CHAPTER 1

REVIEW OF LITERATURE AND INTRODUCTION

1.0 Historical Review

Research is a careful investigation or inquiry specially through search for new facts in any branch of knowledge. It is a systematized effort to gain new knowledge and also a movement from the known to the unknown. According to Clifford Woody, research comprises defining and redefining problems, formulating hypothesis or suggested solutions; collecting, organizing and evaluating data, making deductions and reaching conclusions; and at last carefully testing the conclusions to determine whether they fit the formulating hypothesis. In the Encyclopaedia of Social Sciences, D. Slesinger and M. Stephenson define research as “the manipulation of things, concepts or symbols for the purpose of generalizing to extend, correct or verify knowledge, whether that knowledge aids in construction of theory or in the practice of an art”. Research is, thus, an original contribution to the existing stock of knowledge making for its advancement.

The role of research in several fields of applied economics, whether related to business or to the economy as a whole, has greatly increased in modern times. It provides the basis for nearly all government policies in our economic system. It has its special significance in solving various operational and planning problems of business and industry. It is equally important for social scientists in studying social relationships and in seeking answers to various
social problems. Thus, research is the fountain of knowledge for the sake of knowledge and an important source of providing guidelines for solving different business, governmental and social problems. It is a sort of formal training, which enables one to understand the new developments in one’s field in a better way.

It is possible for a scientist to conduct an investigation without statistics but it is impossible for a statistician to do so without scientific knowledge. However a good scientist becomes a much better one if he used statistical methods.

While experiments can benefit by learning about statistics, the converse is equally true. Statistician can benefit by learning from a genuine interest in practical problems.

Prof. R.A. Fisher’s name is associated with experimental designs. He was responsible for statistics and data analysis at the Rothamstat Agricultural Experiment Station in London, England. As such the study of experimental design has its origin in agricultural research. Prof. R.A. Fisher found that by dividing agricultural fields or plots into different blocks and then by conducting experiments in each of these blocks whatever information is collected and inferences drawn from them, the same happens to be more reliable. This fact inspired him to develop certain experimental designs for testing hypotheses concerning scientific investigations. Fisher developed and first used the analysis of variance as the primary method of statistical analysis in experimental design. Frank Yates worked with Fisher at the experimental station and collaborated on many projects. Many of the
early applications of experimental design methodology were in the agricultural and biological sciences. As a result, much of the terminology of the discipline is derived from this agricultural background.

An agricultural scientist may plant a variety of crops in several plots, then applied different fertilizers or treatments to the plots, and observe the effects of the fertilizers on crop yield. Each plot will produce one observation on yield. The results of experiment are affected not only by the action of the treatments, but also by extraneous variations, which tends to mark the effects of the treatments. The term experimental error is often applied to this variation. Two main sources of experimental errors may be distinguished. The first is inherent variability in the experimental material to which the treatments are applied. The second source of variability is the lack of uniformity in the physical conduct of the experiment. Neither the presence of the experimental errors nor their causes need concern the investigator, provided that his results are sufficiently accurate to permit definite conclusions to be reached. However, with the time and labour that can be given to an experiment the results are greatly influenced by experimental errors that only treatment differences that are large can be detected and even this may be subject to considerable uncertainty. Whatever the source of experimental errors, replication of the experiment steadily decreases the error associated with the difference between the average results for two treatments, provided that randomization have been taken to ensure that one treatment is no more
likely to be favoured in any replicate than another, so that the errors affecting any treatment tends to cancel out as the number of replications is increased. The degrees of freedom for errors are relevant to the comparison of two designs. A change in design that decreases the error degrees of freedom as well as the error variance may not be advantageous.

As Fisher has stressed, striking gains in precision may be achieved by testing different types of treatments in the same experiment, instead of conducting a separate experiment for each type.

In experiments with large number of treatments, numerous attempts have been made to avoid this loss of precision by the formation of groups of experimental units, which do not contain all the treatments and therefore can remain small. These groups are called incomplete blocks. If all the comparisons between pairs of treatments are potentially at equal importance, a different method is used in forming the blocks, when treatments are arranged in incomplete blocks, two treatments which occur in the same block are more precisely compared than two which are placed in different blocks. Therefore the goal is to construct a design such that any pair of treatments occur equally often within same block. The solutions available to date called balanced incomplete block design.

John Stuart Mill was probably the first to give an idea of how to carry out experiment. He separated experiments into "spontaneous" experiments, which are now called observational studies, and "artificial" experiments now called controlled experiments.
In case of experimental design, one does not play with the population, instead first the problem will be created and then on the basis of the problem, data are to be collected by conducting a proper and systematic experiment. While conducting the experiment one has to take care to create the controlled conditions, which is the main important characteristic feature of an experiment. It is the design of an experiment, which specifies the nature of control over the operations in the experiment. Proper designing is necessary also which ensure that the assumptions required for appropriate interpretation of the data are satisfied. However, designing is necessary to increase accuracy and sensitivity of the results.

1.0.1 Design of Experiments

A Design of Experiment is usually characterized by size, number of experimental units, the procedure of allocation of treatments to the experimental units and the nature of blocking etc. in agricultural and biological experiments. For such type of experiments it will not be possible to use the large size of blocks accommodating all the treatments in each block. We generally prefer incomplete block designs. If all the treatments are not used in each block of the plan of a design, then it is called an incomplete block design. Incomplete block design is an arrangement of \( v \) treatments in \( b \) blocks each of size \( k_j \) \((k_j < v)\) such that each of the \( v \) treatments is replicated \( r_i \) times, \( r_i \) and \( k_j \) may be equal or unequal where \( v \), \( b \), \( r_i \) and \( k_j \) be the parameters of an incomplete block design and \( i = 1, 2, \ldots v, j = 1, 2, \ldots b \). When all the pairs of
treatments in a design occur together the same number of times (say, \( \lambda \)) and \( r_1 = r_2 = \ldots = r_v = r, k_1 = k_2 = \ldots = k_b = k \) then the incomplete block design becomes balanced incomplete block design (BIBD). It was first introduced by Yates (1936) for agricultural experiments. For balanced incomplete block design, all elementary contrasts are estimated with the same precision. Constructional problems were developed by Bose (1939). Fisher and Yates (1973) prepared an extensive table of balanced incomplete block designs. Further it has extended by Cochran and Cox (1957). Rao (1961) studied the BIB designs with replications 11 to 15. It was noticed that the balanced incomplete block designs are not available for every parametric combination. To overcome such difficulties, Yates (1936) developed another series of incomplete block designs, which Yates called Lattice designs. Lattice designs have the drawback that they are available only for number of treatments, which are perfect squares or cubes. This limitation was removed by Bose and Nair (1939) by introducing another class of incomplete block designs which they called Partially Balanced Incomplete Block Designs with m associate classes, abbreviated as PBIBD (m). In original definition of Bose and Nair (1939) it was assumed that all \( \lambda_i \) parameters are distinct but this restriction was removed by Nair and Rao (1942). It was also assumed that \( p_{ik}^j = p_{kj}^i \) earlier but Bose and Mesner (1959) showed that this condition is redundant. It is to be remarked that the concept of the association scheme plays the fundamental role in the analysis and classification of PBIB designs. This concept though inherent in Bose and Nair's definition,
though explicitly introduced by Bose and Shimamoto (1952). Bose and Conner (1952), Bose and Shimamoto (1952) defined association scheme between treatments purely from group considerations without relating it to the number of replications of the pair of treatments. Bose and Shimamoto (1952) developed Partially Balanced Incomplete Block Designs with two associate classes, abbreviated as PBIBD (2).

Further Bose and Shimamoto (1952) classified PBIBD (2) into the following five categories on the basis of their association schemes.

(i) Group Divisible (GD) designs
(ii) Simple (S) designs
(iii) Triangular (T) type designs
(iv) Latin square (LS) type designs
(v) Cyclic (C) designs.

Later Bose (1963) introduced another class of PBIB designs with two associate classes and named it partial geometry. Some of PBIB designs with two associate classes which are not covered in above mentioned designs are called miscellaneous PBIB designs.

1.0.2 Optimality of Block designs

Despite the success of the continuous theory Jack Kiefer (1958) found optimum designs for fixed sample size or under more rigid combinatorial restriction. He established mild conditions for optimality in a wide sense of complete block designs and Latin squares.
In block designs the optimality criteria have been defined on some function of the information matrix $C_d$, where $d$ refers to the design. The natural place to look for optimal designs is among partially balanced incomplete block designs with two associate classes or, more particularly group divisible designs. In a series of papers Cheng (1978, 1979, 1980) proved optimality for certain members of a class of designs called regular graph designs. Kiefer (1958) and Cheng (1980) translated their theorems into the language of graph theory gave new results. The different optimality criteria A, D and E in terms of geometric meaning is: An ellipsoidal confidence set instead of a point estimate: the determinant is proportional to the square of the volume; the maximum eigenvalue to the square of the maximum diameter; and the trace to the sum of squares of lengths of the principal axes.

1.1 Summary of the Present Study

This Thesis comprises of six chapters.

In Chapter 1, we explain the basic concepts in Design of Experiments and give a brief summary about our investigation.

In Chapter 2, we have mentioned a few preliminary definitions required in the later chapters. We have developed alternative methods for the construction of Hypercubic Designs (HCD) from symmetrical factorial experiments and explained in Chapter 3. E-optimality of Hypercubic Designs, which are Regular Graph Designs has been discussed in Chapter 4. Construction of Semi-Regular Group Divisible designs from certain types of hypercubic designs and application of
hypercubic designs in the construction of complete, partial and balanced confounded experiments for symmetrical factorial experiments and then to identify the confounded main effects and interaction effects has been discussed in Chapter 5. We have developed a new method of construction of $\lambda$-linked block designs through non-adaptive hypergeometric group testing designs for $v^* \equiv 0 \pmod{6}$ and $v^* \equiv 2 \pmod{6}$ and discussed the Type I optimality of such designs in Chapter 6.

1.1.1 Construction of Hypercubic Designs from Symmetrical Factorial Experiments

Different types of Partially Balanced Incomplete Block Designs (PBIBD) with two or more than two associate classes have been widely discussed by many statisticians like Bose and Nair (1939), Bose and Shimamoto (1952), and are solely characterized by their parameters and association schemes. Some of the most important PBIB association schemes with two classes are group divisible, triangular and Latin square. Schemes with more than two association classes are hierarchical group divisible (Roy (1953)), Cubic (Raghavarao and Chandrasekhararao (1964)) and rectangular (Vartak (1955)) etc.

Hypercubic design was introduced by Shah (1958) and Kusumoto (1965) in which $v = t^m$ treatments are represented by $m$-plets $(x_1, x_2, \ldots, x_m)$ where $1 \leq x_1 \leq x_2 \leq \ldots \leq x_m \leq t$ and two treatments are said to be $i^{th}$ associates if they differ in exactly $i$ components where $i = 1, 2, \ldots m$. It can be verified that for $m = 2$, the hypercubic association scheme reduces to the $L_2$ association scheme. For $m = 3$ one gets the cubic association scheme.
Chang (1989) has constructed some hypercubic designs from symmetrical factorial experiments. In the present work an attempt has been made to discuss some relatively easy methods for constructing hypercubic designs for $t = 2$ and $t = 3$ from symmetrical factorial experiments with $m$ factors and $t$ levels for each factor. Among the two methods of construction, the first method can be used for constructing hypercubic designs for $t = 2$ and $t = 3$, while the second method is used for constructing hypercubic designs for $t = 3$ only. At the end of this chapter tables of hypercubic designs constructed by the two methods have been listed.

1.1.2 Hypercubic Designs and E-optimality

In the early stages of experimentation, only a few experimental designs are at once disposal. But, now-a-days, an experimenter with modern statistical equipments at his disposal, can use a variety of experimental designs as various complete and incomplete block designs, designs with variable replications, factorial designs and others. Hence, whenever the conditions of an experiment allow the existence of a number of experimental designs, the question of selection of an appropriate design, a design which is easy to analyze and which has some optimum properties, naturally arise. A systematic study of the specifications of optimum experimental designs was undertaken by Kiefer (1958, 1959) in a series of papers, where he introduced various optimality criteria viz., A-, D-, E-, L- and M-, discussed the interrelations amongst these and established the optimality property of some well-known designs in some particular problems.
Tocher (1952) has studied various measures of optimality in terms of the variance covariance matrix for the treatment parameters. Designs in Dr, the class of all equi-replicated designs, which minimize the maximum variance of the estimates of all normalized estimable functions of the treatment effects are chosen, to satisfy the optimality criterion. Such designs have C-matrices whose minimum nonzero eigenvalues have maximum size. John and Mitchel (1977) defined the class of Regular Graph Designs (RGD), conjectured to be optimal among all incomplete block designs. Takeuchi (1961), Conniffee and Stone (1975), Shah, Raghavarao and Khatri (1976), Williams, Patterson and John (1976, 1977) and Cheng (1978) have obtained results, which imply the optimality of certain types of Regular Graph Designs, under various criteria. Cheng (1980) investigated the E-optimality of some block designs and established some conditions under which there is a Regular Graph Design. Further Cheng (1980) shown that a group divisible design with $\lambda_1 = \lambda_2 + 1 > 1$ and group size 2 is E-optimal. Jacroux (1980) investigated the E-optimality of Regular Graph Designs, within various classes of proper block designs and several sufficient conditions are given for the existence of an E-optimal RG design, within the classes considered. Duthie (1991) examined the E-optimality of designs, which have the cubic association scheme. Shah (1995) studied the E-optimality of cubic designs whose concurrence matrix $N_dN_d^t$ has all diagonal elements equal and all of its off diagonal elements differing by 0 or 1 or 2.
In this present work, we examine the E-optimality of designs which have hypercubic association scheme and which are Regular Graph Designs. We have shown that dual of such designs also yields E-optimal RG designs. Some sufficient conditions are given for the existence of an E-optimal RG design within the class considered.

A paper based on this chapter entitled “Hypercubic Designs and E-optimality” has been published in “Journal of Statistical Research” (2000), Volume 34, No.1. PP 77 – 81; co-authored with Thannippara, Ghosh and Bagui.

1.1.3 Applications of Hypercubic Designs

In this investigation we have shown that some Semi-Regular Group Divisible designs (SRGD designs) are possible to construct from certain types of Hypercubic designs. However, the class of Semi Regular Group Divisible Designs is known, which can be considered as an application of Hypercubic Designs. The Semi Regular Group Divisible designs we constructed from hypercubic designs with (i) $v = 2^2$, $k = 2$ and (ii) $v = 3^2$, $k = 3$ are already reported in Clatworthy (1973) as SR 1 and SR 23.

Another application of hypercubic design in our investigation is the construction of complete, partial and balanced confounded experiments for symmetrical factorial experiments and then to identify the confounded main effects and interaction effects.

Das and Giri (1986) discussed how to identify a confounded interaction. Kane (1994) has obtained a simple method for identifying
confounded interactions in \(2^n(n = 2, 3)\) factorial experiments when (i) there are only two blocks per replication and (ii) there are more than two blocks per replication. Ghosh and Bagui (1998) added some more conditions to Kane’s (1994) method to identify the confounded interaction effects and they have proposed a new way of identifying the confounded main effects and interaction effects in \(2^n\) and \(3^n\) factorial experiments.

In the present work we have shown that Hypercubic Designs discussed in Section 3.2 of chapter 3 provide some confounding systems for symmetrical factorial experiments. Identification of confounded main effects and interaction effects is discussed using counter-examples.

Cheng (1989) has discussed to obtain partial and balanced confounding factorial experiments only. While in our investigation one can obtain balanced, partial or complete confounding factorial experiment, which depends on the way in which the blocks of the hypercubic design are taken.

A paper based on this chapter and Chapter 3 entitled “Hypercubic designs and applications” is communicated for publication in “Statistical Papers”.

1.1.4 \(\lambda\)-linked Block Designs through Non-adaptive Hypergeometric Group Testing Designs for identifying at most two defectives and their optimalities

A general statement of group testing problem is as follows. Assume that \(n\) varieties \(x_1, x_2, \ldots, x_n\) are given and some of them are
defective or significant. The problem is to determine which of the varieties are defective by testing several factor groups, that is subsets of the set $X = \{x_1, x_2, \ldots, x_n\}$ in a series of $v^*$ or $t$ tests.

The technique of group testing was originally proposed by Dorfman (1943) in the context of blood testing. The group testing problem has been studied by several authors, viz., Sobel and Groll (1959, 1966), Hwang (1978), Hwang and Sos (1981), Wiedeman (1984) and Wiedeman and Raghavarao (1987 a, b). Dorfman's (1943) original procedure has been modified and extended to other screening situations. There are a variety of different fields in which the application of group testing appears. Sobel and Groll (1959) have mentioned many industrial applications of group testing. If the number of defectives $d$ is known prior to testing, a non-adaptive hypergeometric group testing problems arise. This problem was introduced by Hwang and Sos (1981). Wiedeman and Raghavarao introduced a new method of approaching this problem, which is to maximize the number of items that can be tested in a given number of tests. Wiedeman and Raghavarao (1987,b) considered the problem of constructing group testing designs with $n$ items that can identify at most two defectives by performing $v^*$ tests through the dual formation of the designs. Ghosh and Thannippara (1991) discussed some more such designs for hypercubic designs and BIB designs.

Bose (1963, 1975, 1976) studied the problem of designs and multi-graphs. Bose had given the definitions of $\lambda$-linked block designs, block multi-graph designs, treatment multi-graph designs
and geometry designs and their constructions. Jauroux (1985) discussed the Type I optimality criterion for a design.

In this present work we have shown that the non-adaptive hypergeometric group testing designs for identifying at most two defectives for \( v^* = 0 \pmod{6} \) and \( v^* = 2 \pmod{6} \) are \( \lambda \)-linked block designs and at the same time semi-regular graph designs. Here we have also shown that such designs are Type I optimal designs.

A paper based on this chapter entitled "\( \lambda \)-linked block designs through non-adaptive hypergeometric group testing designs for identifying at most two defectives and their optimalities" has been published in "Journal of Applied Statistical Science" (2002), Vol.11, No.1, PP.1-8, co-authored with Thannippara, Ghosh and Bagui.