ANOVA

To test whether the means of samples differ significantly.

#. Anova is developed by R.A. Fisher in 1920.

Assumptions in ANOVA :-

1. Normality
2. Homogenity
3. Independence of error.

Technique of ANOVA -

1. One way & 2- Two way.

One way ANOVA :-

In one way classification the data are classified according to only one criterion.

Null Hypothesis is

$$H_0: \mu_1 = \mu_2 = \mu_3 = --- = \mu_k$$

Alternative Hypothesis is

$$H_1: \mu_1 \neq \mu_2 \neq \mu_3 \neq --- \neq \mu_k.$$

One Way Anova Table :-

Source of Variation	SS (Sum of squares)	ν (Deg of freedom)	MS (Mean square)	F (Variance Ratio)
Between samples	SSC	$\nu_1 = c-1$	$MSC = \frac{SSC}{n-1}$	$F = \frac{MSC}{MSE}$
Within samples	SSE	$\nu_2 = n-c$	$MSE = \frac{SSE}{n-c}$	
Total	SST	$n-1$		

If $|F_{cal}| < |F_{Tab}| \Rightarrow H_0$ accepted, H_1 rejected

If $|F_{cal}| > |F_{tab}| \Rightarrow H_0$ rejected, H_1 accepted.

Working Method :- Let X_1, X_2, X_3 are given.

1) $T = \Sigma X_1 + \Sigma X_2 + \Sigma X_3$

2) Correction Factor $= CF = \frac{T^2}{N}$

where $N =$ total no. of entries in tables.

3) $SST =$ Total Sum of squares of variations

$= \Sigma X_1^2 + \Sigma X_2^2 + \Sigma X_3^2 - CF$

SSC = Sum of squares b/w Samples (col)

$$= \frac{(\Sigma X_1)^2}{\cancel{n}k} + \frac{(\Sigma X_2)^2}{\cancel{n}k} + \frac{(\Sigma X_3)^2}{\cancel{n}k} - CF$$

5) SSE = Sum of squares within samples

$$= SST - SSC.$$

.01 To assess the significance of possible variation in performance in a certain test b/w the convent schools of a city, a common test was given. The results are given below: Make an analysis of variance of data.

Schools

A	B	C	D
8	12	18	13
10	11	12	9
12	9	16	12
8	14	6	16
7	4	8	15.

Soln

	X_1	X_2	X_3	X_4	X_1^2	X_2^2	X_3^2	X_4^2
	8	12	18	13	64	144	324	169
	10	11	12	9	100	121	144	81
	12	9	16	12	144	81	256	144
	8	14	6	16	64	196	36	256
	7	4	8	15	49	16	64	225
Σ	45	50	60	65	421	558	824	875

$N = 20$, $n = 5$, $c = 4$

1. $T = \Sigma X_1 + \Sigma X_2 + \Sigma X_3 + \Sigma X_4 = 45 + 50 + 60 + 65$
$= 220$

2. $CF = \dfrac{T^2}{N} = \dfrac{(220)^2}{20} = 2420$

3. $SST = \Sigma X_1^2 + \Sigma X_2^2 + \Sigma X_3^2 + \Sigma X_4^2 - CF$
$= 421 + 558 + 824 + 875 - 2420$
$= 258$

4. $SSC = \dfrac{(\Sigma X_1)^2}{n} + \dfrac{(\Sigma X_2)^2}{n} + \dfrac{(\Sigma X_3)^2}{n} + \dfrac{(\Sigma X_4)^2}{n} - CF$
$= \dfrac{(45)^2}{5} + \dfrac{(50)^2}{5} + \dfrac{(60)^2}{5} + \dfrac{(65)^2}{5} - 2420$
$= 50.$

5. $SSE = SST - SSC = 258 - 50 = 208.$

6. Anova table –

Sourse of Variation	SS	df	MS	F
Between samples	$SSC = 50$	$c-1 = 3$	$MSC = \dfrac{SSC}{c-1} = 50/3 = 16.7$	$F = \dfrac{MSC}{MSE} = \dfrac{16.7}{13} = 1.287$
Within samples	$SSE = 208$	$n-c = 16$	$MSE = \dfrac{SSE}{n-c} = 208/16 = 13$	
Total	$SST = 258$	$n-1 = 19$		

$F_{cal} = 1.285$

F_{tab} for $\nu_1 = 3$, $\nu_2 = 16$. with LOS 5% is

$$F_{tab} = 3.24$$

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

H_1 : $\mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$

As $|F_{cal}| < |F_{tab}|$

$\Rightarrow$ H_0 accepted

$\Rightarrow$ H_1 rejected.

→ The difference in the mean values of the sample is not significant.

Coding of Data :—

As the final quantity is a ratio and so dimensionless so data can be coded to simplify calculations without the need for any subsequent adjustments of the results.

2 The following data represents the shelf life (in days) based on sample survey of four different health drinks:

A	B	C	D
99	102	102	97
101	97	100	101
99	102	100	100
100	102	100	101
102	100	98	99

At 0.05 los, is there evidence of a difference in the average shelf life of the four health drinks?

Soln $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ $\quad N = 20$

$H_1: \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4.$ $\quad r = 5$

$c = 4$

For simplifying calculations, take 100 as common.

The coded data are given below:-

x_1	x_1^2	x_2	x_2^2	x_3	x_3^2	x_4	x_4^2
−1	1	2	4	2	4	−3	9
+1	1	−3	9	0	0	1	1
−1	01	2	4	0	0	0	0
0	0	2	4	0	0	1	1
2	4	0	0	−2	4	−1	1
1	7	3	21	0	8	−2	12

① $T = \Sigma X_1 + \Sigma X_2 + \Sigma X_3 + \Sigma X_4 = 1 + 3 + 0 - 2 = 2$

② Correction Factor $= CF = \frac{T^2}{N} = \frac{(2)^2}{20} = 0.2.$

$$SST = \Sigma X_1^2 + \Sigma X_2^2 + \Sigma X_3^2 + \Sigma X_4^2 - CF$$

$$= 7 + 21 + 8 + 12 - 0.2$$

$$= 47.8$$

4) $$SSC = \frac{(\Sigma X_1)^2}{n} + \frac{(\Sigma X_2)^2}{n} + \frac{(\Sigma X_3)^2}{n} + \frac{(\Sigma X_4)^2}{n} - CF$$

$$= \frac{1^2}{5} + \frac{3^2}{5} + \frac{0^2}{5} + \frac{(-2)^2}{5} - 0.2$$

$$= 2.6.$$

5) $$SSE = SST - SSC$$

$$= 47.8 - 2.6$$

$$= 45.2$$

6) Anova Table :-

Source of Variation.	SS.	df	MS	F
Between samples	$SSC = 2.6$	$c-1 = 3$	$MSC = \frac{SSC}{c-1} = \frac{2.6}{3} = 0.87$	$F = \frac{MSC}{MSE} = \frac{0.87}{2.83} = 0.31$
Within samples	$SSE = 45.2$	$n-c = 16$	$MSE = \frac{SSE}{n-c} = \frac{45.2}{16} = 2.83$	
Total	$SST = 47.8$	19		

$F_{cal} = 0.31$

F_{tab} for 5% LOS for degree of freedom $\nu_1 = 3$ and $\nu_2 = 16$ is

$$F_{tab} = 3.24$$

As $|F_{cal}| < |F_{tab}|$

$\Rightarrow$ H_0 accepted

H_1 rejected.

$\Rightarrow$ There is no significant difference in the average shelf of four health drinks.

TWO WAY ANOVA

A two way ANOVA tests the effect of two independent variables on a dependent variable.

ANOVA TABLE :-

Source of Variation	Sum of squares SS	Degree of freedom df	Mean sum of squares MS	Ratio of F
Between samples. (C columns)	SSC	$(c-1)$	$MSC = \frac{SSC}{c-1}$	$\frac{MSC}{MSE}$
Between rows.	SSR	$(r-1)$	$MSR = \frac{SSR}{r-1}$	$\frac{MSR}{MSE}$
Residual or error.	SSE	$(c-1)(r-1)$	$MSE = \frac{SSE}{(r-1)(c-1)}$	
Total.	SST	$N-1$		

where.

① $T = \Sigma X_1 + \Sigma X_2 + \Sigma X_3 = \Sigma Y_1 + \Sigma Y_2 + \Sigma Y_3.$

② correction factor $= \frac{T^2}{N} = CF$

$$SSC = \frac{(\Sigma X_1)^2}{r} + \frac{(\Sigma X_2)^2}{r} + \frac{(\Sigma X_3)^2}{r} - CF$$

(sum of squares b/w columns)

(4) $$SSR = \frac{(\Sigma Y_1)^2}{c} + \frac{(\Sigma Y_2)^2}{c} + \frac{(\Sigma Y_3)^2}{c} - CF$$

(sum of squares b/w rows)

(5) SST = Total sum of squares

= sum of squares of all entries − CF

(6) SSE = sum of squares due to error

= SST − SSR − SSC.

Q.1 A tea company appoints 4 salesman A, B, C and D and observes their sales in three seasons - summer, winter and mansoon. The figures (in lakh) are given in the following table -

	Salesman			
Seasons	A	B	C	D.
Summer	36	36	21	36
Winter	28	29	31	31
Mansoon	26	28	29	29.

(i) Do the salesman significantly differ in performance?

(2) Is there significant difference between the seasons?

Soln The above data are classified to criteria salesmen and seasons.
In order to simplify the calculations we code the data by subtracting 30 from each figure. The data in the coded form are given below:—

Seasons	Salesmen A X_1	B X_2	C X_3	D X_4	Total
Summer Y_1	6	6	-9	6	9
Winter Y_2	-2	-1	1	1	-1
Monsoon Y_3	-4	-2	-1	-1	-8
Total	0	3	-9	6	0

$N = 12$

$r = 3$

$c = 4$

Ho for salesmen variance is "There is no significant difference"
H_{01} for salesmen variance is "There is a significant difference".

$T = \Sigma X_1 + \Sigma X_2 + \Sigma X_3 + \Sigma X_4$

$= 0 + 3 - 9 + 6 = 0$

(2) correction factor $= \frac{T^2}{N} = \frac{0}{12} = 0$

(3) $SSC = \frac{(\Sigma X_1)^2}{r} + \frac{(\Sigma X_2)^2}{r} + \frac{(\Sigma X_3)^2}{r} + \frac{(\Sigma X_4)^2}{r} - CF$

$= \frac{0^2}{3} + \frac{3^2}{3} + \frac{(-9)^2}{3} + \frac{6^2}{3} - 0$

$= 42$

(4) $SSR = \frac{(\Sigma Y_1)^2}{c} + \frac{(\Sigma Y_2)^2}{c} + \frac{(\Sigma Y_3)^2}{c} - CF$

$= \frac{9^2}{4} + \frac{(7)^2}{4} + \frac{(-8)^2}{4} - 0$

$= 36.5$

(5) $SST = 6^2 + 6^2 + (-9)^2 + 6^2 + (-2)^2 + (7)^2 + 1^2$
$+ 1^2 + (-4)^2 + (-2)^2 + (7)^2 + (7)^2 - CF$

$= 218.$

(6) $SSE = SST - SSC - SSR$

$= 218 - 42 - 36.5$

$= 139.5$

) Two way ANOVA table -

Source of Variation	SS	df	MS	F
Between samples (Columns) (salesmen)	SSC = 42	$c-1 = 4-1=3$	$MSC = \frac{SSC}{c-1} = \frac{42}{3} = 14$	$F = \frac{MSC}{MSE} = \frac{14}{23.25} = 0.6021$
Between rows. (seasons)	SSR = 36.5	$r-1 = 2$	$MSR = \frac{SSR}{r-1} = \frac{36.5}{2} = 18.25$	$F = \frac{MSR}{MSE} = \frac{18.25}{23.25} = 0.7849$
Residual or Error	SSE = 139.5	$(c-1)(r-1) = 3\times2 = 6.$	$MSE = \frac{SSE}{(r-1)(c-1)} = \frac{139.5}{6} = 23.25$	
Total	SST = 218	$N-1 = 12-1 = 11$		

Now F_{cal} for camparison of salesmen variance

$$|F_{cal}| = 0.6021$$

At 5% LOS, $\nu_1 = 3$, $\nu_2 = 6$. $F_{tab} = 4.76$.

As $|F_{cal}| < |F_{tab}|$

$\Rightarrow$ Ho accepted and H1 rejected.

We conclude that the sales of different salesmen do not differ significantly.

Now F_{cal} for comparison of season variance

$$F_{cal} = 0.7849$$

F_{tab} for LOS 5% and $\nu_1 = 2$ and $\nu_2 = 6$.

$$F_{tab} = 5.14$$

As $F_{cal} < F_{tab}$

$\Rightarrow$ Ho accepted and H1 rejected.

$\Rightarrow$ There is no significant difference in seasons as far as the sales is concernd