Analysis of Variance

ANOVA, or Analysis of Variance, is a statistical method used to compare the means of three or more groups to determine if there is a statistically significant difference among them. Instead of comparing groups one by one (as in a t-test), ANOVA evaluates all groups simultaneously, which reduces the likelihood of making a Type I error (false positive).

Assumption

- **Independence**: The observations in each group must be independent of each other (e.g., one person's result should not affect another's result).
- **Normality**: The data within each group should follow a normal distribution. Small deviations are acceptable if the sample size is large (Central Limit Theorem applies).
- **Homogeneity of Variances**: The variance (spread) of data in all groups should be roughly equal. This is called homoscedasticity. i.e., $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \ldots\ldots = \sigma_k^2 = \sigma^2$ for k populations
- **Dependent Variable is Continuous**: The data being analyzed must be measured on an interval or ratio scale (e.g., height, weight, or test scores).

ANOVA is used when:

- You want to compare means of multiple groups.
- Your data is continuous (e.g., height, weight, test scores).
- You want to determine if variations in the dependent variable are due to differences in the independent variable.

Types of ANOVA

1. **One-Way ANOVA**: Compares the means of groups based on one factor (e.g., testing the effect of different fertilizers on plant growth).
2. **Two-Way ANOVA**: Compares means based on two factors (e.g., testing the effect of fertilizers and watering frequency on plant growth).
3. **Repeated Measures ANOVA**: Used when the same subjects are tested under different conditions (e.g., measuring reaction times before, during, and after consuming caffeine).

One Way ANOVA

One-Way ANOVA (Analysis of Variance) is a statistical method used to determine if there are significant differences between the means of three or more independent groups based on one factor or variable. It helps test whether variations between group means are statistically significant or simply due to random chance.

For example, you might use One-Way ANOVA to compare the effectiveness of three different fertilizers on crop yield.

Steps

1. Hypothesis Set up

H_0: $\mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$ i.e., there is no significant difference of k population means.

H_1: $\mu_1 \neq \mu_2 \neq \mu_3 \neq \dots \neq \mu_k$ i.e., there is significant difference of k population means or at least one of the means of the population is different from the others.

2. Test Statistics

F statistic = $\frac{\text{variance between samples}}{\text{variance within samples}} = \frac{\text{MSB}}{\text{MSE}}$

One Way ANOVA table

Source of variation	Sum of squares	D.f.	Mean sum of square	F-ratio
Between samples(treatments)	SSB	(k-1)	MSB= $\frac{\text{SSB}}{k-1}$	F= $\frac{\text{MSB}}{\text{MSE}}$
Within samples (error)	SSE	(n-k)	MSE= $\frac{\text{SSE}}{n-k}$	
Total	SST	(n-1)		

We calculate the value of SSB, SSE & SST using following formula,

a) Compute the grand total (T) i.e. sum of all observations in all k samples. That is,
   T= $\sum X_1 + \sum X_2 + \cdots + \sum X_k$
b) Compute the correction factor (C.F.) by dividing square of grand total by total number of observations. i.e. C.F. = $\frac{T^2}{n}$
c) Compute the sum of squares of all individual observations of k sample subgroups and subtract the C.F. from it. This gives the total sum of squares (SST). That is,
   SST = $\sum X_1^2 + \sum X_2^2 + \cdots + \sum X_k^2 - \text{C.F.}$

d) Compute the square of the sum of the values of each sample subgroup and divide each such squared value by the corresponding number of values in sample subgroups and then compute the sum of all the resulting values and subtract the C.F. from this sum values. It gives the sum of squares of deviation between the sample subgroups. That is,

$$SSB = \left[\frac{(\sum X_1)^2}{n_1} + \frac{(\sum X_2)^2}{n_2} + \cdots + \frac{(\sum X_k)^2}{n_k}\right] - C.F$$

e) Then the SSE can be computed by subtracting SSB from SST. That is,
SSE = SST – SSB
(Since, SST= SSB+SSE)

3. Level of Significance: α

4. Critical Value: We have to determine the tabulated value of F at α% level of significance for (v_1, v_2) i.e. (k-1, n-k) degree of freedom.

5. Decision:

- If $F_{cal} \leq F_{tab}$, we do not reject H_0
- If $F_{cal} > F_{tab}$, we reject H_0