ANOVA

Two-way classification :-

Suppose n observations are classified into k categories (or classes) say $A_1, A_2, \cdots A_k$. according to some criterion A. Also into h categories say $B_1, B_2, \cdots B_h$ according to some criterian B, having hk combinations.

This scheme of classification is called two way classification, and its analysis is called two way Analysis of variation.

Ex. 1) A farmer applies three types of fertilizers on 4 saparate plots. The figure on yield per acre are tabulated below.

Fertilizers	Yield			
Plots	A	B	C	D
Nitrogen	6	4	8	6
Potash	7	6	6	9
Phosphates	8	5	10	9

Find out if plots are materially different in fertility, as also, if the three fertilizers make any material difference in yields.

Solution :-

A : Fertilizers, represents along the rows of the given table

B : Plots, represents along the column of the given table.

Null Hypothesis :

H_{0A} : $\mu_{1.} = \mu_{2.} = \mu_{3.}$

(i.e No significance difference between the fertilizers)

H_{0B} : $\mu_{.1} = \mu_{.2} = \mu_{.3} = \mu_{.4}$

i.e. There is no significant difference between the plots.

Fertilizers	yield: Plots A	B	C	D	Raw totals (R)	R^2
Nitrogen	6	4	8	6	$R_1 = 24$	$R_1^2 = 576$
Potash	7	6	6	9	$R_2 = 28$	$R_2^2 = 784$
Phosphates	8	5	10	9	$R_3 = 32$	$R_3^2 = 1024$
column totals (C)	$C_1 = 21$	$C_2 = 15$	$C_3 = 24$	$C_4 = 24$	$G = 84$	$\Sigma R_i^2 = 2384$
C^2	$C_1^2 = 441$	$C_2^2 = 225$	$C_3^2 = 576$	$C_4^2 = 576$	$\Sigma C_j^2 = 1818$	

Raw sum of squares $= (RSS) = (6^2 + 4^2 + 8^2 + 6^2)$

$(7^2 + 6^2 + 6^2 + 9^2) + (8^2 + 5^2 + 10^2 + 9^2) = 624$

$RSS = 624$

$G =$ Grand total $= \Sigma\Sigma x_{ij} = 84.$

$n =$ total no. of observations $= hk = 3 \times 4 = 12$

correction factor $= \dfrac{G^2}{n} = \dfrac{(84)^2}{12} = 588$

Total sum of square $= RSS - CF$

$TSS = 624 - 588 = 36.$

sum of squares of fertilizers $= \dfrac{R_1^2}{4} + \dfrac{R_2^2}{4} + \dfrac{R_3^2}{4} - CF$

$SSA = \dfrac{2384}{4} - 588 = 8.$

sum of squares of yield $= \dfrac{C_1^2}{3} + \dfrac{C_2^2}{3} + \dfrac{C_3^2}{3} + \dfrac{C_4^2}{3} - CF$

$SSB = \dfrac{1818}{3} - 588 = 18.$

Error sum of squares $(SSE) = TSS - SSA - SSB$

$$SSE = 36 - 8 - 18 = 10.$$

Source of variation	d.f	Sum of squares	Mean sum of squares	F calculated	F tabulated
Between Rows (Fertilizers)	$k-1 = 3-1 = 2$	$SSA = 8$	$MSSA = \frac{SSA}{k-1} = \frac{8}{2} = 4$	$F_A = \frac{MSSA}{MSSE} = \frac{4}{1.67} = 2.395$	$F_{0.05}(2,6) = 5.14$
Between columns (plots)	$h-1 = 4-1 = 3$	$SSB = 18$	$MSSB = \frac{SSB}{h-1} = \frac{18}{3} = 6$	$F_B = \frac{MSSB}{MSSE} = \frac{6}{1.67} = 3.593$	$F_{0.05}(3,6) = 4.76$
Error	$(h-1)(k-1) = 2.3 = 6$	$SSE = 10$	$MSSE = \frac{SSE}{(h-1)(k-1)} = \frac{10}{6} = 1.67$		
Total		$TSS = 36$			

F_A calculated $= 2.395 < F$ tabulated $= 5.14$.

Hence Accept the null Hypothesis H_{0A}.

F_B calculated $= 3.593 < F$ tabulated $= 4.76$

Hence accept the Null Hypothesis H_{0B}.

Ex. 2. Five doctors each test five treatments for a certain disease and observe the number of days each patient takes to recover. The results are (recovery time in days) given in table.

Doctors	Treatments				
	1	2	3	4	5
1	10	14	23	18	20
2	11	15	24	17	21
3	9	12	20	16	19
4	8	13	17	17	20
5	12	15	19	15	22

Discuss the difference between the doctors and the treatments is significant at $\alpha = 0.05$.

<u>Solution</u> :- Here the factors of variation are

A: Doctors, represented along the rows of given table.

B: Treatements represented along the columns of the given table.

Null Hypothesis :

H_{0A} : $\mu_{1.} = \mu_{2.} = \mu_{3.} = \mu_{4.} = \mu_{5.}$

(ie No significance difference between the Doctors).

H_{0B} : $\mu_{.1} = \mu_{.2} = \mu_{.3} = \mu_{.4} = \mu_{.5}$

(ie There is no significant difference between the treatments)

Doctors	Treatments 1	2	3	4	5	Row totals (R)	R^2
1	10	14	23	18	20	$R_1 = 85$	$R_1^2 = 7225$
2	11	15	24	17	21	$R_2 = 88$	$R_2^2 = 7744$
3	9	12	20	16	19	$R_3 = 76$	$R_3^2 = 5776$
4	8	13	17	17	20	$R_4 = 75$	$R_4^2 = 5625$
5	12	15	19	15	22	$R_5 = 83$	$R_5^2 = 6889$
column totals (C)	$C_1 = 50$	$C_2 = 69$	$C_3 = 103$	$C_4 = 83$	$C_5 = 102$	$G = 407$	$\sum_i R_i^2 = 33259$
C^2	$C_1^2 = 2500$	$C_2^2 = 4761$	$C_3^2 = 10609$	$C_4^2 = 6889$	$C_5^2 = 10404$	$\sum_j C_j^2 = 35163$	

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

$$RSS = 10^2 + 14^2 + 23^2 + \dots + 19^2 + 15^2 + 22^2$$

$$RSS = 7093$$

$G =$ Grand total $= \sum \sum x_{ij} = \sum R_i = \sum C_j = 407$

$n =$ Total number of observations $= h \cdot k = 5 \times 5 = 25$

correction factor $= C.F = \frac{G^2}{n} = \frac{(407)^2}{25}$

$$CF = 6625.96$$

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

$$TSS = 7093 - 6625.96 = 467.04$$

sum of squares of doctors $= \frac{R_1^2}{5} + \frac{R_2^2}{5} + \frac{R_3^2}{5} + \frac{R_4^2}{5} + \frac{R_5^2}{5} - CF$

$$SSA = \frac{33259}{5} - 6625.96 = 25.84$$

Sum of squares of Treatments =

$$\frac{C_1^2}{5} + \frac{C_2^2}{5} + \frac{C_3^2}{5} + \frac{C_4^2}{5} + \frac{C_5^2}{5} - CF$$

$$SSB = \frac{35163}{5} - 6625.96 = 406.64$$

Sum of squares due to error is given by

$$SSE = TSS - SSA - SSB$$

$$SSE = 467.04 - 25.84 - 406.64 = 34.56$$

Sources of Variation	d.f	Sum of squares	Mean sum of squares	Fcal.	F tab.
A Doctors	$k-1=4$	$SSA = 25.84$	$MSSA = \frac{SSA}{k-1} = \frac{25.84}{4} = 6.46$	$F_A = \frac{MSSA}{MSSE} = \frac{6.46}{2.16} = 2.99$	$F(4,16) = 3.01$
B. Treatments	$h-1=4$	$SSB = 406.64$	$MSSB = \frac{SSB}{h-1} = \frac{406.64}{4} = 101.66$	$F_B = \frac{MSSB}{MSSE} = \frac{101.66}{2.16} = 47.06$	
Error	$(h-1)(k-1) = (4)(4) = 16$	$SSE = 34.56$	$MSSE = \frac{SSE}{(k-1)(h-1)} = \frac{34.56}{4\times4} = 2.16$		
Total	$hk-1 = 5\times5-1 = 24$	$TSS = 467.04$			

calculated $F_A = 2.99 <$ F tabulated value $= 3.01$

Hence H_{0A} is Accepted.

calculated $F_B = 47.06 >$ F tabulated value.

Hence H_{0B} is Rejected.