Analysis of variance (ANOVA) :→

→ Why we need ANOVA?

Ans ANOVA is used to Compare the statistical significance b/w the means of three or more Independent groups or samples. Because, when no. of samples are more than two then it is very difficult and time Consuming to apply t-test.

Basic Terms:→

(1) Experimental unit:→ The object on which a measurement is taken.

(2) Factor:→ A factor is an Independent variable whose values are Controlled and varied by the Experimenter.

(3) Level:→ Level is an Intensity setting of a factor.

(4) Treatment:→ A specific Combination of factor ~~and~~ levels.

Ex A group of people is divided into two groups. one group has given the aptitude test after having breakfast while Another group has given the same test without eating breakfast. What are the factors, levels and treatments in this experiment?

Solⁿ:→ Experiment unit - People

factors — meal

Levels < Breakfast / No Breakfast

Treatments

Meal	
Breakfast	No Breakfast
-	-
-	-

Ex:- In Previous Example, If group of people is like 20 men and 20 women for the Experiment. Then what are the factors, Levels, and treatments?

Sol:-

Factors — 2
- Gender
- Meals

Level
- Gender
  - Men
  - Women
- Meal
  - Breakfast
  - No Breakfast.

Factors		Meal: Without Breakfast	With Breakfast
	~~Men~~		
Gender	Men	—	—
	Women	—	—

→ Four Treatments

ANOVA — No. of Samples > 2.
— is of two types.

(1) Completely Randomized design (CRD) :- one way Classification | factors | ~~tests~~ — also Called one way ANOVA

(2) Randomized Block design (RBD) :→ Two way Classification | factors — also Called Two way ANOVA.

→ CRD | One way ANOVA | One Way Factors:→

↳ Involves only one factors.

↳ one factor may have K levels | ~~groups~~ | Samples.

Group/factors	observations				Mean
1	x_{11}	x_{12}	---	x_{1n_1}	$\bar{x}_1$
2	x_{21}	x_{22}	---	x_{2n_2}	$\bar{x}_2$
⋮	|	|		|	|
K	x_{K1}	x_{K2}	---	x_{Kn_K}	$\bar{x}_K$
				Overallmean	$\bar{x}$

Ex

Teaching methods	Observations				
online	12	34	35	43	
offline	21	23	25		
Recorded	23	32	34	39	45

Three Kinds of variations in ANOVA:→

(1) Between groups :- Variation from one group to another

$$\sum n_i(\bar{x}_i - \bar{x})^2$$

(2) Within groups :- Variations among the observations of each specific group.

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$$

(3) Total :- Variations among all the observations. (Sum of between groups & within groups).

$$\sum \sum (x_{ij} - \bar{x})^2$$

Date ______
Page ______

F ratio | ANOVA :→

F ratio | ANOVA is defined as

$$F = \frac{\text{Mean Sum of Squares b/w groups}}{\text{Mean Sum of Squares Within groups}}.$$

Assumptions :- (1) Samples are randomly drawn from normal population

(2) ~~All~~ $\sigma_1^2 = \sigma_2^2 = \_\_\_ = \sigma_K^2$

(3) Total variance should be equal to Sum of Square b/w groups and within groups.

ANOVA Testing

$H_0: \mu_1 = \mu_2 = \_\_\_ = \mu_K$

i.e. there is no significant difference b/w the mean of all groups.

H_1: There is atleast one group means different from others.

One way ANOVA Summary table

Sources of Variation	d.f. (1)	S.S. (2)	MSS. = SS/d.f (2)/(1)	F-ratio
B/W groups	–	–		
within groups	–	–		
Total	–	–		

df – degree of freedom

SS – sum of Squares

MSS – Mean Sum of Squares.

Date
Page

Two method to Calculate Sum of Squares (SS):-

(1) $SS \text{ b/w groups} = \sum n_i(\bar{x}_i - \bar{x})^2$

$SS \text{ Within groups} = \sum\sum (x_{ij} - \bar{x}_i)^2$

$SS \text{ total} = \sum\sum (x_{ij} - \bar{x})^2$

(2)

Groups	Observations				Total
1	x_{11}	x_{12}	—	~~x_{1n}~~	T_1
2	x_{21}	x_{22}			T_2
⋮	⋮	⋮		⋮	⋮
K	x_{K1}	x_{K2}	—		T_K

Grand Total = G

(1) $C = \frac{G^2}{N}$ — N — total no. of elements.

(2) $SS \text{ total} = \sum\sum (x_{ij}^2 - C)$

(3) $SS \text{ b/w group} = \sum_{i=1}^{K} \left(\frac{T_i^2}{n_i} - C\right)$

(4) SS Within group = SS Total − SS b/w group.

Que In an experiment to determine the effect of nutrition on the attention time of elementary school students, a group of 15 students were randomly assigned to each of three meal plans:- no breakfast, light breakfast and full breakfast. Their attention minutes were recorded during a morning reading period and shown as

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

Construct the ANOVA table for this experiment.

Date ______
Page ______

Sol^n

Meal						Total
No Breakfast	8	7	9	13	10	47
Light ~~hto~~ Breakfast	14	16	12	17	11	70
Full Breakfast	10	12	16	15	12	65
						$G = 182$

$$C = \frac{G^2}{N} = \frac{(182)^2}{15} = 2208.26$$

$$SS\ total = \sum\sum x_{ij}^2 - C$$
$$= (8^2 + 7^2 + \_\_ + 12^2) - 2208.26$$
$$= 129.73$$

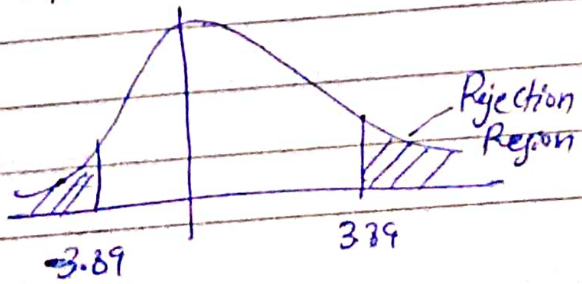
$$SS\ b/w\ groups = \sum_{i=1}^{3}\left(\frac{T_i^2}{n_i}\right) - C$$
$$= \frac{47^2}{5} + \frac{70^2}{5} + \frac{65^2}{5} - 2208.26$$
$$= 58.53.$$

So ANOVA table is

	D.f.	SS	MSS = SS/df	F-ratio.
B/w groups	2 = (3-1)	58.53	58.53/2 = 29.26	29.26/5.93 = $\boxed{4.93}$
Within groups	12	71.20	71.20/12 = 5.93	
Total	14 = (15-1)	129.73		

Table value is $F_{(2,14)}(\alpha) = F_{(2,14)}(0.05) = 3.89$

$\because$ F value > 3.89
$\Rightarrow$ Reject H_0.



Que A truck company wants to test the avg life of each of the 4 brands of tyres. The Company uses all brands on randomly selected trucks. The records showing the lives (thousands of miles) are as given:-

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

Test the hypothesis that the avg life for each brand of tyres is the same at 1% level of significance.

Sol^n Subtract 20 (or any no.) from each observation.

						Total
Brand 1	0	3	−2	−3		−2
Brand 2	−1	−5	−3	0	−4	−13
Brand 3	1	−1	0	−3	−4	−7
Brand 4	−5	−3	−4	−2		−14
						$G = -36$

$$C = \frac{G^2}{N} = \frac{(-36)^2}{18} = 72$$

$$SS\ total = \sum\sum (x_{ij})^2 - c$$
$$= (0^2 + 3^2 + \_\_ + (-2)^2) - 72$$
$$= 82$$

$$SS\ b/w\ groups = \sum_{i=1}^{4} \left(\frac{T_i^2}{n_i}\right) - c$$
$$= \left(\frac{4}{4} + \frac{169}{5} + \frac{49}{5} + \frac{196}{4}\right) - 72$$
$$= 21.6$$

$$SS\ within\ groups = 82 - 21.6 = 60.4$$

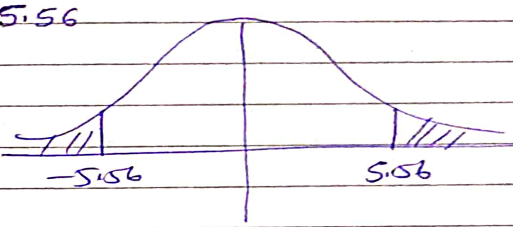
ANOVA table

	d.f.	S.S.	MSS	F-ratio
B/w groups	3 = (4-1)	21.6	7.2	$\boxed{1.67}$
Within groups	14	60.4	4.31	
Total	18-1=17	82		

table value $F_{(3,14)}(0.01) = 5.56$

$\because F_{value} < F_{(3,14)}(0.01)$

$\Rightarrow$ Do not reject H_0.



Que

Brands	(Detergent Brands)					Total
A	77	81	61	76	69	364
B	74	66	58			198
C	73	78	57	69	63	340
D	76	85	77	64		302
					G =	1204

Test the hyp that there is no difference b/w the four brands as regards mean whiteness readings after washing.

Sol^n

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ i.e. $\mu_i = \mu \ \forall i$

$H_1: \mu_i \neq \mu$ for some i

$\alpha = 0.05$

$$C = \frac{G^2}{N} = \frac{(1204)^2}{17} =$$

SS total = 1090.47

SS b/w gps = 216.67

$\Rightarrow$ SS within gps = 873.8

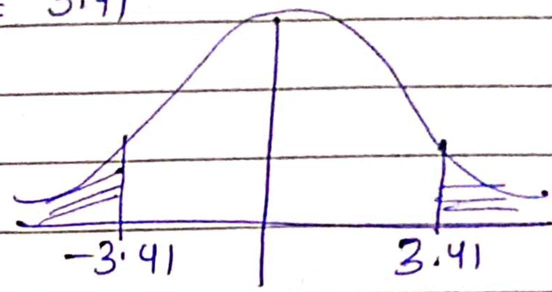
ANOVA table is

	d.f	SS	MSS	F-ratio
B/w groups	4-1=3	216.67	72.22	$\boxed{1.07}$
Within groups	13	873.8	67.21	
Total	17-1=16	1090.47		

$F_{3,13}(0.05) = 3.41$

$1.07 < 3.41$

$\Rightarrow$ Don't Reject H_0.



HW
Que The following data gives the lifetimes in hours of 3 types of batteries.

	I	50.1	49.9	49.8	49.7	50
Type	II	51.0	50.8	50.9	50.9	50.6
	III	49.5	50.1	50.2	49.8	49.3

Analyse this data for a difference b/w mean lifetimes (use 5% level of significance)

HW
Que

Car	27	45	33	31
Bicycle	34	38	43	42
Bus	26	41	35	46

Carry out an ANOVA and test at 5% LOS whether there are differences in the mean journey times for the 3 methods of transport.

Date ______
Page ______

Two way ANOVA / RBD :→ Two factors are Considered.

Eg Factors :- Teaching methods & Intelligence

Levels :- Teaching Methods :- online, offline and Video lecture.
Intelligence :- High, Average and low.

Teaching Methods

	Online	Offline	Video lecture
High	—	—	—
Avg	—	—	—
Low	—	—	—

↳ When Two-way ANOVA Can be used?

Ans When no. of observations in each row are equal and in each Column are equal then only Two-way ANOVA Can be used.

Eg.

A	23	21	24	25	
B	28	29	33		
C	34	45	50	36	64

1-way ANOVA.

Eg

	P	Q	R	S
A	23	21	45	42
B	25	24	36	23
C	28	27	33	40

2-way ANOVA.

↳ As there are two factors, we need 2 hypothesis.

Hypothesis for factor 1 :→

H_0: There is no significant diff b/w the means of the row factors. i.e $\mu_i = \mu$

H_1: there is significant difference b/w the means of row $\mu_i \neq \mu$

Hyp. for factor 2 :→

H_0: There is no significant diff b/w the means of the Column factor. i.e. $\mu_j = \mu$

H_1: There is significant diff b/w the means of the Column factors. i.e. $\mu_j \neq \mu$.

Source table for Two-way ANOVA

Source of Variation	d.f.	SS	MSS = SS/df	F-ratio
B/w rows	$r-1$	SSR	MSSR	MSSR/MSSE
B/w Columns	$C-1$	SSC	MSSC	MSSC/MSSE
Within ~~groups~~ errors	$(r-1)(C-1)$	SSE	MSSE	
Total	$rC-1$	SST		

r – no. of rows ; C – no. of colums

df – deg. of freedom; SS – sum of squares.

MSS – mean sum of squares.

Eg.

Rows \ Columns	P	Q	R	S	
A					T_A
B					T_B
C					T_C
	T_P	T_Q	T_R	T_S	$G = \sum T_i$

$$C = \frac{G^2}{N}$$

$$SSR = \sum_i \frac{T_i^2}{n_i} - C$$

$$SSC = \sum_j \frac{T_j^2}{n_j} - C$$

$$SST_o = \sum\sum x_{ij}^2 - C$$

$$SSE = SST - SSR - SSC.$$

Date ____
Page ____

Ex:→

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

A farmer applies 3 type of fertilizers on 4 plots. The figures on yield per acre are given in table. Test if there is any significant difference in fertility of the plots.

Soln→ $H_0: \mu_1 = \mu_2 = \mu_3$ i.e. three fertilizers are equally Effective
$H_0: \mu_A = \mu_B = \mu_C = \mu_D$ i.e. four plots are equally fertile

	Yield A	B	C	D	
N	6	4	8	6	24
P	7	6	6	9	28
Ph	8	5	10	9	32
	21	15	24	24	84

$$C = \frac{G^2}{N} = \frac{(84)^2}{12} = 588$$

$$SSR = \sum \frac{T_i^2}{n_i} - C = \frac{(24)^2}{4} + \frac{(28)^2}{4} + \frac{(32)^2}{4} - 588$$
$$= 8$$

$$SSC = \frac{(21)^2}{3} + \frac{(15)^2}{3} + \frac{(24)^2}{3} + \frac{(24)^2}{3} - 588 = 18$$

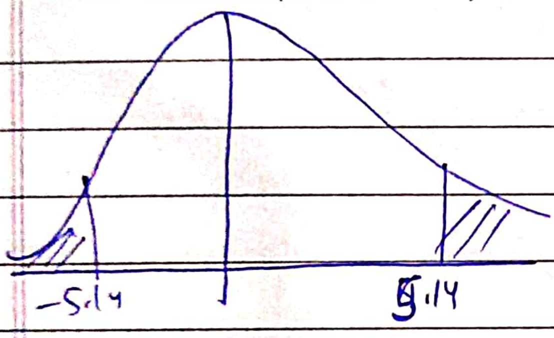
$$SST = \sum\sum x_{ij}^2 - C$$
$$= 6^2 + 4^2 + \_\_ + 9^2 - 588 = 36$$

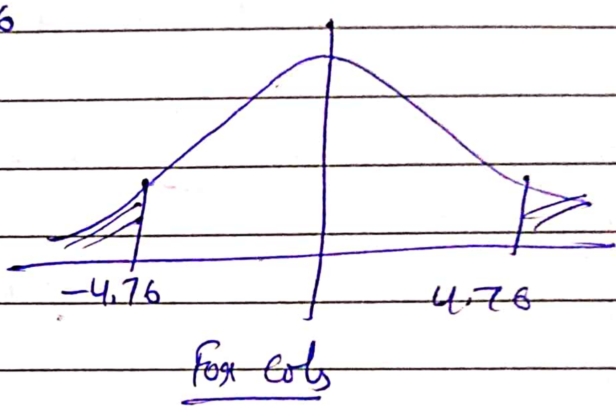
Two way ANOVA table

	df	SS	MSS	F-ratio
B/w Rows	3-1=2	8	8/2 = 4	4/1.67 = 2.39
B/w Cols	4-1=3	18	18/3 = 6	6/1.67 = 3.593
Within error	6	10	10/6 = 1.67	
Total	12-1=11	36		

$F(2,6)\ (0.05) = 5.14$

$F(3,6)\ (0.05) = 4.76$





We fail to reject both the null hypothesis.