

ANOVA (Analysis of Variance)

Theory

ANOVA is a statistical technique used to test whether the means of three or more populations are equal. It works by comparing the variation between groups with the variation within groups. If the between-group variation is sufficiently larger than the within-group variation, we conclude that at least one group mean is different.

Assumptions of ANOVA

1. The sampling distribution of the mean is normal.
2. There are no extreme outliers in the data.
3. Population variances are equal (homogeneity of variance).
4. Observations are independent and randomly selected.

Hypotheses

Null Hypothesis (H_0): $\mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$

Alternative Hypothesis (H_1): At least one population mean is different.

Worked Example (One-Way ANOVA – Fully Elaborated)

Marks obtained by students using three different teaching methods are given below:

Method A: 50, 55, 60

Method B: 60, 65, 70

Method C: 70, 75, 80

Step 1: Calculate Group Means

Mean of Method A = $(50 + 55 + 60) / 3 = 55$

Mean of Method B = $(60 + 65 + 70) / 3 = 65$

Mean of Method C = $(70 + 75 + 80) / 3 = 75$

Step 2: Calculate Grand Mean

Grand Mean = (Sum of all observations) / Total number of observations

Grand Mean = $(50 + 55 + 60 + 60 + 65 + 70 + 70 + 75 + 80) / 9 = 585 / 9 = 65$

Step 3: Between Group Sum of Squares (SSB)

SSB measures variation of group means around the grand mean.

$SSB = n \times [(\text{MeanA} - \text{GrandMean})^2 + (\text{MeanB} - \text{GrandMean})^2 + (\text{MeanC} - \text{GrandMean})^2]$

$SSB = 3 \times [(55 - 65)^2 + (65 - 65)^2 + (75 - 65)^2]$

$SSB = 3 \times [100 + 0 + 100] = 3 \times 200 = 600$

Step 4: Within Group Sum of Squares (SSW)

SSW measures natural variation within each group.

Method A: $(50 - 55)^2 + (55 - 55)^2 + (60 - 55)^2 = 25 + 0 + 25 = 50$

Method B: $(60 - 65)^2 + (65 - 65)^2 + (70 - 65)^2 = 25 + 0 + 25 = 50$

Method C: $(70 - 75)^2 + (75 - 75)^2 + (80 - 75)^2 = 25 + 0 + 25 = 50$

$SSW = 50 + 50 + 50 = 150$

Step 5: Degrees of Freedom

Number of groups (k) = 3

Total observations (N) = 9

df Between = $k - 1 = 2$

df Within = $N - k = 6$

Step 6: Mean Squares

$$MSB = SSB / df \text{ Between} = 600 / 2 = 300$$

$$MSW = SSW / df \text{ Within} = 150 / 6 = 25$$

Step 7: F Statistic

$$F = MSB / MSW = 300 / 25 = 12$$

Step 8: Conclusion

Since the calculated F value is greater than the critical F value at 5% level of significance, we reject the null hypothesis. Hence, at least one teaching method has a significantly different mean.