UNIT - V : Small Sampling

Small sampling:

It is used for testing of hypothesis when the sample size is small and population standard deviation 'σ' is not known.

t- Distribution:

There are 3 kinds of t-distributions.

1. t-test for single mean
2. t-test for difference of means
3. paired t-test

t-test for single mean:

Step 1: Null hypothesis $H_0 : \mu = \mu_0$ (or $\bar{x} = \mu$)

i.e, sample mean is equal to population mean

Step 2: Alternative hypothesis $H_1 : \mu < \mu_0$ (LTT)

$\mu > \mu_0$ (RTT)

$\mu \neq \mu_0$ (TTT)

Step 3: Level of significance (α):

To select appropriate l.o.s 'α' with degrees of freedom $\nu = (n-1)$, then find out tabulated value

Step 4: Test statistic:

To compute the test statistic under the null hypothesis H_0:

Test statistic: $t = \dfrac{\bar{x}-\mu}{s/\sqrt{n}}$

Here s is unknown

(or)

Test statistic : $t = \dfrac{\bar{x}-\mu}{s/\sqrt{n-1}}$

Here $\bar{x}$ = Sample mean

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

s = sample standard deviation

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

μ = population mean

n = sample size

Step 5: Conclusion:

We compare calculated and tabulated values

1. If $|t|_{cal} \leq t_{tab}$, we accept H_0
2. If $|t|_{cal} > t_{tab}$, we reject H_0

Degrees of freedom:

The no. of independent variates in the statistic is called Degrees of freedom and it is denoted by greek letter ν (Nu)

Properties of t-distribution:

1. The shape of T-distribution is bell shape, which is similar to normal distribution and is symmetrical about the mean

2. The T-distribution curve is also asymptotic to the t-axis i.e., the two tails of the curve on both sides of $t = 0$ extends to ∞

3. It is symmetrical about the line, $t = 0$

4. The form of the probability curve varies with degrees of freedom i.e, with sample size,

5. It is unimodal with mean = median = mode

6. The mean of standard normal distribution & as well as t-distribution is zero but the

ν, which is called the degree of freedom

7. The variance of t-distribution exceeds 1 but approaches 1 as $n \to \infty$. Infact the t-distribution with degrees of freedom approaches standard normal distribution as $\nu = (n-1) \to \infty$

Problems:

1. A sample of 26 bulbs gives a mean life of 990h with a S.D of 20 hrs. The manufacturer claims the mean life of bulbs is 1000 h. Is the sample not upto the standard

Sol: Given that,

$$n = 26$$
$$\bar{x} = 990$$
$$s = 20$$
$$\mu = 1000$$
$$\alpha = 0.05$$

St 1: Null hypothesis H_0: $\mu = 1000$ hrs

i.e, The sample is upto the standard

St 2: Alternative hypothesis H_1 : $\mu < 1000$ hrs (LTT)

i.e, The sample is not upto the standard

St 3: L.O.S : $\alpha = 0.05$ & $\nu = (n-1)$ d.f $= (26-1) = 25$ d.o.f

$t_\alpha = 0.05$ & $\nu = 25$ degrees of freedom

$$t_{tab} = 1.708$$

St 4: T.S : $t = \dfrac{\bar{x}-\mu}{s/\sqrt{n-1}} = \dfrac{990-1000}{20/\sqrt{26-1}} = \dfrac{-10 \times 5}{20} = -2.5$

$$|t|_{cal} = 2.5$$

St 5: Conclusion:

$$|t|_{cal}\ (2.5) > t_{tab}\ (1.708)$$

$\therefore$ we reject H_0

Hence the sample is not upto the standard

2. The average breaking strength of the steel rods is specified to be 18.5 Kilo 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.

Sol: Given that,

$n = 14$, $\bar{x} = 17.85$, $S = 1.955$,

$\mu = 18.5$, $\alpha = 0.05$

St 1: Null hypothesis $H_0 : \mu = 18.5$

i.e, The experiment is not significant

St 2: Alternative hypothesis $H_1 : \mu \neq 18.5$ (TTT)

i.e, The experiment is significant

St 3: L.O.S : $\alpha = 0.05$ & $v = (n-1)$ d.f

$\frac{\alpha}{2} = 0.025$ & $v = 14 - 1 = 13$ df

$t_\alpha = 0.025$ & $v = 13$ df

$t_{tab} = 2.160$

St 4: T.S : $t = \dfrac{\bar{x} - \mu}{S/\sqrt{n-1}} = \dfrac{17.85 - 18.5}{1.955/\sqrt{14-1}} = -1.19$

$|t|_{cal} = 1.19$

St 5: Conclusion:

$|t|_{cal}\ (1.19) < t_{tab}\ (2.160)$

$\therefore$ We accept H_0

i.e, The experiment is not significant

3. A random sample of 10 boys had the following IQ's : 70, 120, 110, 101, 88, 83, 95, 98, 107 & 100. Do this data support the assumption of a population mean IQ of 100. Find a reasonable range in which most of the mean IQ values of sample of 10 boys alike

Sol: Given that,

$n = 10$, $\mu = 100$, $\bar{x} = 97.2$

x_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$
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
		$\Sigma(x_i-\bar{x})^2 = 1833.6$

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

$$= \frac{1}{10-1} \times 1833.6$$

$$S^2 = 203.74$$

$$S = \sqrt{203.74}$$

$$S = 14.27$$

i)

St 1: Null hypothesis $H_0 : \mu = 100$

i.e, The data support the average IQ levels of 10 boys is 100

St 2: Alternative hypothesis $H_1 : \mu \neq 100$ (TTT)

i.e, The data doesn't support the average IQ levels of 10 boys is 100

St 3: L.O.S : $\alpha = 0.05$ & $\nu = (n-1)$ d.f

$\alpha/2 = 0.025$ & $\nu = 10-1 = 9$ d.f

$t_{\alpha/2} = 0.025$ & $\nu = 9$ d.f

$t_{tab} = 2.262$

St 4: T.S : $t = \dfrac{\bar{x}-\mu}{S/\sqrt{n}} = \dfrac{97.2-100}{14.27/\sqrt{10}} = \dfrac{-2.8\,(3.1622)}{14.27}$

$= -0.62$

$|t|_{cal} = 0.62$

St 5: Conclusion:

$|t|_{cal}\,(0.62) < t_{tab}\,(2.262)$

$\therefore$ we accept H_0

i.e, The data support the average IQ levels of 10 boys is 100

ii) Range values (or) C.I

C.I $= (\bar{x}-E,\ \bar{x}+E)$

Max

$C.I = (97.2 - 10.2, 97.2 + 10.2)$

$C.I = (87, 107.4)$

4. The heights of 10 males of a given locality are found to be: 70, 67, 62, 68, 61, 68, 70, 64, 64, 66 inches. It is reasonable to believe that, the average height is greater than 64 inches, test at 5% level

Sol: Given that,

$n = 10, \mu = 64, \bar{x} = \frac{660}{10} = 66, \alpha = 0.05$

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
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

$\Sigma(x_i - \bar{x})^2 = 90$

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

$= \frac{1}{10-1} \times 90$

$S^2 = 10$

$S = \sqrt{10}$

$S = 3.16$

St 1: Null hypothesis $H_0: \mu = 64$

i.e., The average height of 10 males is 64 inches

St 2: Alternative hypothesis $H_1: \mu > 64$ (RTT)

i.e., The average height of 10 males is greater than 64 inches

St 3: l.o.s: $\alpha = 0.05$ & $\nu = n-1 = 10-1 = 9$ df

$T_{\alpha = 0.05}$ & $\nu = 9$ df

$t_{tab} = 1.833$

St 4: T.S: $t = \frac{\bar{x} - \mu}{S/\sqrt{n}} = \frac{66 - 64}{3.16/3.162} = 2$

$|t|_{cal} = 2$

St 5: Conclusion:

$|t|_{cal}(2) > t_{tab}(1.833)$

t-test for difference of means:

Let $\bar{x}$ and $\bar{Y}$ be the means of two independent random samples $n_1 < 30$, $n_2 < 30$ drawn from two normal populations having means μ_1 & μ_2. To test whether the two population means are equal:

St 1: Null hypothesis $H_0: \mu_1 = \mu_2$

i.e, the two population means are equal

St 2: Alternative hypothesis $H_1: \mu_1 < \mu_2$ (LTT)

$\mu_1 > \mu_2$ (RTT)

$\mu_1 \neq \mu_2$ (TTT)

St 3: Level of significance (α):

To select appropriate L.O.S α with degrees of freedom $v = (n_1 + n_2 - 2)$ d.f, then find out tabulated value in advance

St 4: Test statistic:

To compute test statistic under the null hypothesis H_0:

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

Here, s^2 = pooled sample variance

$$s^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$$ (where s_1 & s_2 are known)

(or)

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2 + \Sigma(Y_j - \bar{Y})^2}{n_1 + n_2 - 2}$$ (where s_1 & s_2 are unknown)

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

$$s_2^2 = \frac{\Sigma(Y_j - \bar{Y})^2}{n_2 - 1}$$

St 5: Conclusion:

2. If $|t|_{cal} > t_{tab}$, we reject H_0

Problems:

1. Samples of 2 types of electric light bulbs were tested for length of life and following data were obtained

		Type-I	Type-II
Sample	Number	$n_1 = 8$	$n_2 = 7$
"	Mean	$\bar{x} = 1234$	$\bar{y} = 1036$
"	S.D	$s_1 = 36$	$s_2 = 40$

Is the difference in the means sufficient to warrent that type-I is (that) superior to type-II regarding length of life

Sol: Given that,

$n_1 = 8 \quad n_2 = 7$

$\bar{x} = 1234 \quad \bar{y} = 1036$

$s_1 = 36 \quad s_2 = 40$

St 1: Null hypothesis $H_0 : \mu_1 = \mu_2$

i.e, The two types of electric light bulbs are identical

St 2: Alternative hypothesis $H_1 : \mu_1 > \mu_2$ (RTT)

i.e, Type-I is superior to Type-II regarding length of life

St 3: L.O.S : 'α' $= 0.05$ & $\nu = (n_1 + n_2 - 2)$ d.f

$\alpha = 0.05$ & $\nu = 13$ d.f

$t_{tab} = 1.771$

St 4: T.S : $t = \dfrac{\bar{x} - \bar{y}}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ $\quad \because s^2 = \dfrac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$

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

$= 1659.07$

$t = \dfrac{1234 - 1036}{\sqrt{1659.07\left(\frac{1}{8} + \frac{1}{7}\right)}}$

St 5: conclusion:

$t_{cal}\ (9.38) > t_{tab}(1.771)$

$\therefore$ we reject H_0

i.e, The type - I is superior to type - II regarding length of life

$\therefore$ The claim is true

* 2 - A group of 5 patients treated with medicine A weights: 42, 39, 48, 60 & 41 kgs. Second group of 7 patients from the same hospital treated with medicine B weights: 38, 42, 56, 64, 68, 69 & 62 kgs. Do you agree with the claim that medicine B increases the weight significantly.

Sol: Given that, $n_1 = 5,\ n_2 = 7$

$$\bar{x} = \frac{\sum_{i=1}^{n_1} x_i}{n_1} = \frac{42+39+48+60+41}{5} = 46$$

$$\bar{y} = \frac{\sum_{j=1}^{n_2} y_j}{n_2} = \frac{38+42+56+64+68+69+62}{7} = 57$$

x_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$
42	-4	16
39	-7	49
48	2	4
60	14	196
41	-5	25
		$\sum(x_i-\bar{x})^2 = 290$

y_j	$y_j-\bar{y}$	$(y_j-\bar{y})^2$
38	-19	361
42	-15	225
56	-1	1
64	7	49
68	11	121
69	12	144
62	5	25
		$\sum(y_j-\bar{y})^2 = 926$

$$S^2 = \frac{\sum(x_i-\bar{x})^2 + \sum(y_j-\bar{y})^2}{n_1+n_2-2} = \frac{290+926}{5+7-2} = 121.6$$

St 1: Null hypothesis $H_0: \mu_1 = \mu_2$

i.e, There is no significant difference between

St 2: Alternative hypothesis $H_1: \mu_1 < \mu_2$ (LTT)

i.e, Medicine B increase the weight significantly

St 3: l.o.s: 'α' = 0.05 & $\nu = n_1+n_2-2 = 5+7-2 = 10$ d.f

$t_{tab} = 1.812$

St 4: t.s: $t = \dfrac{\bar{x}-\bar{y}}{\sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \dfrac{46-57}{\sqrt{121.6\left(\frac{1}{5}+\frac{1}{7}\right)}} = -1.703$

$|t|_{cal} = 1.703$

St 5: conclusion:

$|t|_{cal}(1.703) < t_{tab}(1.812)$

$\therefore$ We accept H_0

i.e, There is no significant difference between using medicine A and medicine B

$\therefore$ The given statement is false

3. To examine the hypothesis that the husbands are more intelligent than the wives. An investigator took a sample of 10 couples and administrated them a test which measures the IQ. 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 l.o.s 0.05

4. Below are given the gain in weights of pigs feed 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 2 diets differ significantly as regards their effect on increases in weight

Paired t-test (or) Dependent t-test :

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 difference of the 2 situations

St 1: Null hypothesis H_0 : $\mu_1 = \mu_2$ (or) $\mu = 0$

i.e., there is no significant difference b/w the before & after

St 2: Alternative hypothesis H_1 : $\mu_1 < \mu_2$ (LTT)

$\mu_1 > \mu_2$ (RTT)

$\mu_1 \neq \mu_2$ (TTT)

St 3: Level of significance :

To select appropriate l.o.s 'α' with d.o.f v:

$v = n-1$

St 4: Test statistic :

To compute the test statistic under the null hypothesis H_0 :

$$t = \frac{\bar{d}}{S/\sqrt{n}}$$

$d_i = x_i - y_i$ (or) $y_i - x_i$

$$\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n} \quad ; \ n \rightarrow \text{pair of observations}$$

$$S = \sqrt{\frac{1}{n-1} \sum (d_i - \bar{d})^2}$$

St 5: Conclusion :

We compare calculated and tabulated values:

1. If $|t|_{cal} \leq t_{tab}$ → we accept H_0
2. If $|t|_{cal} > t_{tab}$ → we reject H_0 (or) we accept H_1

Problems:

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

$(d_i-\bar{d})^2$:	4	81	16	4	144	64	49	4	16	100	= 482
$d_i-\bar{d}$:	2	-9	4	2	12	-8	-7	-2	-4	10	
(x_i) Before :	67	24	57	55	63	54	56	68	33	43	
(y_i) After :	70	38	58	58	56	67	68	75	42	38	
$d_i = x_i - y_i$:	-3	-14	-1	-3	7	-13	-12	-7	-9	5	

Test whether the intensive training is useful at 0.05 L.O.S.

Sol:

St 1: Null hypothesis $H_0 : \mu_1 = \mu_2$ (or) $\mu = 0$

i.e., The intensive training is not useful

St 2: Alternative hypothesis $H_1 : \mu_1 < \mu_2$ (LTT)

ie., The intensive training is useful

St 3: L.O.S $\alpha = 0.05$ & $\nu = (n-1)$ d.f

$t_\alpha = 0.05$ & $\nu = 9$ d.f

$t_{tab} = 1.833$

Computation:

St 4: T.S :

$$\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n} = \frac{-3-14-1-3+7-13-12-7-9+5}{10} = \frac{-50}{10} = -5$$

$$S = \sqrt{\frac{1}{n-1} \sum (d_i-\bar{d})^2}$$

$$= \sqrt{\frac{1}{9} \sum (d_i-\bar{d})^2} = \sqrt{\frac{482}{9}} = 7.32$$

$\sqrt{10} = 3.16$

$$t = \frac{\bar{d}}{S/\sqrt{n}} = \frac{-5}{7.32/3.16} = -2.16$$

$$|t|_{cal} = 2.16$$

St 5: Conclusion: $|t|_{cal}\ (2.16) > t_{tab}\ (1.833)$

$\therefore$ We reject H_0

2. The blood pressure of 5 women before and after intake of certain drug are given below:

(x_i) Before :	110	120	123	132	125
(y_i) After :	120	118	125	136	121
$d_i = x_i - y_i$:	-10	2	-2	-4	4
$d_i - \bar{d}$:	-8	4	0	-2	6
$(d_i - \bar{d})^2$:	64	16	0	4	36 = 120

Test whether there is significant change in B.P at 1% l.o.s.

Sol:

St 1: Null hypothesis $H_0 : \mu_1 = \mu_2$ (or) $\mu = 0$

i.e, There is no significant change in the BP of the women after using the medicine

St 2: Alternative hypothesis $H_1 : \mu_1 > \mu_2 / \mu_1 < \mu_2$

i.e, There is significant change in the B.P of the women after using the medicine

St 3: l.o.s $\alpha = 0.01$ & $\nu = (n-1)$ d.f

$t_\alpha = 0.01$ & $\nu = 4$ d.f

$t_{tab} = 3.747$

St 4: $\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n} = \frac{-10}{5} = -2$

$$s = \sqrt{\frac{1}{n-1} \sum (d_i - \bar{d})^2} = \sqrt{\frac{1}{4}(120)} = 5.477$$

$$t = \frac{\bar{d}}{s/\sqrt{n}} = \frac{-2 \times 2.236}{5.477} = -0.816$$

$|t|_{cal} = 0.816$

St 5: Conclusion:

$|t|_{cal} (0.816) < t_{tab} (3.747)$

$\therefore$ we accept H_0

i.e, There is no significant change in the B.P of the women after using the medicine

3. 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.

40 & 35, 70 & 65, 45 & 42, 120 & 116, 35 & 33, 55 & 50, 77 & 73

Sol:

Snedecor's F-Test:

Let two independent random samples of sizes n_1 and n_2 be drawn from 2 normal populations.

Step 1: Null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$

i.e, The two population variances are drawn from normal population

Step 2: Alternative hypothesis $H_1 : \sigma_1^2 \neq \sigma_2^2$

i.e, The two population variances are not drawn from same normal population

Step 3: To select appropriate l.o.s 'α' with d.o.f $v_1 = n_1 - 1$, $v_2 = n_2 - 1$ (or) $v_1 = n_2 - 1$, $v_2 = n_1 - 1$

Step 4: Test statistic : $F = \dfrac{\text{Greater variance}}{\text{Smaller Variance}}$

$$F = \frac{S_1^2}{S_2^2} \quad [v_1 = n_1 - 1, v_2 = n_2 - 1] \text{ if } S_1^2 > S_2^2$$

(or)

$$F = \frac{S_2^2}{S_1^2} \quad [v_1 = n_2 - 1, v_2 = n_1 - 1] \text{ if } S_2^2 > S_1^2$$

if s_1 & s_2 are known (or) if s_1 & s_2 are unknown

$$S_1^2 = \frac{n_1 s_1^2}{n_1 - 1} \qquad S_1^2 = \frac{1}{n_1 - 1} \Sigma (x_1 - \bar{x})^2$$

$S_2^2 = \dfrac{n_2 s_2^2}{\ldots}$

Step 5 : Conclusion

We compare calculated and tabulated values :

i) if $F_{cal} \leq F_{tab}$, we accept H_0

ii) if $F_{cal} > F_{tab}$, We reject H_0 or We accept H_1

Problems:

1. In one sample of 8 obs. 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 obs. It was 102.6. Test at 5% level, whether the populations have the same variance

Sol: Given that,

$$n_1 = 8 \qquad n_2 = 10$$

$$\Sigma(X_i - \bar{X})^2 = 84.4$$

$$\Sigma(Y_j - \bar{Y})^2 = 102.6$$

$$S_1^2 = \frac{1}{n_1 - 1} \cdot \Sigma(X_i - \bar{X})^2 = \frac{1}{8-1} \times 84.4 = 12.04$$

$$S_2^2 = \frac{1}{n_2 - 1} \Sigma(Y_j - \bar{Y})^2 = \frac{1}{10-1} \times 102.6 = 11.4$$

St 1: Null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$

i.e., The two sample variances are drawn from Normal population

St 2: Alternative hypothesis $H_1 : \sigma_1^2 \neq \sigma_2^2$

i.e, The two sample variances are not drawn from Normal population

St 3: l.o.s $\alpha = 0.05$ & $\nu_1 = (n_1 - 1)$, $\nu_2 = (n_2 - 1)$ d.f

$\alpha = 0.05$ & $(\nu_1, \nu_2) = (7, 9)$ d.f

$F_{tab} = 3.29$

St 4: T.S: $F = \frac{S_1^2}{S_2^2} = \frac{12.04}{11.4}$

$F_{cal} = 1.056$

St 5: Conclusion:

$F_{cal}(1.056) < F_{tab}(3.29)$

$\therefore$ we accept H_0

i.e, The two sample variances are drawn from Normal population

2. Pumpkins were grown under 2 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 2 variances are equal

Sol: Given that,

$n_1 = 11$ $\quad$ $n_2 = 9$

S_1, S_2 are known $\quad$ $S_1 = 0.8$ $\quad$ $S_2 = 0.5$

*

3. 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, the 2 samples have come from the same normal population.

$\nu_1 = n_2 - 1$

$\nu_2 = n_1 - 1$

χ^2 (chi - square) Test:

If O_i ($i=1,2,\dots,n$) is a set of observed frequencies and E_i ($i=1,2,\dots,n$) is the corresponding set of expected frequencies, then χ^2 is defined as

$$\chi^2 \text{ is } = \Sigma\left[\frac{(O_i - E_i)^2}{E_i}\right] \sim v = (n-1) \text{ d.f}$$

χ^2 is used to test whether difference between observed and expected frequencies are significant or not

Note:

1) In case of binomial distribution, the degrees of freedom $v = n-1$

2) In case of poisson distribution, The d.o.f $v = n-2$

3) In case of normal distribution, The d.o.f $v = n-3$

χ^2 distribution is an important continuous probability distribution and it is used in both large and small test. In χ^2 test is mainly used to test the

i) goodness of fit

ii) independence of attributes

χ^2-test for goodness of fit

We use this test to decide whether the diff. b/w observed and expected frequencies are significant or not

Step 1: Null hypothesis H_0:

There is no significant diff. b/w observed &

Step 2: Alternative hypothesis H_1:

There is a significant diff. b/w observed & expected frequencies.

Step 3: L.O.S:

To select appropriate L.O.S 'α' with d.o.f

$\nu = (n-1)$ d.f

Step 4: Test Statistic:

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

Step 5: Conclusion:

We compare calculated and tabulated values

i) if $\chi^2 cal < \chi^2 tab$, we accept H_0

ii) if $\chi^2 cal > \chi^2 tab$, we reject H_0

Problems:

1. 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	52

Sol:

St 1: Null hypothesis H_0:

i.e, The dice is unbiased

St 2: Alternative hypothesis H_1: The dice is biased

St 3: $l.o.s$ 'α' $= 0.05$ & $\nu = (n-1)$ d.o.f

$\alpha = 0.05$ & $\nu = 6-1 = 5$ d.f

$\chi^2_{tab} = 11.07$

St 4: T.S: Computation

Expected frequency $E_i = \dfrac{\text{Sum of obs}}{\text{No. of obs}}$

$= 40+32+28+58+54+52$

O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
40	44	-4	16	0.36
32	44	-12	144	3.27
28	44	-16	256	5.81
58	44	14	196	4.454
54	44	10	100	2.27
52	44	8	64	1.45

$$\Sigma \frac{(O_i - E_i)^2}{E_i} = 17.614$$

$$\therefore \chi^2_{cal} = 17.6$$

st 5: Conclusion:

$$\chi^2_{cal}\ (17.6) > \chi^2_{tab}\ (11.07)$$

$\therefore$ we reject H_0

i.e, The dice is biased

2. 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.

Sol:

Step 1: Null hypothesis H_0:
The accident conditions were same during this 10 week period

Step 2: Alternative hypothesis H_1:
The accident conditions were not same during this 10 week period

Step 3: l.o.s $\alpha = 0.05$ & $\nu = n-2 = 10-2$

$\alpha = 0.05$ & $\nu = 8$ d.f

$$\chi^2_{tab} = 15.507$$

Step 4: computation:

Expected frequency $E_i = \frac{\text{Sum of Obs}}{}$ 100

O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i-E_i)^2}{E_i}$
12	10	2	4	0.4
8	10	-2	4	0.4
20	10	10	100	10
2	10	-8	64	6.4
14	10	4	16	1.6
10	10	0	0	0
15	10	5	25	2.5
6	10	-4	16	1.6
9	10	-1	1	0.1
4	10	-6	36	3.6

$$\Sigma \frac{(O_i-E_i)^2}{E_i} = 26.6$$

$$\therefore \chi^2_{cal} = 26.6$$

St 5 : Conclusion:

$$\chi^2_{cal}(26.6) > \chi^2_{tab}(15.507)$$

$\therefore$ we reject H_0

i.e., The accident conditions were not same during this 10 week period

3. 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 pairedice on the basis of the χ^2 test at 0.05 level of significance

$P(x)$:	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
Exp freq E_i : $N \cdot P(x)$	10	20	30	40	50	60	50	40	30	20	10
$O_i - E_i$:	-2	4	5	-3	-6	5	1	2	-4	-6	4
$(O_i-E_i)^2$:	4	16	25	9	36	25	1	4	16	36	16

St 1: Null hypothesis H_0

The dice are fair / The dice are unbiased

St 2: Alternative hypothesis H_1

The dice are unfair / The dice are biased

St 3: L.O.S : $\alpha = 0.05$ & $\nu = n-1$ d.f $= 11-1 = 10$ d.f

$\alpha = 0.05$ & $\nu = 10$ d.f

$\chi^2_{tab} = 18.307$

St 4: Computation:

$$T.S \quad \chi^2 = \Sigma \frac{(O_i - E_i)^2}{E_i} = 7.43$$

$$\chi^2_{cal} = 7.43$$

St 5: Conclusion:

$$\chi^2_{cal}\ (7.43) < \chi^2_{tab}\ (18.307)$$

$\therefore$ we accept H_0

i.e, The dice are fair

χ^2 test for independence of attributes:

Attribute:

An attribute means a quality or characteristics.

Ex: beauty, honesty, kindness etc.,

Step 1: Null hypothesis H_0:

The two attributes are independent or There is no association b/w the two attributes

Step 2: Alternative hypothesis H_1:

The two attributes are dependent or There is association b/w the two attributes

Step 3: L.O.S :

To select the appropriate L.O.S 'α' with d.o.f $\nu = (r-1)(c-1)$ d.f

Here $r \rightarrow$ no. of rows

Step 4: T.S :

Let us consider the two attributes A & B is divided into 2 classes and the various self frequencies can be expressed in the following 2x2 contigency table

		A		
		A_1	A_2	
B	B_1	a	b	a+b
	B_2	c	d	c+d
		a+c	b+d	N (or) G = a+b+c+d

$$\text{Expected frequency } E_i = \frac{\text{row total} \times \text{column total}}{\text{Grand total}}$$

$$\text{if } E(a) = \frac{(a+b)(a+c)}{N}$$

$$E(b) = \frac{(a+b)(b+d)}{N}$$

$$E(c) = \frac{(c+d)(a+c)}{N}$$

$$E(d) = \frac{(c+d)(b+d)}{N}$$

$$T.S: \chi^2 = \Sigma\left[\frac{(O_i - E_i)^2}{E_i}\right]$$

Step 5: Conclusion:

We compare calculated & tabulated values

i) if $\chi^2_{cal} \leq \chi^2_{tab}$, we accept Ho

ii) if $\chi^2_{cal} > \chi^2_{tab}$, we reject Ho

Problems:

1. On the basis of information given below about the treatments 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

Sol: St 1: Null hypothesis H_0

The new and conventional treatments are independent

St 2: Alternative hypothesis H_1

The new treatment is superior to the conventional treatment

St 3: l.o.s $\alpha = 0.05$ & $\nu = (r-1)(c-1)$

$= (2-1)(2-1)$

$= 1$

$\chi^2_{tab} = 3.841$

St 4: table:

$E(60) = \frac{90 \times 100}{200} = 45$

$E(30) = \frac{90 \times 100}{200} = 45$

$E(40) = \frac{110 \times 100}{200} = 55$

$E(70) = \frac{110 \times 100}{200} = 55$

O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
60	45	15	225	5
30	45	-15	225	5
40	55	-15	225	4.09
70	55	15	225	4.09
				18.18

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

St 5: conclusion:

$\chi^2_{cal}\ (18.18) > \chi^2_{tab}\ (3.841)$

$\therefore$ we reject H_0

i.e., The new treatment is superior to the conventional treatment.

2. The following table is the classification of 100 workers according to gender and nature of work. Test whether the nature of work is independent of the gender of the worker

	Stable	Unstable	Total
Males	40	20	60
Females	10	30	40
Total	50	50	N = 100

Sol: Step 1 : Null hypothesis H_0:

The nature of work is independent of the gender of the workers.

Step 2: Alternative hypothesis H_1:

The nature of work is dependent of the gender of the workers

Step 3: L.O.S: $\alpha = 0.05$ & $v = (r-1)(c-1)$ d.f $= (2-1)(2-1)$ d.f

$\alpha = 0.05$ & $v = 1$ d.f

$\chi^2_{tab} = 3.841$

Step 4: table

$$\text{Expected frequency } E_i = \frac{\text{row total} \times \text{column total}}{\text{Grand total}}$$

$E(a) = \frac{60 \times 50}{100} = 30$ $\quad$ $E(c) = \frac{40 \times 50}{100} = 20$

$E(b) = \frac{60 \times 50}{100} = 30$ $\quad$ $E(d) = \frac{40 \times 50}{100} = 20$

O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
40	30	10	100	3.33
20	30	-10	100	3.33
10	20	-10	100	5
30	20	10	100	5
				16.66

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

Step 5: Conclusion: $\chi^2_{cal} (16.66) > \chi^2_{tab} (3.841)$

3. From the following data, find whether there is any significant liking in the habit of taking soft drinks among the categories of employees

	Employees			
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	N=350

Sol: St 1: Null hypothesis H_0:

There is no significant liking in the habit of taking soft drinks.

St 2: Alternative hypothesis H_1:

There is a significant liking in the habit of taking soft drinks.

St 3: L.O.S : $\alpha = 0.05$ & $\nu = (r-1)(c-1)$ d.f $= (3-1)(3-1)$ d.f

$\alpha = 0.05$ & $\nu = 4$ d.f

$\chi^2_{tab} = 9.488$

St 4: table

$$E_i = \frac{\text{row total} \times \text{column total}}{\text{Grand total}}$$

$E(a) = \frac{100 \times 75}{350} = 21.42$ $\quad E(b) = \frac{100 \times 115}{350} = 32.85$

$E(c) = \frac{100 \times 160}{350} = 45.71$ $\quad E(d) = \frac{110 \times 75}{350} = 23.57$

$E(e) = \frac{110 \times 115}{350} = 36.14$ $\quad E(f) = \frac{110 \times 160}{350} = 50.28$

$E(g) = \frac{140 \times 75}{350} = 30$ $\quad E(h) = \frac{140 \times 115}{350} = 46$

$E(i) = \frac{140 \times 160}{350} = 64$

O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
10	21.42	-11.42	130.42	6.098
25	32.85	-7.85	61.62	1.875
65	45.71	19.29	372.104	8.14
15	23.57	-8.57	73.44	3.116
30	36.14	-6.14	37.69	1.043
65	50.28	14.72	216.67	4.309
50	30	20	400	13.33
60	46	14	196	4.26
30	64	-34	1156	18.06
				60.23

$$\therefore \chi^2_{cal} = 60.23$$

Step 5: Conclusion:

$$\chi^2_{cal} (60.23) > \chi^2_{tab} (9.488)$$

$\therefore$ we reject H_0

i.e., There is a significant liking in the habit of consuming soft drinks

Analysis of Variance (ANOVA)

• Assumptions.

- The analysis of variance is based on the following assumptions:

1. The samples are independently drawn from the populations
2. Populations from which the samples are selected are normally distributed
3. Each of the population has the same variance

The observations may be classified according to one factor or two factors. The classification according to one factor and 2 factors are called one-way classification & 2-way classification respectively

ANOVA for one-way classification:

In the investigation of Analysis of variance (ANOVA), if we consider the influence of any one factor and more than two sample sets, then it is called one-way classification

Ex: The yields of several plots of land may be classified according to one or more types of fertilizers

Algorithm:

Step1: Null hypothesis H_0 : $\mu_1 = \mu_2 = \mu_3 = \cdots\cdots = \mu_n$

i.e, All means of treatments are homogeneous

Step2: Alternative hypothesis H_1 : $\mu_1 \neq \mu_2 \neq \mu_3 \cdots\cdots \neq \mu_n$

i.e, All means of treatments are not homogeneous or heterogeneous

Step3: Sum of squares:

Total sum of squares = Treatment sum of square + Error sum of square

$$T.S.S = Tr.S.S + E.S.S$$

$T.S.S$ = Raw sum of square − Correction factor

$$= \Sigma x_{ij}^2 - \frac{G^2}{N}$$

$$Tr.S.S = \Sigma \frac{T_{i.}}{n} - \frac{G^2}{N}$$

$$E.S.S = T.S.S - Tr.S.S$$

Here,

Σx_{ij}^2 = Sum of squares of i^{th} row & j^{th} column total

G = Grand Total

N = Entire no. of obs'ns

$T_{i.}$ = i^{th} row total

Step 4: Degrees of freedom

Treatments = $(K-1)$ d.f

Here, K = no. of rows

Error = $(N-K)$ d.f

Total = $(N-1)$ d.f

Step 5: Mean sum of squares

$$M.Tr.S.S = S_t^2 = \frac{Tr.S.S}{K-1} \text{ d.f}$$

$$M.E.S.S = S_E^2 = \frac{E.S.S}{N-K} \text{ d.f}$$

Step 6: F-statistic

Treatment (Tab) : $\alpha\%$ with $\gamma_1 = (K-1), \gamma_2 = (N-K)$ d.f

Treatment (cal) ; $F_{cal} = \frac{S_t^2}{S_E^2}$

Step 7: ANOVA One-way Table

Source of variation	Sum of squares	Degrees of freedom	Mean sum of squares	Fcal	Ftab
Treatments	Tr.S.S	$(K-1)$	$S_t^2 = \frac{Tr.S.S}{K-1}$	S_t^2/S_E^2	$\alpha\%$ & $\nu_1 = (K-1)$ $\nu_2 = (N-K)$ d.f
Error	E.S.S	$(N-K)$	$S_E^2 = \frac{E.S.S}{N-K}$		
Total	T.S.S	$(N-1)$			

Step 8: Conclusion:

We compare calculated and tabulated values

i) if $F_{cal} \leq F_{tab}$, we accept H_0

2) if $F_{cal} > F_{tab}$ we reject H_0

Problems:

1. 3 different machines are used for a production, on the basis of the outputs. Test whether the machines are equally effective

Sol:

						T_j	T_j^2/n
output	Machine 1	10	5	11	10	36	$36^2/4 = 324$
	Machine 2	9	7	5	6	27	$27^2/4 = 182.25$
	Machine 3	20	16	10	4	50	$50^2/4 = 625$
						$G = 113$	$\Sigma \frac{T_j^2}{n} = 1131.25$

Step 1: Null hypothesis H_0:

The 3 machines are equally effective

i.e., $\mu_1 = \mu_2 = \mu_3$

Step 2: Alternative hypothesis (H_1):

The 3 machines are not equally effective

i.e., $\mu_1 \neq \mu_2 \neq = \mu_3$.

Step 3: Sum of squares.

$T.S.S = Tr.S.S + E.S.S$

$Tr.S.S = \Sigma x_{ij}^2 - \frac{G^2}{N}$

$Tr.S.S = \Sigma \frac{\overline{T_{l.}}^2}{n} - \frac{G^2}{N}$

$E.S.S = T.S.S - Tr.S.S$

$\frac{G^2}{N} = \frac{(113)^2}{12} = 1064.08$

$\Sigma x_{ij}^2 = 10^2 + 5^2 + 11^2 + 10^2 + 9^2 + 7^2 + 5^2 + 6^2 + 20^2 + 16^2 + 10^2 + 4^2$

$\Sigma x_{ij}^2 = 1309$

$T.S.S = \Sigma x_{ij}^2 - \frac{G^2}{N}$

$= 1309 - 1064.08$

$$Tr.S.S = \Sigma \frac{T_i^2}{n} - \frac{G^2}{N}$$

$= 1131.25 - 1064.08$

$Tr.S.S = 67.17$

$E.S.S = T.S.S - Tr.S.S = 244.92 - 67.17 = 177.75$

Step 4: Degrees of freedom

Treatments $= (K-1)\ d.f = 3-1 = 2\ d.f$

Error $= (N-K)\ d.f = 12-3 = 9\ d.f$

Total $= (N-1)\ d.f = 12-1 = 11\ d.f$

Step 5: Mean sum of squares

$$M.Tr.S.S = S_t^2 = \frac{Tr.S.S}{K-1} = \frac{67.17}{2} = 33.585$$

$$M.E.S.S = S_E^2 = \frac{E.S.S}{N-K} = \frac{177.75}{9} = 19.75$$

Step 6: F-statistic

Treatment (Tab) $= \alpha = 0.05$ & $v_1 = (K-1), v_2 = (N-K)\ d.f$

$v_1 = 2,\ v_2 = 9\ d.f$

$F_{tab} = 4.26$

Treatments (cal): $F_{cal} = \frac{S_t^2}{S_E^2} = \frac{33.585}{19.75} = 1.70$

Step 7: ANOVA one-way Table:

Source of variation	Sum of squares	Degrees of freedom	Mean sum of squares	Fcal	Ftab
Treatments	67.17	2	33.585	1.7	4.26
Error	177.7	9	19.75		
Total	244.92	11			

Step 8: Conclusion:

$F_{cal}\ (1.7) < F_{tab}\ (4.26)$

$\therefore$ We accept Null hypothesis

* 1. The following table shows the life in hours of 4 batches of electric lamps

Batch 1:	1600	1610	1650	1680	1700	1720	1800	
2:	1580	1640	1640	1700	1750			
3:	1460	1550	1600	1620	1640	1660	1740	1820
4:	1510	1520	1530	1570	1600	1680		

Perform ANOVA of this data show that significant test doesn't reject their homogeniety

Sol: St 1: Null hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$

i.e., all means of batches are homogeneous

St 2: Alternative hypothesis $H_1 : \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$

i.e., all means of batches are not homogeneous

St 3: Sum of squares:

$T.S.S = Tr.S.S + E.S.S$

$T.S.S = \sum x_{ij}^2 - \frac{G^2}{N}$

$Tr.S.S = \sum \frac{T_{i.}^2}{n} - \frac{G^2}{N}$

$E.S.S = T.S.S - Tr.S.S$

The above data can be reduced by using shifting to origin

i.e., $\frac{X-A}{h}$; x = observation

A = Assumed observation

h = common factor

$A = 1640$
$h = 10$

Batch									$T_{i.}$	$\frac{T_{i.}^2}{n}$
1	-4	-3	1	4	6	8	16		28	$28^2/7 = 112$
2	-6	0	0	6	11				11	$11^2/5 = 24.2$
3	-18	9	-4	-2	0	2	10	18	-3	$(-3)^2/8 = 1.125$
4	-13	-12	-11	-7	-4	4			-43	$(-43)^2/6 = 308.16$
									$G = -7$	$\sum \frac{T_{i.}^2}{n} = 445.48$

$\frac{G^2}{N} = \frac{(-7)^2}{26} = 1.88$

$\sum x_{ij}^2 = (-4)^2 + (-3)^2 + 1^2 + 4^2 + 6^2 + 8^2 + 16^2 + (-6)^2 + 6^2 + 11^2 +$

$$T.S.S = \Sigma x_{ij}^2 - \frac{G^2}{N}$$

$$= 1959 - 1.88$$

$$T.S.S = 1957.12$$

$$Tr.S.S = \Sigma \frac{T_i.^2}{n} - \frac{G^2}{N}$$

$$= 445.48 - 1.88$$

$$Tr.S.S = 443.6$$

$$E.S.S = T.S.S - Tr.S.S$$

$$= 1957.12 - 443.6$$

$$E.S.S = 1513.52$$

St 4: Degrees of freedom

Treatments $= (K-1)$ d.f $= 4-1 = 3$ d.f

Error $= (N-K)$ d.f $= 26-4 = 22$ d.f

Total $= (N-1)$ d.f $= 26-1 = 25$ d.f

St 5: Mean sum of squares

$$M.Tr.S.S = S_t^2 = \frac{Tr.S.S}{K-1} = \frac{443.6}{3} = 147.86$$

$$M.E.S.S = S_E^2 = \frac{E.S.S}{N-K} = \frac{1513.52}{22} = 68.79$$

St 6: F statistic

Treatment (Tab) : $\alpha = 0.05$ & $v_1 = (K-1), v_2 = (N-K)$ d.f

$v_1 = 3, \; v_2 = 22$ d.f

$$F_{tab} = 3.05$$

Treatment (cal) : $F_{cal} = \frac{S_t^2}{S_E^2} = \frac{147.86}{68.79} = 2.149$

St 7: ANOVA one-way Table:

S.V	S.S	D.F	M.S.S	Fcal	Ftab
Treatments	443.6	3	$S_t^2 = 147.83$	$F_{cal} = 2.149$	3.05
Error	1513.52	22	$S_E^2 = 68.79$		
Total	1957.12	25			

st 8: Conclusion:

$F_{cal}\ (2.149) < F_{tab}\ (3.05)$

$\therefore$ we accept Null hypothesis

i.e., The (average) means of 4 batches of electric lamps are homogeneous

3. Four salesmen were posted in different areas by a company. The number of units of company sold by them is as follows:

A	20	23	28	29
B	25	32	30	21
C	23	28	35	80
D	15	21	19	25

On the basis of this information, can it be concluded that there is a significant difference in the performance of the four salesmen

ANOVA for two-way classification:

In two-way classification, observations are classified according to two different factors and more than 2 sample sets.
for ex: fertilizers may be tried on different soil textures

Algorithm:

Step 1: Null hypothesis H_0:

All means of treatments are homogeneous
(or)
All means of varieties are homogeneous

Step 2: Alternative hypothesis H_1:

All means of treatments are not homogeneous
(or)
All means of varieties are not homogeneous / heterogeneous

Step 3: Sum of squares:

$T.S.S = Tr.S.S + V.S.S + E.S.S$

$T.S.S = \Sigma x_{ij}^2 - \frac{G^2}{N}$; Σx_{ij}^2 = Sum of squares of ith row & jth column total

G = Grand total

N = Total/entire no. of ob's

$Tr.S.S = \Sigma \frac{T_{i.}^2}{n} - \frac{G^2}{N}$; n = no. of obs'ns corresponding to the each row

$V.S.S = \Sigma \frac{T_{.j}^2}{n} - \frac{G^2}{N}$; T_j = jth column total & n = no. of obs'ns corresponding to the each column

$E.S.S = T.S.S - Tr.S.S - V.S.S$

Step 4: Degrees of freedom:

Treatments $= (r-1)$ d.f

Variety $= (c-1)$ d.f

Error $= (r-1)(c-1)$ d.f

Total $= (N-1)$ d.f

Step 5: Mean sum of square

$$M.Tr.S.S = S_t^2 = \frac{Tr.S.S}{r-1}$$

$$M.V.S.S = S_V^2 = \frac{V.S.S}{c-1}$$

$$M.E.S.S = S_E^2 = \frac{E.S.S}{(r-1)(c-1)}$$

Step 6: F - statistic

Treatments (Tab): $\alpha\%$ & $v_1 = (r-1), v_2 = (r-1)(c-1)$ d.f

Variety (Tab): $\alpha\%$ & $v_1 = (c-1), v_2 = (r-1)(c-1)$ d.f

Treatment (cal): $F_{cal} = \frac{S_t^2}{S_E^2}$

Variety (cal): $F_{cal} = \frac{S_V^2}{S_E^2}$

Step 7: ANOVA Two-way Table:

S.V	S.S	D.f	M.S.S	Fcal	Ftab
Treatments	Tr.S.S	$(r-1)$	$S_t^2 = \frac{Tr.S.S}{r-1}$	$F_{cal} = \frac{S_t^2}{S_E^2}$	$\alpha\%, v_1 = (r-1), v_2 = (r-1)(c-1)$ d.f
Variety	V.S.S	$(c-1)$	$S_V^2 = \frac{V.S.S}{c-1}$	$F_{cal} = \frac{S_V^2}{S_E^2}$	$\alpha\%, v_1 = (c-1), v_2 = (r-1)(c-1)$ d.f
Error	E.S.S	$(r-1)(c-1)$	$S_E^2 = \frac{E.S.S}{(r-1)(c-1)}$		
Total	T.S.S	$(N-1)$			

Step 8: Conclusion:

We compare calculated and tabulated values:

1. Treatments: $F_{cal} \le F_{tab}$, We accept H_0
   o/w we reject H_0

2. Variety: $F_{cal} \le F_{tab}$, we accept H_0

Problems:

*1. A farmer applies 3 types of fertilizers on 4 separate plots. The figure on yields per acre are tabulated below:

plots / Fertilizers	Yield A	B	C	D	$T_{i.}$ Total
Nitrogen	6	4	8	6	24
Potash	7	6	6	9	28
phosphates	8	5	10	9	32
Total	21	15	24	24	84 = G

Find out if the plots are materially different in fertility As also if 3 fertilizers make any material significance in yields

Sol: St 1: Null hypothesis H_0:

All means of fertilizers are homogeneous and all means of yields are homogeneous

St 2: Alternative hypothesis H_1:

All means of fertilizers are not homogeneous and All means of yields are not homogeneous

St 3: Sum of squares:

$$T.S.S = Tr.S.S + V.S.S + E.S.S$$

$$T.S.S = \Sigma x_{ij}^2 - \frac{G^2}{N}$$

$$Tr.S.S = \Sigma \frac{T_{i.}^2}{n} - \frac{G^2}{N}$$

$$V.S.S = \Sigma \frac{T_{.j}^2}{n} - \frac{G^2}{N}$$

$$E.S.S = T.S.S - Tr.S.S - V.S.S$$

$T_{i.}^2/n$

plots / Fertillizers	Yield A	B	C	D	Total $T_{i.}$	$T_{i.}^2/n$
Nit	6	4	8	6	24	$24^2/4 = 144$
pot	7	6	6	9	28	$28^2/4 = 196$
phos	8	5	10	9	32	$32^2/4 = 256$
Total	21	15	24	24	$G = 84$	$\sum T_{i.}^2/n = 596$
$T_{.j}^2/n$	$21^2/3 = 147$	75	192	192	$\sum T_{.j}^2/n = 606$	

$$\frac{G^2}{N} = \frac{84^2}{12} = 588.$$

$$\sum x_{ij}^2 = 6^2+4^2+8^2+6^2+7^2+6^2+6^2+9^2+8^2+5^2+10^2+9^2$$

$$\sum x_{ij}^2 = 624$$

$$T.S.S = \sum x_{ij}^2 - \frac{G^2}{N}$$

$$= 624 - 588$$

$$T.S.S = 36$$

$$Tr.S.S = \sum \frac{T_{i.}^2}{n} - \frac{G^2}{N}$$

$$= 596 - 588$$

$$Tr.S.S = 8$$

$$V.S.S = \sum \frac{T_{.j}^2}{n} - \frac{G^2}{N}$$

$$= 606 - 588$$

$$V.S.S = 18$$

$$E.S.S = T.S.S - Tr.S.S - V.S.S$$

$$= 36 - 8 - 18$$

$$E.S.S = 10$$

St 4: Degrees of freedom

Error = $(r-1)(c-1)$ d.f = 2×3 = 6 d.f

Total = $(N-1)$ d.f = $12-1$ = 11 d.f

St 5: Mean sum of squares:

$$M.Tr.S.S = S_t^2 = \frac{Tr.S.S}{r-1} = \frac{8}{2} = 4$$

$$M.V.S.S = S_V^2 = \frac{V.S.S}{C-1} = \frac{18}{3} = 6$$

$$M.E.S.S = S_E^2 = \frac{E.S.S}{(r-1)(c-1)} = \frac{10}{6} = 1.66$$

St 6: F-statistic:

Treatments (Tab): $\alpha = 0.05$ & $v_1 = (r-1)$, $v_2 = (r-1)(c-1)$ d.f

$v_1 = 2$, $v_2 = 6$ d.f

$F_{tab} = 5.14$

Variety (Tab): $\alpha = 0.05$ & $v_1 = (c-1)$, $v_2 = (r-1)(c-1)$ d.f

$v_1 = 3$, $v_2 = 6$ d.f

$F_{tab} = 4.76$

S. Treatments (cal): $F_{cal} = \frac{S_t^2}{S_E^2} = \frac{4}{1.66} = 2.409$

Variety : $F_{cal} = \frac{S_V^2}{S_E^2} = \frac{6}{1.66} = 3.61$

St 7: ANOVA 2-way Table:

S.V	S.S	D.f	M.S.S	Fcal	Ftab
Treatments	8	2	4	2.409	5.14
Variety	18	3	6	3.614	4.76
Error	10	6	1.66		
Total	36	11			

St 8: Conclusion:

Treatments: F_{cal} (2.409) < F_{tab} (5.14)

$\therefore$ we accept H_0

i.e, All means of fertilizers are homogeneous i.e., All means of yields are homogeneous

2. 3 varieties of charcoal were analyzed by 4 chemists and at the ash content in the varieties are found to be as:

Varieties	Chemists 1	2	3	4
A	8	5	5	7
B	7	6	4	4
C	3	6	5	4

Do the varieties differ significantly in their ash content at 5% level of significance