\chi^2_{0.05} >$ tabulated $\chi^2_{0.05}$, we reject the null hypothesis.

i.e. the observed results are not commensurate with the general examination results.

Example 3 : The following figures show the distribution of digits in numbers chosen at random from a telephone directory.

Digits	0	1	2	3	4	5	6	7	8	9
Frequency	1026	1107	997	966	1075	933	1107	972	964	853

Test whether the digits may be taken to occur equally frequently in the directory.

Solution :

Null Hypothesis H_0 : The digits occur equally frequently in the directory.

Alternative Hypothesis H_1 : The digits do not occur equally frequently.

Under this null hypothesis the expected frequencies for each of the digits 0,1,2,....,9 is $\frac{10,000}{10} = 1000$.

We calculate χ^2 from the following table :

Computation Table for χ^2

Digits	Observed Frequency (O_i)	Expected Frequency (E_i)	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
0	1026	1000	676	0.676
1	1107	1000	11449	11.449
2	997	1000	9	0.009
3	966	1000	1156	1.156
4	1075	1000	5625	5.625
5	933	1000	4489	4.489
6	1107	1000	11449	11.449
7	972	1000	784	0.784
8	964	1000	1296	1.296
9	853	1000	21609	21.609
Total	10000	10,000		58.542

Calculated $\chi^2 = \Sigma \frac{(O_i - E_i)^2}{E_i} = 58.542$

Degrees of freedom $= n - 1 = 10 - 1 = 9$

The tabulated value of χ^2 for 9 d.f. at 5% level $= 16.919$

Since calculated $\chi^2 >$ tabulated χ^2, we reject the null hypothesis H_0 and conclude that the digits do not occur equally frequently in the directory.

Example 4 : A die is thrown 264 times with the following results. Show that the die is biased. [Given $\chi^2_{0.05} = 11.07$ for 5 d.f]

[JNTU (H) Dec. 2011 (Set No. 4)]

No. appeared on the die	1	2	3	4	5	6
Frequency	40	32	28	58	54	52

Solution : **Null Hypothesis H_0 :** The die is unbiased.

The expected frequency of each of the numbers 1, 2, 3, 4, 5, 6 is $\frac{264}{6} = 44$

Computation Table for χ^2

Observed frequency (O_i)	Expected frequency (E_i)	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
40	44	16	0.3636
32	44	144	3.2727
28	44	256	5.8181
58	44	196	4.4545
54	44	100	2.2727
52	44	64	1.4545
264	264		17.6362

$$\therefore \chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 17.6362$$

The number of degrees of freedom $= n - 1 = 5$

The tabulated value of χ^2 for 5 d.f at 5% level = 11.07

Since calculated $\chi^2 >$ tabulated χ^2, we reject the null hypothesis H_0.

i.e., we reject the hypothesis that the die is unbiased. Hence the die is biased.

Example 5 : 200 digits were chosen at random from a set of tables. The frequencies of the digits are shown below :

Digit	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
Frequency	*18*	*19*	*23*	*21*	*16*	*25*	*22*	*20*	*21*	*15*

Use the chi-square test to assess the correctness of the hypothesis that the digits were distributed in equal number in the tables from which these were chosen.

Solution :

Null Hypothesis H_0 : The digits were distributed in equal number in the tables.

The expected frequency of each digit $= \frac{200}{10} = 20$

Computation Table for χ^2

Observed frequency (O_i)	Expected frequency (E_i)	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
18	20	4	0.2
19	20	1	0.05
23	20	9	0.45
21	20	1	0.05
16	20	16	0.8
25	20	25	1.25
22	20	4	0.2
20	20	0	0
21	20	1	0.05
15	20	25	1.25
200	200		4.3

$$\therefore \chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 4.3$$

Thus calculated $\chi^2 = 4.3$

Tabulated χ^2 for 10 – 1 = 9 d.f. at 5% level of significance is 16.919. Since calculated χ^2 < tabulated χ^2, we accept the null hypothesis H_0. *i.e.*, the digits are distributed in equal number in the tables.

Example 6 : A pair of dice are thrown 360 times and the frequency of each sum is indicated below :

Sum	2	3	4	5	6	7	8	9	10	11	12
Frequency	8	24	35	37	44	65	51	42	26	14	14

Would you say that the dice are fair on the basis of the chi-square test at 0.05 level of significance ?

[JNTU 2006S, (A) Nov. 2010 (Set No. 2)]

Solution :

1. **Null Hypothesis H_0 :** The dice are fair
2. **Alternative Hypothesis H_1 :** The dice are not fair
3. **Level of significance :** $\alpha = 0.05$

The probabilities of getting a sum 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12 are

$X = x_i$	2	3	4	5	6	7	8	9	10	11	12
$P(x_i)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Computation Table for χ^2

Sum	Observed Frequency (O_i)	Expected Frequency (E_i) $E_i = 360.\ P(x)$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
2	8	10	4	0.4
3	24	20	16	0.8
4	35	30	25	0.833
5	37	40	9	0.225
6	44	50	36	0.72
7	65	60	25	0.417
8	51	50	1	0.02
9	42	40	4	0.1
10	26	30	16	0.53
11	14	20	36	1.8
12	14	10	16	1.6
	360	360		7.445

$$\therefore \quad \chi^2 = \Sigma \frac{(O_i - E_i)^2}{E_i} = 7.445$$

The number of degrees of freedom $= n - 1 = 11$

The tabulated value of χ^2 for 11 d.f. at 5% level of significance is 18.3

Since calculated $\chi^2 <$ tabulated χ^2, we accept the null hypothesis H_0.

i.e., The dice are fair.

Example 7 : 4 coins were tossed 160 times and the following results were obtained.

No. of Heads :	0	1	2	3	4
Observed Frequencies :	17	52	54	31	6

Under the assumption that coins are balanced, find the expected frequencies of 0, 1, 2, 3 or 4 heads, and test the goodness of fit ($\alpha = 0.05$). **[JNTU 2008 (Set No. 3)]**

Solution : Given N = 160, n = 4

p = probability of getting a head $= \frac{1}{2}$

q = probability of getting a tail $= \frac{1}{2}$

The expected frequencies of getting 0, 1, 2, 3, 4 are the successive terms of the binomial expansion $160\ (p + q)^4$

$$= 160\ [p^4 + 4p^3q + 6p^2q^2 + 4pq^3 + q^4]$$

$$= 160\left[\left(\frac{1}{2}\right)^4 + 4.\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right) + 6.\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2 + 4\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^3 + \left(\frac{1}{2}\right)^4\right]$$

$$= 160\left[\frac{1}{16} + \frac{4}{16} + \frac{6}{16} + \frac{4}{16} + \frac{1}{16}\right] = 10 + 40 + 60 + 40 + 10$$

Thus the theoretical or expected frequencies are 10, 40, 60, 40, 10.

Hence $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$

$$= \frac{(17-10)^2}{10} + \frac{(52-40)^2}{40} + \frac{(54-60)^2}{60} + \frac{(31-40)^2}{40} + \frac{(6-10)^2}{10}$$

$$= \frac{1}{120}(588 + 432 + 72 + 243 + 192) = \frac{1527}{120} = 12.725$$

No. of degrees of freedom = 5 – 1 = 4

$\therefore \chi^2_{0.05} = 9.49$ (From tables)

Since the calculated value of χ^2 is greater than $\chi^2_{0.05}$ the hypothesis that the data follows the binomial law is rejected.

Example 8 : Fit a Poisson distribution to the following data and for its goodness of fit at level of significance 0.05?

x	0	1	2	3	4
f	419	352	154	56	19

[JNTU (K) May 2010 (Set No. 3)]

Solution :

x	f	$f.x$
0	419	0
1	352	352
2	154	308
3	56	168
4	19	76
	N = 1000	$\Sigma fx = 904$

Mean $\lambda = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{904}{1000} = 0.904$

Theoretical distribution is given by $N \times p(x) = 1000 \times \frac{e^{-\lambda}\lambda^x}{x!}$

Hence the theoretical frequencies are given by

$$f(x) = \frac{1000\, e^{-0.904}(0.904)^x}{x!} = \frac{1000 \times 0.4049 \times (0.904)^x}{x!} \quad \text{.... (1)}$$

Putting $x = 0,1,2,3,4$, we get

x	0	1	2	3	4	Total
f	404.9 (406.2)	366	165.4	49.8	11.3 (12.6)	997.4

In order that the total observed and expected frequencies may agree, we take first and last theoretical frequencies as 406.2 and 12.6 instead of 404.9 and 11.3 as shown in brackets.

$$\therefore \chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

$$= \frac{(419-406.2)^2}{406.2} + \frac{(352-366)^2}{366} + \frac{(154-165.4)^2}{165.4} + \frac{(56-49.8)^2}{49.8} + \frac{(19-12.6)^2}{12.6}$$

$$= 0.403 + 0.536 + 0.786 + 0.772 + 3.251$$

$$= 5.748$$

Degrees of freedom $= n - 2 = 5 - 2 = 3$

[$\because$ one d.f. being lost because $\Sigma O = \Sigma E$; 1 d.f., is lost because the parameter λ has been estimated]

For $\nu = 3$, we have, $\chi^2_{0.05} = 7.82$

Since, the calculated value of $\chi^2 < \chi^2_{0.05}$, this agreement between the fact and theory is good and hence the Poisson distribution can be fitted to the data.

Example 9 : A survey of 240 families with 4 children each revealed the following distribution.

Male Births	4	3	2	1	0
Observed freq.	10	55	105	58	12

[JNTU (H) Apr. 2012 (Set No. 3)]

Sol. Let the Null Hypothesis be H_0 that male and female births are equally probable.

i.e., $p = q = \frac{1}{2}$.

The expected frequency x of male births is given by

$f(x) = N \times {}^nC_x\, p^x q^{n-x}$, where $N = 240, n = 4, x = 0,1,2,3,4$

$$\therefore f(0) = 240 \times {}^4C_0 \times \left(\frac{1}{2}\right)^0 \times \left(\frac{1}{2}\right)^{4-0} = 240 \times 1 \times 1 \times \frac{1}{16} = 15$$

$$f(1) = 240 \times {}^4C_1 \times \left(\frac{1}{2}\right)^1 \times \left(\frac{1}{2}\right)^{4-1} = 240 \times 4 \times \frac{1}{2} \times \frac{1}{8} = 60$$

$$f(2) = 240 \times {}^4C_2 \times \left(\frac{1}{2}\right)^2 \times \left(\frac{1}{2}\right)^{4-2} = 240 \times 6 \times \frac{1}{4} \times \frac{1}{4} = 90$$

$$f(3) = 240 \times {}^4C_3 \times \left(\frac{1}{2}\right)^3 \times \left(\frac{1}{2}\right)^{4-3} = 240 \times 4 \times \frac{1}{8} \times \frac{1}{2} = 60$$

$$f(4) = 240 \times {}^4C_4 \times \left(\frac{1}{2}\right)^4 \times \left(\frac{1}{2}\right)^{4-4} = 240 \times 1 \times \frac{1}{16} \times 1 = 15$$

The expected or theoretical (Binomial) frequencies of male births are :

x	4	3	2	1	0
$f(x)$	15	60	90	60	15

Let us now apply χ^2 - test to examine the goodness of fit of the given data to the above Binomial distribution.

Calculations for χ^2

No. of families		$O_i - E_i$	$(O_i - E_i)^2$	$(O_i - E_i)^2 / E_i$
Observed frequency (O)	Expected frequency (E)			
10	15	–5	25	1.67
55	60	–5	25	0.42
105	90	15	225	2.5
58	60	–2	4	0.07
12	15	–3	9	0.6
Total 240	240	–	–	5.26

$$\therefore \chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 5.26$$

Degrees of freedom $= n - 1 = 5 - 1 = 4$

The tabulated value of χ^2 at 5% level with 4 d.f. is 9.488. Since the calculated value of $\chi^2 <$ the tabulated value, we accept the Null Hypothesis and conclude that the male and female births are equally probable.

9.12 CHI-SQUARE TEST FOR INDEPENDENCE OF ATTRIBUTES

Definition : Literally, an attribute means a quality or characteristic. Examples of attributes are drinking, smoking, blindness, honesty, beauty etc.

An attribute may be marked by its presence (position) or absence in a number of a given population. Let the observations be classified according to two attributes and the frequencies O_i in the different categories be shown in a two-way table.

called contigency table. We have to test on the basis of cell frequences whether the two attributes are independent or not. We take the Null - Hypothesis H_0 that there is no association between the attributes *i.e.*, we assume that the two attributes are independent. The expected frequencies (E_i) of any cell $= \dfrac{\text{Row total} \times \text{Column total}}{\text{Grand total}}$

The test statistic $\chi^2 = \sum_i \left[\dfrac{(O_i - E_i)^2}{E_i} \right]$ approximately follows Chi-square distribution with d.f. = (No. of rows -1) $\times$ (No. of columns -1)

If the calculated value of χ^2 is less than the table value at a specified level (generally 5%) of significance, the hypothesis holds good i.e., the attributes are independent and do not bear any association. On the other hand, if the calculated value of χ^2 is greater than the table value at a specified level of significance, we say that the results of the experiment do not support the hypothesis, in other words, the attributes are associated.

Let us consider two attributes A and B. A is divided into two classes and B is divided into two classes. The various cell frequencies can be expressed in the following table known as 2×2 contigency table.

A	a	b
B	c	d

a	b	$a+b$
c	d	$c+d$
$a+c$	$b+d$	$N = a+b+c+d$

The expected frequencies are given by

$E(a) = \dfrac{(a+c)(a+b)}{N}$	$E(b) = \dfrac{(b+d)(a+b)}{N}$	$a+b$
$E(c) = \dfrac{(a+c)(c+d)}{N}$	$E(d) = \dfrac{(b+d)(c+d)}{N}$	$c+d$
$a+c$	$b+d$	$N = a+b+c+d$

The value of χ^2 is given by $\chi^2 = \dfrac{N(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$ where $N = a+b+c+d$ with d.f. $= (2-1)(2-1) = 1$. We use this formula when the expected frequencies are in fractions (or decimals).

SOLVED EXAMPLES

Example 1 : On the basis of information given below about the treatment of 200 patients suffering from a disease, state whether the new treatment is comparatively superior to the conventional treatment.

	Favourable	*Not favourable*	*Total*
New	60	30	90
Conventional	40	70	110

Solution :

Null Hypothesis H_0 : No difference between new and conventional treatment (or) New and conventional treatment are independent.

The number of degrees of freedom is $(2-1)(2-1) = 1$

Expected frequencies are given in the table :

$$\text{Expected frequency} = \frac{\text{Row total} \times \text{Column total}}{\text{Grand total}}$$

$\frac{90 \times 100}{200} = 45$	$\frac{90 \times 100}{200} = 45$	90
$\frac{100 \times 110}{200} = 55$	$\frac{100 \times 110}{200} = 55$	110
100	100	200

Calculation of χ^2 :

Observed Frequency (O_i)	Expected Frequency (E_i)	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
60	45	225	5
30	45	225	5
40	55	225	4.09
70	55	225	4.09
200	200		18.18

$$\therefore \chi^2 = \Sigma \frac{(O-E)^2}{E} = 18.18$$

Tabulated χ^2 for 1 d.f. at 5% level of significance is 3.841.

Since calculated $\chi^2 >$ tabulated χ^2 we reject the null hypothesis H_0 i.e., new and ventional treatment are not independent. The new treatment is comparatively erior to conventional treatment.

Example 2 : The following table gives the classification of 100 workers according to and nature of work. Test whether the nature of work is independent of the sex of the ker.

	Stable	*Unstable*	*Total*
Males	40	20	60
Females	10	30	40
Total	50	50	100

[JNTU (K) May 2013 (Set No. 2)]

Solution :

ll Hypothesis H_0 : The nature of work is independent of the sex of the workers.

pected frequencies are given in the table :

$\frac{50 \times 60}{100} = 30$	$\frac{50 \times 60}{100} = 30$	60
$\frac{50 \times 40}{100} = 20$	$\frac{50 \times 40}{100} = 20$	40
50	50	100

lculation of χ^2 :

O_i	E_i	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
40	30	100	3.333
20	30	100	3.333
10	20	100	5.000
30	20	100	5.000
100	100	$\Sigma \frac{(O_i - E_i)^2}{E_i}$	16.66

$$\therefore \chi^2 = \Sigma \frac{(O_i - E_i)^2}{E_i} = 16.66$$

$\therefore$ Calculated $\chi^2 = 16.66$

Tabulated value of χ^2 for $(2 - 1)(2 - 1) = 1$ d.f. at 5% level of significance .84.

Since calculated χ^2 > tabulated χ^2, we reject the null hypthesis H_0, *i..e*, the nature of work is not independent of the sex of the workers.

i.e., there is difference in the nature of work on the basis of sex.

Example 3 : Given the following contigency table for hair colour and eye colour. Find the value of χ^2. Is there good association between the two ?

		Hair colour			
		Fair	***Brown***	***Black***	***Total***
Eye Colour	*Blue*	*15*	*5*	*20*	*40*
	Grey	*20*	*10*	*20*	*50*
	Brown	*25*	*15*	*20*	*60*
	Total	*60*	*30*	*60*	*150*

Solution :

Null Hypothesis H_0 : The two attributes, hair and eye colour are independent.

Table of expected frequencies :

$\frac{60 \times 40}{150} = 16$	$\frac{30 \times 40}{150} = 8$	$\frac{60 \times 40}{150} = 16$	40
$\frac{60 \times 50}{150} = 20$	$\frac{30 \times 50}{150} = 10$	$\frac{60 \times 50}{150} = 20$	50
$\frac{60 \times 60}{150} = 24$	$\frac{30 \times 60}{150} = 12$	$\frac{60 \times 60}{150} = 24$	60
60	30	60	150

Calculation of χ^2 :

Observed Frequency (O_i)	**Expected Frequency (E_i)**	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
15	16	1	0.0625
5	8	9	1.125
20	16	16	1
20	20	0	0
10	10	0	0
20	20	0	0
25	24	1	0.042
15	12	9	0.75
20	24	16	0.666
Total			3.6458

$$\therefore \chi^2 = \Sigma \frac{(O_i - E_i)^2}{E_i} = 3.6458$$

i.e, Calculated $\chi^2 = 3.6458$

Tabulated χ^2 for $(3-1)(3-1) = 4$ d.f. at 5% level of significance is 9.488

Since calculated $\chi^2 <$ tabulated χ^2, we accept the null hypothesis H_0.

i.e., The hair colour and eye colour are independent

i.e., The hair colour and eye colour are not associated.

Example 4 : Two researchers adopted different sampling techniques while investigating e group of students to find the number of students falling into different intelligence level. e results are as follows :

Researchers	Below Average	Average	Above Average	Genius	Total
X	86	60	44	10	200
Y	40	33	25	2	100
Total	126	93	69	12	300

Would you say that the sampling techniques adopted by the two researchers are gnificantly different. [Given 5% value of χ^2 for 2 degrees and 3 degrees of freedoms e 5.991 and 7.82 respectively].

Solution :

ull Hypothesis H_0 : There is no significant difference between the two researchers.

ble of expected frequencies :

$\frac{126\times200}{300} = 84$	$\frac{93\times200}{300} = 62$	$\frac{69\times200}{300} = 46$	$\frac{12\times200}{300} = 8$	200
$\frac{126\times100}{300} = 42$	$\frac{93\times100}{300} = 31$	$\frac{69\times100}{300} = 23$	$\frac{12\times100}{300} = 4$	100
126	93	69	12	300

alculation of χ^2 :

Observed Frequency (O_i)	Expected Frequency (E_i)	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
86	84	4	0.0476
60	62	4	0.0645
44	46	4	0.0869
10	8	4	0.5
40	42	4	0.0952
33	31	4	0.129
25	23	4	0.1739
2	4	4	1
			2.0971

Now $\chi^2 = \Sigma \dfrac{(O_i - E_i)^2}{E_i} = 2.0971$

i.e., Calculated $\chi^2 = 2.0971$

Tabulated χ^2 for $(2-1)(4-1) = 3$ d.f. at 5% level of significance is 7.82.

Since calculated $\chi^2 <$ tabulated χ^2, we accept the null hypothesis H_0.

i.e., There is no significant difference between the two researchers.

Example 5 : Four methods are under development for making discs of a super conducting material. Fifty discs are made by each method and they are checked for super conductivity when cooled with liquid.

	1st Method	*2nd Method*	*3rd Method*	*4th Method*
Super Conductors	31	42	22	25
Failures	19	8	28	25

Test the significant difference between the proportions of super conductors at 0.05 level. **[JNTU 2006S (Set No. 2)]**

Solution : 1. **Null Hypothesis H_0 :**

$p_1 = p_2 = p_3 = p_4$

2. **Alternative Hypothesis H_1 :**

$p_1 \neq p_2 \neq p_3 \neq p_4$

3. Level of significance, $\alpha = 0.05$

4. Computations.

	1st Method	**2nd Method**	**3rd Method**	**4th Method**	**Total**
Super Conductors	31	42	22	25	120
Failures	19	8	28	25	80
Total	50	50	50	50	200

Table of expected frequencies

$\dfrac{50 \times 120}{200} = 30$	$\dfrac{50 \times 120}{200} = 30$	$\dfrac{50 \times 120}{200} = 30$	$\dfrac{50 \times 120}{200} = 30$	120
$\dfrac{50 \times 80}{200} = 20$	$\dfrac{50 \times 80}{200} = 20$	$\dfrac{50 \times 80}{200} = 20$	$\dfrac{50 \times 80}{200} = 20$	80
50	50	50	50	200

Calculation for χ^2

Observed Frequency (O_i)	Expected Frequency (E_i)	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
31	30	1	1/30
42	30	144	144/30
22	30	64	64/30
25	30	25	25/30
19	20	1	1/20
8	20	144	144/20
28	20	64	64/20
25	20	25	25/20
			19.5

$\therefore$ Calculated $\chi^2 = 19.5$

Tabulated χ^2 for $(4-1)(4-1) = 9$ d.f. at 5% level of significance is 7.815.

Since calculated $\chi^2 >$ tabulated χ^2, we reject the null hypothesis H_0.

i.e., There is a significant difference between the proportions.

Example 6 : From the following data, find whether there is any significant liking in the habit of taking soft drinks among the categories of employees. Use Chi-square distribution test with level of significance 0.05.

Employees

Soft Drinks	*Clerks*	*Teachers*	*Officers*
Pepsi	*10*	*25*	*65*
Thumsup	*15*	*30*	*65*
Fanta	*50*	*60*	*30*

[JNTU 2006S (Set No. 2), (A) Dec. 2015, June 2016, Dec. 2017]

Solution :

Null Hypothesis H_0 : There is a significant liking in the habit of taking soft drinks.

Alternative Hypothesis H_1 : There is no significant liking.

Level of significance : $\alpha = 0.05$ (Given)

4. **Computations.**

Soft Drinks	*Clerks*	*Teachers*	*Officers*	*Total*
Pepsi	*10*	*25*	*65*	*100*
Thumsup	*15*	*30*	*65*	*110*
Fanta	*50*	*60*	*30*	*140*
Total	*75*	*115*	*160*	*350*

Table of Expected Frequencies:

$\frac{75\times100}{350} = 21.4$	$\frac{115\times100}{350} = 32.9$	$\frac{160\times100}{350} = 45.71$
$\frac{75\times110}{350} = 23.6$	$\frac{115\times110}{350} = 36.1$	$\frac{160\times110}{350} = 50.3$
$\frac{75\times140}{350} = 30$	$\frac{115\times140}{350} = 46$	$\frac{160\times140}{350} = 64$

Calculations for χ^2

Observed Frequency (O_i)	Expected Frequency (E_i)	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
10	21.4	129.96	6.073
25	32.9	62.41	1.897
65	45.7	372.49	8.151
15	23.6	73.96	3.134
30	36.1	37.21	1.031
65	50.3	216.09	4.3
50	30	400	13.333
60	46	196	4.261
30	64	1156	18.062
			60.2425

$\therefore$ Calculated $\chi^2 = 60.2425$

Tabulated χ^2 for $(3-1)(3-1) = 4$ d.f at 5% level of significance is 9.488.

Since calculated $\chi^2 >$ tabulated χ^2, we reject the null hypothesis H_0.

Thus we conclude that the habit of taking soft drink depends on categories of employees.

Example 7 : In an investigation on the machine performance, the following results are obtained :

	No. of units inspected	No. of defectives
Machine 1	375	17
Machine 2	450	22

whether there is any significant performance of two machines at $\alpha = 0.05$. (or Can we conclude that both the machines are equally efficient at 0.05 level significance).

[JNTU 2004S, 2005S, (K) May 2013 (Set No. 4)]

Solution :

Null Hypothesis H_0 : There is a significant performance of two machines.

Alternative Hypothesis H_1 : There is no significant performance.

Level of significance, $\alpha = 0.05$

Computations.

	No. of units inspected	No. of defectives	Total
Machine 1	375	17	392
Machine 2	450	22	472
Total	825	39	864 = N

Table of expected frequencies :

$\frac{825 \times 392}{864} = 374.3$	$\frac{39 \times 392}{864} = 17.69$
$\frac{825 \times 472}{864} = 450.69$	$\frac{39 \times 472}{864} = 21.3$

Calculation of χ^2 :

Observed Frequency (O_i)	Expected Frequency (E_i)	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
375	374.3	0.49	0.00131
17	17.69	0.4761	0.0269
450	450.69	0.4761	0.0011
22	21.3	0.49	0.023
			0.05231

$$\text{Now } \chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 0.05231$$

i.e., calculated $\chi^2 = 0.05231$

Tabulated χ^2 for $(2-1)(2-1) = 1$ d.f at 5% level of significance is 3.84.

Since calculated $\chi^2 <$ tabulated χ^2, we accept the null hypothesis H_0.

i.e., There is a significant performance of two machines.

Example 8 : 1000 Students at college level were graded according to their I.Q and the economic conditions of their home. Use χ^2 test to find out whether there is any association between condition at home and I.Q ($\alpha = 0.05$).

Economic Condition	I.Q. High	Low	Total
Rich	460	140	600
Poor	240	160	400
Total	700	300	1000

[JNTU April 2009 (Set No. 3)]

Solution : Set up the null hypothesis H_0 :

There is association between condition at Home and IQ.

Table of Expected Frequencies

700 × 600/1000 = 420	300 × 600/100 = 180	600
700 × 400/1000 = 280	300 × 400/1000 = 120	400
700	300	1000

Calculations for χ^2

Observed frequencies (O_i)	Estimated frequencies (E_i)	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
460	420	1600	1600/420 = 3.809
140	180	1600	1600/180 = 8.880
240	280	1600	1600/280 = 5.714
160	120	1600	1600/120 = 13.33

Degrees of freedom = (2 – 1) (2 – 1) = 1

$$\text{Calculated } \chi^2 = \Sigma \frac{(O-E)^2}{E} = 31.733$$

Tabulated χ^2 for $\alpha = 0.05$ for 1 degree of freedom = 3.84

Calculated χ^2 > Tabulated X^2

Hence we reject the null hypothesis.

There is no association between condition at home and IQ.

9.13 CHI-SQUARE TEST FOR POPULATION VARIANCE

Suppose that a random sample x_i ($i = 1, 2, \ldots, n$) is drawn from a normal population with mean μ and variance σ^2. To test the hypothesis that the population variance σ^2 has a specified value σ_0.

Let the Null Hypothesis be $H_0 : \sigma^2 = \sigma_0^2$ (or $\sigma = \sigma_0$)

Then the Alternative Hypothesis is $H_1 : \sigma^2 \neq \sigma_0^2$ (or $\sigma^2 > \sigma_0^2$)

The test statistic is $\chi^2 = \sum \frac{(x_i - \bar{x})^2}{\sigma_0^2} = \frac{ns^2}{\sigma_0^2}$

where $s^2 =$ sample variance $= \frac{\Sigma(x_i - \bar{x})^2}{n}$ and $ns^2 = (n-1)S^2$

Assuming that H_0 is true, the test statistic χ^2 follows Chi-square distribution with $(n-1)$ d.f.

Conclusion : If the calculated value of $\chi^2 >$ the tabulated value, we reject the Null Hypothesis H_0. Otherwise, we accept H_0.

SOLVED EXAMPLES

Example 1 : A random sample of size 20 from a normal population gives a mean of 42 and a variance of 25. Test the hypothesis that the population standard deviation is 8 at 5% level of significance.

Solution : Given $n =$ sample size $= 20$, $\bar{x} =$ sample mean $= 42$ and $s^2 =$ sample variance $= 25$. $\therefore s = 5$

Let the Null Hypothesis be $H_0 : \sigma = \sigma_0 = 8$

Then the Alternative Hypothesis is $H_1 : \sigma \neq 8$

The test statistic is $\chi^2 = \frac{ns^2}{\sigma_0^2} = \frac{20(25)}{(8)^2} = 7.8125$

$\therefore$ Calculated $\chi^2 = 7.8125$

d.f. $= n - 1 = 20 - 1 = 19$

Tabulated value of χ^2 at 5% level with 19 d.f. is 30.14

Since calculated value of $\chi^2 <$ the tabulated value of χ^2, we accept the Null Hypothesis H_0 and conclude that the population standard deviation is 8 at 5% level for 19 d.f.

Example 2 : A firm manufacturing rivets wants to limit variations in their length as much as possible. The lengths (in cms) of 10 rivets manufactured by a new process are

2.15	1.99	2.05	2.12	2.17
2.01	1.98	2.03	2.25	1.93

Examine whether the new process can be considered superior to the old if the old population has standard deviation 0.145 cm ?

[JNTU 2004, 2005, 2006S, (H) Nov. 2010 (Set No. 4)]

Solution: We have

$$n = 10, \ \bar{x} = \frac{\Sigma x_i}{n} = \frac{20.68}{10} = 2.068$$

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n} = \frac{0.09096}{10} = 0.0091$$

and $\sigma_0 = 0.145$

1. **Null Hypothesis H_0 : $\sigma^2 = \sigma_0^2$**
2. **Alternative Hypothesis H_1 : $\sigma^2 > \sigma_0^2$**
3. Level of significance, $\alpha = 0.05$
4. Assuming that H_0 is true, the test statistic is

$$\chi^2 = \frac{ns^2}{\sigma_0^2} = \frac{0.09096}{(0.145)^2} = 4.326 \quad [\because \sigma_0 = 0.145 \text{ (given)}]$$

$\therefore$ Calculated $\chi^2 = 4.8$

d.f. $= n - 1 = 10 - 1 = 9$

Tabulated χ^2 for 9 d.f. at 5% level of significance is 16.919.

Since calculated $\chi^2 <$ tabulated χ^2, we accept the null hypothesis H_0. *i.e.*, The new process cannot be considered superior to the old process.

REVIEW QUESTIONS

1. Explain briefly the Student's 't' - test
2. What is test statistic, 't' to test the difference of means for small samples. **[JNTU (A) June 2016]**
3. Explain briefly the Variance Ratio test (or F - test)
4. When we apply F-test of significance. **[JNTU (A) Dec. 2017]**
5. (*i*) Explain briefly the χ^2 (Chi-Square) test.
   (*ii*) Explain χ^2 - test as a test of Goodness of fit.

EXERCISE 9 (B)

1. A machine which produces mica insulating washers for use in electric devices is set to turn out washers having a thickness of 10 mm. A sample of 10 washers has an average thickness 9.52 mm with a S.D of 0.6 mm. Calculate student's t.
2. Ten cartons are taken at random from an automatic filling machine. The mean net height of the 10 cartons is 11.802 and S.D is 0.15 OZ. Does the sample mean differ significantly from the intended weight of 12OZ ? Given $t_{0.05}$ for 9 d.f = 2.26.

A soap manufacturing company was distributing a particular brand of soap through a large number of retail shops. Before a heavy advertisement campaign, the mean sales per week per shop was 140 dozens. After the campaign, a sample of 26 shops was taken and the mean sales was found to be 147 dozens with S.D 16. Can you consider the advertisement effective.

A random sample of 16 values from a normal population showed a mean of 41.5 inches and the sum of squares of deviations from this mean equal to 135 square inches. Show that the assumption of a mean of 43.5 inches for the population is not acceptable.

Prices of shares of a company on the different days in a month were found to be 66, 65, 69, 70, 69, 71, 70, 63, 63, 64 and 68. Discuss whether the mean price of the shares in the month is 65. **[JNTU (A) Nov. 2011 (Set No. 1)]**

Experience shows that a fixed dose of a certain drug causes an average increase of pulse rate of 10 beats per minute with a S.D of 4. A group of 9 patients given the same dose showed the following increase : 13, 15, 14, 10, 8, 12, 16, 9, 20. Test at 5% level of significance whether this group is different in response to the drug.

Ten specimens of copper wires drawn from a larger lot have the following breaking strength (in kg unit) 578, 572, 570, 568, 572, 578, 570, 572, 569, 548. Test whether the mean breaking strengths of the lot may be taken to be 578 kg weight.

An I.Q test was administered to 5 persons before and after they were trained. The results are given below.

I.Q before training	110	120	123	132	125
I.Q. after training	120	118	125	136	121

Test whether there is any change in I.Q. after the training programme.

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 two salesmen.

No. of sales	20	18
Average sales (in Rs.)	170	205
Standard deviation (in Rs.)	20	25

Memory capacity of 10 students were tested before and after training. State whether the training was effective or not from the following scores.

Before training	12	14	11	8	7	10	3	0	5	6
After training	15	16	10	7	5	12	10	2	3	8

[JNTU (H) Nov. 2010 (Set No. 2), (A) Dec. 2018]

11. In one sample of 10 observations, the sum of the squares of the deviations of the sample values from sample mean was 120 and in the other sample of 12 observations, it was 314. Test whether the difference in variances is significant at 5% in variances level. **[JNTU (H) Nov. 2010 (Set No. 2)]**

12. The random samples were drawn from two normal populations and the following results were obtained.

Sample I	16	17	18	19	20	21	22	24	26	27	–	–
Sample II	19	22	23	25	26	28	29	30	31	32	35	36

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

13. Two random samples drawn from normal populations are :

I	20	16	26	27	23	22	18	24	25	19	–	–
II	27	33	42	35	32	34	38	28	41	43	39	37

Obtain the estimates of the variances of populations. Test whether the two populations have the same variance.

14. A die is thrown 120 times and the frequencies of various faces are as follows :

Face No.	1	2	3	4	5	6
Frequency	10	15	25	25	18	27

Test whether the die was fair.

15. 12 dice were thrown 4096 times and a throw of 6 was reckoned as success. The observed frequencies were as given below.

Success	0	1	2	3	4	5	6	7 and above
Frequencies	447	1145	118	796	380	115	24	8

Find the value of χ^2 on the hypothesis that the dice were unbiased and hence show that the data are consistent with this hypothesis so far as the χ^2 test is concerned.

16. 1000 students at college level were graded according to their I.Q and the economic conditions of their homes. Use χ^2 test to find out whether there is any association between economic conditions at home and I.Q.

	I.Q		
Economic conditions	**High**	**Low**	**Total**
Rich	460	140	600
Poor	240	160	400
Total	700	300	1000

From the following information, state whether the two attributes i.e., condition of house and condition of child are independent.

Condition of child	Condition of House	
	Clean	Dirty
Clean	69	51
Fairly clean	81	20
Dirty	35	44

Measurements of the fat content of two kinds of ice creams, Brand A and Brand B, yielded the following sample data :

Brand A (%)	13.5	14	13.6	12.9	13
Brand B (%)	12.9	13	12.4	13.5	12.7

Test the null hypothesis $\mu_1 = \mu_2$ (where μ_1 and μ_2 are the respective true average fat contents of the two kinds of ice cream) against the alternative hypothesis $\mu_1 \neq \mu_2$ at the level of significance $\alpha = 0.05$ (or) Test the significant between the means at 0.05 level.

[JNTU 2003, 2003S, 2005S (Set No. 3)]

The following are the samples of skills. Test the significant difference between the means at 0.05 level.

Sample - I	74.1	77.7	74.4	74	73.8	–
Sample - II	70.8	74.9	74.2	70.4	69.2	72.2

[JNTU 2007, 2008S (Set No. 2)]

The two random samples reveal the following data :

SampleNo.	Size	Mean	Variance
I	16	440	40
II	25	460	42

[JNTU (K) Nov. 2012 (Set No. 3)]

Test whether the samples come from the same normal population.

(i) Fit a Poisson distribution to the following data and test the goodness of fit.

x	0	1	2	3	4
f	214	92	20	3	1

(*ii*) Fit a Poisson distribution to the following data Test the goodness of fit at a 0.5 level of significance (LOS).

[JNTU (A) June 2019]

x	0	1	2	3	4	5	6	7
$f(x)$	305	366	210	80	28	9	2	1

22. Fit a Binomial distribution to the following data and test for the goodness of fit at 0.05 level of significance.

No. of Heads (x)	0	1	2	3	4	5
No. of Tosses (f)	38	144	342	287	164	25

23. A certain medicine is given to each of the 9 patients resulted in the following increase of blood pressure. Can it be concluded that the medicine will in general be accompanied by an increase in blood pressure using the data : $7, 3, -1, 4, -3, 5, 6, -4, -1$.

$$\left[\textbf{Hint}: t = \frac{\bar{x} - 0}{S/\sqrt{x-1}} \left(\because H_0 : \mu_1 = \mu_2\right)\right]$$

ANSWERS

1. $t = -2.528$.
2. Calculated $t = 4.0$. The sample mean differs significantly from the mean $\mu = 12$ OZ.
3. Calculated $t = 2.19$. The advertising is effective in increasing the sales
5. Calculated $t = 2.3247$; The mean price of the shares in the month is not 65.
6. $H_0 : \mu = 10$, calculated $t = 2.38$, significant
7. Calculated $t = 1.54$, $\mu = 5.78$
8. $t = 0.816$; no change in I.Q
9. $H_0 : \mu_1 = \mu_2$, $t = -4.65$, significant
10. $H_0 : \mu_1 = \mu_2$, $H_1 : \mu_1 \neq \mu_2$, $t = -1.3646$, not significant
11. Calculated $F = 2.14$, not significant
12. $\sigma_1^2 = \sigma_2^2$, $F = 1.935$, not significant
13. $F = 2.14$
14. $\chi^2 = 11.40$, significant
16. $\chi^2 = 31.74$
17. $\chi^2 = 25.64$
19. $|t| = 4.06$
22. $\chi^2 = 8.92$. Binomial distribution gives a good fit at 5% level.

QUIZ

The t - test is applicable to samples for which n is _____

The main use of χ^2 - test is _____

The main use of F - test is _____

$t_{1-\alpha} =$ _____

(a) t_α (b) $-t_\alpha$ [ ]

(c) $t_{\alpha-1}$ (d) None

$F_{1-\alpha}(\nu_1, \nu_2) =$ _____

(a) $F_\alpha(\nu_2, \nu_1)$ (b) $\frac{1}{F_\alpha(\nu_1, \nu_2)}$ [ ]

(c) $F_\alpha(\nu_1, \nu_2)$ (d) None

The shape of t-distribution is similar to that of

(a) Chi-square distribution (b) F-distribution [ ]

(c) Normal distribution (d) None

If $\bar{x} = 47.5$, $\mu = 42.1$, $S = 8.4$ and $n = 25$ then t is

(a) 3.21 (b) 4.5 [ ]

(d) 3.12 (d) None

Chi-square distribution curve varies from

(a) $-\infty$ to ∞ (b) 0 to ∞ [ ]

(c) $-\infty$ to 0 (d) None

Which distribution is used to test the equality of population means [ ]

(a) Chi-square distribution (b) F-distribution

(c) t-distribution (d) None

If two independent random samples of size $n_1 = 13$ and $n_2 = 7$, are taken from a normal population. The variance of the first sample will be atleast four times as that of a second sample then F is [ ]

(a) $\frac{1}{4}$ (b) 4 (c) 16 (d) None

A random variable X has F - distribution with (m,n) degrees of freedom, then $1/X$ has the same distribution with ___ degrees of freedom. [ ]

(a) (m,n) (b) (n,m)

(c) $\left(\frac{1}{\ }, \frac{1}{\ }\right)$ (d) $\left(\frac{1}{n}, \frac{1}{m}\right)$

12. Range of F - distribution is [ ]
(a) $(-\infty,\infty)$ (b) $(-\infty,0)$
(c) $(0,\infty)$ (d) None

13. Range of t - distribution is [ ]
(a) $(-\infty,\infty)$ (b) $(-\infty,0)$
(c) $(0,\infty)$ (d) None

14. In a t - distribution of sample size n, the degrees of freedom are [ ]
(a) n (b) $n-1$
(c) $n-2$ (d) $n+1$

15. Chi-square distribution is
(a) continuous (b) symmetrical
(c) multimodal (d) None

16. If χ_1^2 and χ_2^2 are two independent distributions with m and n degrees of freedom respectively then $\chi_1^2+\chi_2^2$ will be chi-squared distribution with _____ degrees of freedom. [ ]
(a) mn (b) m/n
(c) $m-n$ (d) $m+n$

17. The mean and variance of a χ^2 distribution with 8 degrees of freedom are ___ and ____ respectively. [ ]
(a) 7,49 (b) 8,64
(c) 8,16 (d) 7,14

18. If the standard deviation of a Chi-square distribution is 10, then its degrees of freedom is [ ]
(a) 25 (b) 49
(c) 50 (d) None

ANSWERS

1. $n<30$
2. To test whether the deviations of the observed frequencies from the expected frequencies are significant are not.
3. To test the equality of population means by comparing sample variances.

4. *b*	5. *b*	6. *c*	7. *a*
8. *b*	9. *b*	10. *b*	11. *b*
12. *c*	13. *a*	14. *b*	15. *a*
16. *d*	17. *c*	18. *c*	