MODULE - 5

ANOVA [Analysis of Variance]

The term analysis of variance was introduced by R.A. Fisher in 1920. It is a powerful statistical tool for Test of Significance. It enables us to compare several population means simultaneously.

Eg: A Hospital might want to compare 3 brands of pain killer before adopting 1.

Eg: An average fuel consumption per km for four popular automobile

Assumption:

1. Random Selection - Samples are drawn randomly
2. Normal Distribution - Population from which samples are drawn follows Normal Distribution
3. Homogenity of Variance - Each one of the population has the same variance.
4. Additivity of Variance - Total Variance should be equal to sum of between variance and within variance.

WORKING PROCEDURE:

group	observation			Total
1	x_{11}	x_{12}	....	$T_1 = x_{11} + x_{12} + \cdots$
2	x_{21}	x_{22}	....	$T_2 = x_{21} + x_{22} + \cdots$
⋮				
K	x_{K1}	x_{K2}	....	$T_K = x_{K1} + x_{K2} + \cdots$

Grand Total = $T_1 + T_2 + \dots + T_k$

(i) Define $H_0 = \mu_1 = \mu_2 = \dots = \mu_k$

(ii) Correction factor = $C = \frac{(G)^2}{N}$ (N -)

$N \to$ Total No of observations

(iii) Sum of Squares Total $= \sum_i \sum_j x_{ij}^2 - C = S_T^2$

(iv) Sum of Squares b/w groups $= \sum_{i=1}^{K} \frac{T_i^2}{n_i} - C = S_t^2$

(v) Sum of Squares of Errors $= S_T^2 - S_t^2 = S_E^2$

Anova Table

Source of Variance	Degree of freedom	Sum of Squares	Mean of (MSS) Sum of Squares	F-r
b/w group	$K-1$ [K - no of grps]	S_t^2	$S_t^2 / K-1$	
Error	$N-K$	$S_E^2 = S_T^2 - S_t^2$		
Total	$N-1$ [N - Total no of observation]	S_T^2	$\frac{S_T^2}{N-K}$	

$$F = \frac{(\text{MSS b/w group})}{(\text{MSS of error})}$$

Conclusion

$F_{(K-1, N-K)}$ Calculated Value < F Tabulated Value $\Rightarrow$ Accept H_0

— " — " — > — " — " — $\Rightarrow$ Reject H_0

PROBLEMS :

1) In an experiment to determine the effect of nutrition on the attention spans of elementary school students. A group of 15 students where randomly assigned to each of 3 meal plants. No breakfast. light breakfast, ~~the~~ Full breakfast. They attention spanas [in minutes] were recorded during a morning reading period and are shown as :

No breakfast	8	7	9	13	10
Light "	14	16	12	17	11
Full "	10	12	16	15	12

Construct the ANOVA table for this experiment

Soln

						Total
No bf	8	7	9	13	10	$T_1 = 47$
light bf	14	16	12	17	11	$T_2 = 70$
Full bf	10	12	16	15	12	$T_3 = 65$

Grand Total (G) $= T_1 + T_2 + T_3$

$G = 180$

Define $H_0 = \mu_1 = \mu_2 = \mu_3$

Correction factor $= C = \frac{(G)^2}{N} = \frac{(182)^2}{15}$

$\boxed{C = 2208.2667}$

Sum of Squares Total $= S_T^2 = \sum_i \sum_j x_{ij}^2 - C$

$= [(8^2 + 7^2 + 9^2 + 13^2 + 10^2) + (14^2 + 16^2 + 12^2 + 17^2 + 11^2) + (10^2 + 12^2 + 16^2 + 15^2 + 12^2)] - 2208.2667$

$463 + 1006 + 869$

$$\boxed{S_T^2 = 129.733}$$

Sum of Squares b/w groups $= S_t^2 = \sum_{i=1}^{3} \frac{T_i^2}{n_i} - C$

$$= \left[\frac{T_1^2}{n_1} + \frac{T_2^2}{n_2} + \frac{T_3^2}{n_3}\right] - C$$

$$= \left[\frac{47^2}{5} + \frac{70^2}{5} + \frac{65^2}{5}\right] - 2208.2667$$

$$\boxed{S_t^2 = 58.533}$$

Sum of Squares of Errors $= S_E^2 = S_T^2 - S_t^2$

$$= 129.733 - 58.533$$

$$\boxed{S_E^2 = 71.2}$$

ANNOVA TABLE

Source of Variance	Degree of freedom	Sum of Squares	Mean of Sum of Squares [MSS]	F ratio
b/w groups	$K-1 = 3-1 = 2$	58.5333	$\frac{58.5333}{2} = 29.267$	$F = \frac{29.267}{5.933}$ $= 4.9326$
Errors	$N-K = 12$	71.2	$\frac{71.2}{12} = 5.933$	
Total	$N-1 = 15-1 = 14$	129.7333	~~$\frac{129.7333}{14}$~~	

$F_{(2.12)}$ Calculated value $>$ $f_{(2.12)}$ Table value

$4.9326 > 3.89$

Reject the H_0

2). A Trial was run to check the effects of dift diets positive numbers indicate weight loss and negative number indicates weight gain. Check if there is an avg difference in the weight of ppl following dift diets using an ANOVA table:

low fat	low calorie	low protein	low carbohydrate
8	2	3	2
9	4	5	2
6	3	4	-1
7	5	2	0
3	1	3	3

Soln.

						Total
low fat	8	9	6	7	3	$T_1 = 33$
low calorie	2	4	3	5	1	$T_2 = 15$
low protein	3	5	4	2	3	$T_3 = 17$
low carbohydrate	2	2	-1	0	3	$T_4 = 6$

Grand Total (T_G)
$= T_1 + T_2 + T_3 + T_4 = 71$

Define $H_0 = \mu_1 = \mu_2 = \mu_3 = \mu_4$

Correction factor $- \; C = \frac{(G)^2}{N} = \frac{(71)^2}{20}$ $\boxed{C = 252}$

Sum of Squares Total $= S_T^2 = \sum_i \sum_j x_{ij}^2 - C$

$= \left[(8^2+9^2+6^2+7^2+3^2) + (2^2+4^2+3^2+5^2+1^2) + (3^2+5^2+4^2+2^2+3^2) + (2^2+2^2+(-1)^2+0^2+3^2)\right] - 252$

$= 239 + 55 + 63 + 18 - 252$ $\boxed{S_T^2 = 123}$

Sum of Squares b/w groups $= S_t^2 = \sum_{i=1}^{4} \frac{T_i^2}{n_i} - C$

$$= \left[\frac{T_1^2}{n_1} + \frac{T_2^2}{n_2} + \frac{T_3^2}{n_3} + \frac{T_4^2}{n_4}\right] - C$$

$$= \left[\frac{(32)^2}{5} + \frac{(15)^2}{5} + \frac{(17)^2}{5} + \frac{6^2}{5}\right] - 252$$

$$\boxed{S_t^2 = 75.8}$$

Sum of Squares of Errors $= S_E^2 = S_T^2 - S_t^2$

$$= 123 - 75.8$$

$$\boxed{S_E^2 = \cancel{60.2}\ 47.2}$$

Source of Variance	Degree of freedom	Sum of Squares	MSS	F-ratio
b/w grps	$K-1=4-1=3$	$S_t^2 = 75.80$	$\frac{75.80}{3} = 25.26$	$F = \frac{25.26}{2.95} = \underline{8.56}$
Errors	$N-K=16$	$S_E^2 = 47.2$	$\frac{47.2}{16} = 2.95$	
Total	$N-1=20-1=19$	$S_T^2 = 123$		

$F_{(3,16)}$ cal $> F_{(3,16)}$ Tab

$8.56 > 3.24$

Reject H_0

TWO WAY ANOVA :

WORKING PROCEDURE :

Groups	Observation A	B	Total
1	x_{11}	x_{12}	$T_1 = x_{11} + x_{12} + \dots$
2	x_{21}	x_{22}	$T_2 = x_{21} + x_{22} + \dots$
3	x_{31}	x_{32}	
⋮	⋮		
K	x_{K1}	x_{K2}	$T_K = x_{K1} + x_{K2} + \dots$
Total	P_1	P_2	Grand Total $= T_1 + T_2 + \dots + T_K$

$$P_1 = x_{11} + x_{21} + x_{31} + \dots x_{K1}$$

$$P_2 = x_{12} + x_{22} + x_{32} + \dots x_{K2}$$

(ii) correction factor $= C = \frac{(G)^2}{N}$

(iii) Sum of Squares of Total $= S_T^2 = \sum_i \sum_j x_{ij}^2 - C$

(iv) Sum of squares b/w groups $= S_t^2$

For rows $= S_r^2 = \sum \frac{T_i^2}{n_i} - C$

For columns $= S_c^2 = \sum \frac{P_i^2}{n_i} - C$

(v) Sum of Square Error $= S_E^2 = S_T^2 - [S_r^2 + S_c^2]$

ANOVA TABLE :

Source of Variation	Degree of freedom	Sum of Squares	Mean Sum of squares [MSS]	F-ratio
b/w groups rows	$r-1$ (r = no. of rows)	S_r^2	$S_r^2/r-1$ $(MSS)_r$	$F_r = \frac{(MSS)_r}{(MSS)_E}$
columns	$c-1$ (c = no. of columns)	S_c^2	$S_c^2/c-1$ $(MSS)_c$	$F_c = \frac{(MSS)_c}{(MSS)_E}$
Error	$(r-1)(c-1)$	S_E^2	$S_E^2/(r-1)(c-1)$ $(MSS)_E$	
Total	$N-1$	S_T^2	$S_T^2/N-1$	

PROBLEMS :

1). Set up ANOVA table for the following Per acre production data for 3 varieties of wheat each grown on 4 plots and state it the variety differences are significant at 5% level of significance.

Per acre Production data			
Plot of land	Variety of Wheat		
	A	B	C
1	6	5	5
2	7	5	4
3	3	3	3
4	8	7	4

Soln.

Per acre Production data				
Plot of land	Variety of Wheat			Total
	A	B	C	
1	6	5	5	$T_1 = 16$
2	7	5	4	$T_2 = 16$
3	3	3	3	$T_3 = 9$
4	8	7	4	$T_4 = 19$
Total	$P_1 = 24$	$P_2 = 20$	$P_3 = 16$	

Grand Total (G) $= 60$

Defining $H_0 = \mu_1 = \mu_2 = \mu_3 = \mu_4$

Correction factor $= C = \frac{(G)^2}{N} = \frac{(60)^2}{12} = 300$.

Sum of Squares Total $= S_T^2 = \sum_i \sum_j x_{ij}^2 - C$

$= \left[(6^2+5^2+5^2)+(7^2+5^2+4^2) + (3^2+3^2+3^2)+(8^2+7^2+4^2)\right] - 300$

$= 86 + 90 + 27 + 129 - 300$

$\boxed{S_T^2 = 32}$

Sum of Squares b/w groups

For Rows $S_r^2 = \sum_{i=1}^{4} \frac{T_i^2}{n_i} - C$

$= \left[\frac{T_1^2}{n_1} + \frac{T_2^2}{n_2} + \frac{T_3^2}{n_3} + \frac{T_4^2}{n_4}\right] - C$

$= \left[\frac{16^2}{3} + \frac{16^2}{3} + \frac{9^2}{3} + \frac{19^2}{3}\right] - 300$

$\boxed{S_r^2 = 18}$

For columns $S_c^2 = \sum_{i=1}^{3} \frac{P_i^2}{n_i} - C$

$= \left[\frac{P_1^2}{n_1} + \frac{P_2^2}{n_2} + \frac{P_3^2}{n_3}\right] - C$

$= \left[\frac{(24)^2}{4} + \frac{(20)^2}{4} + \frac{(16)^2}{4}\right] - 300$

$\boxed{S_c^2 = 8}$

Sum of Squares of Errors $= S_E^2 = S_T^2 - [S_r^2 + S_c^2]$

$= 32 - [18 + 8]$

$\boxed{S_E^2 = 6}$

ANOVA Table :

Source of Variance	Degree of Freedom	Sum of Squares	Mean Sum of Squares (MSS)	F-ratio
b/w groups				
rows	$r-1=4-1=3$	$S_r^2=18$	$\frac{18}{3}=6\ (MSS)_r$	$F_r = \frac{(MSS)_r}{(MSS)_E} = 6$
columns	$c-1=3-1=2$	$S_c^2=8$	$\frac{8}{2}=4\ (MSS)_c$	
Error	$(r-1)(c-1)=6$	$S_E^2=6$	$\frac{6}{6}=1\ (MSS)_E$	$F_c = \frac{(MSS)_c}{(MSS)_E} = 4$
Total	$N-1=12-1=11$	$S_T^2=32$		

Conclusion

F_r : $F_{(3,6)}$ cal Value $>$ $F_{(3,6)}$ Tab Value

$6 > 4.76$

Reject the H_0

F_c : $F_{(2,6)}$ cal value $<$ $F_{(2,6)}$ Tab value

$4 < 19.33$

Accept H_0

2). 3 varieties of coal when analysed by 4 chemist and the ash contents in the varieties was found to be as under :

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

Carry out the analysis of Variance.

Sol:

Varities	Chemists 1	2	3	4	Total
A	8	5	5	7	$T_1 = 25$
B	7	6	4	4	$T_2 = 21$
C	3	6	5	4	$T_3 = 18$
	$P_1 = 18$	$P_2 = 17$	$P_3 = 14$	$P_4 = 15$	

Grand Total = (G)

$= T_1 + T_2 + T_3 = 64$.

Defining $H_0 = \mu_1 = \mu_2 = \mu_3$

Correction Factor $C = \dfrac{(G)^2}{N} = \dfrac{(64)^2}{12} = 341.333$

Sum of Squares of Total $= S_T^2 = \sum_i \sum_j x_{ij}^2 - C$

$= \left[(8^2+5^2+5^2+7^2) + (7^2+6^2+4^2+4^2) + (3^2+6^2+5^2+4^2)\right] - C$

$= [163 + 117 + 86] - 341.333$

$= 366 - 341.333$

$\boxed{S_T^2 = 24.667}$

Sum of Squares b/w groups

For Rows $S_r^2 = \sum_{i=1}^{3} \dfrac{T_i^2}{n_i} - C$

$= \left[\dfrac{T_1^2}{n_1} + \dfrac{T_2^2}{n_2} + \dfrac{T_3^2}{n_3}\right] - C$

$= \left[\dfrac{(25)^2}{4} + \dfrac{(21)^2}{4} + \dfrac{(18)^2}{4}\right] - 341.333$

$\boxed{S_r^2 = 6.167}$