

Unit - 4

TEST OF HYPOTHESIS FOR SMALL SAMPLES.

Topics included:

- Student t-test (or) t-test
- Hypothesis testing concerning for 1 mean and 2 mean.
- F-test
 - χ^2 -test (chi-square) for goodness of Fit
 - Independence of attributes (quality or characteristics)

Degree of Freedom (df):

The no. of independent variables that can be estimated in a statistical analysis, is called degree of Freedom

$$\boxed{df = D = n - 1}$$

Sample Variance (s^2):

If $x_1, x_2, x_3, \dots, x_n$ be any Random sample of size 'n' drawn from a normal population with Mean (μ) and variance (σ^2), then the Sample Variance is,

$$\boxed{s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}}$$

Student t-test (or) t-test / Test of Hypothesis for 1 Mean

If $x_1, x_2, x_3, \dots, x_n$ be any Random sample of size 'n' drawn from a normal population with mean (μ) and variance (σ^2), then the Test statistics for 1 mean in t-test is given by,

★ i)

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

If s is given.

★ ii)

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

If s is not given.

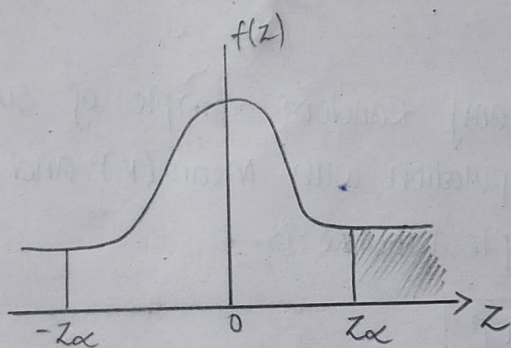
⇒ The Test Statistic for t is a Random variable having t -distribution with degree of Freedom

$$df = v = n-1$$

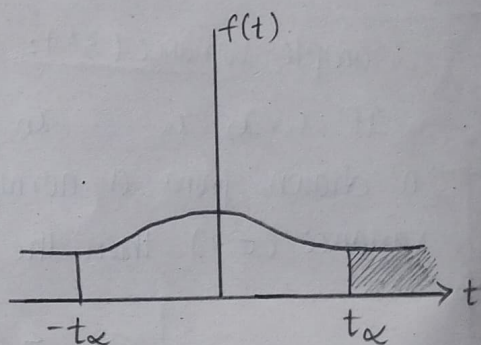
NOTE:

- i) t -distribution is used for testing of hypothesis for small samples and population standard deviation (σ) is not more.
 - For all small samples (s) is required.

2)



Normal Curve



t -distribution Curve

- 3) The Selected values of " t_α " for various values of df can be obtained from the tables of t -distribution, where t_α denotes the area under t -distribution to its right is equal to α . These values are tabulated in tables. In table the left hand column obtains degree of Freedom (df), the column headings are areas α and entries are values of t .

Properties of t-distribution

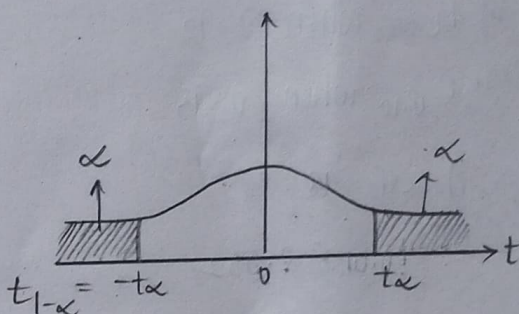
- 1) The shape of t-distribution is bell shaped, which is similar to that of a Normal distribution.
- 2) It is Uni-modal i.e., Mean = Median = Mode
- 3) Total area under the t-distribution curve is 1
- 4) The t-distribution curve is Asymptotic to the t-axis.
- 5) The t-distribution with ν (df) approaches standard Normal distribution as $\boxed{df = \nu = (n-1) \rightarrow \infty}$

NOTE:

- i) a) $-t_{\alpha} = t_{1-\alpha}$
- b) $t_{-\alpha} = -t_{\alpha}$
- c) $t_{-\alpha} = -t_{\alpha} = t_{1-\alpha}$

$$\Rightarrow \boxed{t_{-\alpha} = t_{1-\alpha}}$$

$$\boxed{t_{\alpha} = -t_{1-\alpha}}$$



i) Find

- i) $t_{0.05}$ when $\nu = 16$
- ii) $t_{-0.01}$ when $\nu = 10$
- iii) $t_{0.995}$ when $\nu = 7$

Sol.

i) df $\nu = 16$

$$\boxed{t_{0.05} = 1.746} \text{ from table}$$

ii) df $\nu = 10$

$$t_{-0.01} = -t_{0.01} \rightarrow t_{-\alpha} = -t_{\alpha}$$

$$\boxed{t_{-0.01} = -2.764} \text{ from table.}$$

$$\text{iii) } t_{0.995} =$$

$$df = 7$$

$$\text{when } t_{\alpha} = -t_{1-\alpha}$$

$$\therefore t_{0.995} = -t_{1-0.995}$$

$$= -t_{0.005}$$

$$t_{0.995} = -3.499 \text{ from table.}$$

2) Determine that

$$\text{i) } t_{0.01} \text{ when } D = 18$$

$$\text{ii) } t_{0.05} \text{ when } D = 12$$

$$\text{iii) } t_{-0.10} \text{ when } D = 15$$

Sol. i) $df = 18$

$$t_{0.01} = 2.552$$

$$\text{ii) } df = 12$$

$$t_{0.05} = 1.782$$

$$\text{iii) } df = 15$$

$$t_{-0.10} = -1.341$$

3) Determine

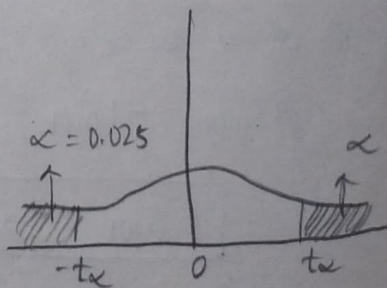
3) Find the t -value with $n=15$ that leads an area 0.025 to the left.

Sol. When $n=15$,

$$df = n-1 \Rightarrow 15-1 = 14$$

$$\alpha = 0.025$$

$$-t_{\alpha} = -t_{0.025} = -2.145 \text{ from table.}$$



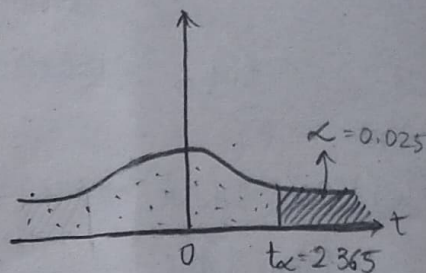
- 4) Find i) $P(t < 2.365)$ when $\nu = 7$
 ii) $P(t > 1.318)$ when $\nu = 24$
 iii) $P(-1.356 < t < 2.179)$ when $\nu = 12$
 iv) $P(t < -2.567)$ when $\nu = 17$

301. i) $P(t < 2.365) =$ Area under t -curve to the left of $t = 2.365$

From table,

When df $\nu = 7$ & $t = 2.365$ then

$$\alpha = 0.025$$



From Figure,

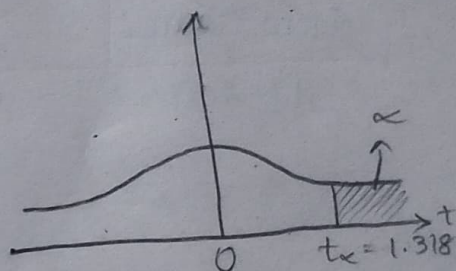
$$P(t < 2.365) = 1 - 0.025 \\ = 0.975 //$$

ii) $P(t > 1.318) =$ Area to the right of $t = 1.318$

From table,

df $\nu = 24$ & $t = 1.318$, then

$$\alpha = 0.1$$



From Figure,

$$P(t > 1.318) = 0.1$$

iii) $P(-1.356 < t < 2.179) =$

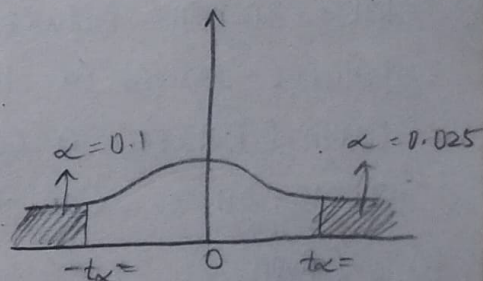
From table,

when df $\nu = 12$, $t = -1.356$, then

$$\alpha = 0.1$$

when df $\nu = 12$, $t = 2.179$, then

$$\alpha = 0.025$$



From figure,

$$P(-1.356 < t < 2.179) = 1 - [0.1 + 0.025] \\ = 0.875$$

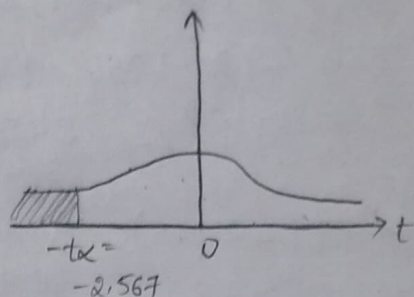
iv) $P(t < -2.567) = \text{Area to the left of } t = -2.567$

From table,

df $\bar{D} = 17$, $t = -2.567$, then

$$\boxed{\alpha = 0.01}$$

$$P(t < -2.567) = 0.01 //$$



5) Find $P(-2.583 \leq t \leq 2.583)$ with 16 df.

From table,

when df $\bar{D} = 16$, $t = -2.583$, then

$$\boxed{\alpha = 0.01}$$

when df $\bar{D} = 16$, $t = 2.583$, then

$$\boxed{\alpha = 0.01}$$

From figure,

$$\begin{aligned} P(-2.583 \leq t \leq 2.583) &= 1 - [0.01 + 0.01] \\ &= 1 - 0.02 \\ &= 0.98 \end{aligned}$$

Problems On 1 mean

- i) An Ambulance service company claims that on an average it takes 20 mins between a call for an ambulance and the patient arrival at the Hospital. If in 6 calls the time taken (between a call and arrival at hospital) are 27, 18, 26, 15, 20, 32. Can the company claim be accepted?

Sol. Given,

$$\mu = 20, n = 6$$

given samples = 27, 18, 26, 15, 20, 32

$$\therefore \bar{x} = \frac{27 + 18 + 26 + 15 + 20 + 32}{6}$$

$$\boxed{\bar{x} = 23}$$

i) N.H.H₀ $\mu = 20$

ii) A.H.H₁ $\mu > 20$ (R.O.T.T)

iii) Choose $\alpha = 0.01 \rightarrow 1 - \alpha = 0.99$

iii) Choose $\alpha = 0.01 = \frac{1}{100} = 0.01$

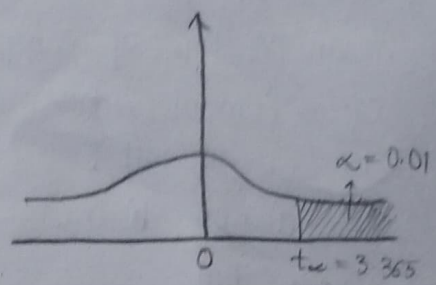
iv) Tabulated t / C.R

$$df = n - 1 = 6 - 1 = 5$$

When $df = 5$, $\alpha = 0.01$

From table

$$t = 3.365$$



v) Test statistics / calculated z

Test statistics for μ mean for small samples =

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \quad (\because s \text{ is not given})$$

sample variance:

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

$$= \frac{(27 - 23)^2 + (18 - 23)^2 + (26 - 23)^2 + (15 - 23)^2 + (20 - 23)^2 + (32 - 23)^2}{6 - 1}$$

$$s^2 = 20.8$$

$$s = 4.560$$

$$\therefore t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{23 - 20}{\frac{4.560}{\sqrt{6}}}$$

$$t_{cal} = 1.611$$

vi) Decision

$$\text{Since } |t_{cal}| = |1.611| = 1.611 < t_{tab} = 3.365$$

Hence N.H.H₀ is Accepted.

2) Mean life time of (mit) computers manufactured by a company is 1120 hrs with S.D of 125 hrs. Test the hypothesis that mean life time of computers has not changed if a sample of 8 computers has a mean life time of 1070 hrs. Is there decrease in mit?

Use a) 0.05 b) 0.01 L.O.S.

Sol. Given that,

$$\mu = 1120, \sigma = 125, n = 8, \bar{x} = 1070$$

a) $0.05 = \alpha$

i) N.H. $H_0 : \mu = 1120$

ii) A.H. $H_1 : \mu < 1120$

iii) L.O.S $\alpha = 0.05$ (L.O.T.T)

iv) C.R / tab t

df $\nu = n - 1 = 8 - 1 = 7$ with $\alpha = 0.05$

Then from table $t = 1.895$

v) Test Statistics / cal t

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}} \quad (\because s \text{ is given})$$

$$= \frac{1070 - 1120}{\frac{125}{\sqrt{8-1}}}$$

$$t_{cal} = -1.058$$

vi) Decision,

Since $|t_{cal}| = |-1.058| = 1.058 < 1.895$

Hence N.H. H_0 is Accepted.

b) 0.01 & 0.05

iv) C.R / tab t

When df $\bar{J} = n-1 = 8-1 = 7$ with $\alpha = 0.01$

Then from table $t = 2.998$

v) Test Statistics / cal t

$$t_{cal} = -1.058$$

vi) Decision,

$$\text{Since } |t_{cal}| = |-1.058| = 1.058 < t_{tab} = 2.998$$

$$1.058 < 2.998$$

Hence N.H.H₀ is Accepted.

T.O.H For 2 Mean / Difference of means

Suppose 2 samples of sizes n_1 and n_2 are drawn from 2 normal populations with same variances $\sigma_1^2 = \sigma_2^2$. We have test the hypothesis that the population means μ_1 and μ_2 are the same.

The Test Statistics for two mean,

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

With df $\bar{J} = n_1 + n_2 - 2$

$$\text{Where } S^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}$$

$$\bar{x}_1 = \frac{\sum x_i}{n_1}, \quad \bar{x}_2 = \frac{\sum x_i}{n_2}$$

$$S_1^2 = \frac{\sum (x_i - \bar{x}_1)^2}{n_1 - 1}, \quad S_2^2 = \frac{\sum (x_i - \bar{x}_2)^2}{n_2 - 1}$$

One Mean

3) Producer of "Gutkha" claims that the Nicotine content in his "Gutkha" on the average is 1.83 mg. Can this claim be accepted if a random sample of 8 "gutkhas" of this type have the Nicotine contents of 2.0, 1.7, 2.1, 1.9, 2.2, 2.1, 2.0, 1.6 mg?

Sol. Given, $\mu = 1.83$
 $n = 8$

$$\bar{x} = \frac{2.0 + 1.7 + 2.1 + 1.9 + 2.2 + 2.1 + 2.0 + 1.6}{8}$$

$$\boxed{\bar{x} = 1.95}$$

i) N.H. H_0 $\mu = 1.83$

ii) A.H. H_1 $\mu > 1.83$ (R.O.T.T)

iii) Choose $\alpha = 1\% = \frac{1}{100} = 0.01$

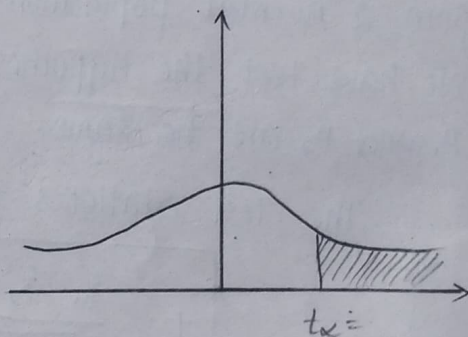
iv) C.R / tab z

$$\text{df } D = n - 1 = 8 - 1 = 7$$

When df $D = 7$ and $\alpha =$

From table

$$t = 1.895$$



v) Test Statistics / cal z :

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \quad (\because s \text{ not given})$$

sample variance:

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

$$\begin{aligned} &= \frac{\begin{matrix} 0.2925 & 0.0625 & 0.0225 \\ (2.0 - 1.95)^2 & (1.7 - 1.95)^2 & (2.1 - 1.95)^2 \end{matrix} + \begin{matrix} (1.9 - 1.95)^2 \\ 0.0025 \end{matrix} + \begin{matrix} (2.2 - 1.95)^2 & (2.1 - 1.95)^2 & (2.0 - 1.95)^2 & (1.6 - 1.95)^2 \\ 0.0625 & 0.0225 & 0.0225 & 0.1225 \end{matrix}}{8-1} \end{aligned}$$

$$s^2 = \frac{3}{70}$$

$$s = \sqrt{\frac{3}{70}}$$

$$s = 0.20702 //$$

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{1.95 - 1.83}{\frac{0.20702}{\sqrt{8}}}$$

$$t = 1.6395 //$$

vi) Decision.

since $|t_{cal}| < t_{tab}$

$$1.6395 < 1.895$$

Hence N.H.H₀ is Accepted.

4) In 1950, in India the mean life expectancy was 50 yrs. If the life expectancies from a random sample of 11 persons are 58.2, 56.6, 54.2, 50.4, 44.2, 61.9, 57.5, 53.4, 49.7, 55.4, 57.0. Does it confirm the expected view.

Given $\mu = 50$

$$n = 11$$

$$\bar{x} = \frac{58.2 + 56.6 + 54.2 + 50.4 + 44.2 + 61.9 + 57.5 + 53.4 + 49.7 + 55.4 + 57.0}{11}$$

$$\bar{x} = 54.40$$

i) N.H.H₀ $\mu = 50$

ii) A.H.H₁ $\mu \neq 50$ (T-T.T)

iii) Choose $\alpha = 0.05$ $\Rightarrow \frac{\alpha}{2} = \frac{0.05}{2}$

iv) C.R $t_{tab} \neq$

$$df = n - 1 \Rightarrow 11 - 1 = 10$$

When $df = 10$ and $\alpha = 0.025$

$$\text{From table } t = 2.228$$

v) Test Statistics / cal t :

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \quad [\because s \text{ is not given}]$$

Sample variance:

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

$$\begin{aligned} \Rightarrow & \frac{(58.2 - 54.40)^2 + (56.6 - 54.40)^2 + (54.2 - 54.40)^2}{11-1} \\ & + (50.4 - 54.40)^2 + (44.2 - 54.40)^2 + (61.9 - 54.40)^2 \\ & + (57.5 - 54.40)^2 + (53.4 - 54.40)^2 + (49.7 - 54.40)^2 \\ & + (55.4 - 54.40)^2 + (57.0 - 54.40)^2 \end{aligned}$$

$$s^2 \Rightarrow 23.607$$

$$\boxed{s = 4.859}$$

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \Rightarrow \frac{54.40 - 50}{\frac{4.859}{\sqrt{11}}}$$

$$\boxed{t = 3.01}$$

vi) Decision:

Since $|t_{cal}| < |t_{tab}|$.

$$3.01 < 2.228 \Rightarrow 3.01 > 2.228$$

Hence N.H.Ho Rejected.

5) An auditor that he takes on an average 10.5 days to file income tax returns (I.T returns). Can this claim be accepted if a random sample shows that he took 13, 19, 15, 10, 12, 11, 14, 18 days to file I.T. returns? Use a) 0.01 b) 0.05 c) 0.5

Given, $\mu = 10.5$
 $n = 8$

$$\bar{x} = \frac{13 + 19 + 15 + 10 + 12 + 11 + 14 + 18}{8}$$

$$\boxed{\bar{x} = 14}$$

i) N.H.Ho $\mu = 10.5$

ii) A.H.H, $\mu \geq 10.5$

iii) L.O.S $\alpha = 0.01 \Rightarrow \frac{0.01}{2} = 0.005$

iv) C.R / tab t

$$\text{df } \bar{D} = n - 1 \Rightarrow 8 - 1 = 7$$

When df $\bar{D} = 7$ and $\alpha = 0.001$

$$\text{From table } \boxed{t = 2.998}$$

v) Test Statistics / cal t:

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \quad [\because s \text{ is not given}]$$

Sample variance:

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

$$= \frac{(13-14)^2 + (19-14)^2 + (15-14)^2 + (10-14)^2 + (12-14)^2 + (11-14)^2 + (14-14)^2 + (18-14)^2}{8-1}$$

$$s^2 = 10.285$$

$$\boxed{s = 3.207}$$

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \Rightarrow \frac{14 - 10.5}{\frac{3.207}{\sqrt{8}}}$$

$$\boxed{t = 3.0869}$$

vi) Decision

Since $|t_{cal}| < t_{tab}$

$$3.0869 < 2.998 \Rightarrow 3.0869 > 2.998$$

Hence $N.H.H_0$ is Rejected.

b) 0.05.

iii) L.O.S $\alpha = 0.05$

iv) C.R / tab t

when df $\bar{D} = n - 1 = 8 - 1 = 7$

df $\bar{D} = 7$ and $\alpha = 0.05$

$$t_{tab} = 1.895$$

v) Test statistics / cal t

$$S = 3.207$$

$$t_{cal} = 3.0869$$

vi) Decision

Since $|t_{cal}| < t_{tab}$

$$3.0869 < 1.895 \Rightarrow 3.0869 > 1.895$$

Hence $N.H.H_0$ is Rejected.

6) If 5 pieces of certain ribbon selected at random have mean breaking strength of 169.5 pounds with S.D of 5.7, do they confirm to the specification mean breaking strength of 180 pounds.

Given, $\mu = 180$

$\mu < 180$

$S = 5.7$

$n = 5$

$\bar{x} = 169.5$

i) N.H. $H_0: \mu = 180$

ii) A.H. $H_1: \mu < 180$

iii) L.O.S $\alpha = 0.05$

iv) C.R / tab t

df $\bar{D} = n - 1 \Rightarrow 5 - 1 = 4$

when $\bar{D} = 4$ and $\alpha = 0.05$.

$$t = 2.015$$

v) T.S / cal z

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}} \quad [S\& \text{ is given}]$$

$$t = \frac{169.5 - 180}{\frac{5.7}{\sqrt{5}}}$$

$$t = -4.119$$

vi) Decision:

Since $|t_{cal}| < t_{tab}$

$$4.119 < 2.015 \Rightarrow 4.119 > 2.015$$

Hence N.H. H_0 is Rejected.

7) In a random sample of 10 bolts produced by a machine the mean length of bolts is 0.53 mm and S.D 0.03 mm. Can we claim from this that the machine is in proper working order if in the past it produced bolts of length 0.50 mm? Use a) 0.005 b) 0.01 L.O.S.

Given, $n = 10$

$S = 0.03$

$\bar{x} = 0.53$

$\mu = 0.50$

a) $\alpha = 0.05$

i) N.H. $H_0: \mu = 0.50$

ii) A.H. $H_1: \mu \neq 0.50$ (T.T.T)

iii) L.O.S $\alpha = 0.05$

iv) tab t

$$\text{df } D = n - 1 \Rightarrow 10 - 1 = 9 \text{ when } \frac{\alpha}{2} = \frac{0.05}{2} \Rightarrow 0.025$$

$$t = 2.262$$

v) T.S / cal t

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}} \Rightarrow \frac{0.53 - 0.50}{\frac{0.03}{\sqrt{10-1}}}$$

$$t = 3$$

vi) Decision

$$\text{Since } |t_{\text{cal}}| < t_{\text{tab}} \Rightarrow 3 < 2.262 \Rightarrow 3 > 2.262$$

Hence N.H.H₀ is Rejected

b) 0.01.

$$\text{iii) } 2 \cdot 0.5 \alpha = 0.01 \Rightarrow \frac{\alpha}{2} = \frac{0.01}{2} = 0.005$$

iv) tab t

$$\text{when df } D = 9 \text{ and } \alpha = 0.005$$

$$t_{\text{tab}} = 3.250$$

v) cal t

$$t_{\text{cal}} = 3$$

vi) Decision

$$\text{Since } |t_{\text{cal}}| < t_{\text{tab}}$$

$$3 < 3.250$$

Hence N.H.H₀ is Accepted.

NOTE:

The two mean t-test cannot be applied if $\sigma_1 \neq \sigma_2$

i) In a mathematics examination 9 students of class A and 6 students of class B obtained the following marks. Test at 0.01 L.O.S whether the performance in mathematics is same or not for the two classes A & B. Assume that the samples are drawn from Normal population having same variance.

A: 44 71 63 59 68 46 69 54 48

B: 52 70 41 62 36 50

Sol. Here $n_1 = 9$, $n_2 = 6$

Mean of A sample = $\frac{\sum x_i}{n_1}$

$$\bar{x}_1 = \frac{44 + 71 + 63 + 59 + 68 + 46 + 69 + 54 + 48}{9}$$

$$\boxed{\bar{x}_1 = 58.4}$$

$$\text{B sample, } \bar{x}_2 = \frac{52 + 70 + 41 + 62 + 36 + 50}{6}$$

$$\boxed{\bar{x}_2 = 51.83}$$

Variance

$$S_1^2 = \frac{\sum (x_i - \bar{x}_1)^2}{n_1 - 1} = \frac{(44 - 58.4)^2 + (71 - 58.4)^2 + \dots + (48 - 58.4)^2}{9 - 1}$$

$$\boxed{S_1^2 = 109}$$

$$S_2^2 = \frac{\sum (x_i - \bar{x}_2)^2}{n_2 - 1} = \frac{(52 - 51.83)^2 + (70 - 51.83)^2 + \dots + (50 - 51.83)^2}{6 - 1}$$

$$\boxed{S_2^2 = 160.96}$$

$$\text{Now } S^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}$$

$$S^2 = \frac{(9 \times 109) + (6 \times 160.96)}{9+6-2}$$

$$S^2 = 149.76$$

$$S = 12.23$$

- i) N.H. Ho $\mu_1 = \mu_2$
- ii) A.H. Ho $\mu_1 \neq \mu_2$ (T.T.T)
- iii) $\alpha = 0.05$ $\alpha = 0.01$
- iv) C.R / tab t

As $\alpha = 0.01$ and $\neq$ type use T.T.T

Since $\alpha = 0.01$

$$\frac{\alpha}{2} = \frac{0.01}{2} = 0.005$$

For the Area 0.005 & df $D = n_1 + n_2 - 2 = 13$
then from table

$$t_{\text{tab}} = 3.012$$

- v) Test Statistics / cal t

Test Statistics for 2 mean is given by,

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

$$t = \frac{58 - 51.83}{12.23 \sqrt{\frac{1}{9} + \frac{1}{6}}}$$

$$t = 0.9566$$

9572

- vi) Decision:

Since $|t_{\text{cal}}| = |0.9566| < t_{\text{tab}} = 3.012$

Hence N.H. Ho is Accepted.

2) Under quality improvement programme some teachers are trained by instruction methodology A and some by methodology B. In a random sample of size 10, taken from a large group of teachers exposed to each of these two methods, the following marks are obtained in an appropriate achievement test.

Method A: 65 69 73 71 75 66 71 68 68 74

Method B: 78 69 72 77 84 70 73 77 75 65.

Sol. Here, $n_1 = 10$, & Assuming that populations sampled are approximately normally -1.8 -1.7 distributed having same variance, Test the claim that method B is more effective, at 0.05 L.O.S.

Mean

Method A

$$\bar{x}_1 = \frac{65 + 69 + 73 + 71 + 75 + 66 + 71 + 68 + 68 + 74}{10}$$

$$\boxed{\bar{x}_1 = 70}$$

Method B.

$$\bar{x}_2 = \frac{78 + 69 + 72 + 77 + 84 + 70 + 73 + 77 + 75 + 65}{10}$$

$$\boxed{\bar{x}_2 = 74}$$

Variance

$$S_1^2 = \frac{\sum (x_i - \bar{x}_1)^2}{n_1 - 1} = \frac{(65-70)^2 + (69-70)^2 + \dots + (74-70)^2}{10-1}$$

$$\boxed{S_1^2 = 11.33}$$

$$S_2^2 = \frac{\sum (x_i - \bar{x}_2)^2}{n_2 - 1} = \frac{(78-74)^2 + (69-74)^2 + \dots + (65-74)^2}{10-1}$$

$$\boxed{S_2^2 = 29.11}$$

$$S^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}$$

$$= \frac{(10 \times 11.33) + (10 \times 29.11)}{10 + 10 - 2}$$

$$S^2 = 22.46$$

$$S = 4.739$$

i) N.H. H_0 $\mu_1 = \mu_2$

ii) A.H. H_0 $\mu_1 \neq \mu_2$

iii) L.O.S $\alpha = 0.05$

iv) C.R / tab $\neq$

$$df \bar{D} = n_1 + n_2 - 2 \Rightarrow 10 + 10 - 2 = 18$$

From table

$$t_{tab} = -1.734$$

v) T.S / cal $\neq$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \Rightarrow \frac{70 - 74}{4.739 \sqrt{\frac{1}{10} + \frac{1}{10}}}$$

$$t_{cal} = -1.8873$$

vi) Decision .

$$\text{since } |cal| < tab \Rightarrow |-1.8873| < -1.734$$

$$1.8873 > -1.734$$

Hence N.H. H_0 is Rejected.

3) Out of random sample of 9 mice suffering with a disease, 5 mice were treated with a new serum while the remaining were not treated. From the time of Commencement of experiment, the following are the survival times.

Treatment : 2.1 5.3 1.4 4.6 0.9

No Treatment : 1.9 0.5 2.8 3.1

Test whether the Serum treatment is effective in curing the disease at 0.05 L.O.S, Assuming that the two distributions are normally distributed with equal variances.

Sol. Here $n_1 = 5$, $n_2 = 4$

Mean of Treatment and No treatment :

$$\bar{x}_1 = \frac{2.1 + 5.3 + 1.4 + 4.6 + 0.9}{5}$$

$$\bar{x}_1 = 2.86$$

$$\bar{x}_2 = \frac{1.9 + 0.5 + 2.8 + 3.1}{4}$$

$$\bar{x}_2 = 2.075$$

Variance of Treatment & No treatment :

$$S_1^2 = \frac{\sum (x_i - \bar{x}_1)^2}{n_1 - 1} = \frac{(2.1 - 2.86)^2 + (5.3 - 2.86)^2 + \dots + (0.9 - 2.86)^2}{5 - 1}$$

$$S_1^2 = 3.883$$

$$S_2^2 = \frac{\sum (x_i - \bar{x}_2)^2}{n_2 - 1} = \frac{(1.9 - 2.075)^2 + (0.5 - 2.075)^2 + (2.8 - 2.075)^2 + (3.1 - 2.075)^2}{4 - 1}$$

$$S_2^2 = 1.3625$$

$$\text{Now } S^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}$$

$$S^2 = \frac{(5 \times 3.883) + (4 \times 1.3625)}{5+4-2}$$

$$S^2 = 3.558$$

$$S = 1.884$$

i) N.H. H_0 $\mu_1 = \mu_2$

ii) A.H. H_1 $\mu_1 > \mu_2$

iii) L.O.S $\alpha = 0.05$

iv) C.R / tab t

$$\text{df } D = n_1 + n_2 - 2 \Rightarrow 5 + 4 - 2 = 7$$

when df $D = 7$ and $\alpha = 0.05$

$$t_{\text{tab}} = 1.895$$

v) T.S / cal t

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \Rightarrow \frac{2.86 - 2.075}{1.884 \sqrt{\frac{1}{5} + \frac{1}{4}}}$$

$$t_{\text{cal}} = 0.6208$$

vi) Decision

Since $|t_{\text{cal}}| < t_{\text{tab}}$

$$0.6208 < 1.90$$

Hence N.H. H_0 is Accepted.

4) 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 L.O.S.

Given, L.O.S = 0.01

$n_1 = 5$, $n_2 = 6$

$$\bar{x}_1 = \frac{8350 + 8070 + 8340 + 8130 + 8260}{5}$$

$$\boxed{\bar{x}_1 = 8230}$$

$$\bar{x}_2 = \frac{7900 + 8140 + 7920 + 7840 + 7890 + 7950}{6}$$

$$\boxed{\bar{x}_2 = 7940}$$

$$S_1^2 = \frac{\begin{matrix} 14400 & 25600 & 12100 \\ (8350 - 8230)^2 + (8070 - 8230)^2 + (8340 - 8230)^2 + \\ 10000 & (8130 - 8230)^2 + (8260 - 8230)^2 \end{matrix}}{5-1}$$

$$\boxed{S_1^2 = 15750}$$

$$S_2^2 = \frac{\begin{matrix} 1600 & 40000 & 400 \\ (7900 - 7940)^2 + (8140 - 7940)^2 + (7920 - 7940)^2 + \\ 10000 & (7840 - 7940)^2 + (7890 - 7940)^2 + (7950 - 7940)^2 \\ & 2500 & 100 \end{matrix}}{6-1}$$

$$\boxed{S_2^2 = 3720}$$

$$S^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}$$

$$= \frac{(5 \times 15750) + (6 \times 3720)}{5+6-2}$$

$$S^2 = 11230$$

$$S = 105.97$$

i) N.H.Ho $\mu_1 = \mu_2$

ii) A.H.H₁ $\mu_1 \neq \mu_2$

iii) L.O.S $\alpha = 0.01 \Rightarrow \frac{\alpha}{2} = \frac{0.01}{2} = 0.005$

iv) tab t

$$\text{df } D = n_1 + n_2 - 2 \Rightarrow 5 + 6 - 2 = 9$$

when df $D = 9$ and $\alpha = 0.005$

$$t_{\text{tab}} = 3.250$$

v) Cal t

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \Rightarrow \frac{8230 - 7940}{105.97 \sqrt{\frac{1}{5} + \frac{1}{6}}}$$

$$t_{\text{cal}} = 4.51$$

vi) Decision

Since $|t_{\text{cal}}| < t_{\text{tab}}$

$$4.51 < 3.25$$

$$4.51 > 3.25$$

Hence N.H.Ho is Rejected.

5) To determine whether vegetarian and Non-vegetarian diets effects significantly on increase in weight a study was conducted yielding the following data of gain in weight

Veg: 34 24 14 32 25 32 30 24 30 31 35 25

Non veg: 22 10 47 31 44 34 22 40 30 32 35 18 21 35 29.

Can we claim that the two diets differ pertaining to weight gain. Assuming that samples are drawn from normal populations with same variance.

Given $n_1 = 12, n_2 = 15$.

$$\bar{x}_1 = \frac{34 + 24 + 14 + 32 + 25 + 32 + 30 + 24 + 30 + 31 + 35 + 25}{12}$$

$$\boxed{\bar{x}_1 = 28}$$

$$\bar{x}_2 = \frac{22 + 10 + 47 + 31 + 44 + 34 + 22 + 40 + 30 + 32 + 35 + 18 + 21 + 35 + 29}{15}$$

$$\boxed{\bar{x}_2 = 30}$$

$$S_1^2 = \frac{(34 - 28)^2 + (24 - 28)^2 + \dots + (25 - 28)^2}{12 - 1}$$

$$S_1^2 = \frac{36 + 16 + 196 + 16 + 9 + 16 + 4 + 16 + 4 + 9 + 49 + 9}{11}$$

$$\boxed{S_1^2 = 34.545}$$

$$S_2^2 = \frac{(22 - 30)^2 + (10 - 30)^2 + \dots + (29 - 30)^2}{14}$$

$$S_2^2 = \frac{64 + 400 + 289 + 1 + 196 + 16 + 64 + 100 + 0 + 4 + 25 + 144 + 81 + 25 + 1}{14}$$

$$\boxed{S_2^2 = 100.714}$$

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

$$s^2 = \frac{(12 \times 34.545) + (15 \times 100.714)}{12 + 15 - 2}$$

$$s^2 = 77.01$$

$$s = 8.775$$

i) N.H.Ho $\mu_1 = \mu_2$

ii) A.H.Ho $\mu_1 \neq \mu_2$

iii) L.O.S $\alpha = 0.01 = \frac{0.01}{2} = \frac{0.05}{2}$

iv) tab t

$$\text{df } D = n_1 + n_2 - 2 \Rightarrow 12 + 15 - 2$$

$$\text{when df } D = 25 \text{ and } \alpha = 0.025$$

$$t_{\text{tab}} = 2.060$$

v) cal t

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \Rightarrow \frac{28 - 30}{8.775 \sqrt{\frac{1}{12} + \frac{1}{15}}}$$

$$t_{\text{cal } t} = -0.5884 \quad (\text{or}) \quad -0.609$$

vi) Decision

$$\text{Since } |t_{\text{cal } t}| < t_{\text{tab } t}$$

$$|-0.5884| < 2.060$$

$$0.609 < 2.060$$

Hence N.H.Ho is Accepted.

F-test :

Suppose if we want to test whether 2 independent samples x_i ($i=1, 2, \dots, n_1$) and x_j ($j=1, 2, \dots, n_2$) have been drawn from the Normal populations with same variance (i.e., $\sigma_1^2 = \sigma_2^2$) (or) Variance differ significantly ($\sigma_1^2 \neq \sigma_2^2$)

For this the Test Statistics,

$$F = \frac{S_1^2}{S_2^2} \quad \text{if } S_1^2 > S_2^2$$

In F-test, N.H.H₀ : $\sigma_1^2 = \sigma_2^2$

Note :

i) We will take greater variance S_1^2 (or) S_2^2 in the Numerator and adjusts the Degrees of Freedom accordingly.

i) If $S_1^2 > S_2^2$ then $F = \frac{S_1^2}{S_2^2}$ with

$$\text{df } \nu_1 = n_1 - 1 \text{ and } \nu_2 = n_2 - 1$$

ii) If $S_2^2 > S_1^2$ then $F = \frac{S_2^2}{S_1^2}$ with

$$\text{df } \nu_1 = n_2 - 1 \text{ and } \nu_2 = n_1 - 1$$

$$2) \quad S_1^2 = \frac{\sum (x_i - \bar{x}_1)^2}{n_1 - 1}$$

$$S_2^2 = \frac{\sum (x_j - \bar{x}_2)^2}{n_2 - 1}$$

Example 1 :

In 1 sample of 8 observations, the sum of the squares of Deviations of the sample values from the sample mean was 84.4 and in another sample of 10 observations it was 102.6. Test whether this difference is significant at 5% level.

Sol.

Given, $n_1 = 8$, $\sum (x_i - \bar{x}_1)^2 = 84.4$

$n_2 = 10$, $\sum (x_j - \bar{x}_2)^2 = 102.6$

$$S_1^2 = \frac{\sum (x_i - \bar{x}_1)^2}{n_1 - 1} \Rightarrow \frac{84.4}{8-1} \Rightarrow 12.057$$

$$S_2^2 = \frac{\sum (x_j - \bar{x}_2)^2}{n_2 - 1} \Rightarrow \frac{102.6}{10-1} \Rightarrow 11.4$$

Here $S_1^2 > S_2^2$

i) N.H.H₀ : $\sigma_1^2 = \sigma_2^2$

ii) A.H.H₀ : $\sigma_1^2 \neq \sigma_2^2$ (T.T.T)

iii) L.O.S $\alpha = 0.05$

iv) tabulated t

As A.H.H₁ is $\neq$ type use T.T.T

Since $\alpha = 0.05$

$$\frac{\alpha}{2} = \frac{0.05}{2} = 0.025$$

For the area 0.025 and df .

$$\begin{matrix} \nu_1 = n_1 - 1 & \text{and} & \nu_2 = n_2 - 1 & (\because S_1^2 > S_2^2) \\ \nu_1 = 7 & & \nu_2 = 9 \end{matrix}$$

then from table

$$\chi^2_{\text{tab}} = 3.29$$

v) Calculated t

Since $S_1^2 > S_2^2$. So

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

$$F_{\text{cal}} = 1.057$$

vi) Decision ,

Since $F_{\text{cal}} = 1.057 < F_{\text{tab}} = 3.29$

Hence N.H.H₀ is Accepted.

Note: In F-test, Use α (given) directly to find tabulated t

8) The Nicotine contents in mg in 2 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 2 samples came from same Normal population. then $\sigma^2 = \sigma^2$

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

$$\bar{x}_1 = \frac{\sum(x_i)}{n_1} = \frac{24 + 27 + 26 + 21 + 25}{5}$$

$$\boxed{\bar{x}_1 = 24.6}$$

$$\bar{x}_2 = \frac{\sum(x_j)}{n_2} = \frac{27 + 30 + 28 + 31 + 22 + 36}{6}$$

$$\boxed{\bar{x}_2 = 29}$$

$$s_1^2 = \frac{\sum(x_i - \bar{x}_1)^2}{n_1 - 1} \Rightarrow \frac{(24 - 24.6)^2 + (27 - 24.6)^2 + \dots + (25 - 24.6)^2}{5 - 1}$$

$$\boxed{s_1^2 = 5.3}$$

$$s_2^2 = \frac{\sum(x_j - \bar{x}_2)^2}{n_2 - 1} \Rightarrow \frac{(27 - 29)^2 + (30 - 29)^2 + \dots + (36 - 29)^2}{6 - 1}$$

$$\boxed{s_2^2 = 21.6}$$

Here $s_2^2 > s_1^2$

i) N.H. H_0 $\sigma_1^2 = \sigma_2^2$

ii) A.H. H $\sigma_1^2 \neq \sigma_2^2$ (T.T.T)

iii) L.O.S $\alpha = 1\% = 0.01$

iv) tabulated t

Since $\alpha = 0.01$

For area 0.01 and for the df

$$d_1 = v_1 = n_2 - 1 = 6 - 1 = 5, \quad d_2 = v_2 = n_1 - 1 = 5 - 1 = 4$$

$$\chi_{\text{tab}} = 15.52$$

v) Calculated t

$$\text{since } S_2^2 > S_1^2$$

$$\text{then } F = \frac{S_2^2}{S_1^2} = \frac{21.6}{5.3}$$

$$F_{\text{cal}} = 4.07$$

vi) Decision

$$\text{Since } F_{\text{cal}} = 4.07 < F_{\text{tab}} = 15.52$$

Hence N.H.H₀ is Accepted.

6) The time taken by 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 variance of Time distribution from population from which these samples are drawn do not differ significantly:

$$\text{Here } n_1 = 6, n_2 = 7$$

$$\bar{x}_1 = \frac{\sum x_1}{n_1} = \frac{20+16+26+27+23+22}{6}$$

$$\bar{x}_1 = 22.33$$

$$\bar{x}_2 = \frac{\sum x_2}{n_2} = \frac{27+33+42+35+32+34+38}{7}$$

$$\bar{x}_2 = 34.428$$

$$S_1^2 = \frac{\sum (x_1 - \bar{x}_1)^2}{n_1 - 1} = \frac{(20-22.33)^2 + (16-22.33)^2 + \dots + (22-22.33)^2}{6-1}$$

$$S_1^2 = 16.266$$

$$S_2^2 = \frac{\sum (x_j - \bar{x}_2)^2}{n_2 - 1} = \frac{(27 - 34.428)^2 + (33 - 34.428)^2 + \dots + (38 - 34.428)^2}{7 - 1}$$

$$S_2^2 = \frac{55.0564 + 2.0164 + 57.4564 + 0.3364 + 5.8564 + 0.1764 + 12.8164}{6}$$

$$S_2^2 = 22.2858$$

$$\text{Here } S_1^2 < S_2^2$$

i) N.H.Ho $\sigma_1^2 = \sigma_2^2$

ii) A.H.H₁ $\sigma_1^2 \neq \sigma_2^2$

iii) L.O.S $\alpha = 0.01$

iv) tab t

$$\text{Since } \alpha = 0.01$$

For area 0.01 and for df

$$d_1 = d_2 = n_1 - 1 = 7 - 1 = 6, \quad d_2 = n_2 - 1 = 6 - 1 = 5$$

$$X_{\text{tab}} = 10.67$$

v) cal t

$$\text{Since } S_2^2 > S_1^2$$

$$\text{then } F = \frac{S_2^2}{S_1^2} = \frac{22.2858}{16.266}$$

$$F_{\text{cal}} = 1.3700$$

vi) Decision

$$\text{Since } F_{\text{cal}} = 1.3700 < F_{\text{tab}} = 10.67$$

Hence N.H.Ho is Accepted.

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

Given $n_1 = 10$, $\sum (x_i - \bar{x}_1)^2 = 102.4$

$n_2 = 12$, $\sum (x_j - \bar{x}_2)^2 = 120.5$

$$S_1^2 = \frac{\sum (x_i - \bar{x}_1)^2}{n_1 - 1} = \frac{102.4}{10 - 1} = 11.37$$

$$S_2^2 = \frac{\sum (x_j - \bar{x}_2)^2}{n_2 - 1} = \frac{120.5}{12 - 1} = 10.95$$

Here $S_1^2 > S_2^2$

i) N.H.H₀ $\sigma^2 = \sigma_2^2$

ii) A.H.H₁ $\sigma^2 \neq \sigma_2^2$

iii) L.O.S $\alpha = 1\% = 0.01$

iv) tab t

As A.H is $\neq$ type use T.T.T

Since $\alpha = 0.01$

For area 0.01 and df

df $\nu_1 = n_1 - 1$ and $\nu_2 = n_2 - 1$

$\nu_1 = 9$ and $\nu_2 = 11$

then from table $\boxed{\chi^2_{\text{tab}} = 4.63}$

v) cal t

$$F = \frac{S_1^2}{S_2^2} \Rightarrow \frac{11.37}{10.95}$$

$$\boxed{F_{\text{cal}} = 1.038}$$

vi) Decision : $F_{\text{cal}} = 1.038 < F_{\text{tab}} = 4.63$

3) Two Random samples gave the following results.

sample	size	sample mean	sum of squares of deviation.
1	10	15	90
2	12	14	108

Test whether the samples came from the same Normal population.

$$n_1 = 10, \quad \sum (x_i - \bar{x}_1)^2 = 90$$

$$n_2 = 12, \quad \sum (x_j - \bar{x}_2)^2 = 108$$

$$S_1^2 = \frac{\sum (x_i - \bar{x}_1)^2}{n_1 - 1} = \frac{90}{10 - 1} = 10$$

$$S_2^2 = \frac{\sum (x_j - \bar{x}_2)^2}{n_2 - 1} = \frac{108}{12 - 1} = 9.81$$

Here $S_1^2 > S_2^2$

i) N.H. $H_0: \sigma_1^2 = \sigma_2^2$

ii) A.H. $H_1: \sigma_1^2 \neq \sigma_2^2$

iii) L.O.S $\alpha = 0.01$

iv) As A.H is $\neq$ type use T.T.T

Since $\alpha = 0.01$

$$\left. \begin{aligned} d_1 = v_1 = n_1 - 1 = 10 - 1 \\ d_1 = 9 \end{aligned} \right\} \begin{aligned} d_2 = v_2 = n_2 - 1 = 12 - 1 \\ d_2 = 11 \end{aligned}$$

From table $\boxed{\chi^2_{\text{tab}} = 4.63}$

v) cal t

$$F = \frac{S_1^2}{S_2^2} = \frac{10}{9.81} = 1.019$$

vi) Decision:

$$F_{\text{cal}} = 1.019 < \chi^2_{\text{tab}} = 4.63$$

Hence N.H. H_0 is Accepted.

4) Pumpkins were grown under two experimental conditions. 2 random samples of 11 and 9 pumpkins. Show the sample S.D 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.

Sol. Given, $n_1 = 11$, $S_1 = 0.8$, $S_1^2 = 0.64$

$n_2 = 9$, $S_2 = 0.5$, $S_2^2 = 0.25$

Here, $S_1^2 > S_2^2$

i) N.H. $H_0 \quad \sigma_1^2 = \sigma_2^2$

ii) A.H. $H_1 \quad \sigma_1^2 \neq \sigma_2^2$

iii) Choose L.O.S $\alpha = 1\%$
 $\alpha = 0.01$

iv) tabulated t

As A.H is $\neq$ type use T.T.T

$$\begin{array}{l} d_1 = \nu_1 = n_1 - 1 \\ d_1 = 11 - 1 \\ d_1 = 10 \end{array} \quad \left\{ \begin{array}{l} d_2 = \nu_2 = n_2 - 1 \\ d_2 = 9 - 1 \\ d_2 = 8 \end{array} \right.$$

From table

$$\boxed{\chi^2_{\text{tab}} = 5.81}$$

v) calculated t

$$F = \frac{S_1^2}{S_2^2} = \frac{0.64}{0.25}$$

$$\boxed{F_{\text{cal}} = 2.56}$$

vi) Decision.

$$\text{Since } F_{\text{cal}} = 2.56 < F_{\text{tab}} = 5.81$$

Hence N.H. H_0 is Accepted.

Chi-Square test for Goodness of Fit

χ^2 -test is used to test whether differences between Observed (O_i) and Expected (E_i) frequencies are significant.

The Test Statistics Formula is given by,

$$\chi^2 = \sum_{i=1}^n \left[\frac{(O_i - E_i)^2}{E_i} \right]$$

with df $= n-1$

Problems

- i) The no. 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 the accident conditions were the same during these 10 week period.

Here $n=10$

Let E_i = Expected Frequency during 10 week period.

$$= \frac{12+8+20+\dots+4}{10}$$

$$E_i = 10$$

O_i	E_i	$(O_i - E_i)^2$	$(O_i - E_i)^2 / E_i$
12	10	+2	0.4
8	10	-2	0.4
20	10	10	10
2	10	-8	6.4
14	10	4	1.6
10	10	0	0
15	10	5	2.5
6	10	-4	1.6
9	10	-1	0.1
4	10	-6	3.6

$$\begin{aligned} & \Rightarrow \frac{\sum (O_i - E_i)^2}{E_i} \\ & = 26.6 \end{aligned}$$

i) $N.H.H_0 \quad O_i = E_i$

ii) $A.H.H_1 \quad O_i \neq E_i$

iii) Choose L.O.S $\alpha = 1\% = 0.01$ (T.T.T)

iv) tabulated χ^2 :

~~As A.H is of + type, use T.T.T~~

Since $\alpha = 0.01$

$$\frac{\alpha}{2} = \frac{0.01}{2} = 0.005$$

For the area ~~0.005~~ ^{0.01} and df $D = n - 1 \Rightarrow 10 - 1 = 9$

From table $\chi^2_{\text{tab}} = \boxed{21.666}$

v) Calculated χ^2 :

$$\chi^2_{\text{cal}} = \sum \left[\frac{(O_i - E_i)^2}{E_i} \right]$$

$$\chi^2_{\text{cal}} = \boxed{26.6}$$

vi) Decision:

$$\text{Since } \chi^2_{\text{cal}} = 26.6 < \chi^2_{\text{tab}} \quad 21.666$$

$$26.6 > 23.589$$

Hence $N.H.H_0$ is Rejected.

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.

Given, $n = 500$

$$\Rightarrow 220, 170, 90, 20$$

Note:

In f-test and chi-square test, Find the directly the tabulated values for given $L.O.S \alpha$.

sol. Total Frequency = 500, Since Expected frequencies $\Rightarrow 4:3:2:1$
The given Observed frequencies are 220, 170, 90, 20.
The corresponding Expected frequencies are given by

$$E_1 = \text{Expected no. of failed} \Rightarrow \frac{4}{10} \times 500 \Rightarrow 200$$

$$E_2 = \text{Expected no. of third class} \Rightarrow \frac{3}{10} \times 500 \Rightarrow 150$$

$$E_3 = \text{Expected no. of 2nd class} \Rightarrow \frac{2}{10} \times 500 \Rightarrow 100$$

$$E_4 = \text{Expected no. of 1st class} \Rightarrow \frac{1}{10} \times 500 \Rightarrow 50$$

Chi-square table

O_i	E_i	$O_i - E_i$	$\left(\frac{(O_i - E_i)^2}{E_i} \right)$
220	200	20	2
170	150	20	2.67
90	100	-10	1
20	50	-30	18

$$\Sigma \left[\frac{(O_i - E_i)^2}{E_i} \right] \Rightarrow 23.67$$

i) N.H. $H_0: O_i = E_i$

ii) A.H. $H_1: O_i \neq E_i$

iii) $L.O.S \alpha = 1\% = 0.01$

iv) tabulated χ^2 :

As $L.O.S \alpha = 0.01$ with df $D = n - 1 \Rightarrow 4 - 1 = 3$

then From table

$$\chi^2_{\text{tab}} = 11.345$$

v) Calculated χ^2 :

$$\chi^2_{cal} = \sum \left[\frac{(O_i - E_i)^2}{E_i} \right]$$

$$\boxed{\chi^2_{cal} = 23.67}$$

vi) Decision.

$$\text{Since } \chi^2_{cal} = 23.67 < \chi^2_{tab} = 11.345$$

$$23.67 > 11.345$$

Hence N.H.H₀ is Rejected.

3) The following figures show the distribution of digits in numbers choosen at Random from a telephone directory digits: 0 1 2 3 4 5 6 7 8 9

Frequency: 1026 1107 997 1075 933 1107 972 964 853

Test whether the digits may be taken to occur equally frequency in the direction.

Here $n=10$

$$df = D = n - 1 = 9$$

Let,

O_i = Observed frequency

E_i = Expected frequency during 10 week period.

$$E_i = \frac{1026 + 1107 + \dots + 853}{10}$$

$$= \frac{10000}{10}$$

Digits	O_i	E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$\left(\frac{(O_i - E_i)^2}{E_i}\right)$
0	1026	1000	26	676	0.676
1	1107	1000	107	11449	11.449
2	997	1000	-3	9	0.009
3	966	1000	-34	1156	1.156
4	1075	1000	75	5625	5.625
5	933	1000	-67	4489	4.489
6	1107	1000	107	11449	11.449
7	972	1000	-28	784	0.784
8	964	1000	-36	1296	1.296
9	853	1000	-147	21609	21.609
					$\Sigma = 58.542$

i) N.H. $H_0: O_i = E_i$

ii) A.H. $H_1: O_i \neq E_i$ (T.T.T)

iii) L.O.S chosen $\alpha = 1\% = 0.01$

iv) tabulated χ^2

As A.H is $\neq$ type use T.T.T

Since $\alpha = 1\% = 0.01$ & $d = n - 1 = 9$

$$\chi^2_{tab} = 21.666$$

v) calculated χ^2

$$\chi^2_{cal} = \Sigma \left[\frac{(O_i - E_i)^2}{E_i} \right]$$

$$\chi^2_{cal} = 58.542$$

vi) Decision:

Since $\chi^2_{cal} < \chi^2_{tab}$

$$58.542 > 21.666$$

Hence N.H. H_0 is Rejected.

4) A die is thrown 264 times with the following results show that the die is biased.

no. appeared on die : 1 2 3 4 5 6
Frequency : 40 32 28 58 54 60.

Given that,

$$E_i = \frac{40 + 32 + 28 + 58 + 54 + 60}{6}$$

$$E_i = 45.333$$

O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\left[\frac{(O_i - E_i)^2}{E_i} \right]$
40	45.333	-5.33	28.40	0.6265
32	45.333	-13.33	177.68	3.9196
28	45.333	-17.33	300.32	6.6251
58	45.333	12.67	160.52	3.5411
54	45.333	8.67	75.16	1.6580
60	45.333	14.67	215.20	4.7474

$$\Sigma = 21.115$$

i) N.H.H₀ $O_i = E_i$

ii) A.H.H₁ $O_i \neq E_i$ (T.T.T)

iii) Choose L.O.S $\alpha = 1\% = 0.01$

iv) tabulated χ^2

As A.H is $\neq$ type use T.T.T

Since $\alpha = 1\% = 0.01$

$$D = n - 1 = 5$$

$$\chi^2_{\text{tab}} = 15.086$$

v) calculated χ^2

$$\chi^2_{\text{cal}} = \Sigma \left[\frac{(O_i - E_i)^2}{E_i} \right]$$

$$\chi^2_{\text{cal}} = 21.115$$

vi) Decision.

Since $\chi^2_{\text{cal}} < \chi^2_{\text{tab}}$

$$21.115 > 15.086$$

Hence N.H.H₀ is Rejected.

5) 200 digits were chosen at random from a set 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 access the correctness of the hypothesis that the digits were distributed in equal no. in the tables from which these were chosen. Given that

$$E_i = \frac{18+19+23+21+16+25+22+20+21+15}{10}$$

$$E_i = 20$$

Digit	O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
0	18	20	-2	4	0.2
1	19	20	-1	1	0.05
2	23	20	3	9	0.45
3	21	20	1	1	0.05
4	16	20	-4	16	0.8
5	25	20	5	25	1.25
6	22	20	2	4	0.2
7	20	20	0	0	0
8	21	20	1	1	0.05
9	15	20	-5	25	1.25
					$\Sigma = 4.3$

i) N.H. H_0 $O_i = E_i$

ii) A.H. H_1 $O_i \neq E_i$

iii) Choose $\alpha = 1\% = 0.01$

iv) tabulated χ^2

As A.H is $\neq$ type use T.T.T

Since $\alpha = 0.01$ when $D = n-1 = 9$

$$\chi^2_{tab} = 21.666$$

v) calculated χ^2

$$\chi^2_{cal} = \Sigma \left[\frac{(O_i - E_i)^2}{E_i} \right] \Rightarrow \chi^2_{cal} = 4.3$$

vi) Decision $\chi^2_{cal} = 4.3 < 21.666 \chi^2_{tab}$

Hence N.H. H_0 is Accepted.

Attributes:

Attribute means a quality or characteristics.

Ex: Drinking, Smoking, Blindness, honesty, Beauty, etc.

Chi-Square test for Independence of Attributes

Let us consider two attributes **A** and **B**. **A** is divided into two classes and similarly **B** is divided into two classes.

The various cell frequencies can be expressed in the following table known as 2×2 Contingency table.

A	a	b
B	c	d

2×2

Expected Frequency Table

pected	frequency	table	total
	$E(a) = \frac{(a+b)(a+c)}{N}$	$E(b) = \frac{(a+b)(b+d)}{N}$	$a+b$
	$E(c) = \frac{(c+d)(a+c)}{N}$	$E(d) = \frac{(c+d)(b+d)}{N}$	$c+d$
total	$a+c$	$b+d$	N

Using chi-square test, we test if two attributes **A** and **B** are independent or not.

Here the H_0 attributes are Independent.

Degrees of Freedom (df) $= (r-1)(s-1)$

where r = no. of rows

s = no. of columns.

- i) 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

Sol.

The Expected Frequencies for given 2×2 contingency table are,

$E(60) = \frac{90 \times 100}{200}$ $= 45$	$E(30) = \frac{90 \times 100}{200}$ $= 45$	90
$E(40) = \frac{110 \times 100}{200}$ $= 55$	$E(70) = \frac{110 \times 100}{200}$ $= 55$	110
100	100	$N = 200$

chi-square test:

O_i	E_i	$(O_i - E_i)$	$\frac{(O_i - E_i)^2}{E_i}$
60	45	15	5
30	45	-15	5
40	55	-15	4.09
70	55	15	4.09

$$\Rightarrow \sum \left[\frac{(O_i - E_i)^2}{E_i} \right]$$

$$= 18.18$$

i) N.H.H₀ $O_i = E_i$

ii) A.H.H₁ $O_i \neq E_i$

iii) Choose L.O.S $\alpha = 1\%$

$$\alpha = 0.01$$

iv) tabulated χ^2

for L.O.S $\alpha = 0.01$ with df $J = (r-1)(s-1)$
 $J = (2-1)(2-1) = 1$

then from table

$$\chi^2_{\text{tab}} = 6.6349$$

v) calculated χ^2

$$\sum \left[\frac{(O_i - E_i)^2}{E_i} \right] = 18.18$$

$$\chi^2_{\text{cal}} = 18.18$$

vi) decision.

Since $\chi^2_{\text{cal}} = 18.18 > \chi^2_{\text{tab}} = 6.6349$.

Hence N.H.Ho is Rejected.

4) 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

Sol.

Here $r = 3$, $s = 2$

Expected frequency table: Given 3×2 contingency Table

$E(69) = \frac{120 \times 185}{300}$ $= 74$	$E(51) = \frac{120 \times 115}{300}$ $= 46$	120
$E(81) = \frac{101 \times 185}{300}$ $= 62.28$	$E(20) = \frac{101 \times 115}{300}$ $= 38.71$	101
$E(35) = \frac{79 \times 185}{300}$ $= 48.71$	$E(44) = \frac{79 \times 115}{300}$ $= 30.28$	79
total 185	115	$N = 300$

χ^2 table

O_i	E_i	$(O_i - E_i)$	$\left(\frac{(O_i - E_i)^2}{E_i}\right)$
69	74	-5	0.33
51	46	5	0.54
81	62.28	18.72	5.62
20	38.71	-18.71	9.04
35	48.71	-13.71	3.85
44	30.28	13.72	6.21

$$\Rightarrow \sum \left[\frac{(O_i - E_i)^2}{E_i} \right] = 25.64$$

i) N.H.Ho $O_i = E_i$

ii) A.H.Hi $O_i \neq E_i$

iii) Choose L.O.S $\alpha = 0.01$

iv) tabulated χ^2

For L.O.S $\alpha = 0.01$ with df $\hat{v} = (r-1)(c-1) \Rightarrow (3-1)(2-1)$
 $\hat{v} = 2$

then from table,

$$\boxed{\chi^2_{tab} = 9.2103}$$

v) calculated χ^2

$$\sum \left[\frac{(O_i - E_i)^2}{E_i} \right] = 25.64$$

$$\boxed{\chi^2_{cal} = 25.64}$$

vi) Decision.

Since $\chi^2_{cal} = 25.64 > \chi^2_{tab} = 9.2103$

Hence, N.H.Ho is Rejected. (Dependent)

2) The following table gives guess the classification of 100 workness according to gender and nature of work. Test whether the nature of work is independent of the gender.

	Stable	Unstable	Total
males	40	20	60
Females	10	30	40
	50	50	100

Sol. Given $r=2$
 $n=2$

Expected Frequency for given 2×2 Contingency table are;

$E(40) = \frac{60 \times 50}{100}$ $\Rightarrow 30$	$E(20) = \frac{60 \times 50}{100}$ $= 30$	60
$E(10) = \frac{40 \times 50}{100}$ $= 20$	$E(30) = \frac{40 \times 50}{100}$ $= 20$	40
50	50	100

χ^2 table:

O_i	E_i	$O_i - E_i$	$\frac{(O_i - E_i)^2}{E_i}$
40	30	10	3.33
20	30	-10	3.33
10	20	-10	5
30	20	10	5

$$\sum \left[\frac{(O_i - E_i)^2}{E_i} \right] \Rightarrow 16.66$$

i) N.H.H₀ $O_i = E_i$

ii) A.H.H₁ $O_i \neq E_i$

iii) Choose L.O.S $\alpha = 1\% = 0.01$

iv) tabulated χ^2 :

As $\alpha = 0.01$ with df $= (r-1)(s-1) \Rightarrow (2-1)(2-1)$
 $df = 1$

from table $\chi^2_{tab} = 6.6349$

v) calculated χ^2 :

$$\sum \left[\frac{(O_1 - E_1)^2}{E_1} \right]$$

$$\boxed{\chi^2_{\text{cal}} = 16.66}$$

vi) Decision:

$$\text{Since } \chi^2_{\text{cal}} = 16.66 < \chi^2_{\text{tab}} = 6.6349$$

$$16.66 > 6.6349$$

Hence N.H.H₀ is Rejected.

3) Given the following contingency table for hair colour and eye colour. Find the value of χ^2 . Is there good associate on b/w the two.

		Hair Colour			
	Fair	Brown	Black	Total	
Eye Colour	Blue	15	5	20	40
	Gray	20	10	20	50
	Brown	25	15	20	60
	Total	60	30	60	150

Sol. Given $r = 3, s = 3$.

Expected frequency for given 3x3 Contingency Table

$E(15) = \frac{40 \times 60}{150}$ $= 16$	$E(5) = \frac{40 \times 30}{150}$ $= 8$	$E(20) = \frac{40 \times 60}{150}$ $= 16$	40
$E(20) = \frac{50 \times 60}{150}$ $= 20$	$E(10) = \frac{50 \times 30}{150}$ $= 10$	$E(20) = \frac{50 \times 60}{150}$ $= 20$	50
$E(25) = \frac{60 \times 60}{150}$ $= 24$	$E(15) = \frac{60 \times 30}{150}$ $= 12$	$E(20) = \frac{60 \times 60}{150}$ $= 24$	60
60	30	60	150

χ^2 table:

O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2 / E_i$
15	16	-1	0.062
5	8	-3	1.125
20	16	4	1
20	20	0	0
10	10	0	0
20	20	0	0
25	24	1	0.041
15	12	3	0.75
20	24	-4	0.66

$$\sum \left[\frac{(O_i - E_i)^2}{E_i} \right] = 3.63$$

i) N.H.H₀ $O_i = E_i$

ii) A.H.H₁ $O_i \neq E_i$

iii) Choose L.O.S $\alpha = 1\% = 0.01$

iv) As $\alpha = 0.01$ with df $= (r-1)(s-1) \Rightarrow (3-1)(3-1) = 2 \times 2 \Rightarrow 4$

From table

$$\chi^2_{\text{tab}} = 13.277$$

v) calculated χ^2

$$\chi^2_{\text{cal}} = 3.63$$

vi) Decision:

$$\chi^2_{\text{cal}} = 3.63 < \chi^2_{\text{tab}} = 13.277$$

Hence N.H.H₀ is Accepted.

Formulas

S.No.	Test Statistics for	Formula	d.o.f
1.	One mean	i) $t = \frac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{n}}\right)}$ if s is not given ii) $t = \frac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{n-1}}\right)}$ if s is given.	$D = (n-1)$
2.	Two Mean	$t = \frac{\bar{x}_1 - \bar{x}_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$	$D = n_1 + n_2 - 2$
3.	F-test	$F = \frac{S_1^2}{S_2^2}$ (or) $F = \frac{S_2^2}{S_1^2}$	$d_1 = n_1 - 1$ $d_2 = n_2 - 1$ (or) $d_1 = n_2 - 1, d_2 = n_1 - 1$
4.	χ^2 test	$\chi^2 = \sum \left[\frac{(O_i - E_i)^2}{E_i} \right]$	$D = n - 1$
5.	Expected Frequency.	$E(a) = \frac{(a+b)(a+c)}{N}$ $E(b) = \frac{(a+b)(b+d)}{N}$ $E(c) = \frac{(c+d)(a+c)}{N}$ $E(d) = \frac{(c+d)(b+d)}{N}$	$D = (r-1)(s-1)$ $r = \text{no. of rows}$ $s = \text{no. of columns}$