Analysis of Variance (ANOVA)

Dr. Kalyan N

October 14, 2024

What is ANOVA?

- **Analysis of Variance (ANOVA)** is a statistical technique used to compare the means of three or more groups.
- It tests whether there are any statistically significant differences between the means of independent (unrelated) groups.
- ANOVA works by comparing the variation within groups to the variation between groups.
- It is an extension of the two-sample t-test, which compares the means of two groups, to more than two groups.

The Basic Principle of ANOVA

- ANOVA decomposes the total variability in the data into two parts:
  - **Between-group variability:** Variation due to the differences between the group means.
  - **Within-group variability:** Variation due to differences within individual groups.
- The ratio of these two sources of variability is used to determine if the group means are significantly different from each other.
- Test statistic for ANOVA:

$$F = \frac{\text{Between-group variance}}{\text{Within-group variance}}$$

- If the F-value is large, it suggests that the group means are not all equal, and the null hypothesis can be rejected.

Hypotheses in ANOVA

- **Null hypothesis ($\mathbf{H}_0$):** The means of all groups are equal.

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$$

- **Alternative hypothesis ($\mathbf{H}_A$):** At least one group mean is different.

$$H_A : \text{At least one } \mu_i \neq \mu_j$$

What is One-Way ANOVA?

- **One-Way ANOVA** is used when we are testing for differences in one factor (independent variable) across multiple groups.
- Example: Testing whether the mean test scores of students in three different classrooms are significantly different.
- It compares the means of three or more unrelated groups based on one factor.
- Assumptions of One-Way ANOVA:
  - The observations are independent.
  - The groups are normally distributed.
  - The groups have homogeneity of variances (equal variances).

ANOVA Formula

The total variation in the data can be divided into two components:

- **Total Sum of Squares (SST):**

$$SST = \sum_{i=1}^{N} (x_i - \bar{x})^2$$

  where N is the total number of observations, x_i is each individual observation, and $\bar{x}$ is the grand mean.

- **Between-Group Sum of Squares (SSB):**

$$SSB = \sum_{j=1}^{k} n_j (\bar{x}_j - \bar{x})^2$$

  where n_j is the number of observations in group j, $\bar{x}_j$ is the mean of group j, and $\bar{x}$ is the grand mean.

- **Within-Group Sum of Squares (SSW):**

$$SSW = \sum^{k} \sum^{n_j} (x_{ij} - \bar{x}_j)^2$$

ANOVA Test Statistic

The F-statistic in ANOVA is calculated as follows:

$$F = \frac{\text{Between-group variance}}{\text{Within-group variance}} = \frac{SSB/(k-1)}{SSW/(N-k)}$$

where:

- k is the number of groups.
- N is the total number of observations.
- SSB is the between-group sum of squares.
- SSW is the within-group sum of squares.

Compare the calculated F-statistic with the critical value from the F-distribution table to make the decision.

Example 1: Comparing Exam Scores Across Three Classes

Three classes took the same exam, and we want to know if the mean exam scores are different across the three classes.
The data is as follows:

Class	Scores
1	$85, 88, 90, 92$
2	$78, 80, 85, 87$
3	$82, 84, 88, 90$

Test if there is a significant difference in the mean scores at $\alpha = 0.05$.

Solution: Step-by-Step Calculation

Step 1: Calculate Group Means and Grand Mean

$$\bar{x}_1 = \frac{85 + 88 + 90 + 92}{4} = 88.75$$

$$\bar{x}_2 = \frac{78 + 80 + 85 + 87}{4} = 82.5$$

$$\bar{x}_3 = \frac{82 + 84 + 88 + 90}{4} = 86$$

The grand mean is:

$$\bar{x} = \frac{88.75 + 82.5 + 86}{3} = 85.75$$

Solution: Step-by-Step Calculation (Cont.)

Step 2: Calculate SSB (Between-group sum of squares)

$$SSB = 4 \times \left((88.75 - 85.75)^2 + (82.5 - 85.75)^2 + (86 - 85.75)^2\right)$$

$$SSB = 4 \times (9 + 10.5625 + 0.0625) = 4 \times 19.625 = 78.5$$

Step 3: Calculate SSW (Within-group sum of squares)

$$SSW = (85 - 88.75)^2 + (88 - 88.75)^2 + \cdots + (90 - 86)^2$$

After calculations:

$$SSW = 39.5$$

Solution: Step-by-Step Calculation (Final Steps)

Step 4: Calculate the F-Statistic

$$F = \frac{SSB/(k-1)}{SSW/(N-k)} = \frac{78.5/2}{39.5/9} = \frac{39.25}{4.39} = 8.94$$

Step 5: Decision

- From the F-distribution table with $df_1 = 2$ and $df_2 = 9$, the critical value at $\alpha = 0.05$ is 4.26.
- Since $F = 8.94 > 4.26$, we reject the null hypothesis.
- Conclusion: There is a significant difference in the mean exam scores between the three classes.

Example 2: Comparing Crop Yields

A farmer wants to test whether different fertilizers affect the yield of crops. Three different fertilizers were applied to 3 plots each, and the yields (in kg) were recorded as follows:

Fertilizer	**Plot 1**	**Plot 2**	**Plot 3**
A	50	55	52
B	48	53	50
C	54	56	58

Test if there is a significant difference in the crop yields due to different fertilizers at $\alpha = 0.05$.