

1 Main Idea

ANOVA stands for analysis of variance. It is used to compare the sample means across different treatment levels.

2 Basic Notations

$\bar{Y}_{..}$ is the total average across all treatment levels. We know that there is going to be variability within each group. This variability within our groups will be represented by an error term ϵ_{ij} . Across all notations, i represents the i^{th} group or treatment level and j represents the j^{th} observation within each group.

n is the number of samples within each group

N is the number of samples across all groups (total number of observations)

Note: The equations and formulas are simpler if sample sizes are the same across groups.

$$y_{..} = \sum_{i=1}^a \sum_{j=1}^n y_{ij}$$

$$\bar{y}_{..} = \frac{y_{..}}{N}$$

$$\bar{y}_i = \frac{\sum_{j=1}^n y_{ij}}{n}$$

This last equation is the important one. The goal of ANOVA is to see if these \bar{y}_i , for $i = 1, 2, \dots, a$ where a is the number of treatment levels, are the same or different across all treatment levels.

This leads us to our hypothesis test.

3 Hypothesis Testing

Since our goal is to see if different treatments have an impact on the mean, our hypothesis test is:

$$H_0 : \bar{y}_1 = \bar{y}_2 = \dots = \bar{y}_a.$$
$$H_a : \bar{y}_i \neq \bar{y}_j.$$