CHAPTER NO. 1.

ELEMENTARY CONCEPTS

In this chapter, we discuss about the "Experimental Design" in general. In the section 1.1, we define certain terms such as 'experiment','experimental unit', 'treatment' etc. and explain these with examples. Further we see as to why planning of the experiment is essential and how it is come. In the section 1.2, we discuss three principles of 'Design of Experiment' in brief. Section 1.3 and 1.4 give the information on types of.designs and some standard designs respectively. After carrying out the experiment we have to 'analyse' the collected data resulted from the experiment. In section 1.5, we describe the technique of 'Analysis of Variance 'and its general structure. For the sake of completion of dissertation some definitions which are used later on are given at the end of this chapter.

1.1 INTRODUCTION:-

There are a lot of problems in our daily life. And we think that the best way to get rid of the problem is to solve it. And to solve a problem we have to carry out the experiments regarding it. An experiment is, "a test or trial carried out carefully in order to study what happens and achieve new knowledge". After a statistical problem has been set up, the next step is to perform experiments for collecting information on the basis of which inferences can be made in the best possible manner. For this purpose a set of 'expermental units' and adequate 'experimental material' are required. Equal sized plots of land, a single or a group of plants etc. are used as experimental units for

3.

agricultural experiments. for animal husbandary experiments animals, animal organs etc. serve as the experimental units. Again in industrial experiments, machines, ovens and other similar objects form the experimental units.

In the literature, a general name, 'treatment' is given to denote 'experimental material' among which comparison is desired by utilising the effects which are produced when the experimental material is applied to the experimental units. The problems are usually in the form of comparisons among a set of treatments in respect of their effects. For example, 'in agricultural experiments, different varieties of a crop, different fertiliser doses, different levels of irrigation, different combinations of levels of two or more of the above factors, viz. variety of a crop, irrigation, fertilisers, date of sowing etc. may constitute the 'treatments'. In chemical experiments. In industrial experiments, different operating temperatures, different brands of tyres etc. may constitute treatments.

With a given set of treatments, in order to carry out experiment scientifically, certain planning is essential. Such a plan is called as, 'Design of Experiment'. It specifies, 'the size and number of experimental units, the manner in which treatments are to be allotted to the experimental units and also the appropriate type of grouping of the experimental units. These requirements of a design ensure [Bicking (2,P.22)] validity, interpretability and accuracy of the results from an analysis of the observations.

For instance, suppose a chemical engineer hopes to improve the yield of some petrochemical in an oil refinary plant by comp-

aring several catalysts. Crude oil is fed into the plant which is charged with the catalyst; some of the crude oil or feedstock passes through the plant unchanged; some is converted into the petrochemical or product. The liquid that comes out of the plant is separated into product and unconverted feedstock, and the yield or response, is the percentage of feedstock converted into product.

An obvious procedure is to make one or more plant repetitions using each of the catalysts and to compare the average yields on each catalysts. There are, however, some other considerations that enter in the picture viz. How many catalysts? How many repetitions? How co we compare the ave**rage**s after obtaining them? Taking into consideration all these facts we have to design the experiment. Before discussing the various types of designs we give the historical development of it.

The theory of experimental design was first developed in agricultural field at 'Rothamsted Experimental Station' in England. Fisher was the first to develope it and to use the technique of 'Analysis of Variance' as the method of statistical analysis in experimental design. The first general account of the results of this research work was given by him in his book, titled 'The Design of Experiments' which originally appeared in the year 1935. Since then a number of books and papers have come out which helped for further development of this branch. Yates worked with Fisher and they collaborated on many projects. Yates also became a primary contributor to the literature of experimental design. In addition to these two, there are many other statisticians who are responsible for the further development of this

field. Amongst these, Kempthrone, Cochran, Bose, Ogawa, Das, Vartak, Khatri etc. are the major contributors. And excellent books due to the authors such as Kempthrone (1952), Cochran and Cox (1950), Fedrer (1955), Scheffee (1959), Das and Giri (1979), John (1971), Raktoe; Hidayat and Fedrer (1981), Ogawa (1974) etc. are avaliable on the literature of design of experiments.

As already stated the design of experiment was used in early stages in agricultural and biological sciences. As a result, much of the terminology is derived from this agricultural background. Now-a-days, the experimental design methods and the technique of `analysis of variance´are widely employed.in all fields of enquiry such as agricultural, biological sciences, social sciences, medical sciences, engineering sciences etc. 1.2 : THE THREE PRINCIPLES OF DESIGNS OF EXPERIMENTS:-

As already stated, the design of experiments [Bicking (2)] ensures validity, interpretability and accuracy of the results obtainable from the analysis of observations. These purposes can be achieved by the following three principles, viz.

[1]. Randomisation ,

1

[2]. Replication and

[3]. Local Control.

We discuss below each of the above, with their roles in data collection and interpretation.

[1]. Randomisation:- It defines the manner of allocation of the treatments to the experimental units. The treatments are allotted to the experimental units at random to avoide any type of personal of subjective bias. This ensures validity of the results and independance of the observations. Every design has its

own way of randomisation. It has been discussed in detail in Fisher (1942), Kemthrone (1952) , Ogawa (1974), among others.

[2]. REPLICATION :- If a treatment is allotted to `r' experimental units in an experiment, it is said to be replicated r times. If in a design, each treatment is replicated r times, the design is said to have r replications. By replication we can increase the accuracy of estimates of treatment effects. It also provides an estimate of the error variance. Though the more the number of replications better is the design so far as precision of estimates is concerned, it cannot be increased indefinitely as it increases cost of experimentation. Moreover, due to limited availability of experimental resources, too many replications cannot be taken. Therefore, the number of replications are decided keeping in view the permissible expenditure and the required degree of precision. Usually the precision of estimates is measured in terms of error variance. For a given measure of error variance based on a set of experimental units and desired level of accuracy, the number of replications needed are obtained from

(1.2.1)

Where,

t is the critical value of t distribution at the desired level of significance,

d is the difference between averages of two treatment
effects;

2 s is a measure of error variance. E

2

Given d and s , we can determine the value of 'r'. E Using randomisation and replication we can achieve the desired precision but one cannot reduce experimental error. To reduce the magnitude of experimental error we use the principle of 'error control'.

[3]. LOCAL CONTROL :- It increases the precision by choosing appropriate type of experimental units and also their groupings. The standard error of estimate of a treatment effects is

reduce this standard error of treatment effects. But, only by taking a large value of r, we cannot reduce the error variance. However, there are some other measures of reducing the error variance. Such measures are called, 'error control'. One such measure is to make experimental units homogeneous. Another method is to form the units into several homogeneous groups, usually called as 'blocks' while allowing variation between the groups. Also by the technique of 'confounding' and 'analysis of covariance ' experimental error can be reduced.

With the help of above principles, design of experiments can be classified as follows --

1.3. CLASSIFICATION OF EXPERIMENTAL DESIGN

All experimental designs [Fedrer (1955)] may be divided in to two parts.

1. Systematic Designs and

ć

2. Randonised Designs.

We discuss them below--

1.3 Systematic Designs:- Prior to the development of modern experimental designs, experimenters used various arrangements which are not subject to the laws of 'chance '. Systematic schemes of arranging the treatments in the various repetitions have been devised. One such scheme is to arrange all duplicates, triplicates etc. of the treatments together. Suppose the experimenter wishes to test three treatments A, B and C and he decides to have four replications of each treatment. The arrangment of three treatments over the experimental area could be one of the following :

|____A___A___A___| BBBB Α Α A <u>B_B_B</u>_ В C C C C В С Α 1 С A 1 В С Α В в C в В С С Α A В В С С

In the above different arrangments, the geometrical strucure of the field is also considered.

Before, Fisher proposed the concept of randomisation, a sysematic ordering of treatment in each block or repetition seemed natural. One of the more common types of systematic arrangements in which the treatments are repeated several times is the

following---

Replicate I	Replicate II	Replicate III
:_A_;_B_;_C_;	: <u>A;B;C</u> ;	<u> _A_;_B_;_C_;</u>

Advantages of Systematic Designs

Following are the main advantages of Systematic Designs:-1. Many experimenters feel that planting and harvesting in agrronomic trials are faciliated by using systematic arrangements. 2. The systematic design provides `adequate ' sampling of the experimental area i.e. it allows for "intelligent placement" of the various treatments.

 Varieties amy be arranged in order of maturity. For instance, fertilisers can be arranged in order of increasing fertility.
 It may be desirable to alternate dissimilar varieties so that natural crossing or mechanical mixtures can be detected subsequently.

The Disadvantages of The Systematic Designs.

Following are the some of the disadvantages of systematic designs:-

1. There is no valid estimate of the variance of treatments effects.

2. The correlation between adjacent plots may lead to systematic errors in assessing treatment differences.

2. RANDOMISED DESIGNS:

In this design, treatments are randomly allotted to the experimental units. The use of randomisation is the keystone of the application of statistical theory to the design of experiments

and the validity of results depend upon it.

For example, an agronomist comparing two varieties of crop would not rationally assign to one variety all the plots that were in the shade and to the other all the plots that were in the sunlight. If one does so, he would not be able to tell after the experiment whether any apparent difference in yields resulted from varital differences or from the fact that one variety had received more sunlight.

In order to eliminate the element of subjectivity such as occurd in the above example, it is essential to follow the principle of randomisation. In addition to this, as pointed out by Fisher (1947) we get an adequate basis for obtaining the tests of significance and confidence intervals.

We achieve the randomisation by some standard procedure, such as : Lottery method and Use of Random Number Tables. Fisher's Random Number Tables and Tipett's Random Number Tables are most commonly used.

Some standard randomised design are discussed in the next section. Henceforth we refer, 'randomised designs' by simply the word, 'designs'.

1.4 . SOME STANDARD DESIGNS .

Designs are usually characterised by the nature of grouping of experimental units and the procedure of random allocation of treatments to the experimental units.

Following are the some of the standard dsigns :-

- 1. Completely Randomised Design (CRD)
- 2. Randomised Block Design (RBD) and

3. Latin Square Design (LSD) .

 ϕ

Below we discuss each of these in brief:

1. Completely Randomised Design (CRD) :- It is the simplest randomised design. In this design the experimental units are taken in a single group. As far as possible the units forming a group should be homogeneous. CRD is one, in which a group of 'v' treatments are randomly allocated to the whole set of experimental units, without making any effort to group the experimental units in any way for more homogeneity. There is no restriction upon the number of replications of a treatment.

Layout Of The Design:-

By layout we mean the placement of experimental treatments on the experimental site whether it be over space, time or type of material. The entire homogeneous experimental area is divided into number of experimental units, say N. A random selection of 'r ' experimental units is made and one of the 'v' treatments is 1 applied to these units. A random selection of 'r ' of remaining 2 'N - r 'experimental units is made and one of the remaining 1 'v - 1' treatments is applied to these particular units. Continue this process until all treatments have been applied. And after the experimentation we observe the response.

2. Randomised Block Design (RED) :- We have seen that CRD is useful for small number of treatments and homogeneous experimental material. When there are large number of treatments 'v' to be tested and experimental material is not homogeneous, CRD is not useful. If the experimental material is not homogeneous, it may be possible to group the material into blocks of v - units each, as- the blocks are homogeneous within themselves and heterogene-

ous between themselves. Then 'v' treatments are allocated on each block randomly. Such a design is called as, 'Randomised Block Design (RBD).

In RBD, the randomisation is resulted and treatments are randomly allocated with in each block. If there are 'b' blocks. then to test 'v' treatments we need, N = b.v total number of experimental units.

This design was originally developed by Fisher (1926) and has become popular in a large number of field experiments.

Layout Of RBD :

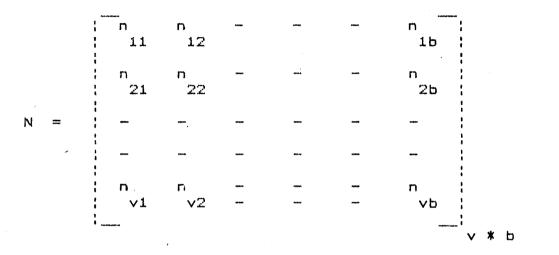
Suppose we have v = 5 treatments and b = 4 blocks each of size 5. Let Y denote the response on the experimental unit ij in the j th block receiving i th treatment and let A, B, C, D, E denote the treatments, then we have following layout of RBD :-

Block	I	A Y 11	C Y 31	B Y 21	E Y 51	D Y 41	8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8
Block	II	 В ү 22	D Y 42	с ү 32	A Y 12	E Y 52	• • • • • • • • • • • • • • • • • • •
Block	III	 D Y 34	E Y 35	в Ү 32	C Y 33	A Y 31	t t t t t t t t t t t t
Block	IV	E Y 54	D Y 44	C Y 34	В Ү 24	A Y 14	

The same plan can be generalised for different values of b and v.

In RBD, it is essential to occur every treatment once and

only once in each block. But there are many practical situations in which this restrication cannot be satisfied. In some experimments,we may have to repeat a certain treatment at least once in each block. Let n denote the number of times i th treatment i_j occurs in the j th block ; where n >/ 1. And also every i_j treatment may not occur same number of times. Such a block design is called as, 'General Block Design '. And a matrix -





Let,

r denote number of replication of i th treatment;
i i=1,2,- - v.

and k denote number of plots in j th block; j= 1,2,- - 5. j Then,

and

 $\sum_{i=1}^{V} r = \sum_{i=1}^{b} k$

And a design, for which n > 1, is called a complete block ij design'.

When the number of treatments 'v' in an experiment is large, it may not be possible due to various reasons to use large size blocks to accomodate all treatments at least once in each block. In such cases we think that it is not necessary for every treatment to occur in each block. Some treatments occur and remaining will not occur which implies that block, size is less than total nubmer of treatments. Such a design is called as, 'Incomplete Block Design '.

Further, if n takes either the values 0 or 1 then the ij incomplete block design is called a, `binary design '. Such a type of design is common in practice.

3. Latin Square Design (LSD) :- As we have seen, to eliminate fertility gradint occuring in one direction only, we use R.B.D. But when fertility gradient is in two directions which are prependicular to each other, we use Latin Square Design (LSD). -For this, we divide the given field into different rows and columns each having same number of experimental units, then we allocate the treatments to experimental units in such a way that each treatment occurs once and only once in each row and in each column.

A Latin square design is an incomplete 3 -way layout in which, each of three factors viz. rows, columns and treatments 2 is at v `levels ' and only v possible treatment combinations are taken.

Latin square designs were originated for agricultural experimentation by Fisher (1926). At present they are useful in industry, laboratory, greenhouse, medical, marketing, sociological experiments etc.

Layout of L.S.D.

Suppose there are four treatments A,B,C and D to be tested in L.S.D. Then we will have four rows and four columns.

And the layout will be --

A	в	С	D
в	С	D	Α
С	Ð	A	В
D	A	в	С

In the next section we discuss about the `analysis of observations ', obtained as a result of the experiment.

1.5 . ANALYSIS OF VARIANCE :

As a result of carrying out experiment, we get observations. After the observations are collected they are statistical analysed to get relvent information regarding the objective of the experiment. As we know, the the objective is usually to make comparisons among the effects of the several treatments when the observations are subject to variation. Such comparisons are made by the technique of 'analysis of variance 'which is due to Fisher. According to him, "the analysis of variance technique essentially consists of partitioning the total variation in an experiment into components ascribable to different sources of variation due to 'the controlled ' factors and 'uncontrolled ' sources of variation, called 'error '.

Symbolically, it can be written as,

 $\sigma_{\rm T}^2 = \sigma_{\rm C}^2 + \sigma_{\rm P}^2$

Where.

 \mathbb{Z} 2

is the total variation ,

is the variation due to `controlled ´ factors c

and

is the variation due to error

In order to faciliate the analysis and to simplify the tests significance, we have to make some assumptions about the nature of observations or responses obtained from an experiment.

Assumptions Involved In The Analysis Of Variance :-

Following are the three important assumptions commonly made in the, analysis of variance.[John and Quenouille (1953)]

1. The uncontrolled variation or error in different measurments follow a normal distribution.

2. Different measurements are independent.

3. The relative sizes of errors in different measurements are unrelated to any factor of the experiment.

The implication of the above assumptions will be made clear in the further discussion.

Now we discuss the general structure of `Analysis of Varian-

Consider the linear model -

Y = XB + E -----(1.5.1)

Where,

Y is a column vector of observations (y, y, ---y), 1 2 n

eta is a column vector of the parameters eta , j

X is a matrix of known coeffecients known as `design matrix'

and E is a column vector of error components.

The above assumptions are equivalent to E(E) = 0 and $V(E) = 6^2 I$; where I is an identity matrix.

By the method of least squares, the least square estimate of parameter B is obtained by solving the equations,

$$x'x_{\beta} = x'y$$
 ----- (1.5.2)

These equations are called, 'normal equations '.

Let $\hat{\beta}$ be the solution of equation (1.5.2), and when $\beta = \hat{\beta}$, E'E attains minimum value and is unique. And this minimum value,

denoted by R is given by -

$$R^{2} = (Y - X\hat{\beta})' (Y - X\hat{\beta}). ----(1.5.3)$$

In analysis of variance, each component of variation is associated with another quantity, called as, 'degrees of freedom ', (d.f.), which is defined as follows ---

Definition : 1.5.1 Degrees Of Freedom (d.f.) :-

The "degrees of freedom" associated with any component are the number of independent parameters required to describe that component in the model. [Cochran and Cox (1959) pp.57].

In the case of treatments, this always equal to one less than the number of treatments and similarly for blocks.

Suppose R carries v d.f. Now suppose we wish to test 0 1 the linear hypothesis

$$H : H' \beta = 0$$

Under this hypothesis we get the residual sum of squares, denoted 2 by R ; where,

$$R = m_{10}, \quad (Y - XB), \quad (Y - XB), \quad (-1, -XB), \quad (-1, -2, -4)$$

$$I = H^{2}B = \bigcup_{1}^{2} M^{2} M^{2}$$

1.7

under H , this F follows Snedecores F -- distribution n

$$F(v-v, v)$$

The computation leading to the F statistic may be presented in tabular form, called 'Analysis Of Variance Table ' or simply, ` ANO√A ′ Table.

Table No. 1.5.1 .

	ANOVA	Table
--	-------	-------

Source Of Variation	d.f.	Sum Of Squares
Deviation, from hypot- thess Ηβ : (9 Ο	y - y 2 1	2 2 R - R ^ 1 0 ;
Residual	ア 1	2 . R . O
Total	V 2	2 R 1

The entry marked by \uparrow is obtained by substration .

Models for the most of designs discussed earlier can be expressed in the general set up

Where the terms have similar meanings as explained earlier. So, for the 'analysis of design ' we follow the above technique of `analysis of variance '.

1.6 . SOME DEFINITIONS :

Below we give some definitions and result which are useful in further discussions .

C -- matrix :--Definition :- 1.6.1

For the binary design, the matrix

C = D(r,r, --r, --r) - N D(k, k - k, k) N, $1 2 \qquad i \qquad v \qquad 1 2 \qquad j \qquad b$ where, $D(\Theta,\Theta, - -, \Theta)$ is a k * k diagonal matrix with dia-1 2 k gonal elements δ , δ , -, δ ; r is the number of replcations of i th treatment , i = 1,2,- - - \vee and k , the size of j th block; j = 1, 2, - - - b; is called a C --- matrix of the incomplete block design. Linear Parametric Function :-Definition 1.6.2 :-A parametric function is said to be a linear parametric function of $\emptyset = (0, 0, ---, 0)$ if it is of the form $\psi = c' \Theta = \sum_{i=1}^{k} c \Theta$ where the vector C' = (c, c, --, c) is a vector of $1 \quad 2 \quad k$ known coefficients. Definition 1.6.3 : Two linear parametric functions $\Psi = C'_{0}$ and $\psi = C'_{O}$ are said to be (algebrically) independent if C cannot be written as a scalar multiple of C Definition 1.6.4 : Constrast :-Suppose y , y , - - y are k observations, then a linear 1 - 2 - kfunction C = c y + c y + - - - , c y, is called an observat-11 2 2 k k $\frac{k}{1000}$ ional contrast if 2 c = 0 . And a contrast is called a i=1 i normalised contrast if

Definition 1.6.5 : Orthogonal Contrast :-

Two contrasts
$$_k$$
 $_k$
 $C = \ge c y$ and $C = \ge c y$ `are said
1 i=1 1 i 2 i=1 2i i

to be orthogonal contrast to each other if ,

$$\sum_{i=1}^{2} c c = 0$$

i=1 1i 2i

E(c) = 0, and v(y) = $\frac{2}{6}$, and y's are normally distributed, i
2
k 2
then (C / >_c) is distributed as x with 1 d.f.

Definition 1.6.6 : Unbiased Estimator :-

A function t(Y) of the observations Y is said to be an unbiased estimataor of parametric function if E [t (Y)] is equal to the parametric function.

Definition 1.6.7 : Elementary Contrast :-

A contrast C't is called an elementary contrast if the vector C has only two nomzero entires 1 and -1, the other entries being zero.

Now we will present some of the properties of Block Designs. The proof of various results are available in Raghar Rao (1971). Definition 1.6.8 : Connected Design :-

" A design where in all elementary contrast are estimable is called a connected design ".

Theorem 1.6.1 : An incomplete block design with v treatments is connected if and only if the rank of it's C - matrix is v-1. Theorem 1.6.2 :- (Chakarbarti 1963) :- In a connected design all the diagonal elements of it's C - matrix are positive and the principal minors of all orders of it's C - matrix are positive. The idea of connected design is due to Bose. Definition 1.6.9 : Balanced Design :-- "A connected design is said to be balanced if all elementary contrasts in the treatment effects can be estimated with the same precision (inverse of the variance of the estimator)".

This definition does not hold for the disconnected design, as all elementary contrasts are not estimable in this design. To over come this difficulty , Vartak (1963) defined " a design (not necessarily connected) to be balanced if every estimable normalised contrast in the treatment effects can be estimated with the same precision".

Theorem 1.6.3 : (Rao 1958) :- A connected design is balanced if and only if all characteristic roots of it's C -matrix are equal. Definition 1.6.10 : Orthogonality :--

Yates defined orthogonality of a design as follows, " orthogonality of a design is the property which ensures that the different effects will be capable of seperate estimation and testing with any entanglement ".

By the nature of design here we can say that RBD and LSD are connected, balanced and orthogonal. CRD is also orthogonal and connected but it is not balanced.

In section 1.4 we have discussed CRD, RBD and LSD, which are designs of simple experiments. A simple experiment takes into account the different levels of only one factor at a time.

21

But this procedure is not always desirable or practicable. Many times we have to consider more than one combinations of different levels of different factors, at a time. Such an experiment, involving different factors at different levels is called a, 'factorial experiment '. Such experiments are in existance from several decades.

In the subesquent chapters , we will discuss about `factorial experiments ' in detail.



