UNIT 13: ANALYSIS OF VARIANCE

UNIT STRUCTURE

13.1 Learning Objectives
13.2 Introduction
13.3 Introduction of Analysis of Variance
13.4 One-way ANOVA
13.5 Two-way ANOVA
13.6 Design of Experiments
13.7 Principles of Design of Experiments
13.8 Completely Randomized Design (CRD)
13.9 Randomized Block Design (RBD)
13.10 Factorial Design
13.11 Let Us Sum Up
13.12 Further Reading
13.13 Answers to Check Your Progress
13.14 Model Questions

13.1 LEARNING OBJECTIVES

After going through this unit, you will be able to

- learn the meaning of ANOVA
- discuss One-way ANOVA and Two-way ANOVA
- discuss Principles of Design of Experiments
- learn Completely Randomised Design(CRD)
- learn Randomised Block Design(RBD)
- describe Factorial Design

13.2 INTRODUCTION

In this unit we are going to discuss about ANOVA and its various concepts. ANOVA is a statistical method that stands for analysis of variance. This unit will give us a fair idea about the One-way ANOVA and Two-way ANOVA, the principles of Design of Experiments.

At the end of this unit we will discuss the Completely Randomised Design(CRD) and Randomised Block Design(RBD).

13.3 INTRODUCTION OF ANALYSIS OF VARIANCE

ANOVA was developed by Ronald Fisher in 1918 and is the extension of the t and the z test. Before the use of ANOVA, the t-test and z-test were commonly used. But the problem with the T-test is that it cannot be applied for more than two groups. In 1918, Ronald Fisher developed a test called the analysis of variance. This test is also called the Fisher analysis of variance, which is used to do the analysis of variance between and within the groups whenever the groups are more than two.

To use the ANOVA we made the following assumptions:

(i) Each group sample is drawn from a normally distributed population.
(ii) All populations have a common variance,
(iii) All samples are drawn independently of each other.

ANOVA is a method of splitting the total variation of a data into integral parts which measures different sources of variations. The total variations split up into two components:

(a) Variations within the subgroups of samples
(b) Variation between the subgroups of samples

These two variations are tested for their significance by F test which is given by

$$F = \frac{\text{Variance between the samples}}{\text{Variance within the samples}}$$

General Purpose of ANOVA: Researchers are using ANOVA in many ways.

The use of ANOVA depends on the research design. Commonly, researchers are using ANOVA in three ways: one-way ANOVA, two-way ANOVA, and N-way Multivariate ANOVA.

Let us discuss these concepts in the following sections:

13.4 ONE-WAY ANOVA

When we compare more than two groups, based on one factor (independent variable), this is called one way ANOVA. For example, it is used if a

manufacturing company wants to compare the productivity of three or more employees based on working hours. This is called one way ANOVA.

The one-way analysis of variance (ANOVA) is used to determine whether there are any statistically significant differences between the means of two or more independent (unrelated) groups.

The one-way ANOVA compares the means between the groups you are interested in and determines whether any of those means are statistically significantly different from each other. Specifically, it tests the null hypothesis:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = ... = \mu_k$$

where μ = group mean and k = number of groups.

Under one-way ANOVA, we consider only one factor and then observe that the reason for said factor to be is that several possible types of samples can occur within that factor. We then determine if there are differences within that factor. The technique involves the following steps:

(i) Obtain the mean of each sample i.e., obtain

$$\bar{x}_1, \bar{x}_2, \bar{x}_3, ...\bar{x}_k$$

when there are k samples.

(ii) Determine the mean of the sample means as follows:

$$\bar{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \bar{x}_3 + ... + \bar{x}_k}{\text{No. of samples (k)}}$$

(iii) Find the Sum of Squares for variance between the samples (or SS between). Symbolically this can be written as:

$$\text{SS between} = n_1(\bar{x}_1 - \bar{\bar{x}})^2 + n_2(\bar{x}_2 - \bar{\bar{x}})^2 + ... + n_k(\bar{x}_k - \bar{\bar{x}})^2$$

(iv) Obtain mean square between samples. This is given by

$$\text{MS between} = \frac{\text{SS between}}{(k - 1)}$$

where (k-1) represents the degrees of freedom between samples.

(v) Obtain sum of squares for variance within samples which is given by

$$\text{SS within} = \sum (x_{1i} - \bar{x}_1)^2 + \sum (\bar{x}_{2i} - \bar{x}_2)^2 + ... + \sum (\bar{x}_{ki} - \bar{x}_k)^2$$

i=1,2,3,...

(vi) Determine mean square within samples which is given by

$$\text{MS within} = \frac{\text{SS within}}{(n - k)}$$

where (n-k) represents the degrees of freedom within samples.

(vii) Finally F-ratio may be worked out as under:

$$\text{F - ratio} = \frac{\text{MS between}}{\text{MS within}}$$

This F-ratio worked as the test statistic and follows F- distribution with (k-1), (n-k) degrees of freedom.

Table 13.1 Analysis of Variance Table for One-way ANOVA

(There are k samples having in all n items)

Source of Variation	Sum of squares (SS)	Degrees of freedom(df)	Mean Square (MS)	F-ratio
Between samples	SS between	(k-1)	$\text{MS between} = \frac{\text{SS between}}{(k - 1)}$	$\frac{\text{MS between}}{\text{MS within}}$
Within Samples	SS within	(n-k)	$\text{MS within} = \frac{\text{SS within}}{(n - k)}$	
Total	SS Total	(n-1)		

13.5 TWO-WAY ANOVA

Two way ANOVA technique is used when the data are classified on the basis of two factors. For example, based on the working hours and working conditions, if a company wants to compare employee productivity, it can do that through two way ANOVA. Two-way ANOVA can be used to see the effect of one of the factors after controlling for the other, or it can be used to see the interaction between the two factors.

Analysis of two-way classified data with one observation per cell

We can plan an experiment in such a way to study the effects of two factors in the same experiment. Let the factors be A and B and the levels $A_1, A_2, \ldots, A_p$ and $B_1, B_2, \ldots B_q$. Let y_{ij} be the observation under the i^{th} level of A and the j^{th} level of B.

Here, the mathematical model may be represented as

$$y_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij},$$

where μ is a constant general effect, present in all observatio ns, α_i is an effect due to the i^{th} level of the factor A, β_j is an effect due to the j^{th} level of the factor B and γ_{ij} is called the interactio n between i^{th} level of the factor A and the j^{th} level of the factor B.

The null hypotheses to be tested here are:

$$H_A : \alpha_1 = \alpha_2 = ... = \alpha_p = 0$$
$$H_B : \beta_1 = \beta_2 = ... = \beta_q = 0$$

for the equality of the effects of the different levels of A and of the different levels of B, respectively. To derive appropriate test for the hypotheses we prepare the following ANOVA Table:

Table 13.2 Analysis of Variance Table for Two-way Classified Data with One Observation per cell

Source of Variation	Sum of squares (SS)	Degrees of freedom (df)	Mean Square (MS)	F-ratio
Between the leveles of A	SSA	(p-1)	$MSA = \frac{SSA}{(p-1)}$	$F = \frac{MSA}{MSE}$
Between the leveles of B	SSB	(q-1)	$MSB = \frac{SSB}{(q-1)}$	$F = \frac{MSB}{MSE}$
Error	SSE	(p-1)(q-1)	$MSE = \frac{SSE}{(p-1)(q-1)}$	
Total	TSS	(pq-1)		

If

$$H_A : \alpha_1 = \alpha_2 = ... = \alpha_p = 0$$

is true, E(MSA) = E(MSE), and hence F =MSA/MSE will give the test of H_A. So a test for the hypothesis of equality of the effects of the different levels of A is provided by this F, which follows the F-distribution with df = (p-1), (p-1)(q-1). The null hypothesis will be rejected at the α level of significance if(and only if)

$$F = \frac{MSA}{MSE} \rangle F_{\alpha;(p-1),(p-1)(q-1)}$$

Similarly,

$$H_B : \beta_1 = \beta_2 = ... = \beta_q = 0$$

for the equality of the effects of the different levels of B, is rejected at the α level of significance if

$$F = \frac{MSB}{MSE} \rangle F_{\alpha;(q-1),(p-1)(q-1)}$$



CHECK YOUR PROGRESS

Q.1. What is ANOVA?

.....

.....

Q.2. What are the assumptions of ANOVA?

.....

.....

Q.3. Which test statistic is used in ANOVA? Write down the test statistic.

.....

.....

13.6 DESIGN OF EXPERIMENTS

Experimental Design is a blueprint of the procedure that enables the researcher to test his hypothesis by reaching valid conclusions about relationships between independent and dependent variables. It refers to the conceptual framework within which the experiment is conducted. One of the main objectives of designing an experiment is how to verify the hypothesis in an efficient and economical way. In the contest of the null hypothesis of equality of several means of normal populations having same variances, the analysis of variance technique can be used. Note that such techniques are based on certain statistical assumptions. If these assumptions are violated, the outcome of the test of hypothesis then may also be faulty and the analysis of data may be meaningless. So the main question is how to obtain the data such that the assumptions are met and the data is readily available for the application of tools like analysis of variance. The designing of such mechanism to obtain such data is achieved by the design of experiment.

13.7 PRINCIPLES OF EXPERIMENTAL DESIGN

There are three basic principles of design which were developed by Sir R.A. Fisher.

(i) Randomization

(ii) Replication

(iii) Local control

(i) Randomization

The principle of randomization involves the allocation of treatment to experimental units at random to avoid any bias in the experiment resulting from the influence of some extraneous unknown factor that may affect the experiment. In the development of analysis of variance, we assume that the errors are random and independent. In turn, the observations also become random. The principle of randomization ensures this.

(ii) Replication:

In the replication principle, any treatment is repeated a number of times to obtain a valid and

more reliable estimate than which is possible with one observation only. Replication

provides an efficient way of increasing the precision of an experiment. The precision

increases with the increase in the number of observations.

(ii) Local control (error control)

The replication is used with local control to reduce the experimental error. For example, if the experimental units are divided into different groups such that they are homogeneous within the blocks, then the variation among the blocks is eliminated and ideally the error component will contain the variation due to the treatments only. This will in turn increase the efficiency.

13.8 COMPLETELY RANDOMISED DESIGN (CRD)

The CRD is the simplest design. Suppose there are *v* treatments to be compared.

- All experimental units are considered the same and no division or grouping among them exist.
- In CRD, the *v* treatments are allocated randomly to the whole set of experimental units, without making any effort to group the experimental units in any way for more homogeneity.

⑩ Design is entirely flexible in the sense that any number of treatments or replications may be used.

Example: Suppose there are 4 treatments and 20 experimental units, then

- the treatment 1 is replicated, say 3 times and is given to 3 experimental units,
- the treatment 2 is replicated, say 5 times and is given to 5 experimental units,
- the treatment 3 is replicated, say 6 times and is given to 6 experimental units

and

- finally, the treatment 4 is replicated [20-(6+5+3)=]6 times and is given to the remaining 6 experimental units

- All the variability among the experimental units goes into experimented error.
- CRD is used when the experimental material is homogeneous.
- CRD is more useful when the experiments are conducted inside the lab.
- CRD is well suited for the small number of treatments and for the homogeneous experimental material.

Layout of CRD

Following steps are needed to design a CRD:

- Divide the entire experimental material or area into a number of experimental units, say *n*.
- Fix the number of replications for different treatments in advance (for given total number of available experimental units).
- No local control measure is provided as such except that the error variance can be reduced by choosing a homogeneous set of experimental units.

Analysis

There is only one factor which is affecting the outcome – treatment effect. So the set-up of one-way analysis of variance is to be used.

13.9 RANDOMISED BLOCK DESIGN (RBD)

The completely randomized design (CRD) will seldom be used if the experimental units are not alike. For in that case the variation among the units will vitiate the test of significance of the treatment effects. The simplest design which enables us to take care of the variability among the units is the RBD. This is also the simplest design using all the three principles enunciated by Fisher.

Suppose we want to compare the effects of p treatments, each treatment being replicated an equal number o times, say q times. Then we used n = pq experimental units, and these units are not homogeneous. The RBD consists of two steps. The first step is to divide the units into p more or less homogeneous groups. In each group or block, we take as many units as there are treatments. Thus the number of blocks is the same as the common replication number (p). The same technique should be applied to the units of a block.

The second step is to assign the treatments at random to units of a block. This randomisation has to be done afresh for each block. This is the difference between of an RBD from a CRD. In an RBD randomisation is restricted within a homogeneous block.

Layout and analysis of RBD follows the analysis of variance for a two-way classified data.

13.10 FACTORIAL DESIGN

Experiments where the effects of more than one factor, say for example variety, manure, etc., each at two or more levels, are considered together are called factorial experiments, while experiments with one factor at varying levels, say only variety or manure, may be called the simple experiments. A Factorial Design is an experimental setup that consists of multiple factors and their separate and conjoined influence on the subject of interest in the experiment. A factor is an independent variable in the experiment and a level is a subdivision of a factor. Factors and levels are different conditions that the experimental subjects are exposed to. A study or experiment is used to see if any of the conditions influence the subject and in what ways

they are influential. The benefit of a factorial design is that it allows the researchers to look at multiple levels at a time and how they influence the subjects in the study.



13.11 LETS US SUM UP

ANOVA is a method of splitting the total variation of a data into integral parts which measures different sources of variations.

It is used to do the analysis of variance between and within the groups whenever the groups are more than two.

In the contest of the null hypothesis of equality of several means of normal populations having same variances, the analysis of variance technique can be used.

The use of ANOVA depends on the research design. Commonly, researchers are using ANOVA in three ways: one-way ANOVA, two-way ANOVA, and N-way Multivariate ANOVA.

Experimental Design is a blueprint of the procedure that enables the researcher to test his hypothesis by reaching valid conclusions about relationships between independent and dependent variables.

There are three basic principles of design which were developed by Sir R.A. Fisher.

(i) Randomization
(ii) Replication
(iii) Local control

Based on the application of the principles of design, many design are being developed, viz.

(a) Completely Randomized Design (CRD)
(b) Randomised Block Design (RBD)
(c) Latin Square Design (LSD)
(d) Graco Latin Square Design
(e) Factorial Design, etc.



13.12 FURTHER READING

1. Kothari, C.R. and Garg, G.(2014), '*Research Methodology - Methods and Techniques'*, New Age International Publishers
2. Malhotra. N.K.(2011), *Marketing Research – An Applied Orientation*, Pearson



13.13 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q.1. ANOVA is a statistical method that stands for analysis of variance. ANOVA was developed by Ronald Fisher in 1918 and is the extension of the t and the z test. Before the use of ANOVA, the t-test and z-test were commonly used. But the problem with the T-test is that it cannot be applied for more than two groups. In 1918, Ronald Fisher developed a test called the analysis of variance. This test is also called the Fisher analysis of variance, which is used to do the analysis of variance between and within the groups whenever the groups are more than two.

Ans to Q.2. To use the ANOVA we made the following assumptions:

(i) Each group sample is drawn from a normally distributed population.

(ii) All populations have a common variance,

(iii) All samples are drawn independently of each other.

Ans to Q.3. The test statistic used in ANOVA technique is F-test. It is given by

$$F = \frac{\text{Variance between the samples}}{\text{Variance within the samples}}$$



13.14 MODEL QUESTIONS

Q.1. What are the assumptions of ANOVA technique?

Q.2. What is Design of Experiment?

Q.3. What are the principles of Design of Experiment laid by R.A. Fisher?

*****