ADDIS ABABA UNIVERSITY

OFFICE OF GRADUATE PROGRAMS

COLLEGE OF NATURAL SCIENCES

DEPARTEMENT OF STATISTICS

OPTIMAL INCOMPLETE BLOCK DESIGNS FOR COMPLETE DIALLEL CROSS SYSTEM IV

By

Birhanu W/Yohanness

A Thesis submitted to the Office of Graduate Programs of Addis Ababa University in Partial fulfillment of the requirement for the Degree of Master of Science in Statistics

JUNE 2010

ADDIS ABABA UNIVERSITY

OFFICE OF GRADUATE PROGRAMS

COLLEGE OF NATURAL SCIENCES

DEPARTEMENT OF STATISTICS

OPTIMAL INCOMPLETE BLOCK DESIGNS FOR COMPLETE DIALLEL CROSS SYSTEM IV

By

Birhanu W/Yohanness

Approved by the board of examiners:

_____	_____
Department Head	**signature**
_____	_____
Examiner	**signature**
_____	_____
Examiner	**signature**

Tel. 25-1-11-23-94-65. **P.O. Box 1176 Addis Ababa, Ethiopia.**

Declaration

I hereby declare that this thesis was submitted by me to the Office of Graduate Programs of Addis Ababa University is my original work and has not previously in its entirety or part has been submitted to any other University. All sources of materials used for the study have been duly acknowledged.

Declared by

Name:-Birhanu Wolde Yohanness

Signature:-_____

Date:-26-07-2010

Confirmed by the Advisor

Name:-Dr.M.K .Sharma

Signature:-_____

Date:-26-07-2010

Acknowledgements

At the moment I am writing these lines, feeling like going back in time and remember all the steps I had to undertake from the very beginning of my graduate studies in Statistics Department at the Addis Ababa University, to the very end of this thesis. During all this time I have received valuable support and advice from many people – colleagues, friends, and most importantly my advisor – as well as the permanent encouragement I received from my family. Here I would like to acknowledge those, whose assistance and contribution was decisive to the realization of this thesis. I am very grateful to all of them

My deepest appreciation goes to Dr. M.K Sharma for his patience and perseverance. He helped in bringing out the intuition and motivation behind my research. He showed me how to break away from the cobwebs of details and see the bigger picture. He helped me tremendously in ensuring that the document has become clear, understandable and well presented. Most importantly, I really appreciate all the support and encouragement he provided to help me finish my thesis on time. Many thanks to him for his understanding and for his generosity in keeping his door open all the time.

I would like to thank Ato Mekonen Taddesse for being so generous with time spent to read my work and for the valuable suggestion he made.

I would also like to express my gratitude to Almaz Mesert for her unconditional support, patience, and for staying close to me in the difficult moments.

I would also like to take this opportunity to express my gratitude to Legesse Abate, Zemene Yohannes and Tamiru Ayana for their help.

I would also like to thank my present employer, Central Statistical Agency (CSA), for giving me the desirable freedom to complete my thesis and allowing me to leverage the computational resources for my thesis.

Finally, I would like to express my deep gratitude to my parents and brothers, who in spite of geographical distance made me feel them close to me all the time, and to my aunt Zewde Habte Mariam for her moral support, understanding and love. And, I cannot end these lines without dedicating a very special thought to the memory of my father and mother. My father is not dead. He has gone to heaven. Mami and Dad! This is dedicated to you.

ABSTRACT

In this study we used triangular partially balanced incomplete block design with two associate classes to construct optimal incomplete block designs for complete diallel cross system IV. A method is proposed to analyze Mating-Environment designs for estimating general combining ability and specific combining ability effects of lines. The designs derived are all variance-balanced in the sense that the variances of the best linear unbiased estimators of elementary contrasts among general combining ability effects are all the same. In this study we report some triangular designs which are found optimal in the sense of Kempthorne (1956). These designs are not listed by Dey and Midha (1996), Parsad and Gupta (2004) and Sharma and Fanta (2010). The analysis of these designs is presented with help of a numerical data.

TABLE OF CONTENTS

Declaration i

Acknowledgements ii

Abstract iv

List of Tables vii

List of Acronyms vii

Chapter one

1. Introduction 1
   - 1.1 General 1
   - 1.2 Motivation/statement of the problem 2
   - 1.3 Objectives of the study 3
   - 1.4 Organization of the document 4

Chapter two

2. LITERATURE REVIEW 5
   - 2.1 Why use experimental designs 5
   - 2.2 The general objective of genetical experiments 5
     - 2.2.1 Comparative diallel experiments 5
     - 2.2.2 Exploratory diallel experiments 6
   - 2.3 Achievement of objectives 7
     - 2.3.1 Mating designs 7
       - 2.3.1.1 Diallel and multi- cross designs 8
     - 2.3.2 Environmental designs 10
       - 2.3.2.1 Incomplete Block Designs 10
         - 2.3.2.1.1 Balanced Incomplete Block Designs 11
         - 2.3.2.1.2 Partially Balanced Incomplete Block Designs 11
           - 2.3.2.1.2.1 Triangular partially balanced incomplete block designs 12
   - 2.4 Analysis of incomplete block design intrablock analysis 13

2.5 Efficiency factor of an incomplete block design 21
2.5.1 Definition of efficiency factor 22
2.6 Optimal designs 26
2.6.1 Information function 26
2.6.2 Optimality criteria 27
2.6.3 Optimal symmetric designs 28
2.7 Combining ability 29
2.8 Variance-balanced designs 31

Chapter three

3. Construction and analysis of optimal complete diallel cross system IV 32
3.1 Method of construction 32
3.2 Analysis and optimality 35
3.2.1 Model and estimation of cross effects 35
3.2.2 Estimation of gca and sca 36
3.2.3 Variance-balance design for CDC system IV 38
3.2.4 Efficiency factor and optimality 38
3.2.5 List of optimal designs for diallel crosses from catalogue of Clatworthy (1973) 39

Chapter four

4. Application 41

Chapter five

5. Discussion, conclusion and recommendation 46
5.1 Discussion 46
5.2 Conclusion 47
5.3 Recommendation 47

References 48

Appendices 54

LIST OF TABLES

Table 2.1. T| B - ANOVA for incomplete block designs.....20

Table 2.2. B|T - ANOVA for incomplete block designs 20

Table 3.1. Analysis of variance for the proposed design.....37

Table 3.2 . Optimal designs for diallel crosses 39

Table 4.1. Design and observations for Example 1 41

Table 4.2. Analysis of variance for the data……. 43

Table 4.3. The estimate of gca and their estimated standard error 43

Table 4.4. The estimate of sca effects and their estimated standard error 44

LIST OF ACRONYMS

ANOVA	**Analysis of Variance**
AD	**Anthesis Date**
BIB	**Balanced Incomplete Block**
CDC	**Complete Diallel Cross**
CRD	**Completely Randomized Design**
gca	**General Combining Ability**
GDPBIB	**Group Divisible Partially Balanced Incomplete Block**
IBD	**Incomplete Block Design**
M-E	**Mating–Environment**
NBB	**Nested Balanced Block**
NBIB	**Nested Balanced Incomplete Block**
NE	**Normal Equations**
NRE	**Normal Reduced Equations**
PBIB	**Partially Balanced Incomplete Block**
RCBD	**Randomized Complete Block Design**
sca	**Specific Combining Ability**

Chapter one

1. INTRODUCTION

1.1 General

Various forms of diallel mating designs (for a description see, e.g. Hinkelmann, 1977) are commonly used in plant and animal breeding to investigate the genetic properties and potentials of inbred lines or individuals (for application in plant breeding see, e.g. Hallauer and Miranda, 1988). These mating designs consist of all possible crosses or a subset of all crosses among a set of inbred lines or individuals. This, however, is only one component of the actual experiment since the offspring of these crosses have to be reared (grown) by using what we shall refer to as an environment design (see, e.g. Hinkelmann, 1975). In other words, the offspring have to be grown in the "field" in order to obtain observations, which can be analyzed.

Most commonly, diallel cross experiments have been evaluated using completely randomized designs or randomized complete block designs as the environment design. In most cases, however, the number of crosses is too large leading to, an overall inefficient experiment. It is for this reason that the use of incomplete block designs as environment designs has been advocated by, for example, Braten (1965), Aggrawal (1974), Ceranka and Kielezewska (1986), Ceranka and Mejza (1988), Agarwal and Das (1987, 1990), Divecha and Ghosh (1994) and Sharma (1996).

To increase the precision of experiments for variety selection, Yates (1936) introduced Balanced Incomplete Block designs. As these designs consume too many resources, owing to the large number of replications of such designs, Bose and Nair (1939) came out with the notion of partially balanced incomplete block designs that are available for small numbers of replications for almost all numbers of varieties. In these designs, v varieties are arranged in b blocks, each of size k units such that each variety occurs in r blocks.

We shall be concerned here with PBIB designs for diallel crosses referred to by Griffing (1956) as experimental method IV. Among the four types of diallel crosses discussed by Griffing, method IV is the most commonly used diallel in plant breeding. Specifically, we have p lines, denoted by $i = 1, 2, \ldots, p$. We then consider crosses of the form ixj with $i < j$.

With all possible p(p-1)/2 crosses this is sometimes also referred to as the modified or half diallel. We shall refer to it as a complete diallel cross (CDC).

1.2 Motivation/ Statement of problem

The diallel cross is a type of mating design used in plant breeding and animal breeding to study the genetic properties and potential of inbred lines or individuals. Suppose there are p inbred lines and it is desired to perform a diallel crossing experiment involving p(p-1)/2 crosses of the type (ixj) for $i < j$, $i, j = 1,2, \ldots, p$. This is the Type IV mating design of Griffing (1956), who studied the detailed analysis of such mating designs laid out in a randomized complete block design. The number of crosses in such a mating design increases rapidly with increase in the number of lines, p; for $p = 5$, there are only 10 crosses while, with $p = 10$ lines, the number of crosses is 45. If there is heterogeneity in the experimental material in one direction, adoption of a randomised complete block design with crosses as treatments would result in a large error variance, even with a moderate number of inbred lines.

In order to control the error, one would therefore look for an appropriate incomplete block design in preference to a complete block design. One possibility is to use available incomplete block designs, for instance a balanced incomplete block design, for the experiment, identifying crosses with treatments of the incomplete block design. Such an approach has been followed, for example by Ceranka and Mejza (1988). However, this approach is not quite appropriate if one is interested in statistical properties like optimality of the design. Even an optimal incomplete block design may turn out to be of poor efficiency when used for a diallel cross experiment. This is because the analysis of a diallel cross experiment in incomplete blocks depends on the incidence of line blocks, rather than on that of treatments, or crosses, in blocks. It is therefore apparent that special techniques are required to obtain ‘good ’design for diallel cross experiments.

Gupta and Kageyama (1994), Dey and Midha (1996), Mukerjee (1997), Das *et al.(*1998), Parsad *et al.*(1999), Sharma (2004) addressed the problem of finding optimal designs by using nested incomplete block designs (NBIB), triangular partially balanced incomplete

block design (PBIB), nested balanced block (NBB) designs, GD PBIB designs and circular designs, etc. Gupta and Kageyama (1994) and Sharma (2004) reported optimal designs in which every cross is replicated once but their designs differ in their parametric values with proposed designs. Das *et al*. (1998) and Parsad *et al.* (1999) reported optimal designs for a single as well as more replications. Dey and Midha (1996) reported optimal and efficient designs in which the crosses are replicated in the range $3 \leq r \leq 10$.

These designs also differ in parametric values of the design proposed in this study. We derive additional incomplete block designs for the same mating design, using triangular partially balanced incomplete block designs with two associate classes such that one of the λ's is zero.

1.3 The objectives of the study

The main objective of the study is to construct optimal incomplete block designs for complete diallel cross system IV using triangular partially balanced incomplete block designs with two associate classes.

The specific objectives of the study are:

- To present a method to analyze the proposed designs.
- To see if the proposed designs are all variance-balanced in the sense that the variances of the best linear unbiased estimators of elementary contrasts among the general combining ability effects are all the same.
- To see if the designs derived from triangular designs are optimal in sense of Kempthorne (1956).
- To derive efficiency factor for proposed design as compared to randomised block design having the same number of crosses.
- To report some optimal designs for diallel crosses.

1.4 Organization of the document

This chapter provided an overview of the research undertaken in this thesis. The remainder of the document is organized as follows. Chapter 2 summarizes research in the area of environmental and mating designs, analysis of incomplete designs, efficiency factor of incomplete block designs, optimal designs combining ability and variance-balanced design. Chapter 3 describes how we construct the optimal incomplete block design for complete diallel cross and presents two stage procedures for estimating general combining ability and specific combining ability effects for proposed design. Chapter 4 shows the essential steps of analysis of diallel cross experiment using an incomplete block design proposed in this study and Chapter 5 concludes the discussion by highlighting the salient contribution of the research making recommendation for future research.

Chapter two

2. LITERATURE REVIEW

This chapter reviews the literature in the fields that are relevant to this research. It touches on all topics right from the use of experimental designs to variance-balanced design.

2.1 Why use experimental designs?

Experimental designs are used so that the treatments may be assigned in an organized manner to allow valid statistical analysis to be carried out on the resulting data. Different designs identify different known or suspected sources of variation so that the treatments effects can be evaluated free of extraneous environmental or other influences. Statistical theory also requires that certain conditions be met during the execution of the experiment to permit valid probability statements to be made (Sharma, 2000).

2.2 The General objectives of genetical experiments

In order to discuss what type of inference can be made from diallel experiment, we shall classify experiments in quantitative genetics into two groups:
(i) Comparative diallel experiments, (ii) Exploratory diallel experiment (Hinkelmann, 1975)

2.2.1 Comparative diallel experiments

In this case we usually wish to make comparisons of some sort among the lines included in the experiment, such as comparisons among their average performance in crosses, or comparisons among the crosses themselves. This is fairly straight-forward and concepts of quantitative genetics do not really enter at this point except such statistical-genetical concepts as general and specific combining ability, concepts very much related to main effects and interactions in factorial experiments (Hinkelmann, 1975). We shall illustrate this in statistical term for the simplest situation. Suppose we consider the cross between individuals i (male) and j (female) and observe the performance (with regard to some quantitative trait) of their offspring. For the k^{th} offspring this observed value will be denoted by y_{ijk} and since it is a phenotypic value, it is composed additively of two parts i.e. a genetic component and an environmental component (for the present time we shall ignore the possible interaction

between genotype and environment). The genetical component, in turn is broken down into four parts: a contribution due to (i) male i, s_i (ii) female j, d_j (iii) interaction between i and j, $(sd)_{ij}$ and (iv) random component associated with k^{th} offspring of $(i, j)^{th}$ cross. We thus have (see Hinkelmann, 1975).

$$y_{ijk} = \mu + s_i + d_j + (sd)_{ij} + e_{ijk}$$

where e_{ijk} is the environmental effects, s_i and d_j are also known as the general combining ability of the i^{th} male and j^{th} female, respectively, and $(sd)_{ij}$ as specific combining ability associated with these two individuals. One is then interested in comparisons of the form $s_i - s_{i'}$ or $d_j - d_{j'}$.

2.2.2 Exploratory diallel experiments

In this case we generally wish to make much wider inferences and obtain genetic information of a different type, namely information about various types of gene action like additive effects, dominance deviations, and different epistatic effects (Hinkelmann, 1975). The main difference between a comparative and an exploratory experiment can be best explained in term of the underlying linear models. The genetical components of a comparative model are considered to be fixed effects whereas those for an exploratory experiment are considered to be random effects. In other words, the inferences to be drawn from those two types of experiments are quite different and aimed at answering different questions. An exploratory experiment is design to answer question such as: what is the genetic structure of a population; i.e. what type of gene action are present in the population; or. Is there any interaction between genotypes and environment present and if so, what type this interaction with regard to this question, the example discussed under comparative diallel experiments can be used as an illustration. Now i.e. s_i, d_j and $(sd)_{ij}$ are considered to be independent random variables with zero means and variance / covariance σ_s^2, σ_d^2 and σ_{sd}^2 respectively. In order to obtain genetic information from such model, variance components, one must express them in terms of genetic variance components and then estimate and interpret Hinkelmann, 1975).

2.3 Achievement of objectives

Having defined, in very general terms, objectives of genetical experiments, the next question is: How can these objectives be achieved? This brings us to the actual design aspects for genetical experiments, which means finding ways to "generate" data that are appropriate for answering relevant questions of a comparative or exploratory nature (Hinkelmann, 1975). There are two design aspects involved in this problem of generating data:
(i) the mating design, (ii) the environmental design.

2.3.1 The mating designs

The "treatments" in genetical experiments are genetic entities or, more precisely, crosses among genetic entities. Inferences about these entities are made by evaluating the offspring from these crosses (matings). More generally, offspring of various degrees are used in genetical experiments. The offspring are obtained by following a certain propagation scheme, which shall be referred to as mating design. "Simple" mating designs usually allow one to investigate a limited number of questions or hypotheses, relying furthermore on a number of assumptions. There is, therefore, a need for more elaborate and specific mating designs. The choice of the mating is sometimes severely limited by the genetic material one is working with and by practical considerations. Whereas complicated mating designs can be handled for many plants and small animals, they become impractical or even impossible for large animals. The limitations and other difficulties, usually beyond the control of the experimenter, may not always enable us to use the mating design that fully achieves the objectives of the experiments: to estimate the desired comparisons or to estimate genetic variance components such that they are unconfounded with other variance components (Hinkelmann, 1975).

2.3.1.1 Diallel and multi-cross designs

Basic to experiments in quantitative genetics is the underlying mating design as it determines what kinds of genetic information can be obtained, at least theoretically, from such an experiment (e.g. Hinkelmann, 1975). Among the mating designs that have been used most often are diallel crosses. The concepts of diallel were introduced by Schimidt to denote all possible crosses among a collection of male and female animals. More generally now, a diallel cross refers to a set of all possible crosses among a collection of genetic entities, such as (monoecious or dioecious) individuals or (inbred) lines.

The various types of diallel crosses can be classified as follows:

1) Diallel mating design type I (Hinkelmann and Stern, 1960), also referred to as factorial mating design (Cockerham, 1963) or North Carolina design II (Comstock and Robinson, 1952), Consists of all crosses among "a" males and "d" females;
2) Diallel mating designs type II involving p monoecious individuals or (inbred) lines;
3) Partial diallel crosses of types I and II (Hinkelmann, 1966, Curnow, 1963);
4) Two-level diallel crosses (Hinkelmann, 1974); and
5) Variations of the pure forms mentioned above (e.g. top cross designs)

With regard to multi-cross experiments, the concept of the type II diallel cross has been extended to the

1) Three-way mating design consisting of all p(p-1)(p-2)/6 crosses among p lines (Rawlings and Cockergam, 1962);
2) Four-way mating design consisting of all p(p-1)(p-2)(p-3)/12 double crosses among p lines (Rawlings and Cockerham, 1962) ; and
3) Partial three-way and four-way mating designs (Hinkelmann, 1965, 1968).

Diallel crossing is a very useful technique for conducting plant and animal breeding experiments, especially for estimating combining ability effects of lines. It is a useful technique to identify higher yielding combinations and to compare combining abilities of the parents and compute appropriate standard errors of difference between effects.

The modern use of the method started with the concepts of general and specific combining ability given by Sprague and Tatum (1942). The general combining ability (gca) is the

average performance of a line in hybrid combination, while specific combining ability (sca) is the interaction between the i^{th} and j^{th} line.

The estimation of general and specific combining ability is important in plant and animal breeding. It is especially useful in connection with testing procedures in which it is desired to study and compare the performance of the lines in hybrid combination. A complete diallel cross (CDC) system means one in which a set of p inbred lines is chosen and all possible crosses among p lines are considered. It is a widely used method in plant-breeding experiments. Depending on whether or not the parental lines/ reciprocal cross are included, or both, the CDC system can be classified into four types (Griffing, 1956):

- **CDC system I**-parents, one set of F1 terms, reciprocal F1 terms; total of p^2 Combinations;
- **CDC system II**- parents, one set of F1 terms; total of p(p+1)/2 combinations;
- **CDC system III**- one set F1 terms, reciprocal F1 terms; total of p(p-1)/2 combinations;
- **CDC system IV**- one set of F1 terms; total p(p-1)/2 combinations.

Each system necessitates a different form of analysis, when experimented in a randomized block design or a completely randomized design, taking treatments as crosses (genotypes). A detailed analysis of all CDC systems in a randomized block design is discussed by Griffing (1956). Incomplete block designs were suggested to tackle problems arising from larger block sizes. We shall be concerned here with appropriate incomplete block design for the diallel referred to by Griffing (1956) as experimental method IV. Among the four types of diallels discussed by Griffing (1956), method IV is the most commonly used in plant-breeding.

2.3.2 The environmental designs

Just as a factorial experiment has to be embedded in an experimental design, a mating design represents only a part of a genetical experiment. The genetic crosses, as specified by mating design, can be observed only under certain environmental conditions. That is to say one has to superimpose on the mating design an environmental design. To assure that proper inferences can be made the environmental design must have certain properties and must be feasible from an experimental points of view. It must assure that genetic parameters can be estimated free from environmental effects. Within the framework of the combined underlying linear model and unbiased estimation, this means that the expected value of the estimator for genetic parameters must be free from parameters representing the environmental design. In many cases the environmental designs are familiar experimental designs, such as completely randomized designs, complete and incomplete block designs, split plot designs, etc. In some cases the mating designs lead quite naturally to certain known environmental designs. In other cases specific environmental design form an integral part and can hardly be separated. In other words, the mating design is drawn up with the specific environmental conditions in mind (Hinkelmann, 1975).

2.3.2.1 Incomplete Block Designs (IBD)

These designs were introduced by Yates in order to eliminate heterogeneity to a greater extent than is possible with randomized blocks and Latin squares when the number of treatments is large. The precision of the estimate of a treatment effect depends on the number of replications of the treatment - the larger the number of replications, the more is the precision. Similar is the case for the precision of estimate of the difference between two treatment effects. If a pair of treatment occurs together a large number of times in the design, the difference between these two treatment effects can be estimated with more precision. To ensure equal or nearly equal precision of comparisons of different pairs of treatment effects, the treatments are allocated to the experimental units in different blocks of equal sizes such that treatment occurs at most once in a block and it has an equal number of replications and each pair of treatments has the same or nearly the same number of replications. When the number of replications of all pairs of treatments in a design is the same, then we have an important class of designs called **Balanced Incomplete Block** (BIB) designs and when there

are unequal number of replications for different pairs of treatments, then the designs are called as **Partial Balanced Incomplete Block (**PBIB**)** designs (Sharma, 2000).

2.3.2.1.1 Balanced Incomplete Block (BIB) Designs

A BIB design is an arrangement of v treatments in b blocks each of size k (<v) such that

(i) each treatment occurs at most once in a block

(ii) each treatment occurs in exactly r blocks

(iii) each pair of treatments occurs together in exactly λ blocks.

The symbol v, b, r, k, λ are called the parameters of the design. These parameters satisfy the relations $vr = bk$ and $\lambda(v-1) = r(k-1)$. A BIB design cannot exist unless this relation satisfied (Sharma, 2000).

2.3.2.1.2 Partially Balanced Incomplete Block (PBIB) Designs

BIB designs may not fit well to many experimental situations as these designs require a large number of replications. Moreover, these designs are not available for all numbers of treatments and block sizes. To overcome these difficulties PBIB designs were introduced. In BIB designs the variance of every estimated elementary contrast among treatment effects is not the same (Sharma, 2000). The definition of PBIB designs is based on the association scheme.

Association scheme

According to (Sharma, 2000), given v treatment symbols 1, 2, …, v, a relation satisfying the following conditions is called an m-class association scheme ($m \geq 2$):

(i) Any two symbols are either 1^{st}, 2^{nd} ,…, or m^{th} associates; the relation of association being symmetric, i.e., if the symbol α is the i^{th} associate of β, then β is the i^{th} associate of α.

(ii) Each symbol α has n_i i^{th} associates, the number n_i being independent of α,

(iii) If any two symbols α and β are i^{th} associates, then the number of symbols that are j^{th} associates of α and k^{th} associate of β is p_{jk}^{i} and is independent of the pair of i^{th} associates α and β.

The numbers v, n_i and p^i_{jk} (i,j,k =1, 2,…, m) are called the parameters of the association scheme and satisfying the following relations:

$$\sum_{i=1}^{m} n_i = \text{v-1},$$

$$\sum_{k=1}^{m} p^i_{jk} = \text{nj -1, if i = j}$$

$$= n_j, \quad \text{if } i \neq j$$

$$n_i p^i_{jk} = n_j p^j_{ik}$$

Given an association scheme with m classes ($m \geq 2$) we have a PBIB design with m associate classes based on the association scheme, if the v treatment symbols can be arranged into b blocks, such that

(i) every symbol occurs at most once in block.

(ii) every symbol occurs in exactly r blocks.

(iii) if two symbols are i^{th} associates, then they occur together in λ_i blocks, the number λ_i being independent of the particular pair of i^{th} associates α and β.

The numbers v, b, r, k, λ_i (i =1, 2, …, m) are called the parameters of the design. It can be easily seen that vr = bk and $\sum_{i=1}^{m} n_i \lambda_i = r(k-1)$

Two–class association schemes and the two–associate PBIB designs have been extensively studied in literature and are simple to use (Sharma, 2000).

2.3.2.1.2.1 Triangular partially balanced incomplete block designs

According to (Sharma, 2000), a partially balanced with two associate classes is said to be triangular partially balanced incomplete block designs if the number of treatments is v = n (n-1)/2 and the associate scheme is an array of n rows and n columns with the following properties:

1) The positions in the main diagonal (running from upper left-hand to lower right-hand corner) are left blank.

2) The n (n-1)/2 positions above the main diagonal are filled by the numbers 1, 2, 3, …., n (n-1)/2 corresponding to the treatments.
3) The n (n-1)/2 positions below the main diagonal are filled so that the array is symmetrical about the main diagonal.
4) For any treatment i the first associates are exactly those treatments which lie in the same rows (or in the same columns) as i.

The following relations clearly hold:

(a) $v = n(n-1)/2$, $n_1 = 2(n-2)$, $n_2 = (n-2)(n-3)/2$

(b) $p_1 = \begin{pmatrix} n-2 & n-3 \\ n-3 & (n-3)(n-4)/2 \end{pmatrix}, p_2 = \begin{pmatrix} 4 & 2n-8 \\ 2n-8 & (n-4)(n-5)/2 \end{pmatrix}$

where $n \geq 5$. The case n = 4 has $p_{12}^2 = 0$ so that it is also of the group divisible type and is classified in this way. We note further that for n = 2, or 3, $n_2 = 0$ so that there is only one associate class and the, if it exists, degenerate to the D (1) design.

2.4 Analysis of incomplete block designs and intrablock analysis.

Notations, discussion and results used in this section are taken from (Hinkelmann and Kempthorne, 2005). The analysis of incomplete block designs is different from the analysis of complete block designs in that comparisons among treatment effects and comparisons among block effects are no longer orthogonal to each other. This is referred to usually by simply saying that treatments and blocks are not orthogonal. This nonorthogonality leads to an analysis analogous to that of the two - way classification with unequal subclass numbers. However, this is only partly true and applies only to the analysis that has come to be known as the intra block analysis

The name of the analysis is derived from the fact that contrasts in the treatment effects are estimated as linear combinations of comparisons of observations in the same block. In this way the block effects are eliminated and the estimates are functions of treatment effects and error (intra block error) only. Coupled with the theory of least squares and the Gauss-Markov theorem, this procedure will give rise to the best linear unbiased intrablock estimators for treatment comparisons. Historically, this has been the method first used for analyzing incomplete block design Yates (1936).

Based upon considerations of efficiency, Yates (1939) argued that the intra-block analysis ignores part of the information about treatment comparisons, namely that information contained in the comparison of block totals. This analysis has been called recovery of inter block information or inter block analysis. Yates (1939, 1940) showed for certain types of lattice designs and for the balanced incomplete block design how these two types of analyses can be combined to yield more efficient estimators of treatment comparisons. Nair (1944) extended these results to partially balanced incomplete block designs, and Rao (1947) gave the analysis for any incomplete block design showing the similarity between the intra block analysis and the combined intra- and inter block analysis.

The intrablock analysis, as it is usually presented, is best understood by assuming that the block effects in the underlying linear model are fixed effects. But for the recovery of inter block information the block effects are then considered to be random effects. This leads sometimes to confusion with regard to the assumptions in the combined analysis, although it should be clear from the previous remark that the block effects have to be considered random effects for both the intra- and interblock analysis. To emphasize, we can talk about intrablock analysis under the assumption of either fixed or random block effects. In the first case ordinary least squares (OLS) will lead to best linear unbiased estimators for treatment contrasts. This will, at least theoretically, not be true in the second case, which is the reason for considering the inter block information in the first place and using the Aitken equation, which is also referred to as generalized (weighted) least squares block designs.

We shall be concerned here only with the intrablock analysis for the general incomplete block designs and discuss below models and equations developed by (Hinkelmann and Kempthorne, 2005).

Suppose we have t treatments replicated $r_1, r_2, \ldots, r_t$ times, respectively, and b blocks with $k_1, k_2, \ldots, k_b$ units, respectively. We then have

$$\sum_{i=1}^{t} r_i = \sum_{j=1}^{b} k_j = n$$

where n is the total number of observations.

Following the derivation of a linear model for observations from a randomized complete block design (RCBD), using the assumption of additivity in the broad sense, an appropriate linear model for observations from an incomplete block design is

$$y_{ij\ell} = \mu + \tau_i + \beta_j + e_{ij\ell} \tag{2.1}$$

(i = 1, 2…,t; j =1, 2,…, b; ℓ = 0, 1,…, n_{ij}), where τ_i is the effect of the i^{th} treatment, β_j the effect of the j^{th} block, and $e_{ij\ell}$ the error associated with the observation $y_{ij\ell}$. As usual, $e_{ij\ell}$ contains both experimental and observational (sampling) error.

$$e_{i\ell j} = \in_{ij\ell} + \eta_{ij\ell}$$

with $\in_{ij\ell}$ representing an experimental error and $\eta_{ij\ell}$ representing observational error. We can treat the $e_{ij\ell}$ as a random error with mean zero and variance $\sigma_e^2 = \sigma_\varepsilon^2 + \sigma_\eta^2$.

Model (2.1) can also be written in matrix notation as

$$\mathbf{y} = \mu\mathbf{J} + \mathbf{X}_\tau\boldsymbol{\tau} + \mathbf{X}_\beta\boldsymbol{\beta} + \mathbf{e} \tag{2.2}$$

where **J** is a column vector consisting of *n* ones, $\mathbf{X}_\beta$ is the observation-block incidence matrix

$$\mathbf{X}_\beta = \begin{bmatrix} \mathbf{J}_{k1} & & & & \\ & \mathbf{J}_{k2} & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & \mathbf{J}_{kb} \end{bmatrix}$$

with J_{kj} denoting a coulumn vector of k_j unity elements (j = 1, 2,…, b) and

$\mathbf{X}_\tau$ is an observation-treatment incidence matrix which takes specific forms depending on the design considered, where $\mathbf{x_i}$ is a column vector with r_i unity elements and $(n\text{-}r_i)$ zero elements such that

$\mathbf{x}_i'\mathbf{x}_i = r_i$ and $\mathbf{x}'x_{i'} = 0$ for $i \neq i'(i, i' = 1,2,....,t)$.

The normal equations (NE) for μ, τ_I, and β_j are then

$$n\hat{\mu} + \sum_{i=1}^{t} r_i \hat{\tau}_i + \sum_{j}^{b} k_j \hat{\beta}_j = G$$

$$r_i\hat{\mu}+r_i\hat{\tau}_i+\sum_{j=1}^{b}n_{ij}\hat{\beta}_j=T_i \qquad (i = 1, 2,\ldots, t) \qquad (2.3)$$

$$k_j\hat{\mu}+\sum_{i=1}^{t}n_{ij}\hat{\tau}_i+k_j\hat{\beta}_j=B_j \quad (j = 1, 2,\ldots, b)$$

where

$T_i = \sum_{j\ell} y_{ij\ell}$ = i^{th} treatment total

$B_j = \sum_{i\ell} y_{ij\ell}$ = j^{th} block total

$G = \sum_i T_i = \sum_j B_j$ = over all total

Equations (2.3) can be written in matrix notation as

$$\begin{pmatrix} \mathbf{J}'_n\mathbf{J}_n & \mathbf{J}'_n\mathbf{X}_\tau & \mathbf{J}'_n\mathbf{X}_\beta \\ \mathbf{X}'_\tau\mathbf{J}_n & \mathbf{X}'_\tau\mathbf{X}_\tau & \mathbf{X}'_\tau\mathbf{X}_\beta \\ \mathbf{X}'_\beta\mathbf{J}_n & \mathbf{X}'_\beta\mathbf{X}_\tau & \mathbf{X}'_\beta\mathbf{X}_\beta \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\mu}} \\ \hat{\boldsymbol{\tau}} \\ \hat{\boldsymbol{\beta}} \end{pmatrix} = \begin{pmatrix} \mathbf{J}'_\beta\mathbf{y} \\ \mathbf{X}'_\tau\mathbf{y} \\ \mathbf{X}'_\beta\mathbf{y} \end{pmatrix} \qquad (2.4)$$

which, using the properties of $\mathbf{J}$, $\mathbf{X}_\tau$, $\mathbf{X}_\beta$, can be written as

$$\begin{bmatrix} \mathbf{J}'_n\mathbf{J}_n & \mathbf{J}'_\tau\mathbf{R} & \mathbf{J}'_b\mathbf{K} \\ \mathbf{R}\mathbf{J}_t & \mathbf{R} & \mathbf{N} \\ \mathbf{K}\mathbf{J}_b & \mathbf{N}' & \mathbf{K} \end{bmatrix} \cdot \begin{bmatrix} \hat{\boldsymbol{\mu}} \\ \hat{\boldsymbol{\tau}} \\ \hat{\boldsymbol{\beta}} \end{bmatrix} = \begin{bmatrix} \mathbf{G} \\ \mathbf{T} \\ \mathbf{B} \end{bmatrix} \qquad (2.5)$$

where $\mathbf{R} = \text{diag}(r_i)$ txt

$\mathbf{K} = \text{diag}(k_j)$ bxb

$\mathbf{N} = (n_{ij})$ txb (the incidence matrix)

$\mathbf{T}' = (T_1, T_2, \ldots, T_t)$

$\mathbf{B}' = (B_1, B_2, \ldots, B_b)$

$\boldsymbol{\tau}' = (\tau_1, \tau_2, \ldots, \tau_t)$

$\boldsymbol{\beta}' = (\beta_1, \beta_2, \ldots, \beta_b)$

and the **J**'s are column vectors of ones with dimensions indicated by the subscripts.

From the third set of equations in (2.5) we obtain

$$\hat{\mu}\mathbf{J}_b + \hat{\boldsymbol{\beta}} = \mathbf{K}^{-1}(\mathbf{B} - \mathbf{N}'\hat{\boldsymbol{\tau}}) \tag{2.6}$$

Substituting (2.6) into the second set of (2.5), which can also be expressed as

$\mathbf{NJ}_b\hat{\mu} + \mathbf{N}\hat{\boldsymbol{\beta}} + \mathbf{R}\hat{\boldsymbol{\tau}} = \mathbf{T}$ (since $\mathbf{NJ}_b = \mathbf{RJ}_t$) leads to the reduced normal equations (RNE) for $\boldsymbol{\tau}$

$$(\boldsymbol{R} - \boldsymbol{NK}^{-1}\boldsymbol{N})\hat{\boldsymbol{\tau}} = \boldsymbol{T} - \boldsymbol{NK}^{-1}\boldsymbol{B} \tag{2.7}$$

The standard notation for (2.7) is

$$\mathbf{C}\hat{\boldsymbol{\tau}} = \mathbf{Q} \tag{2.8}$$

where $$\mathbf{C} = \mathbf{R} - \mathbf{NK}^{-1}\mathbf{N}' \tag{2.9}$$

$$\mathbf{Q} = \mathbf{T} - \mathbf{NK}^{-1}\mathbf{B} \tag{2.10}$$

According to Hinkelmann and Kempthorne (2005) the C matrix and estimable functions can be presented in the following ways.

We note that the matrix **C** of (2.9) is determined entirely by the specific design, that is, by the incidence matrix **N**. It is, therefore, referred to as the **C** matrix (sometimes also as the information matrix) of that design. The **C** matrix is symmetric, and the elements in any row or any column of **C** add to zero, that is, $\mathbf{CJ} = \mathbf{0}$, which implies that rank(**C**) $\leq$ t − 1. Therefore, **C** does not have an inverse and hence (2.8) cannot be solved uniquely. Instead we write a solution to (2.8) as

$$\hat{\boldsymbol{\tau}} = \mathbf{C}^{-}\mathbf{Q} \tag{2.11}$$

where $\mathbf{C}^{-}$ is a generalized inverse for **C.** If we write $\boldsymbol{C} = (\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_t)$, where $\boldsymbol{c}_i$ is the i^{th} column of $\boldsymbol{C}$, then the set of linear functions

$$\mathbf{c}_i'\boldsymbol{\tau} \qquad \text{where i = 1, 2,…, t}$$

which span the totality of estimable functions of the treatment effects, has dimension $r(\boldsymbol{C})$. Let $\mathbf{c}'\boldsymbol{\tau}$ be an estimable function and $\mathbf{c}'\hat{\boldsymbol{\tau}}$ its estimator, with $\hat{\boldsymbol{\tau}}$ from (2.11).Then

$$\begin{aligned} E(\mathbf{C}'\hat{\boldsymbol{\tau}}) &= E(\mathbf{c}'\mathbf{C}^{-}\mathbf{Q}) \\ &= \mathbf{c}'\mathbf{C}^{-}E(\mathbf{Q}) \\ &= \mathbf{c}'\mathbf{C}^{-}\mathbf{C}\boldsymbol{\tau} \end{aligned}$$

For $\mathbf{c}'\hat{\boldsymbol{\tau}}$ to be an unbiased estimator for $\mathbf{c}'\boldsymbol{\tau}$ for any $\boldsymbol{\tau}$, we then must have

$$\mathbf{c}'\mathbf{C}^{-}\mathbf{C} = \mathbf{c}' \tag{2.12}$$

Since $\mathbf{CJ} = \mathbf{0}$, it follows from (2.12) that $\mathbf{C}'\mathbf{J} = \mathbf{0}$. Hence, only treatment contrasts are estimable. If $r(\boldsymbol{C}) = t - 1$, then all treatment contrasts are estimable. In particular, all differences $\boldsymbol{\tau}_i - \boldsymbol{\tau}_{i'}$ $(i \neq i')$ are estimable, there being $t - 1$ linearly independent estimable functions of this type. Then the design is called a connected design

Following work done by (Hinkelmann and Kempthorne, 2005) to solve the reduced normal equations in what follows, we shall assume that the design is connected; that is, r(**C**) = t − 1. This means that **C** has t − 1 nonzero (positive) eigenvalues and one zero eigenvalue.

From $$\mathbf{C}\begin{bmatrix}1\\1\\.\\.\\.\\1\end{bmatrix} = \mathbf{0} = 0\begin{bmatrix}1\\1\\.\\.\\.\\1\end{bmatrix}$$

it follows then that $(1, 1, \ldots, 1)'$ is an eigenvector corresponding to the zero eigenvalue. If we denote the nonzero eigenvalues of $\boldsymbol{C}$ by $d_1, d_2, \ldots, d_{t-1}$ and the corresponding eigenvectors by $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{t-1}$ with $\mathbf{e}_i'\mathbf{e}_i = 1$ $(i = 1, 2, \ldots, t - 1)$ and $\mathbf{e}_i'\mathbf{e}_{i'} = 0$ $(i \neq i')$, then we can write $\boldsymbol{C}$ in its spectral decomposition

$$\mathbf{C} = \sum_{i=1}^{t-1} d_i \mathbf{e}_i \mathbf{e}_i' \tag{2.13}$$

or with $d_t = 0$ and $\mathbf{e}_t = \frac{1}{\sqrt{t}}(1, 1, \ldots, 1)$ alternatively as

$$\mathbf{C} = \sum_{i=1}^{t} d_i \mathbf{e}_i \mathbf{e}_i' \tag{2.14}$$

We note that $\mathbf{e}_t' \mathbf{e}_t = 1$ and $\mathbf{e}_i' \mathbf{e}_t = 0$ for i =1, 2, …, t-1.

We now return to (2.8) and consider a solution to these equations of the form given by (2.11). Although there are many methods of finding generalized inverses, we shall consider here one particular method, which is most useful in connection with incomplete block designs, especially balanced and partially balanced incomplete block designs.
This method is based on the following theorem, which is essentially due to (Shah, 1959).

Theorem 2.1 Let **C** be a t × t matrix as given by (2.9) with r(**C**) = t − 1.

Then $\tilde{\mathbf{C}} = \mathbf{C} + a\mathbf{J}\mathbf{J}'$, where $a \neq 0$ is a real number, admits an inverse $\tilde{\mathbf{C}}^{-1}$, and $\tilde{\mathbf{C}}^{-1}$ is a generalized inverse for **C**.
We give the theorem with out proof.

Substituting $\tilde{\mathbf{C}}^{-1}$ into (2.13) yields a solution of the RNE (2.8); that is,

$$\hat{\boldsymbol{\tau}} = \tilde{\mathbf{C}}^{-1}\mathbf{Q} \tag{2.15}$$

We know from the Gauss–Markov theorem that for any linear estimable function of the treatment effects, say $\mathbf{c}'\boldsymbol{\tau}$,

$$E(\mathbf{c}'\hat{\boldsymbol{\tau}}) = \mathbf{c}'\boldsymbol{\tau} \tag{2.16}$$

is independent of the solution to the NE.
We have further

$$\text{var}(\mathbf{c}'\hat{\boldsymbol{\tau}}) = \mathbf{c}'\,\tilde{\mathbf{C}}^{-1}\mathbf{c}\,\sigma_e^2 \tag{2.17}$$

Hinkelmann and Kempthorn (2005) gave the two forms of analysis of variance as given in Tables 2.1 and 2.2. They referred to the analysis of variance in Table 1.1 as the treatment-after-block ANOVA or T|B - ANOVA as it is associated with the ordered model

$$\mathbf{y} = \boldsymbol{\mu}\mathbf{J} + \mathbf{X}_{\beta}\boldsymbol{\beta} + \mathbf{X}_{\tau}\boldsymbol{\tau} + \mathbf{e}$$

whereas the analysis of variance in Table 1.2 is associated with the ordered model

$$\mathbf{y} = \boldsymbol{\mu}\mathbf{J} + \mathbf{X}_{\tau}\boldsymbol{\tau} + \mathbf{X}_{\beta}\boldsymbol{\beta} + \mathbf{e}$$

and hence they referred to as the block-after-treatment ANOVA or B |T- ANOVA.

Table 2.1 T|B - ANOVA for incomplete block design

Source	Degree of freedom	Sum of squares (SS)	E(mean SS)
$\mathbf{X}_{\beta}\|\mathbf{J}$	b-1	$\sum \frac{B_j^2}{k_j} - \frac{G^2}{n}$	
$\mathbf{X}\tau \|\mathbf{J}, \mathbf{X}_{\beta}$	t-1	$\sum_{i=1}^{t} \hat{\boldsymbol{\tau}}_i Q_i$	$\sigma_e^2 + \frac{\boldsymbol{\tau}'\mathbf{C}\boldsymbol{\tau}}{\text{t - 1}} +$
$\mathbf{I}\|\mathbf{J}, \mathbf{X}_{\beta}, \mathbf{X}_{\tau}$	n-b-t+1	difference	σ_e^2
Total	n-1	$\sum_{ij\ell} y^2_{ij\ell} - \frac{G^2}{n}$	

Table 2.2 B|T-ANOVA for incomplete block design

Source	Degree of freedom	Sum of squares
$\mathbf{X}_{\tau}\|\mathbf{J}$	t-1	$\sum_{1=1}^{t} \frac{T_i^2}{ri} - \frac{G^2}{n}$
$\mathbf{X}_{\beta}\|\mathbf{J},\mathbf{X}_{\tau}$	b-1	Difference
$\mathbf{I}\|\mathbf{J},\mathbf{X}_{\beta},\mathbf{X}_{\tau}$	n-b-t+1	From Table2.1
Total	n-1	$\sum_{ij\ell} y^2_{ij\ell} - \frac{G^2}{n}$

2.5 Efficiency factor of an incomplete block design

Using the results of (Hinkelmann and Kempthorne, 2005), we can compare different error control designs with each other by using the notion of relative efficiency. In this case, we compare two error control designs after we have preformed the experiment using a particular error control design. For example, after we have used an RCBD we might ask: How would we have done with a corresponding CRD? In other cases, however, we may want to compare error control designs before we begin an experiment. In particular, we may want to compare an incomplete block design (IBD) with either a CRD or an RCBD, or we may want to compare competing IBDs with each other. For this purpose we shall use a quantity that is referred to as efficiency factor of the IBD. It compares, apart from the residual variance, σ_e^2, the average variance of simple treatment comparisons for the two competing.

Based on (Hinkelmann and Kempthorne, 2005) results, average variance for treatment comparisons for an IBD is given by

$$\underset{i\neq i'}{av}.\text{var}(\hat{\tau}_i - \hat{\tau}_{i'}) \tag{2.18}$$

where av.var stands for average variance for connected IBD. We know, of course, (2.18) is a function of $\mathbf{C}^-$, a generalized inverse of the **C** matrix. Suppose now that all the block sizes are equal to k.

Then we have $\mathbf{C} = \mathbf{R} - \frac{1}{k}\mathbf{NN}'$

and we know that **C** has one zero $d_t = 0$ say, with the associated normalized eigenvector $\mathbf{e}_t = (1/\sqrt{t})\mathbf{J}$. Let the other roots be $d_1, d_2, \ldots, d_{t-1}$ with the associated orthonormal eigenvectors $\mathbf{e}_1, e_2, \ldots, e_{t-1}$. Then

$$e'_i\,\mathbf{C} = d_i\mathbf{e}'_i \qquad (i = 1, 2\ldots., t-1)$$

and from $\mathbf{e}'_i\,C\boldsymbol{\tau} = d_i\mathbf{e}'\tau$

it follows that $\mathbf{e}_i'\hat{\boldsymbol{\tau}} = \frac{1}{d_i}\mathbf{e}'_i\mathbf{Q}$

and $\text{var}(\mathbf{e}_i'\hat{\boldsymbol{\tau}}) = \frac{1}{d_i^2}\mathbf{e}_i'C\mathbf{e}_i = \frac{1}{d_i}\sigma_e^2$ (2.19)

using the fact that $e_t = (1/\sqrt{t})\mathbf{J}$ that $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{t-1}$ are mutually perpendicular and are perpendicular to $\mathbf{e}_1$,

and that $$\sum_{i=1}^{t} e_i e_i' = \mathbf{I}$$

We have with $z' = (z_1, z_2, \ldots, z_t)$

$$\begin{aligned} & z'(\mathrm{I} - \mathbf{e}_t \mathbf{e}'_t) z \\ &= \sum_{i=1}^{t-1} (e_i' z)^2 = z' \left(\sum_{i=1}^{t-1} e_i e_i' \right) z \; \sum_{i=1}^{t} z_i^2 - \frac{1}{t} \left(\sum_{i=1}^{t} z_i \right)^2 \\ &= \sum_{i=1}^{t} (z_i - \bar{z})^2 \end{aligned} \tag{2.20}$$

It is also to verify that

$$\frac{1}{t(t-1)} \sum_{i \neq i'} (z_i - \bar{z})^2 = \frac{2}{t-1} \sum_{i=1}^{t} (z_i - \bar{z})^2 \tag{2.21}$$

Taking $z_i = \hat{\tau}_i - \tau_i$ and substituting into (2.21) using (2.20) and then taking expectation and using (2.19), yields for (2.18)

$$\underset{i \neq 1}{av} \cdot \mathrm{var}\ (\hat{\tau}_i - \hat{\tau}_{i'}) = \frac{2}{t-1} \sum_{i=1}^{t-1} \frac{1}{d_i} \sigma_e^2 \tag{2.22}$$

where av.var stands for average variance for connected IBD.

2.5.2 Definition of efficiency factor

Hinkelmann and Kempthorne (2005) evaluated the efficiency of an IBD to compare it with a CRD since this is always a possible competing design. For a CRD with r_i replications for treatment i, the average variance of treatment differences is

$$\underset{i \neq 1}{av} \cdot \mathrm{var}(\hat{\tau}_i - \hat{\tau}_i') = \underset{i \neq 1}{av} \cdot \mathrm{var} \left(\frac{1}{r_i} + \frac{1}{r_i'} \right) \sigma^2_{e(CRD)} = \frac{1}{\bar{r}_h} \sigma^2_{e(CRD)} \tag{2.23}$$

where $\bar{r}_h$ is the harmonic mean of r_i, that is, $\frac{1}{\bar{r}_h} = \frac{1}{t} \sum_i \frac{1}{r_i}$.

We shall digress here for a moment and show that the best CRD is the one with all $r_i = r$, and that is the design with which we shall compare the IBD. For this and later derivations we need the "old" result that the harmonic mean of a set of positive numbers is not greater than the arithmetic mean (see, Hinkelmann and Kempthorne, 2005).

It seems useful to give an elementary proof of this.

Let the set of numbers be $\{x_i, i = 1, 2 \ldots, m\}$ consider the quadratic form

$$q(\beta) = \sum_{i=1}^{m} \left(\sqrt{x_i} - \beta \frac{1}{\sqrt{x_i}} \right)^2$$

Clearly $q(\beta) \geq 0$ for all β. The minimum value of β is obtained by using least squares which gives the NE

$$\sum_i \frac{1}{x_i} \tilde{\beta} = m$$

The minimum sum of squares is

$$\sum_i x_i - \tilde{\beta} m$$

Hence

$$\sum_i x_i - \frac{m^2}{\sum_i \frac{1}{x_i}} \geq 0$$

or

$$\left(\frac{1}{m} \sum_i x_i \right) \quad \left(\frac{1}{m} \sum_i \frac{1}{x_i} \right) \geq 1$$

or

$$\frac{\bar{x}}{\bar{x}_h} \geq 1$$

with equality if and only if $x_i = x$ for all i.

This result implies that the best CRD will have $r_i = r$ and $r = n/t$ where n is the total numbers of experimental units. This happen, of course, only if n/t is an integer. If n/t is not an integer so that $n = pt + q$ ($o<q<t$), then the best CRD will have q treatments replicated p+1 times (Hinkelmann and Kempthorne, 2005).

Consider now the case of an IBD with b blocks of size k and r_i replications for the i^{th} treatment. Then the total number of experimental units is n = bk = $\sum r_i$. Suppose also that n = rt, so that an equireplicate CRD is possible. The average variance for such a design is $\sigma_e^{2}{}_{(CRD)} / r$, whereas the average variance for the IBD is $\sigma^2_{e(IBD)} / c$, as shown in (2.22), where c is harmonic mean of positive eigenvalues of (**R** - $\frac{1}{k}$**N N'**) (see Kempthore, 1956). It is natural to write c = rE, so that with $\sigma^2_{e(CRD)} = \sigma^2_{e(IBD)}$ and we have

$$\frac{av.\mathrm{var}(\hat{\tau}_i - \hat{\tau}_{i'})CRD}{av.\mathrm{var}(\hat{\tau}_i - \hat{\tau}_{i'})IBD} = \frac{2/r}{2/rE} = E \tag{2.24}$$

The quantity E thus defined is called the efficiency factor of the IBD. It is clearly a numerical property of the treatment-block configuration only and hence a characteristic of a given IBD (Hinkelmann and Kempthorne, 2005).

We add the following remarks (see, Hinkelmann and Kempthorne, 2005). :

1. The same definition of E in (2.24) could have been obtained by using the average variance for an CRD with b = r blocks instead of the average variance for an equireplicate CRD assuming that $\sigma^2_{e(CRD)} = \sigma^2_{e(IBD)}$.
2. Although E is a useful quantity to compare designs, it does not, of course, give the full story. It compares average variances only under the assumption of equality of residual variances, whereas we typically expect $\sigma^2_{e(IBD)} < \sigma^2_{e(CRD)}$ and $\sigma^2_{e(IBD)} < \sigma^2_{e(RCBD)}$.
3. The efficiency factor pertains only to the intrablock analysis and ignores the interblock information.
4. Each IBD will have associated with it an efficiency factor E. In order to compare two competing IBDs with the same n and with efficiency factors $E_1$and E_2, respectively, we would typically choose the one with the higher E value.

Hinkelmann and Kempthorne (2005) provided upper bound for the efficiency factor using the fact that the harmonic mean of positive numbers is not greater than the arithmetic mean. They derived the upper bound for efficiency factor as

$$(t-1)\,c \leq \sum_{i=1}^{t-1} d_i = \text{trace}\left(\boldsymbol{R} - \tfrac{1}{k}\boldsymbol{NN}'\right)$$

$$= \sum_{i=1}^{t} r_i - \frac{1}{k}\sum_{ij} n_{ij}^2$$

$$= n - \frac{1}{k}\sum_{ij} n_{ij}^2 \qquad (2.25)$$

The largest value of the right-hand side of (2.25) is obtained for the smallest value of $\sum_{ij} n_{ij}^2$. Since n_{ij} is one of the numbers 0, 1, 2,…, k, the minimum value of $\sum n_{ij}^2$ will be achieved when n of the n_{ij}'s are 1 and the remaining are zero. Since the $n_{ij}^2 = n_{ij}$ and $\sum_j n_{ij} = r_i$, it follows from (2.25) that

$$(t-1)\,c \leq \sum_i r_i - \frac{1}{k}\sum_i r_i = \frac{k-1}{k} t\bar{r}$$

Or, since c = rE,

$$E \leq \frac{(K-1)t/(t-1)k}{\bar{r}/r}$$

But since $t\bar{r} = n = tr$, we have finally

$$E \leq \frac{(k-1)t}{(t-1)k} \qquad (2.26)$$

Since for an IBD k < t, we can write further

$$E \leq \frac{(k-1)t}{(t-1)k} < 1 \qquad (2.27)$$

The upper bound given in (2.26) will be achieved for the balanced incomplete block design. Sharper upper bounds for certain classes of IBDs are given by Jacroux (1984), Paterson (1983), and Tjur (1990).

2.6 Optimal designs

Optimal designs are a class of experimental designs that are optimal with respect to some statistical criterion. In the design of experiments for estimating statistical models, optimal designs allow parameters to be estimated without bias and with minimum variance. A non-optimal design requires a greater number of experimental runs to estimate the parameters with the same precision as an optimal design. In practical terms, optimal experiments can reduce the costs of experimentation (Hinkelmann and Kempthorne, 2005).

We have argued in the previous section that in order to compare two designs, d_1 and d_2 say, we may consider their efficiency factors E_1 and E_2, respectively, and choose the design with the higher efficiency factor. In particular, if the efficiency factor of one of those designs achieves the upper bound for that class of designs, we would consider that design to be optimal in sense Kempthorne (1956). Such considerations have led to the development of the notion of optimal designs and to various criteria for optimality (Hinkelmann and Kempthorne, 2005). We shall describe briefly some of these criteria.

2.6.1 Information function

According to (Hinkelmann and Kempthorne, 2005), Initial contributions to the formal discussion of optimal designs were made by Wald (1943) and Ehrenfeld (1953). Extending their results, Kiefer (1958) provided a systematic account of different optimality criteria. These can be discussed either in terms of maximizing a suitable function of the information matrix or minimizing a corresponding function of the dispersion matrix of a maximal set of orthonormal treatment contrast estimates.

In the context of our discussion the information matrix is given by **C** in (2.9).

Let $\mathbf{P}'\boldsymbol{\tau}$ represent a set of t − 1 orthonormal contrasts of the treatment effects. Using intrablock information from a connected design d, the estimator for $\mathbf{P}'\boldsymbol{\tau}$ is given by $\mathbf{P}'\hat{\boldsymbol{\tau}}$ with $\hat{\boldsymbol{\tau}}$ from (2.15). The dispersion matrix for $\mathbf{P}'\hat{\boldsymbol{\tau}}$ is then given

$$\mathbf{Var}(\mathbf{P}'\hat{\boldsymbol{\tau}}) = \mathbf{P}'\mathbf{C}_d^{-}\mathbf{P}\sigma_e^2 = (\mathbf{P}'\mathbf{C}_d\mathbf{P})^{-1}\sigma_e^2 \tag{2.28}$$

where $\mathbf{C}_d$ is the information matrix to the specific design d used, $\mathbf{C}_d^-$ is the generalized inverse of $\mathbf{C}_d$, σ_e^2 is the residual variance and $\mathbf{P}'$ is some vector The information matrix for the design d is then defined as

$$\mathbf{C}^*_d = (\mathbf{P}'\ \mathbf{C}_d^-\mathbf{P})^{-1} \tag{2.29}$$

which shows, of course, the connection between $\mathbf{C}^*_d$ and $\mathbf{C}_d$.

An information function or optimality criterion is then a real-valued function ϕ that has the following properties (see Pukelsheim, 1993):

1. The function ϕ is a monotonic function; that is information matrix $\mathbf{C}^*$ is at least as good as another information matrix $\mathbf{D}^*$ if $\phi(\mathbf{C}^*) \geq \phi(\mathbf{D}^*)$;
2. The function ϕ is concave function, that is, $\phi[(1-\alpha)\mathbf{C}^*+\alpha\mathbf{D}^*] = (1-\alpha)\phi(\mathbf{C}^*)+\alpha\phi(\mathbf{D}^*)$ for $\alpha \in (0, 1)$;
3. The function ϕ is positively homogenous, that is, $\phi(\delta\mathbf{C}^*) = \delta\phi(\mathbf{C}^*)$.

Condition (2) says that information cannot be increased by interpolation. And condition (3) says that even if we defined the information matrix to be directly proportional to the number of observations, n, and inversely proportional to σ_e^2, that is, the information matrix is the form $(n/\sigma_e^2)\mathbf{C}^*$, we need to consider only $\mathbf{C}^*$. Let D be the set of competing designs. The problem of finding an optimal design d in D can be then be reduced to finding a design that maximizes $\phi(\mathbf{C}^*_d)$ over d in D (see Cheng, 1996). Such a design is ϕ-optimal. As indicated above, an alternative, and historically original, approach to finding an optimal design is to consider minimization of some convex and non increasing function Φ of dispersion matrices, as indicated by the relationship between (2.28) and (2.29). Accordingly, we shall then talk about a Φ-optimal design.

2.6.2 Optimality criteria

According to (Hinkelmann and Kempthorne, 2005), there are several optimality criteria, that is, several functions ϕ or Φ have been considered for studying optimal designs. These criteria

can be expressed conveniently in terms of the eigenvalues of $\mathbf{C}^*_d$ or, equivalently, the non zero eigenvalues of $\mathbf{C}_d$, say $\mu_{d1} \geq \mu_{d2} \geq \ldots \geq \mu_{d,t-1}$.

The most commonly used optimality criteria are D-, A-, and E-optimality, which maximize the following information functions:

1. D-optimality: Determinant criterion or

$$\phi(\mathbf{C}^*_d) = \prod_{i=1}^{t-1} \mu_{di}$$

2. A-optimality: Average variance criterion or

$$\phi(\mathbf{C}^*_d) = \left(\frac{1}{t-1} \sum_{i=1}^{t-1} \mu_{di}^{-1} \right)^{-1}$$

3. E-optimality: Smallest eigenvalue criterion or

$$\phi(\mathbf{C}^*_d) = \mu_{d,\,t-1}$$

In terms of the corresponding Φ function, these optimality criteria can be expressed as minimizing

$$1.\ \Phi(\mathbf{V}_d) = \det \mathbf{V}_d = \prod_{i=1}^{t-1} \mu_{di}^{-1} \tag{2.30}$$

$$2.\ \Phi(\mathbf{V}_d) = \mathrm{tr} V_d = \sum_{i=1}^{t-1} \mu_{di}^{-1} \tag{2.31}$$

$$3.\ \Phi(\mathbf{V}_d) = \text{maximum eigenvalue of } \mathbf{V_d} = \mu_{d,t-1}^{-1} \tag{2.32}$$

The statistical meaning of these criteria is that minimizing (2.30) minimizes the generalized variance of $P'\hat{\tau}$, (2.31) minimizes the average variance of the set $P'\hat{\tau}$, and (2.32) minimizes the maximum variance of a single normalized contrast.

2.6.3 Optimal symmetric designs

There exist special classes of designs for which all the nonzero eigenvalues of the information matrix $\mathbf{C}_d$ are equal. Such designs are called symmetric design*s*. Examples of symmetric designs are balanced incomplete block designs, Latin square designs, Youden squares, and so forth. The information matrix of a symmetric design is of the form a**I** +b**JJ**'

for some constants a and b, and it is referred to as a completely symmetric matrix (Hinkelmann and Kempthorne, 2005).

In general, if a design is Φ_1 optimal, it may not be Φ_2 optimal for two different optimality criteria Φ_1 and Φ_2. However, for symmetric designs Kiefer (1958) showed that they are A-, D- and E-optimal. This leads to the definition of universal optimality (Kiefer, 1975) or Kiefer optimality (Pukelsheim, 1993).

Let $B_{t,o}$ be the set of all txt nonnegative definite matrices with zero row and column sums, and let Φ be a real- valued function on $B_{t,o}$ such that

(a) Φ is concave,

(b) $(\delta \mathbf{C})$ is non increasing in the scalar $\delta \geq 0$, and

(c) Φ is invariant under each simultaneous permutation of rows and columns

A design d* is called universally optimal in D if d* minimizes designs $\Phi(C_d)$ for every Φ satisfying conditions (a), (b), and (c) (Kiefer, 1975).

To help identify universal optimal designs we have the following fundamental theorem.

Theorem2.2 (Kiefer, 1975) suppose a class $C = \{\mathbf{C}_d , d \in D\}$ of matrices in $B_{t,0}$

Contains a $\mathbf{C}_{d*}$ for which

(a) $\mathbf{C}_{d*}$ is completely symmetric, and

(b) $\text{tr } \mathbf{C}_{d*} = \max_{d \in D} \text{tr } \mathbf{C}_d$,

Then d* is universally optimal in D.

2.7 Combining ability

Combining ability of inbred lines is the ultimate factor determining future usefulness of the lines for hybrids (Hallauer and Miranda, 1988). The concept was refined by Sprague and Tatum (1942) to produce two expressions, general combining ability and specific combining ability. They called the additive portion of genotypic variance general combining ability (gca), determined by mean hybrid performance of a determined line. The non-additive portion was the specific combining ability (sca), a measure for cases where some hybrid combinations are better, or worse, than expected based on mean performance of the lines

evaluated. They defined sca as those instances in which certain hybrid combinations are either better or poorer than would be expected on the average performance of the parent inbred lines included in the crosses. Specific combining ability used to indicate the value of superior genotype combinations. General combining ability is also defined as the average performance of a line in a hybrid combination, when expressed as a deviation from the overall mean of all its crosses (Falconer, 1989).

These deviations can be positive or negative. A positive deviation can be favorable or unfavorable, depending on the trait under consideration.

Positive deviation for yield is desirable as this indicates high yielding potential. On the contrary, positive high values on ear rots and foliar disease ratings would not be desirable. Negative gca values on anthesis date (AD) are more desirable for selection of early maturing combinations. General combining ability tests are used for preliminary screening of lines from a large number of lines in a breeding program.

Lines with poor gca are discarded. gca estimates can also be used in genetic studies to identify the type of gene action governing traits of interest. A high gca estimate is indicative of additive gene action (Hallauer and Miranda, 1988). Any particular cross has an expected value, which is the sum of the general combining abilities of its two parental lines. The cross may deviate from the expected value to a greater or lesser extent and this deviation is called the specific combining ability (sca) of the two lines in combination (Falconer, 1989). sca is used to indicate the value of superior genotype combinations. The sca measurement represents the final stage in the selection of inbred lines as it identifies specific inbred combinations to use in hybrid formation (Hallauer and Miranda, 1988).

Specific combining ability estimates are also used in genetic studies to identify the type of gene action governing the traits of interest. A high sca measure indicates non-additive gene action. In addition, sca estimates can be used to determine heterotic relationships among different genotypes. As an example, if a line, A, gives a large positive sca estimate for yield, when crossed to line B, but a large negative sca estimate, when crossed to line C, line A is in the same heterotic group with line C but different group with line B. Lines from different heterotic groups which give high positive sca estimates are said to be complementary to each

other (Hallauer and Miranda, 1988). General combining ability and specific combining ability estimates are dependent on the particular set of materials (inbred lines, populations or varieties) included in the test, and therefore any new germplasm introduced in a breeding program have to be tested for gca and sca (Hallauer and Miranda, 1988).

2.8. Variance-balanced designs

A design d is said to be variance-balanced if every normalized treatment contrast is estimated with the same variance. Variance-balanced designs give results that are particularly easy to interpret and allow for simple implementation of technique such as decomposition of the treatment sum of squares via orthogonal contrasts. This fact is particularly attractive to the experimenter, as it enables one to carry out the analysis of the experiment in an extremely simple manner**.** For this reason variance-balance is itself a desirable quality in a design (Morgan and Uddin, 1995).

Chapter three

3. CONSTRUCTION AND ANALYSIS OF OPTIMAL COMPLETE DIALLEL CROSS SYSTEM IV

For enabling breeders to make use of CDC system IV in their experimentation, it is necessary to develop a method of construction and analysis.

3.1 Method of construction

Our method of construction of IBD for CDC experiment is essentially the same as that of Agarwal and Das (1987) and Dey and Midha (1996) but with some difference.

Assume that there are p inbred lines and it is desired to find an IBD design for a mating design involving p(p-1)/2 crosses. Consider a two associate triangular PBIB design d_1 as auxiliary design with parameters v = p(p-1)/2, b, r, k , b, λ_1, λ_2, n_1, n_2, p^i_{jk} (i, j, k = 1, 2) such that one of the λ's is zero. The treatments of the design d_1 are indexed by a pair (ixj) for i < j, i,j, = 1, 2,…, p. From the design d_1, we derive a block design, say, d_2 by replacing a treatment in d_1 by the cross (ixj) for i < j, i, j = 1, 2, …, p. Clearly d_2 involves p (p-1)/2 crosses, b blocks each of size k. Each cross is replicated r times. Now in d_2 we consider block size k as number of blocks and the number of blocks b as block size. This make our design optimal because every line occurs in a block equal number of times. Then we obtain design d_3 for CDC experiment with parameters v = p (p-1)/2, b′ = k, k′ = b, r. We shall refer the derived design d_3 as Mating-Environment, or simply the (M-E) design. The method of construction is illustrated below.

Example 1: Let p = 5, we take the following ordered 10 crosses indexed as given below.

1	2	3	4	5	6	7	8	9	10
1x2	1x3	1x4	1x5	2x3	2x4	2x5	3x4	3x5	4x5

Now we consider a triangular PBIB design d_1 (T_{28}, Clatworthy, 1973) with parameters v = 10, b = 5, k = 4, r = 2, λ_1 = 1, λ_2 = 0. From design d_1, a mating design d_2 for CDC can be obtained by replacing each treatment with the cross corresponding to i^{th} treatment, we get the required mating design for CDC.

d_1

B_1	1	2	3	4
B_2	5	6	7	1
B_3	8	9	2	5
B_4	10	3	6	8
B_5	4	7	9	10

d_2

B_1	1x2	1x3	1x4	1x5
B_2	2x3	2x4	2x5	1x2
B_3	3x4	3x5	1x3	2x3
B_4	4x5	1x4	2x4	3x4
B_5	1x5	2x5	3x5	4x5

Now consider columns of d_2 as blocks, we get design d_3 for CDC experiment system

d_3

B_1	B_2	B_3	B_4
1x2	1x3	1x4	1x5
2x3	2x4	2x5	1x2
3x4	3x5	1x3	2x3
4x5	1x4	2x4	3x4
1x5	2x5	3x5	4x5

Thus we get IBD for CDC experiment with parameters v =10, b′ = 4, k′ = 5, r = 2.

Example 2: Let p = 6, we take the following ordered 15 crosses

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1x2	1x3	1x4	1x5	1x6	2x3	2x4	2x5	2x6	3x4	3x5	3x6	4x5	4x6	5x6

Now we take triangular design d_1 (T_{48}, Clatworthy, 1973) with parameters v = 15, b = 6, r = 2, k = 5, $\lambda_1 = 1$, $\lambda_2 = 0$. From d_1 design, a mating design d_2 for CDC can be obtained by replacing each i^{th} treatment corresponding i^{th} cross. We get the required mating design for

CDC. Now consider columns of mating design as blocks, we get design d_3 for CDC system IV.

d1

B_1	1	2	3	4	5
B_2	6	7	8	9	1
B_3	10	11	12	2	6
B_4	13	14	3	7	10
B_5	15	4	8	11	13
B_6	5	9	12	14	15

d2

B_1	1X2	1X3	1X4	1X5	1X6
B_2	2X3	2X4	2X5	2X6	1X2
B_3	3X4	3X5	3X6	1X3	2X3
B_4	4X5	4X6	1X4	2X4	3X4
B_5	5X6	1X5	2X5	3X5	4X5
B_6	1X6	2X6	3X6	4X6	5X6

Now consider columns of design d_2 as blocks, we get design d_3 for CDC system IV.

d_3

B_1	B_2	B_3	B_4	B_5
1X2	1X3	1X4	1X5	1X6
2X3	2X4	2X5	2X6	1X2
3X4	3X5	3X6	1X3	2X3
4X5	4X6	1X4	2X4	3X4
5X6	1X5	2X5	3X5	4X5
1X6	2X6	3X6	4X6	5X6

Thus we get IBD for CDC experiment with parameters $v = 15$, $b' = 5$, $k' = 6$, $r = 2$.

For convenience in further discussion, we will denote design d_3 as d.

3.2 Analysis and optimality

3.2.1 Model and estimation of cross effects

For the analysis of data obtained from design d, we will follow Singh and Hinkelmann (1998) two stage procedures for estimating gca and sca effects. The first stage is to consider the proposed designs to estimate cross effects, say,

$\mathbf{T} = (\tau_{12}, \tau_{13}, \ldots, \tau_{p(p-1)/2})'$ for design d by the following linear additive fixed effect model.

$$\mathbf{y} = \mu\mathbf{1} + \mathbf{X}\boldsymbol{\tau} + \mathbf{D}\boldsymbol{\beta} + \mathbf{e} \qquad (3.1)$$

where **y** is an nx1 vector of observations, **1** is the nx1 vector of ones**, X** is the n x v design matrix for treatments and **D** is the n x b design matrix for blocks, μ is a general mean, **τ** is a v x 1 vector of treatment parameters, **β** is a b x 1 vector of block parameters and **e** is an n x 1 vector of errors. All except **e** are considered as fixed effects and **e** is normally distributed with $E(\mathbf{e}) = \mathbf{0}$, $V(\mathbf{e}) = \sigma^2\mathbf{I}$ and $Cov(\boldsymbol{\beta}, \mathbf{e}') = \mathbf{0}$.

Following Tocher (1952), Raghava Rao (1971) and Dey (1986), the least square method for the analysis of proposed design leads to the following reduced normal equations for the crosses for model (3.1).

$$\mathbf{C}_d\boldsymbol{\tau} = \mathbf{Q}_d \qquad (3.2)$$

where $\mathbf{C}_d = \mathbf{r}^{\delta} - \mathbf{N}\mathbf{k}^{-}\mathbf{N}'$ and $\mathbf{Q}_d = (Q_{1d}, \ldots, Q_{vd})' = \mathbf{T} - \mathbf{N}\mathbf{k}^{-\delta}\mathbf{B}$ with $\mathbf{C}_d\mathbf{1}_p = \mathbf{0}$, $Q_d\mathbf{1}_p = \mathbf{0}$

In the above expressions $\mathbf{r}^{\delta}$ and $\mathbf{k}^{\delta}$ are diagonal matrices of order v x v and bxb with elements r and p in the diagonal, respectively of designs d. $\mathbf{N} = \mathbf{X}'\mathbf{D}$ is the v x b incidence matrix of the designs d; $\mathbf{T} = \mathbf{X}'\mathbf{Y}$ and $\mathbf{B} = \mathbf{D}'\mathbf{Y}$ are vectors of cross totals and block totals of order vx1 and b x 1 respectively for designs d.

Hence a solution for (3.2) is given by

$$\hat{\boldsymbol{\tau}} = \mathbf{C}_d^{-}\mathbf{Q}_d \qquad (3.3)$$

where $\mathbf{C}_d^{-}$ is the generalized inverses of $\mathbf{C}_d$ with property $\mathbf{C}\mathbf{C}^{-}\mathbf{C} = \mathbf{C}$.

The sum of squares due to crosses is $\mathbf{Q}'_d\,\mathbf{C}_d^{-}\,\mathbf{Q}_d$ with d.f. = rank ($\mathbf{C}_d$) for design d and expectation and variance $\mathbf{Q}_d$ is as $E(\mathbf{Q}_d) = \mathbf{C}_d\boldsymbol{\tau}$ and

$$V(\mathbf{Q}_d) = \sigma^2\mathbf{C}_d\,. \qquad (3.4)$$

3.2.2 **Estimation of gca and sca**

Now we will utilize equations (3.2), (3.3) and (3.4) to estimate the genetic parameters in the proposed design. The second stage is to utilize the fact that the cross effects can be expressed in terms of gca and sca effects. So we can write

$$\tau_{ij} = g_i + g_j + s_{ij} \qquad (3.5)$$

where g_i (g_j) is the gca for the i^{th} (j^{th}) parent, s_{ij} ($s_{ij} = s_{ji}$) is the sca for the cross between the i^{th} and the j^{th} parent (i< j =1, 2, … , p). Furthermore, the restriction $\sum_{i=1}^{p} g_i = 0$ and $\sum_{i=1}^{p} s_{ij} = 0$, for each j = 1, …, p are imposed on the combining ability elements for solving normal equation for general and specific combining ability effects.

In matrix notation equation (3.5) can be written as

$$\boldsymbol{\tau} = \mathbf{Z}\,\mathbf{g} + \mathbf{s} \qquad (3.6)$$

where $Z = (z_{ui})$ (u = 1, 2, …, n : i = 1, 2, …, p) is the cross and gca relation matrix and

z_{ui} = 2, if the u^{th} cross has both parent i,

= 1, if the u^{th} cross has only one parent i,

= 0, otherwise.

Following the approach used in Kempthrone and Curnow (1961), the equation (3.2) can then be written as

$$\mathbf{C_d}\boldsymbol{\tau} = \mathbf{C_d}\,\mathbf{Z}\,\mathbf{g} + \mathbf{C_d}\,\mathbf{s}$$

where $$E(\mathbf{Q_d}) = \mathbf{C_d}\mathbf{Z}\mathbf{g} + \mathbf{C_d}\,\mathbf{s} \qquad (3.7)$$

Since the matrix C is singular, we use the unified theory of least square due to Rao (1973, p 300).

So we get a solution for **g** as

$$\hat{\mathbf{g}} = (\mathbf{Z}'\,\mathbf{C_d}\,\mathbf{C_d}^{-}\,\mathbf{C_d}\,\mathbf{Z})^{-}\;\mathbf{Z}'\,\mathbf{Q_d} = (\mathbf{Z}'\,\mathbf{C_d}\,\mathbf{Z})^{-}\,\mathbf{Z}'\,\mathbf{Q_d} \qquad (3.8)$$

Here the matrix $(\mathbf{Z}'\,\mathbf{C_d}\,\mathbf{Z}) = r(p\text{-}2)\,[\,\mathbf{I}_p - \frac{1}{p}\mathbf{1}_p\,\mathbf{1}'_p\,]$, $\mathbf{I}_p$ is the pxp identity matrix, $\mathbf{1}_p$ is column vector of ones.

So $\hat{\mathbf{g}} = (\mathbf{Z}' \mathbf{C}_d \mathbf{Z})^{-} \mathbf{Z}' \mathbf{C}_d \boldsymbol{\tau}$ (3.9)

$\hat{\mathbf{g}} = \mathbf{H}_1 \boldsymbol{\tau}$, where $\mathbf{H}_1 = (\mathbf{Z}' \mathbf{C}_d \mathbf{Z})^{-} \mathbf{Z}'\mathbf{C}_d$

Hence $\text{Var}(\hat{\mathbf{g}}) = \mathbf{H}_1 \mathbf{C}_d \mathbf{H}'_1 \sigma^2 = \sigma^2 \frac{1}{r(p-2)} \mathbf{I}_p$ (3.10)

Substituting the estimate of **g** in equation (3.6), we obtain a solution for **s**

$\hat{\mathbf{s}} = (\mathbf{C}_d^{-} - \frac{1}{r(p-2)} \mathbf{Z}\mathbf{Z}') \mathbf{Q}_d = (\mathbf{C}_d^{-} - \frac{1}{r(p-2)} \mathbf{Z}\mathbf{Z}') \mathbf{C}_d \boldsymbol{\tau}$ (3.11)

$= \mathbf{H}_2 \boldsymbol{\tau}$, Where $\mathbf{H}_2 = (\mathbf{C}_d^{-} - \mathbf{Z}(\mathbf{Z}' \mathbf{C}_d \mathbf{Z})^{-} \mathbf{Z}') \mathbf{C}_d$ and is symmetric idempotent.

$\text{Var}(\hat{\mathbf{s}}) = \mathbf{H}_2 \mathbf{C}_d \mathbf{H}'_2 \sigma^2$ (3.12)

because $\mathbf{H}_1 \mathbf{1}_v = \mathbf{0}$, $\mathbf{H}_2 \mathbf{1}_v = \mathbf{0}$, $\mathbf{H}_1 \mathbf{H}_2' = \mathbf{0}$, rank $(\mathbf{H}_1)$ = p-1 and rank $(\mathbf{H}_2)$ = v-p.

Since $\mathbf{H}_1 \mathbf{H}_2' = \mathbf{0}$, it follows that **g** and **s** represented by treatment contrasts that carry p-1 and v-p degrees of freedom, respectively, and that contrasts representing **g** are orthogonal to those representing **s**. This means the proposed design d allows for gca and sca effects to be estimated independently.

The sum of squares due to gca and sca for d are given by

SS (gca) = $\mathbf{Q}'_d \mathbf{Z} (\mathbf{Z}' \mathbf{C}_d \mathbf{Z})^{-} \mathbf{Z}' \mathbf{Q}_d$ (3.13)

SS (sca) = $\mathbf{Q}_d' \hat{\mathbf{S}} \mathbf{Q}_d$

The ANOVA is given in Table 3.1

Table 3.1 ANOVA for the proposed design

Source of variation	Degrees of Freedom	Sum of squares
Block	k-1	$\mathbf{B}'\mathbf{B}/p - g^2/p(p-1)$
Crosses(adjusted for blocks)	rank $(\mathbf{C}_d)$	$\mathbf{Q}'_d \mathbf{C}_d^{-} \mathbf{Q}_d$
gca	rank $(\mathbf{H}_1)$	$\mathbf{Q}'_d \mathbf{Z} (\mathbf{Z}' \mathbf{C}_d \mathbf{Z})^{-} \mathbf{Z}' \mathbf{Q}_d$
sca	rank $(\mathbf{H}_2)$	$\mathbf{Q}_d' (\mathbf{C}_d^{-} - \mathbf{Z}(\mathbf{Z}' \mathbf{C}_d \mathbf{Z})^{-} \mathbf{Z}') \mathbf{Q}_d$
Residual	by subtraction	by subtraction
Total	n-1	$\mathbf{y}'\mathbf{y} - g^2/p(p-1)$

Where g = grand total of all n observations

3.2.3 Variance-balanced design for CDC system IV

From Equation (3.10) we have Var ($\hat{\mathbf{g}}$) = $\mathbf{H}_1 \mathbf{C}_d \mathbf{H}'_1 \sigma^2 = \sigma^2 \frac{1}{r(p-2)} \mathbf{I}_p$ which shows the covariance matrix of $\hat{\mathbf{g}}$ is a constant times $\mathbf{I}_p$. Therefore, the proposed design is variance-balanced for gca effects.

3.2.4 Efficiency factor and optimality

It is clear from (3.10) that using design d each elementary contrast among gca effects is estimated with variance $\sigma^2 \frac{1}{r(p-2)}$ and under randomized block design with equal replications, the variance of any elementary contrast among gca effects is $\sigma_1^2 \frac{1}{r(p-2)}$,

where σ_1^2 is the per observation variance in the case of randomized block experiment. Hence the efficiency factor of design d as compared to randomized block design having the same number of crosses under the assumption of equal intrablock variances $(\sigma_1^2 = \sigma^2)$ is given by

$$E = \frac{\textit{Average variance for RBD}}{\textit{Average variance for d}} = 1.$$

Hence the proposed design d is optimal in the sense of Kempthorne (1956) and the design uses minimum possible number of experimental units and at the same time this design allows the same precision for the estimation of general and specific combining ability of participating lines as compared to RCD.

We thus have the following result.

The proposed design d is variance-balanced and optimal for estimating gca effects.

3.2.5 List of optimal designs for diallel crosses from catalogue of Clatworthy (1973).

We investigated 88 triangular designs in catalogue of Clatworthy (1973) leaving the 12 optimal designs of Dey and Midha (1996). There are 59 triangular deigns with $5 \leq p \leq 10$ satisfying one of λ's is zero. The remaining two being left out, as for these deigns p(p-1)/2 exceeds 100. Optimal designs derived from these triangular designs are provided in Table 3.2 in comparisons to the Dey and Midha designs.

Table 3.2 Optimal designs for diallel crosses.

No.	p	Ref.	v	r	k	b	E_d	E_{DM}	Remark
1	5	T1	10	6	2	30	1.00		N
2	5	T2	10	3	2	15	1.00	0.83	R
3	5	T3	10	6	2	30	1.00	0.83	R
4	5	T4	10	9	2	45	1.00	0.83	R
5	6	T5	15	8	2	60	1.00	-	N
6	6	T6	15	6	2	45	1.00	0.75	R
7	7	T7	21	10	2	105	1.00	-	N
8	7	T8	21	10	2	105	1.00	0.7	R
9	5	T9	10	3	3	10	1.00	-	N
10	5	T10	10	6	3	20	1.00	-	N
11	5	T11	10	9	3	30	1.00	-	N
12	6	T14	15	4	3	20	1.00	-	N
13	6	T15	15	8	3	40	1.00	-	N
14	6	T16	15	3	3	15	1.00	1.00	R
15	6	T17	15	6	3	30	1.00	1.00	R
16	6	T19	15	9	3	45	1.00	1.00	R
17	7	T20	21	5	3	35	1.00	-	N
18	7	T21	21	10	3	70	1.00	-	N
19	7	T22	21	10	3	70	1.00	0.93	R
20	8	T23	28	6	3	56	1.00	-	N
21	9	T24	36	7	3	84	1.00	-	N
22	10	T25	45	8	3	120	1.00	-	N
23	5	T28	10	2	4	5	1.00	-	N
24	5	T29	10	4	4	10	1.00	-	N
25	5	T30	10	6	4	15	1.00	-	N
26	5	T31	10	8	4	20	1.00	-	N

27	5	T32	10	10	4	25	1.00	-	N
28	6	T38	15	8	4	30	1.00	-	N
29	8	T39	28	8	4	56	1.00	-	N
30	8	T40	28	10	4	70	1.00	1.00	R
31	9	T41	36	7	4	63	1.00	0.96	R
32	6	T48	15	2	5	6	1.00	-	N
33	6	T49	15	4	5	12	1.00	-	N
34	6	T50	15	6	5	18	1.00	-	N
35	6	T51	15	8	5	24	1.00	-	N
36	6	T52	15	10	5	30	1.00	-	N
37	7	T53	21	10	5	42	1.00	-	N
38	10	T54	45	7	5	63	1.00	1.00	R
39	7	T65	21	2	6	7	1.00	-	N
40	7	T66	21	4	6	14	1.00	-	N
41	7	T67	21	6	6	21	1.00	-	N
42	7	T68	21	8	6	28	1.00	-	N
43	7	T69	21	10	6	35	1.00	-	N
44	8	T72	28	2	7	8	1.00	-	N
45	8	T73	28	4	7	16	1.00	-	N
46	8	T74	28	6	7	24	1.00	-	N
47	8	T75	28	8	7	32	1.00	-	N
48	8	T76	28	10	7	40	1.00	-	N
49	9	T78	36	2	8	9	1.00	-	N
50	9	T79	36	4	8	18	1.00	-	N
51	9	T80	36	6	8	27	1.00	-	N
52	9	T81	36	8	8	36	1.00	-	N
53	9	T82	36	10	8	45	1.00	-	N
54	10	T86	45	2	9	10	1.00	-	N
55	10	T87	45	4	9	20	1.00	-	N
56	10	T88	45	6	9	30	1.00	-	N
57	10	T89	45	8	9	40	1.00	-	N
58	10	T90	45	10	9	50	1.00	-	N
59	10	T90	45	10	9	50	1.00	-	N

Note Ref. means the design numbers from the catalogue of Clateworthy (1973), DM denotes Dey and Midha, N denotes optimal design not reported earlier while R denotes design reported by Dey and Midha. E_{DM} and E_d are the efficiencies of DM and design d, respectively.

Chapter four

4. APPLICATION

We show the essential steps of analysis of a diallel cross experiment, using an incomplete block design proposed in this study. For this purpose, we take data from an experiment conducted by Terumi Mukai on Drosophila melanogaster Cockerham and Weir (1977) on page number 203. For the purpose of illustration, we take the data of relevant crosses from this experiment. The design chosen is the one given in Example 1of Chapter three. There are 10 crosses and the design has 4 blocks, each of size 5. Each cross replicated twice.

The layout and observations in parentheses are given in Table 4.1.

Table 4.1 Design and observations for Example 1

B_1	B_2	B_3	B_4
1x2 (15.4)	1x3 (31.8)	1x4 (16.2)	1x5 (14.6)
2x3 (21)	2x4 (11.4)	2x5 (12.2)	1x2 (16.5)
3x4 (16.8)	3x5 (15.2)	1x3 (30.4)	2x3 (23)
4x5 (15.2)	1x4 (17 .8)	2x4 (13)	3x4 (16.3)
1x5 (18.8)	2x5 (13.6)	3x5 (15.4)	4x5 (13.8)

The following are the vectors of treatments total, block total, and adjusted treatment total respectively.

T = (31.9, 62.2, 34, 33.4, 44, 24.4, 25.8, 33.1, 30.6, 29)′

B = (87.2, 89.8, 87.2, 84.2)′

Q = (-2.38, 26.8, -1.4, -0.88, 9.72, -11, -9.6, -1.18, -4.8, -5.28)′

C = R - 1/5NN′ =

$$\begin{bmatrix} 1.6 & 0.0 & 0.0 & -0.4 & -0.4 & 0.0 & 0.0 & -0.4 & 0.0 & -0.4 \\ 0.0 & 1.6 & -0.4 & 0.0 & 0.0 & -0.4 & -0.4 & 0.0 & -0.4 & 0.0 \\ 0.0 & -0.4 & 1.6 & 0.0 & 0.0 & -0.4 & -0.4 & 0.0 & -0.4 & 0.0 \\ -0.4 & 0.0 & 0.0 & 1.6 & -0.4 & 0.0 & 0.0 & -0.4 & 0.0 & -0.4 \\ -0.4 & 0.0 & 0.0 & -0.4 & 1.6 & 0.0 & 0.0 & -0.4 & 0.0 & -0.4 \\ 0.0 & -0.4 & -0.4 & 0.0 & 0.0 & 1.6 & -0.4 & 0.0 & -0.4 & 0.0 \\ 0.0 & -0.4 & -0.4 & 0.0 & 0.0 & -0.4 & 1.6 & 0.0 & -0.4 & 0.0 \\ -0.4 & 0.0 & 0.0 & -0.4 & -0.4 & 0.0 & 0.0 & 1.6 & 0.0 & -0.4 \\ 0.0 & -0.4 & -0.4 & 0.0 & 0.0 & -0.4 & -0.4 & 0.0 & 1.6 & 0.0 \\ -0.4 & 0.0 & 0.0 & -0.4 & -0.4 & 0.0 & 0.0 & -0.4 & 0.0 & 1.6 \end{bmatrix}$$

We note that the matrix C is determined entirely by the specific design, that is, by the incidence matrix N. the C matrix is symmetric, and the sum of any row or any column of **C** add to zero, that is, **CJ = 0**, which implies that the matrix **C** is not of full rank. Hence Cdoes not have an inverse but the generalized inverse, which is given by

Ginv(**C**) =

$$\begin{bmatrix} 0.4 & 0 & 0.0 & -0.1 & -0.1 & 0.0 & 0.0 & -0.1 & 0.0 & -0.1 \\ 0.0 & 0.4 & -0.1 & 0.0 & 0.0 & -0.1 & -0.1 & 0.0 & -0.1 & 0.0 \\ 0.0 & -0.1 & 0.40 & 0.0 & 0.0 & -0.1 & -0.1 & 0.0 & -0.1 & 0.0 \\ -0.1 & 0.0 & 0.0 & 0.4 & -0.1 & 0.0 & 0.0 & -0.1 & 0.0 & -0.1 \\ -0.1 & 0.0 & 0.0 & -0.1 & 0.4 & 0.0 & 0.0 & -0.1 & 0.0 & -0.1 \\ 0.0 & -0.1 & -0.1 & 0.0 & 0.0 & 0.4 & -0.1 & 0.0 & -0.1 & 0.0 \\ 0.0 & -0.1 & -0.1 & 0.0 & 0.0 & -0.1 & 0.4 & 0.0 & -0.1 & 0.0 \\ -0.1 & 0.0 & 0.0 & -0.1 & -0.1 & 0.0 & 0.0 & 0.4 & 0.0 & -0.1 \\ 0.0 & -0.1 & -0.1 & 0.0 & 0.0 & -0.1 & -0.1 & 0.0 & 0.4 & 0.0 \\ -0.1 & 0.0 & 0.0 & -0.1 & -0.1 & 0.0 & 0.0 & -0.1 & 0.0 & 0.4 \end{bmatrix}$$

Table 4.2 Analysis of variance of the data

Source	D.F	Sum of squares	Mean sum of square	F
Blocks	3	3.144	1.0467	
Crosses	9	543.294	60.3660	27.276141
gca	4	394.6006	98.6502	44.574707
sca	5	148.6934	29.7387	13.437315
Intrablock error	7	15.492	2.213143	
Total	19	561.932		

Tabulated valued of $F_{9,7}$ at $\alpha = 0.05$ is 3.637. Since, $F_{cal} = 27.276$ is greater than 3.637, there is a significant difference among genotypes (crosses). Hence, it is appropriate to proceed further for genetic analysis. For testing H_0: all gca effects are equal against H_1: at least two of the gca effects are not equal, we make use of the statistic $F = \frac{98.6502}{2.213143} = 44.574707.$

The tabulated valued of $F_{4,7}$ at $\alpha = 0.05$ is 3.36. Hence, we may conclude that gca effects are significantly different. Hence in present case, the pair wise comparisons of gca effects is possible. Similarly sca effects are significant at $\alpha = 0.05$.

Table 4.3 The estimate of gca and their estimated standard error

Parent	Estimate of (gca)	$\pm$S.E
1	3.67524	. 3688572
2	-2.20116	. 3688572
3	5.06964	. 3688572
4	-3.13076	. 3688572
5	-3.41296	. 3688572

The standard errors of lines can be utilized to construct confidence interval for gca effects of each line and to test the significance of the individual lines by t-test. For instance, t value for g_1 is $t_1 = \frac{3.67524}{.3688572} = 9.9637475$

Tabulated value of $t_{.025,\ 19} = 2.132$, which is less than t-calculated. Hence we conclude that g_1 is significant. Similarly g_2, g_3, g_4 and g_5 are significant at $\alpha = 0.05$.

In the same manner we can perform t-test for testing the significance of sca effects

Table 4.4 The estimate of sca effects and their estimated standard error

S.C.A	Estimate of (sca)	S.E	S.C.A	Estimate of (sca)	$\pm$S.E
S_{12}	-2.66408	1.3570	S_{24}	-0.16808	1.3570
S_{13}	4.65512	1.3570	S_{25}	0.81412	1.3570
S_{14}	4.65512	1.3570	S_{34}	-2.52888	1.3570
S_{15}	-0.70228	1.3570	S_{35}	-4.05668	1.3570
S_{23}	1.99152	1.3570	S_{45}	3.90372	1.3570

Chapter five

5 . DISCUSSION, CONCLUSION AND RECOMMENDATION

5.1 Discussion

Experiments for CDC system IV can be carried out in randomized block design, considering each cross as a treatment. However, if the number of crosses is large, accommodating all the crosses in a single block may result in large intrablock variance. Incomplete block designs may be used to reduce the block size (Agarwal and Das, 1990). The method of accommodating the crosses from a CDC in balanced incomplete block design, as environmental design requires, in general, a large number of replications. Alternatively, a two associate classes partially balanced incomplete block design with smaller number of blocks may be used (Singh and Hinklemann, 1998). It is for this reason we constructed optimal incomplete block design for CDC system IV using partially balanced incomplete block design with two associate classes. The properties of our designs are (i) block size small (ii) the designs are optimal because every line occurs in a block equal number of times.

From the result chapter 4 we observe that the information matrix (**C**) is symmetric, and the sum of any row or any column of **C** add to zero, that is, $\mathbf{CJ} = \mathbf{0}$, which implies that the matrix **C** is not of full rank. In fact the rank of C is equal to 9, which is essential condition for the design to be connected. Therefore the design is connected for estimating treatment contrasts.

The proposed design is connected; therefore, the sum of square due to crosses can further be splitted up into two components, namely, S.S due to gca and S.S due to sca. It is easy to see from table 4.2 that 543.294 =394.6006 + 148.6934.

As it can be seen from table 4.2, both gca and sca effects are significant at $\alpha = 0.05$. If the gca effects are significant, it is concluded that there are additive effects having practical relevance. At the same time, significance of sca shows the relevance of dominant effects. Selection of crosses for best additive effects and dominance effects is made on basis of g_i and s_{ij} values. The breeder should look for the significant values of g_i and s_{ij}. As a general guidance, one should pick a cross having highest g_i value if the trait under consideration is to be increased. For example, to increase the yield of crop, so is true for s_{ij}. But one should pick

up a cross having lowest negative value of g_i and / or s_{ij} if the trait to be decreased. For example, if the breeder's interest lies in dwarfing.

5.2 Conclusion

We have listed 59 designs which can be used for $5 \leq p \leq 10$. When information on general and specific combining ability for set specific lines is desired in connection with plant or animal breeding problems. In plant breeding material (seed) if it can be assumed that there will be no genotype reciprocal effects (this is usually the case) CDC system IV is the most suitable. So in this situation optimal designs proposed in this study can be used.

5.3 Recommendation

Instead of conducting complete diallel cross experiment in randomized block design one can use our proposed designs to conduct the complete diallel cross experiments because our designs consume less experimental units in comparison to randomized block design. So the cost of conducting the experiment will be reduced.

REFERNCES

Agrawal, S. C and Das, M. N. (1987). A note on construction and application of balanced n-ary design , *Sankhya, Series* **B 49**(**2**), 192-196.

Agrawal, S. C and Das, M. N. (1990). Incomplete block design for partial diallel cross. *Sankhya,Series* **B 52**(**2**), 75-81.

Aggrawal, K. R. (1974). Analysis of Li(s) and triangular design, *journal of the Indian Society for Agricultural Statistics*, **26**, 3-13.

Bose, R. C. and Nair, K.R. (1936). (1939). Partially balanced incomplete block design. *Sankhya*. **4**, 337-372.

Braten , O.B.(1965).The union of partial diallel mating design and incomplete environmental design. *Dessertation*, North Carolina State University.

Ceranka, B.and Kielezewska,H.(1986). Analysis of diallel table for experiment in block design. *Biometrical Journal* **28**, 529-638.

Ceranka, B. and Mejza, S. (1988). Analysis of diallel table for experiments in balanced incomplete block mixed model. *Biometrical Journal*, **30**, 3-16.

Chang, Li-Chien (1959). The uniqueness and non uniqueness of the triangular association scheme. *Science Record, New Series* **3:12**,604-613.

Cheng, C.S. (1996). *Optimal designs*. *Exact theory. Hand book of Statistics*, 13, Eds. S. Ghosh and C. R. Rao. 977-1006.

Clateworthy, W. H. (1973). Tables of Two-Associates-Class Partially Balanced Designs. *Applied. Maths. Ser.* No.63. Washington D. C. National Bureau of Standards

.

Cockerham, C. C. (1963). Estimation of genetic variance components. *Statistical Genetics and plant breeding.* Natl. Acad. sci. Natl. Res. Council Publ. **982**, 53-94.

Cockerham, C. C. and Weir B. S. (1977). Quadratic analysis of reciprocal crosses. *Biometrics* **33**, 187-203.

Comstock, R. E. and Robinson, H. F. (1952). Estimation of average dominance of genes. *Heterosis*, Iowa State College Press, 494-516.

Connor, W. S. (1958). The uniqueness of the triangular association scheme. *Ann. Math. Statist.* **29**, 262-266.

Curnow, R. N. (1963). Sampling the diallel cross. *Biometrics* **19**, 287-306.

Das, A, Dean, M. Angela (1998). On optimality of some partial diallel cross designs. *Sankhya,* vol. **60***, Series* **B**, 511-524.

Dey, A. (1986). *Theory of block designs*. Wiley Eastern, New, Delhi

Dey, A. and Midha, C. K. (1996).Optimal complete diallel crosses. *Biometrika*, **83**(**2**), 484-489

Divecha, J and Ghosh, D. K. (1994). Incomplete block designs for complete diallel crosses and their analysis. *Journal of Applied Statistics* **21**, 359-408.

Ehrenfeld, S. (1953). On efficiency of experimental designs. *Ann.Math.Stat.***26**, 247-255.

Falconer, D. S. (1989). *Introduction to quantitative genetics*. 3rd ed.John Wiley and Sons Inc, New York.

Griffing, B. (1956). Concept of general specific combining ability in relation to diallel crossing systems. *Aust.J. Biol. Sci.* ***9***, 463- 93.

Gupta, S. and Kageyama, S. (1994). Optimal complete diallel cross. *Biometrika* ***86*** (***2***) 420-424.

Hallauer, A.R. and Miranda, J.B. (1988). *Quantitative genetics in maize breeding*. Iowa State University Press. Iowa.

Hinkelmann, K. (1965). Partial triallel crosses. *Sankhya* **A¸ 27**, 173-196.

Hinkelmann, K. (1968). Partial tetra-allel crosses. *Theo. Apll. Genetics* **38**, 242-265.

Hinkelmann, K. (1974). Two – level diallel cross experiments. *Silv. Genet*. **23**, 18-22.

Hinkelmann, K. (1977). Design and Multi-cross designs: what do they achieve? In *proceedings of the international conference on quantitative Genetics*. Iowa State Unversity Press. Ames, Iowa.

Hinkelmann, K. (1966). Unvollstandige diallel kreuzungsplane. *Bio.Zeit*. **8**, 242-265.

Hinkelmann, K. and Stern, K. (1960). Kreuzunesplane zur selektionszuchtung bei walbdaden., *silv. Genet.***9**, 121-133

.

Hinkelmann, K. (1975). Design of genetical experiments. *In A Survey of Statistical Design and linear models* J. N. Srivastava(ed), 243-269. Amestrdam: North Holland.

Hinkelmann, K. and Kempthorne, O. (2005). *Design and Analysis of Experiment Volume 2: Advanced Experimental Design*. John Wiley and Sons, Inc, New York.

Jacroux, M. (1984).Upper bounds for efficiency factors of block designs. *Sankhya. series* ***B*** **46**, 263-274.

Kemphthorne, O. and Curnow, R. N. (1961). The partial diallel cross. *Biometrics* **17,** 229-250.

Kempthorne, O. (1956). The efficiency factor of an incomplete block designs. *Ann. Math. Stat*. **27,** 846-849.

Kiefer, J. (1975). Construction and optimality of generalized youden designs. *In survey of Statistical designs and linear models.*333-353 (J.N.Srivastava ed).North Holland Amestrdam.

Kiefer, J. (1958). On non randomized optimality and randomized non-optimality of symmetric designs. *Ann.Math.Stat.***29**, 675-699

Morgan,J.P. and Uddin, N. (1995). Optimal, non-binary, variance balanced design. *Statistica Sinica* **5**, 535-546.

Mukerjee, R. (1997). Optimal partial diallel crosses. *Biometrika* **84**(**4**): 937-946.

Nair, K.R. (1944). The recovery of inter-block information in incomplete block designs. *Sankhya* **6**, 383.

Parsad, R, Gupta, VK, and Srivastava, R. (1999). Optimal designs for diallel crosses. *J.Soc. Stat .Comput. Appl.* **1:**35-52.

Paterson, L. (1983). An upper bound for the minimum canonical efficiency factor in incomplete block designs. *Biometrika* **70**(**2**), 441-446.

Parsad, R.; and Gupta, V.K. (2004): Optimal designs for diallel cross experiments:
An overview. (Chapter 1), Dhanpat Rai &Co. (P)Ltd., N. Delhi.

Pukelsheim, F. (1996). *Optimal design of experiments.* Wiley and Sons, New York.

Raghavar R. D. (1971).*Construction and Combinatorial Problems in Design of Experiments.* J. Wiley and Sons, New York.

Rao, C. R. (1973). Linear statistical inference and its application (2^{nd} ed.) Wiley, New York

Rawilings, J. O.and Cocerham, C.C. (1962). Triallel analysis. *Crop sci*. **2**, 228-231.

Rawilings, J.O.and Cocerham, C.C. (1962). Analysis of double cross hybrid populations. *Biometrics* **18,** 229-244.

Singh, M. and Hinkelmann, K. (1998)**.** Analysis of Partial diallel crosses in incomplete

blocks. *Biometrical Journal*, **40**, 165-181.

Shah, B.V. (1959). A generalization of partially balanced incomplete block designs. *Ann.Math.Statist*.**30**, 1041-1050.

Sharma, M. K. (1996). Blocking of complete diallel crossing plans using balanced lattice design. *Sankhya Series* **B**, **58**(**3**), 427-430.

Sharma, M. K. (2004). Optimal diallel cross. *Recent advances in mating designs*. Dhanpat Rai and Company (p) Ltd. New Delhi.

Sharma, V.K. (2000). *In Balanced Incomplete Block Design .Design And Analysis Of Agricultural Experiments, IASRI,* New Delhi.

Sharma, M.K. and Fanta, S. (2010). Optimal block designs for diallel crosses. *Metrika*, **71**, 361-372.

Shrikhande, S.S. (1959). On characterization of triangular association scheme. *Ann. Math.Statist*.**30**, 39-40.

Sprague, G. F. and Tatum, L.A. (1942). General versus specific combining ability in single cross of corn. *Journal of American Society of Agronomy* **34***:* 923-932.

Tjur, T. (1990). A new upper bound for efficiency factor of a block design. *Austral. J. Statist*. **32**, 231-237.

\

Tocher, K. D. (1952). The design and analysis of block experiments (with discussion*).* J. *Royal Stat. Soc.* **B. 14***: 45-100*

Wald, A. (1943). Test of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Amer. Math. Soc*. **54**, 426-482.

Yates, F. (1936). Incomplete randomized blocks. *AnA.Eugenics* **7**, 121-140.

Yates, F. (1940a). Lattice squares. *J. Agri. Sci.* **30**,672-687.

Yates, F. (1939). The recovery of inter-block information in variety trials: Quasi-Latin squares. *Ann. Eugenics* **7**,319-331.

APPENDICES

Auxiliary design ($d_{1)}$ and mating-environment design(d_3)

Design T_{28}, V = 10, r =2, k = 4, b = 5, $n_1 = 6$, $n_2 = 3$, $\lambda_1=1$, $\lambda_2 = 0$

$$\mathbf{P_1}= \begin{bmatrix} 3 & 2 \\ & 1 \end{bmatrix} \qquad , \mathbf{P}_2 = \begin{bmatrix} 4 & 2 \\ & 0 \end{bmatrix}$$

$C_1=\frac{2}{5}$, $C_2=\frac{-1}{5}$, $H=\frac{13}{4}$, $\Delta = \frac{5}{2}$, $E_1=.83$, $E_2=.71$, $E=.79$

Solution unique; non-resolve

Complement: design T57

Dual : The BIB design (5, 10, 4, 2, 1), Roy and Laha (1956)

Association scheme T (5) (see)

Plan

1	2	3	4
5	6	7	1
8	9	2	5
10	3	6	8
4	7	9	10

Source: Table of Clathworthy (1973)

P = 5

1	2	3	4	5	6	7	8	9	10
1x2	1x3	1x4	1x5	2x3	2x4	2x5	3x4	3x5	4x5

B_1	B_2	B_3	B_4
1x2	1x3	1x4	1x5
2x3	2x4	2x5	1x2
3x4	3x5	1x3	2x3
4x5	1x4	2x4	3x4
1x5	2x5	3x5	4x5

Source : own computation..

Incidence matrices

$$\mathbf{N} = \begin{bmatrix} 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 \end{bmatrix} \qquad \mathrm{R} = \begin{bmatrix} 8 & 2 & 2 & 2 & 2 \\ 2 & 8 & 2 & 2 & 2 \\ 2 & 2 & 8 & 2 & 2 \\ 2 & 2 & 2 & 8 & 2 \\ 2 & 2 & 2 & 2 & 8 \end{bmatrix}$$

$$\mathbf{NN'} = \begin{bmatrix} 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 \end{bmatrix}$$

$$\mathbf{C_1 = R - 1/5*NN'} = \begin{bmatrix} 4.8 & -1.2 & -1.2 & -1.2 & -1.2 \\ -1.2 & 4.8 & -1.2 & -1.2 & -1.2 \\ -1.2 & -1.2 & 4.8 & -1.2 & -1.2 \\ -1.2 & -1.2 & -1.2 & 4.8 & -1.2 \\ -1.2 & -1.2 & -1.2 & -1.2 & 4.8 \end{bmatrix}$$

$$= 1/5[\ 30\mathbf{I} - \mathbf{J}\ ]$$

$$= 6[\mathbf{I} - 1/5\mathbf{J}]$$

$\mathbf{C_1}$ $= r(p-2)\ [\mathbf{I} - 1/5\mathbf{J}]$, here r = 2 and p = 5

The above incidence matrices are incidence matrices for RCD.

Source : own computation.

$$\mathbf{R^*} = \begin{bmatrix} 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix}$$

$$\mathbf{N^*} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{N^*N^{*\prime}} = \begin{bmatrix} 2 & 0 & 0 & 2 & 2 & 0 & 0 & 2 & 0 & 2 \\ 0 & 2 & 2 & 0 & 0 & 2 & 2 & 0 & 2 & 0 \\ 0 & 2 & 2 & 0 & 0 & 2 & 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 & 2 & 0 & 0 & 2 & 0 & 2 \\ 2 & 0 & 0 & 2 & 2 & 0 & 0 & 2 & 0 & 2 \\ 0 & 2 & 2 & 0 & 0 & 2 & 2 & 0 & 2 & 0 \\ 0 & 2 & 2 & 0 & 0 & 2 & 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 & 2 & 0 & 0 & 2 & 0 & 2 \\ 0 & 2 & 2 & 0 & 0 & 2 & 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 & 2 & 0 & 0 & 2 & 0 & 2 \end{bmatrix}$$

The above incidence matrices are incidence matrixes for IBD.

Source: own computation.

Information matrix(C)

C* = R*-1/5N*N* =

$$\begin{bmatrix} 1.6 & 0.0 & 0.0 & -0.4 & -0.4 & 0.0 & 0.0 & -0.4 & 0.0 & -0.4 \\ 0.0 & 1.6 & -0.4 & 0.0 & 0.0 & -0.4 & -0.4 & 0.0 & -0.4 & 0.0 \\ 0.0 & -0.4 & 1.6 & 0.0 & 0.0 & -0.4 & -0.4 & 0.0 & -0.4 & 0.0 \\ -0.4 & 0.0 & 0.0 & 1.6 & -0.4 & 0.0 & 0.0 & -0.4 & 0.0 & -0.4 \\ -0.4 & 0.0 & 0.0 & -0.4 & 1.6 & 0.0 & 0.0 & -0.4 & 0.0 & -0.4 \\ 0.0 & -0.4 & -0.4 & 0.0 & 0.0 & 1.6 & -0.4 & 0.0 & -0.4 & 0.0 \\ 0.0 & -0.4 & -0.4 & 0.0 & 0.0 & -0.4 & 1.6 & 0.0 & -0.4 & 0.0 \\ -0.4 & 0.0 & 0.0 & -0.4 & -0.4 & 0.0 & 0.0 & 1.6 & 0.0 & -0.4 \\ 0.0 & -0.4 & -0.4 & 0.0 & 0.0 & -0.4 & -0.4 & 0.0 & 1.6 & 0.0 \\ -0.4 & 0.0 & 0.0 & -0.4 & -0.4 & 0.0 & 0.0 & -0.4 & 0.0 & 1.6 \end{bmatrix}$$

Ginv(**C**) =

$$\begin{bmatrix} 0.4 & 0 & 0.0 & -0.1 & -0.1 & 0.0 & 0.0 & -0.1 & 0.0 & -0.1 \\ 0.0 & 0.4 & -0.1 & 0.0 & 0.0 & -0.1 & -0.1 & 0.0 & -0.1 & 0.0 \\ 0.0 & -0.1 & 0.40 & 0.0 & 0.0 & -0.1 & -0.1 & 0.0 & -0.1 & 0.0 \\ -0.1 & 0.0 & 0.0 & 0.4 & -0.1 & 0.0 & 0.0 & -0.1 & 0.0 & -0.1 \\ -0.1 & 0.0 & 0.0 & -0.1 & 0.4 & 0.0 & 0.0 & -0.1 & 0.0 & -0.1 \\ 0.0 & -0.1 & -0.1 & 0.0 & 0.0 & 0.4 & -0.1 & 0.0 & -0.1 & 0.0 \\ 0.0 & -0.1 & -0.1 & 0.0 & 0.0 & -0.1 & 0.4 & 0.0 & -0.1 & 0.0 \\ -0.1 & 0.0 & 0.0 & -0.1 & -0.1 & 0.0 & 0.0 & 0.4 & 0.0 & -0.1 \\ 0.0 & -0.1 & -0.1 & 0.0 & 0.0 & -0.1 & -0.1 & 0.0 & 0.4 & 0.0 \\ -0.1 & 0.0 & 0.0 & -0.1 & -0.1 & 0.0 & 0.0 & -0.1 & 0.0 & 0.4 \end{bmatrix}$$

The cross and gca relation matrix

$$\mathbf{Z} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

$$\mathbf{Z'} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

$$\mathbf{C}_2 = (Z'C^*Z) = \begin{bmatrix} 4.8 & -1.2 & -1.2 & -1.2 & -1.2 \\ -1.2 & 4.8 & -1.2 & -1.2 & -1.2 \\ -1.2 & -1.2 & 4.8 & -1.2 & -1.2 \\ -1.2 & -1.2 & -1.2 & 4.8 & -1.2 \\ -1.2 & -1.2 & -1.2 & -1.2 & 4.8 \end{bmatrix}$$

We can see that $\mathbf{C_1}$and $\mathbf{C_2}$ are equal.

Source: own computation.

OUTPUTS

tread=c(8.32,-8.58,15.72,-7.38,-8.08)
Q=matrix (tread, 1, 5)
Q=
Vector of Block total= (87.2, 89.8, 87.2, 84.2)
Bl
ock sum of square=BB'=30361.36

Treatment total

T12=31.9
T13=62.2
T14=34.0
T15=33.4
T23=44.0
T24=24.4
T25=25.8
T34=33.10
T35=30.60
T45=29.0

Sum treatment total = T = T12+T13+T14+T15+T23+T24+T25+T34+T35+T45= 348.4

Sum of block total = B= B1+B2+B3+B4= 348.4

Q1= T12-.2*(B1+B4) = -2.38
Q2=T13-.2*(B2+B3) = 26.8
Q3= T14-.2*(B3+B2) = -1.4
Q4=T15-.2*(B4+B1) = -0.88
Q5= T23-.2*(B1+B4) = 9.72
Q6= T24-.2*(B2+B3) = -11
Q7=T25-.2*(B3+B2) = -9.6
Q8= T34-.2*(B1+B4) = -1.18
Q9= T35-.2*(B2+B3) = -4.8
Q10=T45-.2*(B1+B4) =-5.28

Hence, adjusted treatment totaL= Q

(-2.38, 26.8,-1.4,-0.88, 9.72, -11,-9.6, -1.18,-4.8,-5.28)

Q1+Q2+Q3+Q4+Q5+Q6+Q7+Q8+Q9+Q10=0

-2.38+ 26.8+-1.4+-0.88+ 9.72+ -11+-9.6+ -1.18+-4.8+-5.28=0
Crosses (adjusted for blocks=543.294

gca= 394.6006
sca= 67.8716
Total sum of square y'y=13227.98

R code for the above results

```
bole=c(2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2)
N=matrix ( bole,5,4)
N
Ntr=t (N)
Ntr
NNtr=N%*%Ntr
NNtr
fast=c(8,2,2,2,2,2,8,2,2,2,2,2,8,2,2,2,2,2,8,2,2,2,2,2,8)
R=matrix(fast,5,5)
R
c1=R-1/5*NNtr
c1
fast=c(1,0,0,1,1,0,0,1,0,1,0,1,1,0,0,1 ,1
,0,1,0,0,1,1,0,0,1,1,0,1,0,1,0,0,1,1,0,0,1,0,1)
N2=matrix(fast,10,4)
N2
N2tr=t(N2)
N2tr
NNtr=N2%*%N2tr
NNtr
second=c(2,0,0,0,0,0,0,0,0,0,0,2,0,0,0,0,0,0,0,0,0,0,2,0,0,0,0,0,0,0,0,
0,0,2,0,0,0,0,0,0,0,0,0,0,2,0,0,0,0,0,0,0,0,0,0,2,0,0,0,0,0,0,0,0,0,0,2
,0,0,0,0,0,0,0,0,0,0,2,0,0,0,0,0,0,0,0,0,0,2,0,0,0,0,0,0,0,0,0,0,2)
R2=matrix(second,10,10)
R2
c2=R2-1/5*NNtr
c2
c2inv=ginv(c2)
c2inv
d1=round(c2inv,digits=3)
d1
ged=c(1,1,1,1,0,0,0,0,0,0,1,0,0,0,1,1,1,0,0,0,0,1,0,0,1,0,0,1,1,0,0,0,1
,0,0,1,0,1,0,1,0,0,0,1,0,0,1,0,1,1)
z=matrix(ged,10,5)
z
ztr=t(z)
ztr
c3=ztr%*%c2%*%z
c3
f=ginv(c3)
f
d2=round(f,digits=3)
d2
f2=z%*%d2%*%ztr
f2
s=c2inv-f2
s
```

```
block=c(87.2,89.8,87.2,84.2)
b=matrix(block,1,4)
b
btr=t(b)
btr
btrb=btr%*%b
btrb
bbtr=b%*%btr
bbtr
T12=31.9
T13=62.2
T14=34.0
T15=33.4
T23=44.0
T24=24.4
T25=25.8
T34=33.10
T35=30.60
T45=29.0
B1=87.2
B2=89.8
B3=87.2
B4=84.2
T12-.2*(B1+B4)
T13-.2*(B2+B3)
T14-.2*(B3+B2)
T15-.2*(B4+B1)
T23-.2*(B1+B4)
T24-.2*(B2+B3)
T25-.2*(B3+B2)
T34-.2*(B1+B4)
T35-.2*(B2+B3)
T45-.2*(B1+B4)
T12+T13+T14+T15+T23+T24+T25+T34+T35+T45
B1+B2+B3+B4
x=-2.38+ 26.8+-1.4+-0.88+ 9.72+ -11+-9.6+ -1.18+-4.8+-5.28
x
round(x,digits=3)
adjtr=c(-2.38, 26.8,-1.4,-0.88, 9.72, -11,-9.6, -1.18,-4.8,-5.28)
Q=matrix(adjtr,1,10)
Q
Qtr=t(Q)
Qtr
A=Q%*%d1%*%Qtr (cross sum of square)
A
G=Q%*%f2%*%Qtr (ss.due to gca)
G
s3=Q%*%(d1-.20*z%*%ztr)%*%Qtr (ss due to sca)
s3
tot=c(31.9,62.2,34,33.4,44,24.4,25.8,33.1,30.6,29)
tot^2

 1017.61 +3868.84 +1156.00 +1115.56 +1936.00 + 595.36 +665.64+ 1095.61
+ 936.36 = 841.00
```