ANOVA

- The ANalysis Of VAriance (ANOVA) is used to determine whether there is any statistical significant difference 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 —

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 (homoscedastic)

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.