ANOVA

Analysis of Variance :-

When we have three or more samples to consider at a time a procedure is needed for testing the hypothesis that all the samples are drawn from the populations with the same mean, such procedure is ANOVA.

The basic purpose of the analysis of variance is to test the homogeneity of several means.

Formulas for computation of variu. - sum of squares.

1) Grand total $= G = \sum_i \sum_j x_{ij}$

$=$ total of all observations

2) Correction Factor (C.F.) $= \frac{G^2}{n}$

where $n = n_1 + n_2 + \dots + n_k$

3) Raw sum of squares (RSS) $= \sum_i \sum_j x_{ij}^2$

$=$ sum of the squares of all observations.

4) Total sum of squares $=$ RSS $-$ C.F.

5) $T_i = \sum_{j=1} x_{ij}$ $=$ sum of observations in i^{th} class.

6) Between classes sum of squares (or) Treatement

$$\text{Sum of squares} = TrSS = \frac{T_1^2}{n_1} + \frac{T_2^2}{n_2} + \cdots + \frac{T_k^2}{n_k} - C.F.$$

7) With in classes sum of squares (or) Error sum

$$\text{of squares} = ErSS = TSS - TrSS.$$

Source of variation	degrees of freedom	Sum of squares	mean sum of square	F-calculated.	F-tab.
Between classes (or) Treatements:	$K-1$	$TrSS$	$MTr = \frac{TrSS}{K-1}$	$F_{cal} = \frac{MTr}{MEr}$	$F(K-1, n-K)$ degrees of Freedom
With in classes, (or) Error	$n-K$	$ErSS$	$MEr = \frac{ErSS}{n-K}$		

Null Hypothesis H_0: $\mu_1 = \mu_2 = \mu_3 = \cdots = \mu_k$.

(ie The population means are same)

Test statistic: $F = \frac{MTr}{MEr} = \left(\frac{\text{Between class M.S.S}}{\text{with in class M.S.S.}}\right)$

<u>conclusion</u>:

If F computed is greater than F tabulated at $(K-1, n-K)$ degrees of freedom then we reject the Null hypothesis. otherwise we Accept the null Hypothesis (ie $F_{cal} < F_{tab}$ accept)

Ex 1) A trucking company wishes to test the average life of each of the four brands of tyres. The company uses all brands on randomly selected trucks. The records showing the lives (thousands of miles) of tyres are as given in the table

Brand 1	Brand 2	Brand 3	Brand 4
20	19	21	15
23	15	19	17
18	17	20	16
17	20	17	18
	16	16	

Test the hypothesis that the average life for brand of tyre is the same. Assume $\alpha = 0.01$.

Solution :

Null Hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$.

(ie The mean life of the tyres of all brands is same.)

	Brand 1	Brand 2	Brand 3	Brand 4
	20	19	21	15
	23	15	19	17
	18	17	20	16
	17	20	17	18
		16	16	
Total :	$T_1 = 78$	$T_2 = 87$	$T_3 = 93$	$T_4 = 66$

n = total no. of observations = 18 (4+5+5+4)

G = Grand total = 78 + 87 + 93 + 66 = 324.

Correction Factor = CF = $\frac{G^2}{n} = \frac{(324)^2}{18} = 5832.$

Raw sum of square (RSS) = $\sum\sum x_{ij}^2$

$= [(20)^2 + (23)^2 + (18)^2 + (17)^2] + [(19)^2 + (15)^2 + (17)^2 + (20)^2 + (16)^2] + [(21)^2 + (19)^2 + (20)^2 + (17)^2 + (16)^2] + [(15)^2 + (17)^2 + (16)^2 + (18)^2]$

RSS = 5914.

Total sum of square (TSS) = RSS − CF

TSS = 5914 − 5832 = 82

Between samples sum of squares = $\frac{T_1^2}{n_1} + \frac{T_2^2}{n_2} + \frac{T_3^2}{n_3} + \frac{T_4^2}{n_4} - CF$

$TrSS = \frac{(78)^2}{4} + \frac{(87)^2}{5} + \frac{(93)^2}{5} + \frac{(66)^2}{4} - 5832$

TrSS = 21.6.

Error sum of squares = TSS − TrSS

ErSS = 82 − 21.6 = 60.4

Sources of variation	d.f	sum of squares	mean sum of squares	F cal	F-table
Between Brands of tyres	4−1=3	TrSS = 21.6	$M_{TY} = \frac{TrSS}{3} = \frac{21.6}{3} = 7.2$	$F = \frac{M_{TY}}{M_{ER}} = \frac{7.2}{4.31} = 1.67$	at (3, 14) d.f & 1% level. $F_{tab} = 5.56$
Error	18−4=14	ErSS = 60.4	$M_{ER} = \frac{ErSS}{14} = \frac{60.4}{14} = 4.31$		

∴

F calculate $= 1.67 < 5.56$ (F table value)

Hence we accept the null Hypothesis.

i.e. There is no significant difference between the average lives of the four brands of tyres 1, 2, 3 and 4.

Ex 2) To test the hypothesis that the average number of days a patient is kept in the three local hospital say A, B and C is the same, a random check on the number of days that seven patients stayed in each hospital reveals the following.

Hospital A	8	5	9	2	7	8	2
Hospital B	4	3	8	7	7	1	5
Hospital C	1	4	9	8	7	2	3

Test the hypothesis at $\alpha = 0.05$.

Solution :- Null Hypothesis H_0 : $\mu_1 = \mu_2 = \mu_3$.

	A	B	C
	8	4	1
	5	3	4
	9	8	9
	2	7	8
	7	7	7
	8	1	2
	2	5	3
Total	$T_1 = 41$	$T_2 = 35$	$T_3 = 34$

n = Total number of observations $= 7 + 7 + 7 = 21$.

Grand Total $(G) = \Sigma\Sigma x_{ij} = (8+5+9+2+7+8+2) + (4+3+8+7+7+1+5) + (1+4+9+8+7+2+3)$

$G = 41 + 35 + 34 = 110$

C.F = correction Factor $= \frac{(G)^2}{n} = \frac{(110)^2}{21} = 576.1905$

Raw sum of square (RSS) $= \sum_i \sum_j x_{ij}^2$

$= (8^2+5^2+9^2+2^2+7^2+8^2+2^2) +$
$(4^2+3^2+8^2+7^2+7^2+1^2+5^2) + (1^2+4^2+9^2+8^2+7^2+2^2+3^2)$

$= 291 + 213 + 224 = 728$

RSS = 728.

Total sum of square (TSS) = RSS − C.F.

$= 728 - 576.1905$

TSS = 151.8095

Between classes sum of squares $= \frac{T_1^2}{n_1} + \frac{T_2^2}{n_2} + \frac{T_3^2}{n_3} - C.F$
(or) correction Factor
(Tr SS) (Treatment sum of square)

$TrSS = \frac{(41)^2}{7} + \frac{(35)^2}{7} + \frac{(34)^2}{7} - 576.1905$

Tr SS = 4.0952.

within classes sum of square (or) Error sum of square = TSS − Tr SS

(Error s.s) = Er SS $= 151.8095 - 4.0952$

Er SS $= 147.7138$

Source of variation	degrees of freedom	Sum of squares	Mean sum of square	Fcal	F-tab
1	2	3	$\frac{3}{2}$		
Between sum of squares (or) Treatement sum of square	$3-1=2$	TrSS $= 4.10$	$MTr = \frac{TrSS}{K-1}$ $= \frac{4.10}{2}$ $= 2.05$	$F_{cal} = \frac{MEr}{MTr}$ $= \frac{8.21}{2.05}$ $= 2.0049$	at $(18,2)$ degrees of freedom and 5% L.S. $F_{tab}(18,2)$ $= 19.4$.
With in classes sum of square (or) Error sum of square	$21-3=$ 18	ErSS $= 147.71$	$MEr = \frac{ErSS}{n-K}$ $= \frac{147.71}{18}$ $= 8.21$		

F calculated $= 2.0049 < F_{tab} = 19.4$.

Hence we accept the null hypothesis.

i.e The average no. of days the patient kept in the three hospitals A, B & C is same.