ANLYSIS OF VARIANCE (ANOVA)

Variance is described as mean of square of deviations taken from the mean of the series of data. It is frequently used as measure of variation.

So far we have discussed that the significance of the difference between two means can be judged by either Z-test or t-test. But problems arises when one had to test the significance of difference for more than two sample means at the same time. At that time by using analysis of variance (ANOVA) technique one can draw inferences whether the means of number of populations are equal. This power full technique was developed by R.A. fisher for separation of the experimentally observed variance into a number of components traceable to specific services. He defines it as the separation of the variance ascribable to other groups. In other words, ANOVA is a method of splitting total variation into components which measures different sources of variations. In an experiment generally these is several factors at work, each one of which may cause a certain amount of variability in the observation made.

The aim of ANOVA is to find how much of the total variability is due to each factor and by comparing these contributory amount of variation, we can test the homogeneity of the observations.

VARIANCE WITHIN AND BETWEEN CLASSES – Suppose a random sample is taken from a normal distributed population. We divide this sample into a number of sub-samples (classes) on the basis of a set of conditions, then the first step in the ANOVA is to separate the total variation in the whole number of observations into two **components**

(i) Variation within classes (groups)

(ii) Variation between classes (groups)

The variation between classes is due to assignable causes whereas the variation within classes is due to various chance causes. The variation due to assignable causes can be detected and measured whereas the variation due to chance is beyond control of human hand and can not be traced separately.

Measure of the variability used in the ANOVA is called mean square. This is similar to variance and defined by

$$\text{Mean Square} = \frac{Sum\ of\ square\ of\ deviation\ from\ mean}{degrees\ of\ freedom}$$

Classes of Models in ANOVA –In the simplest form, ANOVA have a dependent variable **t**hat is metric and one or more independent variables. The independent variables must be all categorical (non metric). Categorical independent variables are also called **factors**. A particular combination of factor levels or categories is called **treatment.**

The following three models are used in ANOVA :

(1) Fixed effect experimenter-Categories of independent variables i.e. treatments are assumed to be fixed. This model applies to the situations in which experimenter applies one or more treatments to the subject of the experiment to see whether the response variable values change .This allows the experimenter to estimate the ranges of response variable values that the treatment would generate in the population as a whole.

(2) Random effect Model : It is used when the treatments are not fixed. This occurs when the various factor levels are sampled from a large population. Because these levels themselves are random variables.

(3) Mixed effect Model : This contains experimental factors of both fixed and random effect types, with appropriately different interpretations and analysis of two types.

ASSUMPTIONS IN ANOVA - ANOVA is based on the following assumptions:

1. Normality i.e. the values in each class are normally distributed.

2. Homogeneity ,i.e. the variance within each class should be equal for all classes.

3. Independence of error ,It states that the error (variation of each value around its own class mean) should be independent for each value.

It may be noted that theoretically speaking, whenever any of these assumptions is not met, the analysis of variance technique cannot be employed to yield valid inferences.

In practice it has been observed that one or more of these assumptions can be bent without appreciable loss in the adequacy of the F-test

Classification of ANOVA - ANOVA is carried out by two ways :

(1) One way ANOVA or One way classification

(2) Two way ANOVA or Two way classification

ANOVA IN One way classification – In one way classification the data are classified

according to only one criterion.

Test Procedure-

(I) Set up the Null hypothesis H_0 : $\mu_1 = \mu_2 = \mu_3 = ... = \mu_k$

And the Alternative hypothesis : $\mu_1 \neq \mu_2 \neq \mu_3 \neq . . . . . \neq \mu_k$

(**II**) **C**onstruct the following table : Suppose a sample of N values of a given variate x is subdivided into k classes according to some criterion of classification. Let ith class consist of n_i members and let jth member of the ith class be denoted by x_{ij} . Then $\sum n_i = N$

Class	Values	n_i	S_i	$\frac{S_i^2}{n}$	$\sum x_{ij}^2$
1	x_{11} x_{12} - - - x_{1j} - - - x_{1n_1}	n_1			
2	x_{21} x_{22} - - - x_{2j} - - - x_{2n_2}	n_2			
3	x_{31} x_{32} - - - x_{3j} - - - x_{3n_3}	n_3			
4					
i	x_{i1} x_{i2} - - - x_{ij} - - - x_{in_i}	n_i			
K	x_{K1} x_{K2} - - - x_{Kn_K}	n_K			
Total	Correction factor $= \frac{S^2}{N}$	$N = \sum_{i=1}^{K} n_i$	$S = \sum S_i$	$\sum \frac{S_i^2}{n_i}$	$\sum_i \sum_j x_{ij}^2$

where $S_i = \sum_j x_{ij}$ and $S = \sum_i S_i = \sum_i \sum_j x_{ij}$.

(III) Calculate the following sum of squares

(i) Total sum of squares (SST)

$$= \sum_i \sum_j (x_{ij} - \bar{x})^2 = \sum_i \sum_j x_{ij}^2 - \frac{S^2}{N}$$

with $N-1$ degrees of freedom.

(ii) Sum of square within the classes (SSE)

$$= \sum\sum (x_{ij} - \bar{x}_i)^2 = \sum\sum x_{ij}^2 - \sum\left(\frac{S_i^2}{n_i}\right)$$

with $N-K$ degrees of freedom

(iii) Sum of square between the classes (SSC)

$$= SST - SSE = \sum_i \left(\frac{S_i^2}{n_i}\right) - \frac{S^2}{N}$$

with $K-1$ degrees of freedom

(IV) Now we construct ANOVA table for one way classification model

Sources of variation	Sum of Squares	Degrees of freedom	Mean Squares	Variance ratio of F
Between classes (samples)	SSC	$K-1$	$MSC = \frac{SSC}{K-1}$	$F = \frac{MSC}{MSE}$
Within classes	SSE	$N-K$	$MSE = \frac{SSE}{N-K}$	
Total	SST	$N-1$		

NOTE- Variance ratio F is the ratio b/w greater variance and smaller variance so if MSE is more than MSC then numerator and denominator should be interchanged and degrees of freedom adjusted accordingly.

(V) Obtain critical value of F from the table for degrees of freedom v_1 & v_2, where $v_1 = K-1$ & $v_2 = N-K$ at 5% level of significance ($\alpha = 0.05$)

(VI) Inference

If calculated F value < table value of F, then Null hypothesis is accepted at α level to significance i.e. there is no significance difference b/w the means.

If the calculated F value > table value of F we reject our null hypothesis i.e. difference b/w means is significant.