

STUDY MATERIAL FOR

BA/BSc 4th Semester

STATISTICS (Gen)

Unit 2: Design of Experiments.

by: D.N. Adhikary.

Analysis of Variance (स्टैटिस्टिक्स वरिएन्स) ?

The analysis of variance (ANOVA or AOV) is a systematic procedure of analysing the variation present in an observed set of classified data. This method helps us in ^(or partitioning) separating the total variation into a number of components according to the nature of classification. We can test the significance of the variation of the observations over different classes or groups.

For example, let us suppose that we have conducted a test of mathematics among the students of class X ~~from~~ selected from different schools of a town and want to know whether their performance varies with respect to the schools. By applying AOV technique to the set of marks (or scores) secured by the students ~~in~~ in the test we can examine if there exists significant difference between the average marks scored by the students of different schools.

Linear models in analysis of variance.:

Each observation in AOV is expressed as a linear combination of some variables and some constants or parameters (combination).

If y_1, y_2, \dots, y_n are n observed values of an experiment then we suppose that

$$y_i = \mu_i + e_i \quad \text{--- (1)}$$

where μ_i = True part

$$= \alpha_{i1}T_1 + \alpha_{i2}T_2 + \dots + \alpha_{ik}T_k$$

$\left\{ \begin{array}{l} \alpha_{ij} = \text{known constants} \\ \tau_j = \text{unknown parameters or effects} \end{array} \right.$

e_i = Error part

Thus the linear combination ~~are~~ is

$$\begin{aligned} Y_{ij} &= \mu_i + e_i \\ &= \alpha_{i1}T_1 + \alpha_{i2}T_2 + \dots + \alpha_{ik}T_k + e_i \end{aligned} \quad \text{--- (2)}$$

This is known as linear model in AOV.

T_j are called effects which are due to ~~one~~ assignable causes of variation and e_i is usually a random variable called error or random effect. It arises due to the chance causes of variation.

There are three types of linear models. These are

1. Fixed effect model :- A model is called fixed effect model if all the effects are fixed ~~one~~ param. -eters except the error effect.
2. Random effect model :- A model is called random effect model if all the effects except the additive constant are random variables.
3. Mixed effect model :- A model is known as mixed effect model if some of the effects are constants and some are random variables.

Assumptions in analysis of variance :

There are some basic assumptions in AOV. These are :

1. ~~The~~ Observations are independently selected from a normal population.
2. The effects of the model are additive in nature.
3. The error or random effects are independently distributed random variables each following the normal distribution with zero mean and equal variance.
4. The error components have a common variance which is the population variance.

Analysis of Variance for one way classified data:-

Let us suppose that we have a set of n observations

- n is classified according to p classes.

If y_{ij} denote the j th value in the i th class then the linear model will be

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij} \quad \dots \quad (1)$$

$$(i=1, 2, \dots, p; j=1, 2, \dots, n_i)$$

where

μ = General mean effect

α_i = Additional effect
of i th class

ϵ_{ij} = Error or random effect

Assumptions are:

1) μ and α_i are constants.

2) $\sum_i n_i \alpha_i = 0$ where n_i = No. of observations
in the i th class.

3). ϵ_{ij} are independent random variables each.

following normal distribution with zero mean and common variance σ^2 .

The unknown constants in (1) are estimated by using the least square method. The estimates are

$$\hat{\mu} = \bar{y}_{..} \text{ where } \bar{y}_{..} = \frac{1}{n} \sum_{i,j} y_{ij} = \frac{G}{n} = \text{Grand mean}$$

$$\hat{\alpha}_i = \bar{y}_{i..} - \bar{y}_{..} \text{ where } \bar{y}_{i..} = \frac{1}{n_i} \sum_j y_{ij} = \frac{C_i}{n_i} = \text{Mean of } i\text{th class.}$$

Using these estimates in (1) we get

$$y_{ij} - \bar{y}_{..} = (\bar{y}_{i..} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i..})$$

$$\Rightarrow \sum_i \sum_j (y_{ij} - \bar{y}_{..})^2 = \sum_i n_i (\bar{y}_{i..} - \bar{y}_{..})^2 + \sum_i \sum_j (y_{ij} - \bar{y}_{i..})^2$$

Thus Total sum of squares = Sum of squares due to Class

+ Sum of squares due to error

In short

$$TSS = SSC + SSE$$

The total SS (TSS) has $n-1$ degrees of freedom (d.f.)
 The SS due to Class has $p-1$ d.f. and
 the SS due to Error has $n-p$ d.f.

Dividing an SS by its d.f we get the corresponding

MS Mean Sum of Squares (MS). They give estimates of population variance (σ^2).

Thus

$$\text{MS due to Class (MSC)} = \frac{\text{SSC}}{p-1}$$

$$\text{and MS due to Error} = \frac{\text{SSE}}{n-p} \\ (\text{MSE})$$

Here we can test the null hypothesis

$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_p = 0$ i.e. there is no significant differences between the class means.

This null hypothesis is tested by using the test statistic

$F = \frac{\text{MSC}}{\text{MSE}}$ which follows the F distribution when H_0 is true.

The results of the analysis is shown in the following table known as ANOVA or AOV table

ANOVA Table for one-way classified data

Source of variation	d.f.	SS	MS	F	Table value of F
Between Classes	$p-1$	$\text{SSC} = \sum n_i (\bar{y}_i - \bar{y}_.)^2$	$\text{MSC} = \frac{\text{SSC}}{p-1}$	$F = \frac{\text{MSC}}{\text{MSE}}$	$F_{0.05, p-1, n-p}$
Within Classes or Error	$n-p$	$\text{SSE} = TSS - \text{SSC}$	$\text{MSE} = \frac{\text{SSE}}{n-p}$		
Total	$n-1$	$TSS = \sum \sum (y_{ij} - \bar{y}_{..})^2$		—	

Q. Discuss in detail the analysis of variance of one-way classified data.