B.Sc (Statistics) Sem. IV CC 409 (Linear Model and Regression Analysis)

Topic- Analysis of Variance

Dr. Saurabh
Asstt. Professor
Dept. of Statistics
(email- shaurabh.bhu@gmail.com

ANOVA

ANOVA – Analysis Of Variance

Analysis of variance (**ANOVA**) is a collection of statistical models used to analyze the differences among group means.

- The Student's *t*-test cannot be used for comparison of three or more groups.
- The purpose of ANOVA is to test if there is any significant difference between the means of two or more groups.
- The analysis of variance is the systematic algebraic procedure of decomposing the overall variation in the responses observed in an experiment into variation.
- Two variances – (a) between-group variability and (b) within-group variability that is variation existing between the samples and variations existing within the sample.
- The within-**group** variability (error variance) is the variation that cannot be accounted for in the study design.
- The between-group (or effect variance) is the result of treatment

ASSUMPTIONS OF ANOVA

- The population in which samples are drawn should be normal
- The sample observations are independent of each other
- The samples are selected at random
- The samples are drawn from population having equal variance
- The sample size should not differ widely
- The various effects(treatment and errors) are additive in nature
- The experimental error are normally and independently distributed with mean Zero

ONE WAY ANOVA

- It compares three or more unmatched groups when data are categorized in one way.

Total sum of Square(TSS) = treatment sum of square(SST) + error sum of square(SSE)

Example.

- 1. Compare control group with three different doses of aspirin in rats
- 2. Effect of supplementation of vit C in each subject before, during and after the treatment.

One – Way ANOVA

- The mathematical model that describes the relationship between the response and treatment for the one way ANOVA is given by

$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$, i = 1,2....k, j = 1,2,..... n_i

where, Y_{ij} : j^{th} observation on i^{th} treatment,

μ : common effect for the whole experiment,

α_i : i^{th} treatment effect,

ε_{ij} : random error.

Hypothesis of one – way ANOVA

H_0: The means of all the groups are equal.

Vs

H_1: Not all of the group means are the same.

- A simplified formula for the *F* statistic is

$$F = \frac{MST}{MSE}$$

where *MST* is the mean squares between the groups and *MSE* is the mean squares within groups

One way ANOVA table

Source	d.f	Sum of Square	Mean Square	F
Between treatment	t-1	SST	MST = SST/t-1	F= MST/MSE $\sim F_{(t-1),(n-t)}$
Within treatment error	n-t	SSE	MSE = SSE/n-t	
Total	n-1	TSS		

TWOWAY ANOVA

- It is used to determine the effect of two nominal predictor variables on a continuous outcome variable.
- A two-way ANOVA test analyzes the effect of the independent variables on the expected outcome along with their relationship to the outcome itself.

Example :

- Effect of two antihypertensive drugs in two different doses
- Comparing the employee productivity based on the working hours and working conditions

Two – Way ANOVA

- Two-way ANOVA is a type of study design with one numerical outcome variable and two categorical explanatory variables.
- Mathematical model of two –way ANOVA is as follows

$Y_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$

where μ is overall mean effect

α_i is effect due to i^{th} level of first factor

β_j is effect due to j^{th} level of second factor

γ_{ij} is effect due to interaction between i^{th} level of first factor and j^{th} level of second factor

Assumptions:

- The populations from which the samples were obtained must be normally or approximately normally distributed.
- The samples must be independent.
- The variances of the populations must be equal.
- The groups must have the same sample size.

Two way ANOVA table

Source	d.f	Sum of Square	Mean Square	F
Treatment Blocks Error	t-1 r-1 (r-1)(t-1)	SST SSB SSE	MST = SST/t-1 MSB = SSB/r-1 MSE = SSE/(r-1)(t-1)	F_t= MST/MSE $\sim F_{(t-1),(t-1)(r-1)}$ F_b= MSB/MSE $\sim F_{(r-1),(t-1)(r-1)}$
Total	rt-1	TSS		

Difference between one & two way ANOVA

- a one-way ANOVA is used to determine if there is a difference in the mean height of stalks of three different types of seeds.
- Since only 1 factor that could be making the heights different.
- if three different types of seeds, and then add the possibility that three different types of fertilizer is used
- The mean height of the stalks could be different for a combination of several reasons.
- Two factors (type of seed and type of fertilizer), use a two-way ANOVA.