ANOVA

- The ANalysis Of VAriance (ANOVA) is used to determine whether there is any statistical significant differences between the means of three or more independent (unrelated) groups.
- Developed by R.A Fisher in 1920.
- Also known as f-test which is based on f distribution.
- Anova test determines whether all group are taken from common population or not. If the mean is similar then they are taken from same population and if the mean is very different, then they are taken from different population respectively.
- Anova is a ratio between "Mean Sum of Square between (MSS_B) and "Mean Sum of Square Within (MSS_W)".

$$F = \frac{MSS_B}{MSS_W} = \frac{\text{Between Variance}}{\text{Within Variance}} = \frac{\text{Between 2 groups or more}}{\text{Within each group}}$$

- The variation among the observation of each specific group is called its internal variation and the totality of internal variation is called variability with group.
- The totality of variations from one group to another, i.e, variation dues to group is called variability between Groups.
- Anova test the null hypothesis is —

H_0: There is no significant difference between means of all groups. All groups are same.

$H_0 = \mu_1 = \mu_2 = \mu_3 = \dots = \mu_K$ where μ = group mean, K = no of groups.

Alternate hypothesis, H_A: There are atleast two groups means that are statistically significant different from each other. $H_0: \mu_1 \neq \mu_2 \neq \dots \neq \mu_K$.

Types of ANOVA — i) One way ANOVA ii) Two way ANOVA. iii) N way ANOVA.

Assumptions — i) Random Selection: Samples should be randomly selected.

ii) Normal distribution: Independent variable should be normally distributed.

iii) Homogenity of Variance: All sub population have the same variance (homsceclasti's)

iv) Additivity of variance: Total variance should be equal to sum of between variance & within variance.

Summary table for One way Anova

K - Number of groups, N → Total number of members present in one group.
SSB → Sum of square between, SSW → Sum of square Within, TSS → Total sum of square
MSS → Mean Sum of square, B → Between, W → Within

Source of Variation	df	SS	MSS	F Calculated	F tabulated at 5% and 1% level
Between (factors)	K-1	SSB	MSSB = SSB/(K-1)	= MSSB / MSSW	
Within (factors)	N-K	SSW	MSSW = SSW/(N-K)		
Total	N-1	TSS			

Suppose $F(2, 14) = 5.758$, $p = 0.05$

↑ degree of freedom; ↑ f ratio/value; ↑ p-value.

2 → df for variance between groups.
14 → df for variance within groups.

Two WAY ANOVA - To check Interaction effect of 2 or more Independent categorical variable to one dependent continuous variable

What does "One-Way" or "Two-Way" anova Mean?

→ A Two way ANOVA is an extension of One Way ANOVA.

→ One way or two way refers to numbers of Independent Variables (IVs) in ANNOVA.

→ One-way has one independent variables (with 2 or more Groups / levels)

Eg - IV (Teaching Variable) → Levels (1. Lecture Method, 2. Lecture-cum-Demonstration 3. Video based learning)

→ Two-way has two independent variables (with 2 or more Group/levels).

Eg - IVs (Teaching Method & Intelligence) → levels (1. Lecture Method 2. Lecture-Demonstration 3. Video based learning) (1. High, 2. Average, 3. Low)

→ With a One-way, we have one independent variable affecting a dependent variable. It means by which means student acheive highest marks in exam.

Independent variable → way of teaching, Dependent Variable → Marks scored in exam.

Lecture Method	Lecture cum demonstration	Video base learning
84	80	100
61	92	91
73	91	68
⋮	⋮	⋮

(Main effect)

→ Two-way anova, there are two independent variables.

(Interaction effect)

	Lecture Method	Lecture cum demonstration	Video based learning
High	95, 51, 78, 36	89, 59, 23.	74, 96.
Average	76, 25, 47, 88, 14	59, 44, 75, 69.	52, 71, 63.
Low	35, 74, 19	41, 77, 45, 27	61, 75

H_0: There is no significant Independent effect of IV_A on DV. (Main effect) Teaching effect → Exam Score. (Independent var) (Dependent var)

H_0: There is no significant independent effect of IV_B on DV. (Main effect) Intelligence → Exam score.

H_0: There is no significant interaction effect of IV_A & IV_B on DV. (Interaction effect) Teaching method + Intelligence → Exam score.

Summary table for Two-way ANNOVA

Source of variation	df	SS	MSS	F-calculated
IV A.	Column - 1 (C-1)	SSB_A	$MSSB_A$	F1
IV B.	Rows - 1 (R-1)	SSB_B	$MSSB_B$	F2
Interaction A × B.	(C-1)(R-1)	SSB_{int}	$MSSB_{int}$	F3.
Cells	No of cell - 1	SS_{cell}		
Within	N - No of cell	SSW	MSS W	
Total	N-1	SST		

From F1, F2 and F3 we can conclude which null hypothesis to accept.

(Teaching, Intelligence)

Assumptions - i) Dependent Variable should be continuous (Marks) ii) Independent variable should be categorical. Independent groups.

iii) Sample Independence - Each sample has been drawn independent of other sample

iv) Variance equality - Variance of each group should be same. v) Normally distributed population