EFFICIENCIES OF REINFORCED BIB AND AUGMENTED BLOCK DESIGNS IN SUGARCANE TEST FAMILIES VS CONTROLS EXPERIMENTS

BY

MAINA PETER WACHIRA

A Thesis submitted in partial fulfillment of the requirements for the Degree of Master of Science in Applied Statistics

Faculty of Science

Maseno University

ABSTRACT

The designs used for experimentation generally require making all the possible paired comparisons among the treatments but in plant breeding selection programmes the comparison of interest is usually a subset of all the possible paired comparisons. These comparisons are usually between the new varieties and commercial varieties called control varieties. These designs for test versus control experiments when the test treatments contain homogeneous material, such as mass selection in sugarcane breeding, have received adequate attention. Breeders in sugarcane breeding programmes have shifted from mass selection to family selection where the test treatments are more heterogeneous. This shift has created a need for efficient experimental designs to evaluate hybridized sugarcane families. In this study we evaluate two designs, Augmented Block Design (ABD) and Reinforced Block Incomplete Block Design (RBIBD), which have been proposed for test versus control experiments though their efficiencies in test families versus control experiments are not known. To evaluate the designs, we simulated data for five families and two controls through Monte-Carlo simulation framework. RBIBD and ABD designs were constructed and data fitted by inclusion of block effects and random errors. The fitted data was then analysed and compared the Randomized Complete Block Design (RCBD). More concrete results in this area could improve the efficiency of sugarcane selection process which would be of great benefit to the stakeholders in the sugar industry.

1 CHAPTER ONE - INTRODUCTION

1.1 BACKGROUND

Experimentation is an integral component of any research investigation. The designs used for experimentation generally require making all the possible paired comparisons among the treatments. But this is not always the case. There exist situations when the interest of the experimenter is only in the subset of all the possible paired comparisons. This is usually the case in plant breeding selection programmes where the comparison of interest is between the new (test) varieties and the commercial varieties called check or control varieties. In test versus control experiments the test treatments are either replicated or not replicated. The two designs proposed for test versus control experiments are the augmented block design (ABD) and reinforced balanced incomplete block design (RBIBD)

Plant breeders frequently evaluate large numbers of entries in field trials for selection. Generally, the tested entries are related by pedigree. The simplest case is a nested treatment structure, where entries fall into groups or families such that entries within groups are more closely related than between groups. Some plant breeders prefer to plant close relatives next to each other in the field. This contrasts with common experimental designs such as the α-design, where entries are fully randomized. Another design option is to randomize in such a way that entries of the same group are separated as much as possible. This has been a problem to sugarcane breeders especially since sugarcane selection relies on individual selection.

Sugarcane breeders have traditionally employed intensive selection of individual seedlings or seedling bunches (mass selection) to select seedling in seedlings stage. Selection is usually subjective, based on visual appraisal for cane yield. Although satisfactory gains have been achieved using individual seedling selection, it is not efficient according to Hogarth et al. (1997) and Skinner (1971). The lack of replications, competition effects among seedlings and, because individual clone selection is labor intensive and expensive, all contribute to reduce selection efficiency.

Research carried out in some sugarcane breeding programmes in Australia and Mauritius suggests that family selection would be superior to mass selection at this stage. Family selection is particularly useful for traits with low heritability because, unlike clones, families can be replicated across years and sites, thereby improving estimates of family means as well as aiding in the identification of stable families. Because sugarcane is exploited commercially as a clone, the rationale for family selection is not to produce superior families with commercial value but rather to identify families with a higher frequency of superior clones. Family selection makes it possible to focus selection for superior clones (mass selection) on the best families, because the probability of finding superior clones at later stages of the program is highest within these families.

In this study we are interested in evaluating the experimental designs that would be suitable for family selection in sugarcane breeding thus increasing the efficiency.

1.2 STATEMENT OF THE PROBLEM

During seedlings stage of plant breeding large populations of genotypes of which few would be eventually released for commercial cultivation are tested against commercial varieties. Due to large numbers of seedlings involved, vast tracks of land are required to evaluate the seedlings. It is impossible to ensure soil and nutrient homogeneity. Since only a small percentage of the seedlings are passed to the next stages, a lot of superior genotypes are lost and mediocre ones passed on due to inefficient discrimination against the environmental effect and genetic effect.

Many plant breeding programmes have adopted the use of family selection, i.e. large numbers of individuals are picked from best performing families than from worst performing families. Construction of experimental design for such scenarios which goes against the α-design construction criteria is not known. Augmented block design and reinforced incomplete block design have been proposed for test versus control experiments though their efficiencies especially in test families versus control experiments have not been tested.

In this study we are going to evaluate the relative efficiencies of the two proposed designs, i.e reinforced balanced incomplete block design and augmented block design, against randomized complete block design.

1.3 RESEARCH OBJECTIVES

1. To determine the relative efficiency of reinforced balanced incomplete block design against randomized complete block design.
2. To determine the relative efficiency of augmented block design against randomized complete block design
3. To compare merits and demerits of reinforced balanced incomplete block design and augmented block design

1.4 RESEARCH HYPOTHESES

Let

μ_i be the true mean for the i^{th} treatment

μ_{Ri} be the mean of the i^{th} treatment in RBIBD

μ_{Ai} be the mean for the i^{th} treatment in ABD

μ_{c4i} be the mean for the i^{th} treatment in RCBD with 4 reps

μ_{c3i} be the mean for the i^{th} treatment in RCBD with 3 reps

The Hypothesis we will test are:

H_0: $\mu_{Ri} = \mu_{c4i}$ against H_1: $\mu_{Ri} \neq \mu_{c4i}$

H_0: $\mu_{Ai} = \mu_{c3i}$ against H_1: $\mu_{Ai} \neq \mu_{c3i}$

H_0: $\mu_{Ri} = \mu_{Ai}$ against H_1: $\mu_{Ri} \neq \mu_{Ai}$

1.5 CONCEPTUAL FRAMEWORK

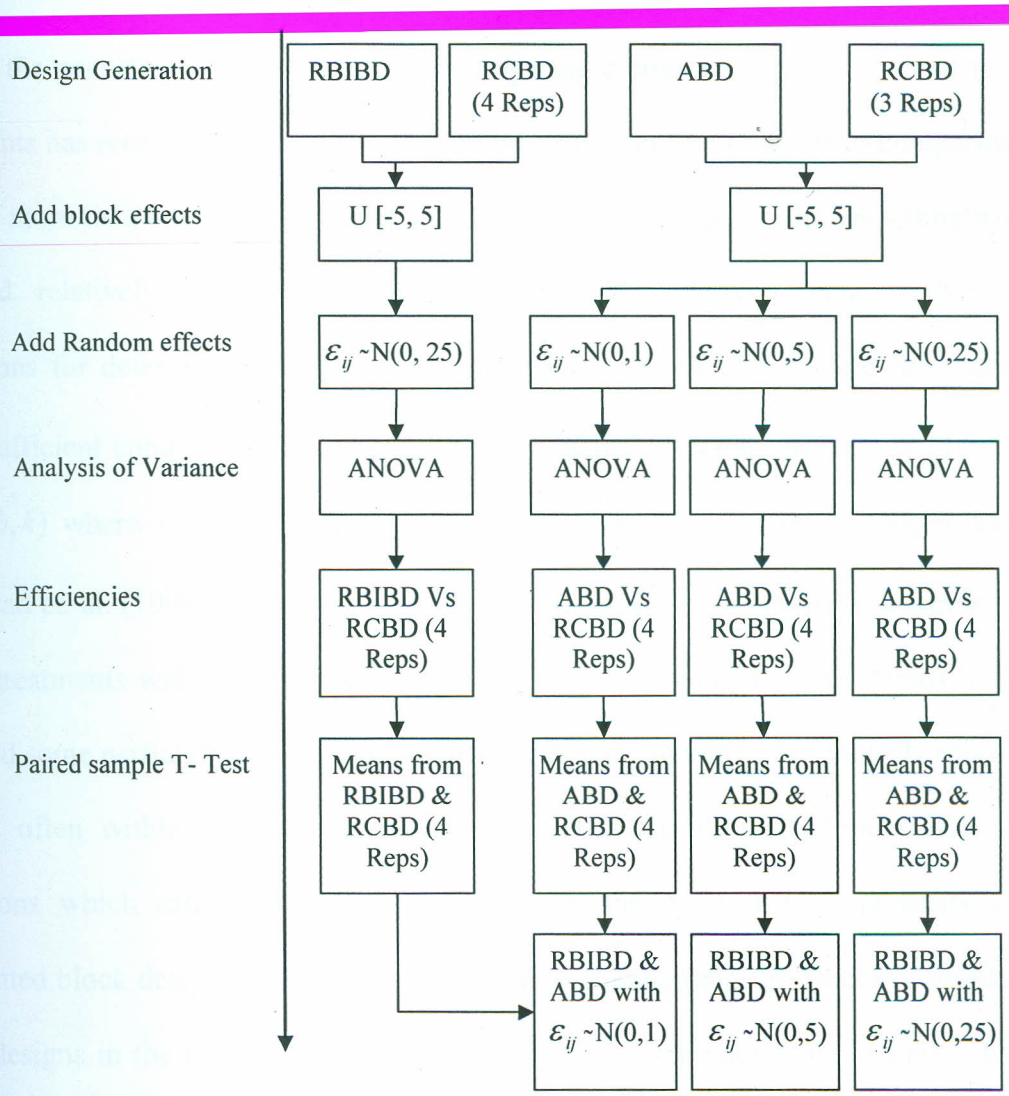

Figure 1: Conceptual framework for comparison of efficiencies between ABD and RBIBD

2 CHAPTER TWO - LIRERATURE REVIEW

During the past 25 years the problem of comparing a single control to a set of test treatments has received a good deal of attention. However, the problem of comparing a set of more than one control to a set of test treatments in a block design setting has received relatively little attention. Majumdar (1986) derived some sufficient conditions for determining A-optimal designs in classes $D(s,t;b,k)$ and then used these sufficient conditions to establish the A-optimality of certain designs in classes $D(s,t;b,k)$ where $s=t$ and in classes $D(s,t;b,k)$ where k is substantially larger than v. Parsad et. al. (1996) studied the optimality of designs for comparing s controls to t test treatments within classes of binary designs whereas Jaggi et. al. (1996) have obtained some partial results on A-optimality for designs in which the controls appear equally often within blocks. Jacroux (2000, 2002) has derived some sufficient conditions which can often be used to establish the A- and MV-optimality of augmented block designs which can be obtained by augmenting the blocks of regular block designs in the test treatments with replications of the control treatments. Two additional papers which provide more recent overviews of known results for comparing test treatments to controls are Majumdar (1996) and Gupta and Parsad (2001). In this study, we are interested in determining the efficiencies of augmented and reinforced balanced incomplete block designs in comparison of test families against a set of controls.

Augmented (Hoonuiaku) designs were introduced by Federer (1956) to fill a need arising in screening new strains of sugarcane at Experimental Station of Hawaii Sugarcane Planters Association. These designs are used in an experiment in which the experimental material for new (test) treatments is just enough for a single replication. However, the connectedness property of the design is ensured by augmenting any standard connected design in control treatments with new (test) treatments and replications of the control provide the estimate of error. More precisely, an augmented experimental design is any standard experimental design in standard treatments to which additional (new) treatments have been added. The additional treatments require enlargement of the complete block, incomplete block, row - column designs, etc. The groupings in an augmented design may be of unequal sizes. The most important of the augmented designs is the augmented block design where the standard design is the randomized complete block design. The blocks are expanded to accommodate the test treatments.

Das (1958) introduced a series of incomplete block designs which are obtainable by including any number of control treatments, say, p in every block of an existing incomplete block design with v treatments b blocks each of size k and r replications. The resulting design will have $v + p$ treatments distributed in b blocks each of size $k + p$ such that each of the p newly introduced treatments is replicated b times and the original v treatments r times each. These designs were called reinforced incomplete block designs. They have no fresh constructional problems

and can be made available for any numbers of varieties by suitably choosing v and p.

Yong-Bi et. al (1999) studied family and clonal trials to evaluate family and clone performance. They simulated test site with various patterns of environmental variation to examine the effectiveness of randomized incomplete block designs under different design parameters (constant block size, variable block size, block shape and orientation, and family size). Simulations showed that blocks of fixed size 5-20 were effective in removing most site variation in a test with 120 full-sib (both parents in common) families of three seedlings per family.

2.1 BLOCK DESIGNS

The most basic type of statistical design for making inferences about treatment means is the completely randomized design (CRD), where all treatments under investigation are randomly allocated to the experimental units. The CRD is appropriate for testing the equality of treatment effects when the experimental units are relatively homogeneous with respect to the response variable. When the experimental units are heterogeneous, the notion of blocking is used to control the extraneous sources of variability. The major criteria of blocking are characteristics associated with the experimental material and the experimental setting. The purpose of blocking is to sort experimental units into blocks, so that the variation within a block is minimized while the variation among blocks is maximized. An effective blocking not only yields more precise results than an experimental design of comparable size without blocking, but also increases the range of validity of the experimental results. In this chapter we will

concentrate on randomized complete block design, incomplete block design and balanced incomplete block design.

2.1.1 RANDOMIZED COMPLETE BLOCK DESIGN

Any experimental design in which the randomization of treatments is restricted to groups of experimental units within a predefined block of units assumed to be internally homogeneous is called a randomized block design. Blocks of units are created to control known sources of variation in expected (mean) response among experimental units. There are two classifications or factors in a randomized complete block design: block effects and treatment effects.

For instance, a typical block in an agricultural experiment is a field - fields differ substantially in soil quality, etc., and the same experimental treatment might produce different means in different fields. Formally, the design is as follows: within each of the blocks, assign 1 experimental unit at random to each of v treatments. Thus, all treatments appear within each block, and each block-treatment combination receives 1 experimental unit, which produces the observed response.

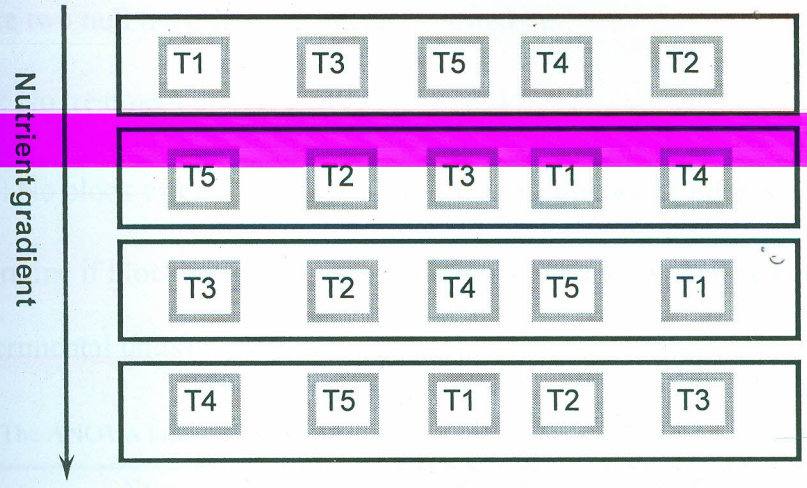

Figure 2: An example of RCBD on fertility trial

To analyse the results of RCBD, we assume additive and homoscedastic linear model with fixed effects and without interactions:

$$Y_{ij} = \mu + t_i + b_j + \varepsilon_{ij}$$

Hence,

$$E(Y_{ij}) = \mu + t_i + b_j = \mu_{ij}$$

Table 1: Table of means for RCBD

	Block					mean
Treatment	1	2	3	...	b	
1	μ_{11}	μ_{12}	μ_{13}	...	μ_{1b}	$Y_{1.}=\mu+t_1$
2	μ_{21}	μ_{22}	μ_{23}	...	μ_{2b}	$Y_{2.}=\mu+t_2$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
v	μ_{v1}	μ_{v2}	μ_{v3}	...	μ_{vb}	$Y_{v.}=\mu+t_v$
mean	$Y_{.1}=\mu+b_1$	$Y_{.2}=\mu+b_2$	$Y_{.3}=\mu+b_3$	...	$Y_{.b}=\mu+b_b$	

Treatment effects are relative such as

$$E(Y_{1\bullet} - Y_{2\bullet}) = (\mu + t_1) - (\mu + t_2) = t_1 - t_2$$

There are two null hypotheses that can be tested i.e.

H_0^T : No treatment effect: $t_1 = t_2 = \cdots t_v = 0$

H_0^B : No block effect: $b_1 = b_2 = \cdots b_b = 0$ (Usually not of interest but assessed to determine if blocking was successful in reducing the variability in the experimental units)

Table 2: The ANOVA table for RCBD

Source	Sum of Squares	Degrees of Freedom	Mean Sum of Square	F - Statistic
Treatments	SST	v-1	MST=SST/(v-1)	MST/MSE
Blocks	SSB	b-1	MSB=SSB/(b-1)	MSB/MSE
Error	SSE	(b-1)(v-1)	MSE=SSE/(b-1)(v-1)	
Total	**TSS**	**bv-1**		

2.1.2 INCOMPLETE BLOCK DESIGN

If in a block the number of plots is smaller than the number of treatments, then the blocks are said to be incomplete and a design constituted of such blocks is called an incomplete block design. Let v denote the number of treatments in an experiment and k, where $k < v$, denote the number of plots in each of the blocks. In order to ensure equal or nearly equal precision of comparisons of different pairs of treatments, the treatments are allotted to the different blocks such that each pair of treatments has the same or nearly the same number of replications and each treatment has an equal number of replications, say, r.

Different patterns of values of the number of replications of different pairs of treatments in a design, have given rise to different types of incomplete block designs.

When the number of replications of all pairs of treatments in a design is the same, then an important series of designs known as balanced incomplete block (BIB) design is obtained. This series of designs ensures equal precisions of the estimates of all pairs of treatment effects. It was first devised by Yates (1936) for agricultural experiments. These designs have evidently some constructional problems because the allotment of k of the v treatments in different blocks so that each pair of treatments is replicated a constant number of times is not straight forward. The constructional problems were solved by the joint efforts of Fisher, Yates and Bose (1939) among others using combinatorial mathematics.

For incomplete block design we get two-way data classified according to blocks and treatments. If there are b blocks in a design with v treatments and k as the block size, then there are $b \, x \, v$ cells in the two-way table with frequencies 0 or 1. Since k of the v treatments occur in a block, the frequencies in the k corresponding cells in the row for the block, are unity and those in the remaining cells of the row are zero. The data obtained from such designs are, therefore, non-orthogonal. The $b \, x \, v$ cells frequency table is called the incidence matrix of the corresponding design usually denoted by N. As the cells take two values 0 or 1 these designs are called binary designs.

2.1.3 BALANCED INCOMPLETE BLOCK DESIGNS

An incomplete block design with v treatments distributed over b blocks, each of size k, where $k < v$ such that each treatment occur in r blocks, no treatment occur more

than once in a block and each pair of treatments occurs together in λ blocks, is called a balanced incomplete block design (BIBD). The symbols v, b, k, r and λ are called the parameters of the design. The five parameters are not independent, but satisfy the two relations

1). $vr = bk$

2). $\lambda(v-1) = r(k-1)$

A BIBD is therefore commonly written as simply (v, k, λ), since b and r are given in terms of v, k, and λ by

$$b = \frac{v(v-1)\lambda}{k(k-1)}$$

$$r = \frac{\lambda(v-1)}{k-1}$$

2.1.4 BIB CONSTRUCTION

There is no single method of constructing all BIB designs. Solutions of many designs are still unknown. Solutions of all known designs indicating impossible designs have been tabulated by Fisher (1940), Rao (1961) and Sprott (1962). We describe below a few well-known series of BIB designs as detailed by Sharma (2000)

2.1.4.1 Unreduced BIB Designs

These designs are obtained by taking all combinations of the v treatments k at a time. Therefore, the parameters of all unreduced BIB designs are:

$$v, k, b = {}^{v}C_{k}, r = {}^{v-1}C_{k-1}, \lambda = {}^{v-2}C_{k-2}$$

These unreduced designs usually require a large number of blocks and replications so

that the resulting designs will often be too large for practical purposes.

2.1.4.2 BIB Designs with Parameters

$$v=s^2, b=s^2+s, k=s, r=s+1, \lambda=1 \qquad 2.1$$

Before we describe the method, we explain the concept of mutually orthogonal Latin squares which will be used in the construction of BIB designs.

A Latin square of order s is an arrangement of s symbols in an $s \times s$ array such that each symbol occurs once in each row and once in each column of the array. For example, the following are 4x4 Latin squares of order 4 in symbols A, B, C, and D:

A B C D	A B C D	A B C D
B A D C	C D A B	D C B A
C D A B	D C B A	B A D C
D C B A	B A D C	C D A B

Two Latin squares are pair-wise orthogonal if, when one square is superimposed on the other, each symbol of one Latin square occurs once with each symbol of the other square. Three or more squares are mutually orthogonal if they are pair-wise orthogonal. The three 4x4 Latin squares above are mutually orthogonal.

A complete set of $s-1$ mutually orthogonal Latin squares is known to exist for any $s = p^n$, where p is a prime number. Tables can be found in Fisher and Yates (1963). Now we describe the methods of constructing BIB designs with parameters given at 2.1

Suppose $v = s^2$ treatments are set out in an $s \times s$ array. A group of s blocks each of

size s is obtained by letting the rows of the array represent blocks. Another group of s blocks is given by taking the columns of the array as blocks. Now suppose one of the orthogonal Latin squares is superimposed on to the array of treatments. A further group of s blocks is obtained if all treatments common to a particular symbol in the square are placed in a block. Each of the $s-1$ orthogonal squares produces a set of s blocks in this manner. The resulting design is a BIB design with parameters in 2.1.

Example: For $v = 3^2 = 9$ treatments a 3 x 3 array and a complete set of mutually orthogonal Latin squares of order 3x3 are:

1 2 3	A B C	A B C
4 5 6	C A B	B C A
7 8 9	B C A	C A B

Four groups of 3 blocks are obtained from the rows, columns and the symbols of the two squares, as follows:

Rows	(1 2 3)	Columns	(1, 4, 7)
	(4 5 6)		(2, 5, 8)
	(7 8 9)		(3, 6, 9)
First square	(1, 5, 9)	Second square	(1, 6, 8)
	(2, 6, 7)		(2, 4, 9)
	(3, 4, 8)		(3, 5, 7)

This is a BIB design with parameters v = 9, b = 12, r = 4, k = 3, and $\lambda = 1$.

2.1.4.3 Complementary Design

The complement of the design in Example above obtained by replacing treatments in a block by those which do not occur in the block, is the following

(4, 5, 6, 7, 8, 9)	(2, 3, 5, 6, 8, 9)
(1, 2, 3, 7, 8, 9)	(1, 3, 4, 6, 7, 9)
(1, 2, 3, 4, 5, 6)	(1, 2, 4, 5, 7, 8)
(2, 3, 4, 6, 7, 8)	(2, 3, 4, 5, 7, 9)
(1, 3, 4, 5, 8, 9)	(1, 3, 5, 6, 7, 8)
(1, 2, 5, 6, 7, 9)	(1, 2, 4, 6, 8, 9)

The complementary design is also a BIB design with parameters v = 9, b = 12, r = 8, k =6, $\lambda = 5$.

In general if, we have a BIB design with parameters v, b, r, k, λ then its complementary design is a BIB design with parameters $v'= v$, $b'=b$, $r'= b-r$, $k'. = v -k$, $\lambda' = b-2r + \lambda$.

The complementary design of the design with parameters in (2.1) will be a balanced incomplete block design with parameters:

$$v = s^2, b = s(s+1), r = s^2-1, k = s(s-1), \lambda = s^2 - s -1$$

The s(s+1) blocks of the design for $v=s^2$ treatments have been arranged in s+1 groups of s blocks each. Now suppose a new treatment is added to all the blocks in a particular group and that the treatment added is different for each group; also, that one further block is added which consists entirely of these s+1 new treatments. This method produces a second series of BIB designs with parameters

$$v = b = s^2 + s + 1, \quad r = k = s + 1, \lambda = 1.$$

Its complement is also a BIB design with parameters

$$v = b = s^2+s+1, r = k = s^2, \lambda = s(s-1)$$

2.1.4.4 Symmetric BIB Designs

A BIB design in which v = b or r = k is called a symmetric BIB design. In symmetric BIB designs any two blocks have λ treatments in common.

2.1.4.5 α-Resolvable and Affine α-Resolvable Designs

It has been seen above that the blocks of the designs $v = s^2$, $b = s^2 + s$, $r = s+1$, $k = s$, $\lambda = 1$ can be divided into (s+1) groups, each consisting of s blocks such that in each group each of the treatments is replicated once. That is, each group is a complete replicate. Such designs are called resolvable designs or 1-resolvable designs.

In general a BIB design is called α-Resolvable if its blocks can be divided into t groups each consisting of m blocks such that in each group every treatment appears exactly α times.

In addition to this, if any two blocks of the design belonging to the same group have a constant number of treatments in common, say q1, and any two blocks belonging to different groups have a constant number of treatments in common, say q2, then the design is called affine α-resolvable BIB design.

2.1.4.6 Dual Design

The dual of a BIB design with parameters v, b, r, k,λ is obtained by interchanging the treatment and block symbols in the original design. The parameters of the dual design are $v' = b$, $b' = v$, $r' = k$, $k'=r$. The dual of a BIB design is not always a BIB design. If the original design is a symmetrical BIB design, then its dual is also a BIB

design with the same parameters.

2.1.4.7 Residual Design

In a symmetric BIB design with parameters $v = b$, $r = k$, λ delete one block and also those treatments which appear in this (deleted) block from the remaining (b-1) blocks, the design so obtained is known as the residual design. The residual design is also a BIB design with parameters $v^* = v-k$, $b^* = b-1$, $r^* = r$, $k^* = k-\lambda$, λ.

2.1.4.8 Derived Design

By deleting any block of a symmetric BIB design with parameters $v = b$, $r = k$, λ and retaining all the treatments in b-1 blocks that appear in the deleted block, we obtain a BIB design which is called the derived design.

The parameters of the derived design are:

$$v'' = k,\ b'' = b-1,\ r = r-1,\ k = \lambda,\ \lambda = \lambda-1.$$

2.2 ANALYSIS OF NON-ORTHOGONAL TWO WAY DATA

When the number of observations in a 2-way classified data is constant, the data are called orthogonal; otherwise the data are called non-orthogonal. Data from incomplete block design are non-orthogonal and require special analysis detailed below (Das, 1979).

Let the two factors of classification be denoted by A and B at p and q levels respectively. Let n_{ij} denote the number of observations in the (i,j)th cell. Denoting the k^{th} observation in the (i,j)th cell by y_{ijk}

$(k = 1, 2, \ldots, n_{ij})$, $(j = 1, 2, \ldots, q)$, $(i = 1, 2, \ldots, p)$

We make the following substitutions:

i. $\sum_{k=1}^{n_{ij}} y_{ijk} = T_{ij}$, the cell total

ii. $\sum_{j=1}^{q} T_{ij} = A_i$, the marginal total corresponding to the i^{th} level of A

iii. $\sum_{i=1}^{p} T_{ij} = B_j$, the marginal total of B

iv. $\sum_{i=1}^{p} A_i = \sum_{j=1}^{q} B_j = G$, grand total

v. $\sum_{j=1}^{q} n_{ij} = n_{i.}$, $\sum_{i=1}^{p} n_{ij} = n_{.j}$ and $\sum_{j=1}^{q} n_{i.} = \sum_{i=1}^{p} n_{.j} = n_{..}$

When the data structure is non-orthogonal the analysis is not simple as straight solution of the normal equations are not always available. Taking the following model:

$$y_{ijk} = \mu + t_i + b_j + \varepsilon_{ijk}$$

Where y_{ijk} is the random variable corresponding to the observation Y_{ijk} ; μ , t_i , b_j $(i = 1,2,\ldots,k,\ j = 1,2,\ldots,r)$ are fixed effects. We assume that the levels of B

represent the different columns in the two-way table and those of A, the different rows and G is the grand total.

Since ε_{ijk} has zero mean and μ, t_i and b_j are constants, we have

$$E(y_{ijk}) = \mu + t_i + b_j$$

Hence μ, t_i $(i = 1,2,\ldots w)$ and b_j $(j = 1,2,\ldots k)$ can be estimated by the least squares method, that is, by minimizing the sum of squares.

$$E = \sum_{ijk} (y_{ijk} - \mu - t_i - b_j)^2$$

The normal equations are

$$\frac{\partial E}{\partial \mu} = -2\sum_{ijk} (y_{ijk} - \mu - t_i - b_j) = 0$$

$$\frac{\partial E}{\partial t_i} = -2\sum_{jk} (y_{ijk} - \mu - t_i - b_j) = 0 \ (i = 1,2,\ldots, p)$$

$$\frac{\partial E}{\partial b_j} = -2\sum_{jk} (y_{ijk} - \mu - t_i - b_j) = 0 \ (j = 1,2,\ldots, q)$$

The above when simplified reduce to

$$n\hat{\mu} + \sum_i n_{i.}\hat{t}_i + \sum n_{.j}\hat{b}_j = G \qquad 2.2$$

$$n_{i.}\hat{\mu} + n_{i.}\hat{t}_i + \sum_j n_{ij}\hat{b}_j = A_i \ (i = 1,2,\ldots,k) \qquad 2.3$$

$$n_{.j}\hat{\mu} + n_{.j}\hat{b}_j + \sum_i n_{ij}\hat{t}_i = B_j \ (j = 1,2,\ldots,r) \qquad 2.4$$

Obtaining $\hat{b}_j$ from 2.4 and putting it in 2.3 we get

$$n_{..}\hat{\mu}+n_{i.}\hat{t}_i+\sum_j n_{ij}\left(\frac{B_j}{n_{.j}}-\hat{\mu}-\frac{1}{n_{.j}}\sum_m n_{mj}\hat{t}_m\right)=A_i \qquad 2.5$$

$$\left(A_i-\sum_j\frac{n_{ij}B_j}{n_{.j}}\right)=\hat{t}_i\left(n_{i.}-\sum_j\frac{n_{ij}^2}{n_{.j}}\right)-\sum_{m\neq i}\hat{t}_m\left(\sum_j\frac{n_{ij}n_{mj}}{n_{.j}}\right) \qquad 2.6$$

Substituting

$$Q_i \text{ for } \left(A_i-\sum_j\frac{n_{ij}B_j}{n_{.j}}\right) \qquad 2.7$$

$$C_{ii} \text{ for } n_{i.}-\sum_j\frac{n_{ij}^2}{n_{.j}} \qquad 2.8$$

$$C_{im} \text{ for } -\sum_j\frac{n_{ij}n_{mj}}{n_{.j}}, \text{ thus } C_{im}=C_{mi} \qquad 2.9$$

Normal equations (2.6) are written as

$$C_{ii}\hat{t}_i+\sum_{m\neq i}C_{im}\hat{t}_m=Q_i\ (i=1,2,\ldots,k) \qquad 2.10$$

These equations are called reduced normal equations. Q_i is called the adjusted total of the of the i^{th} level of A. The k equations at 2.10 are not independent, because when these equations are added, both the left hand and the right hand sides vanish. That is, $\sum Q_i=0$ and $\sum C_{im}=0$. This can be proved easily by writing down their actual expressions and adding.

As $C_{im} = C_{mi}$, it also follows that the sum of the coefficients of the t_i's in each equation at 2.10 is zero. Obviously then, if $\hat{t}_i (i = 1, 2, \ldots, k)$ is a set of solutions of 2.10, then $\hat{t}_i + \theta (i = 1, 2, \ldots, k)$ where θ is a constant, is also a set of solutions.

The equation at (2.10) has thus no unique solutions since we have k equations and 2k parameters to be estimated. To get unique solutions we impose the restriction $\sum \hat{t}_i = 0$. This implies that t_i's are estimated as deviates from the means. As a matter of fact, the restriction need not be $\sum \hat{t}_i = 0$ always. It can be any linear function of t_i's other than their contrasts. Such restriction change θ only.

When a set of solutions of (2.10) is obtained, we can get the solutions of b_j's from (2.4) if so required. The error sum of squares,

$$E = \sum_{ijk} (y_{ijk} - \mu - t_i - b_j)^2$$

$$= \sum_{ijk} y_{ijk}^2 - \mu G - \sum_i t_i A_i - \sum_j b_j B_j$$

$$= \sum_{ijk} y_{ijk}^2 - \mu G - \sum_i t_i A_i - \sum_j B_j \left(\frac{B_j}{n_{.j}} - \mu - \frac{1}{n_{.j}} \sum n_{mj} t_m \right)$$

$$= \sum_{ijk} y_{ijk}^2 - \sum_j \frac{B_j^2}{n_{.j}} - \sum_i t_i Q_i \qquad 2.11$$

Instead of eliminating $\hat{b}_j$ to get the reduced normal equations in $\hat{t}_i$'s, we could have eliminated $\hat{t}_i$'s and got the reduced normal equations in $\hat{b}_j$'s and finally the error s.s. as function of $\hat{b}_j$'s. In this situation

$$E = \sum_{ijk} y_{ijk}^2 - \sum_i \frac{A_i^2}{n_{i.}} - \sum_j b_j R_j \qquad 2.12$$

Where $R_j = B_j - \sum_i \frac{n_{ij} A_j}{n_{i.}}$

As 2.11 and 2.12 are both the same error sum of squares, we have

$$\sum \frac{A_i^2}{n_{i.}} - \sum_j - \frac{B_j^2}{n_{.j}} = \sum_i t_i Q_i - \sum_j b_j R_j \qquad 2.13$$

Next, for getting the sum of squares of A we make the hypothesis $t_1 = t_2 = \ldots = t_k = 0$ and get the reduced model $y_{ijk} = \mu + b_j + e_{ijk}$

The error sum of square on this model is

$$E_1 = \sum_{ijk} y_{ijk}^2 - \sum_j \frac{B_j^2}{n_{.j}}$$

So, the sum of squares of $A = E_1 - E = \sum_i t_i Q_i$.

2.3 ANALYSIS OF BIB DESIGN

Writing $V = \{x_i\}_{i=1}^v$ and $A = \{A_j\}_{j=1}^b$, then the incidence matrix of the BIBD is given by the $v \times b$ matrix $\mathbf{M} = ((n_{ij}))$ defined by

$$n_{ij} = \begin{cases} 1 & if\ x_i \in A_j \\ 0 & otherwise \end{cases}$$

This Matrix satisfies the equation

$$MM^T = (r-\lambda)I + \lambda J$$

Where I is a $v \times v$ identity matrix and J is the $v \times v$ unit matrix according to Fisher (1938)

Table 3: An example of tabular incidence Matrix for a BIBD with parameters (v, k, λ)

Block	Treatments					Treatment block size n_{ij}	Block total
	1	2	3	⋯	v		
1	1	0	1	⋯	0	k	B_1
2	0	0	1	⋯	1	k	B_2
3	1	1	1	⋯	0	k	B_3
.	.	.	.	⋯	.	k	.
.	.	.	.	⋯	.	k	.
b	0	1	0	⋯	1	k	B_b
Total							
No. of rep.	r	r	⋯	r			
Observational total	A_1	A_2	⋯	A_v			
Adjusted total	Q_1	Q_2	⋯	Q_v			

Thus from equation 2.7

$$Q_i = A_i - \frac{1}{k}\sum_j n_{ij} B_j \quad (i = 1,2,\ldots,v)$$

where n_{ij} denotes the entry in the above table in the cell defined by the i^{th} treatment and j^{th} block.

Using the model

$$Y_{ij} = \mu + t_i + b_j + e_{ij}$$

The Total Sum of Squares is given by

$$TSS = \sum_i \sum_j Y_{ij}^2 - \frac{\left(\sum_i \sum_j Y_{ij}\right)^2}{rk}$$

The Block Sum of Squares unadjusted is given by

$$SSB_U = \frac{\sum_j b_j}{k} - \frac{\left(\sum_i \sum_j Y_{ij}\right)^2}{rk}$$

The Treatment Sum of Squares unadjusted is given by

$$SST_U = \frac{\sum t_i^2}{r} - \frac{\left(\sum_i \sum_j Y_{ij}\right)^2}{rk}$$

Since treatments appear in different blocks we need to adjust the treatment sum of squares to block effect. To get the Adjusted Treatment Sum of Squares (SST_A) we use the reduced normal equations obtainable from equation 2.6 as follows

$$\left(r - \frac{r}{k}\right)t_i - \frac{\lambda}{k}\sum_{m \neq i} t_m = Q_i \quad (i = 1, 2, \ldots, v)$$

Noting that:

$$\sum_{m=i} t_i = \sum_{m \neq i} t_m + t_i$$

Thus

$$\left(r - \frac{r}{k}\right)t_i - \frac{\lambda}{k}\left(\sum_{m=i} t_i - t_i\right) = Q_i \quad (i = 1, 2, \ldots, v)$$

Taking the restriction $\sum_{m=i} t_i = 0$ then the solutions of the treatment effects are as follows:

$$\left(r-\frac{r}{k}\right)t_i+\frac{\lambda}{k}t_i=Q_i$$

Thus

$$t_i=\frac{Q_i}{rE}, \text{ where } E=\frac{v\lambda}{rk}\ (i=1,2,\ldots,v)$$

The adjusted Treatment Sum of Squares is

$$SST_A=\sum_i t_iQ_i=\frac{1}{rE}\sum_i Q_i^2$$

The adjusted Block Sum of Squares is given by

$$SSB_A=SST_A+SSB_U-SST_U$$

The Error Sum of Squares is given by

$$SSE=TSS-SSB_U-SST_A$$

Table 4: The ANOVA table for BIBD

Source	Degrees of Freedom	Sum of Squares	Mean Sum of Square	F - Statistic
Treatment (unadjusted)	v-1	SST_U		
Blocks (unadjusted)	b-1	SSB_U		
Treatment (adjusted)	v-1	SST_A	MST	MST/MSE
Blocks (adjusted)	b-1	SSB_A	MSB	MSB/MSE
Error	rk-b-v+1	SSE	MSE	
Total	**rk-1**	**TSS**		

2.4 REINFORCED BALANCED INCOMPLETE BLOCK DESIGN

The technique of reinforcement was initiated to control large intra-block variances but if looked from another angle, this may also be used to compare a set of tests with a set of controls. The reinforced designs were developed to meet the need of experimental situations, where it may not be possible to use balanced incomplete block (BIB) design, a lattice design or a partially balanced incomplete block design. Since BIB designs are used to make all the possible paired comparisons among the treatments but in this case we are interested only in the subset of all the possible paired comparisons. Inadequate experimental resources may also not allow equal replication of all treatments. To deal with this situation, Das (1954) suggested a design obtainable by adding some extra treatments to each of the blocks of a BIB design. With a given number of treatments w, the p extra treatments can always be so adjusted such that $v\,(=w-p)$ treatments form a BIB design. The resulting design will have $v+p$ treatments distributed in b blocks each of size $k+p$ such that each of the p-newly introduced treatment is replicated b times and the original v treatments are replicated r times each. These designs have no fresh constructional problems and can be made available for any number of varieties by suitably choosing v and p.

Using the homoscedastic and fixed effects mode of the form

$$y_{ij} = \mu + t_i + b_j + \varepsilon_{ij}$$

Das (1958) stated the reduced normal equations for reinforced incomplete block designs to the following:

$$\left(r-\frac{r}{k+p}\right)t_i-\frac{r}{k+p}\sum_{m=v+1}^{v+p}t_m-\frac{\lambda}{k+p}\sum_{i'\neq i}^{v}t_{i'}=Q_i \quad (i=1,2,\cdots,v) \qquad 2.14$$

$$\left(b-\frac{b}{k+p}\right)t_m-\frac{b}{k+p}\sum_{m'\neq m}^{v+p}t_{m'}-\frac{r}{k+p}\sum_{i=1}^{v}t_i=Q_m \quad (m=v+1,v+2,\cdots,v+p) \qquad 2.15$$

Where t_i and $t_{i'}$ stand for the original treatments replicated r times and t_m and $t_{m'}$ stand for the newly introduced treatments each replicated b times.

Das (1978) showed that equations 2.14 and 2.15 are derived through the use a two-way table below (Table 5) and Equation 2.10 through;

Writing $X=\{x_i\}_{i=1}^{v+p}$ and $A=\{A_j\}_{j=1}^{b}$, the incidence matrix of the RBIBD is given by the $b\times(v+p)$ matrix $\mathbf{M}=(\mathbf{N}_1:\mathbf{N}_2)$ where;

1. $\mathbf{N}_1$ is $b\times v$ matrix with elements

$$n_{ij}=\begin{cases}1 & if\ x_i\in A_j\\ 0 & otherwise\end{cases}$$

2. $\mathbf{N}_2$ is $b\times p$ matrix of 1's i.e. J_{bp}

Below is an example of a tabular incidence matrix for RBIBD design with $v+p$ treatments and b blocks of size $k+p$.

Table 5: An example of tabular incidence Matrix for a RBIB design

	Treatments (i)						Block Size ($n_{.j}$)
Block (j)	1	2	. .	v	$v+1$	$v+p$	
1	1	1	. .	1	1	1	$k+p$
2	0	1	. .	1	1	1	$k+p$
.	1	1	. .	0	1	1	$k+p$
.	1	0	. .	1	1	1	$k+p$
b	1	1	. .	1	1	1	$k+p$
Replication ($n_{i.}$)	r	r	. .	r	b	b	

Putting equation 2.10 in our scenario where we have $v+p$ treatments, we get

$$C_{ii}\hat{t}_i + \sum_{n\neq i} C_{in}\hat{t}_n = Q_i \quad (i=1,2,\ldots,v+p) \qquad 2.16$$

This can be split into two equations

$$C_{ii}\hat{t}_i + \sum_{n\neq i} C_{in}\hat{t}_n = Q_i \quad (i=1,2,\ldots\ldots,v) \qquad 2.17$$

$$C_{mm}\hat{t}_m + \sum_{m\neq n} C_{mn}\hat{t}_n = Q_m \quad (m=v+1,\ldots,v+p) \qquad 2.18$$

Evaluating $C_{ii} = n_{i.} - \sum_j \frac{n_{ij}^2}{n_{.j}}$ in equation 2.16 then $C_{ii} = r - \frac{r}{k+p}$

Since from Table 5 $n_{i.} = r$, $\sum_{j=i}^{b} n_{ij} = r$ and $n_{.j} = k+p \quad \forall\, j$

Evaluating $C_{in} = -\sum_j \frac{n_{ij}n_{nj}}{n_{.j}}$ still in equation 2.16 we break it into two parts

i) Newly introduced treatments not equal to i which becomes $C_{in} = -\frac{r}{k+p}$ and,

ii) Original treatments not equal i and concurs with i in λ blocks which becomes

$$C_{in} = -\frac{\lambda}{k+p}$$

For equation 2.15 the same procedure can be applied.

The equation 2.15 can also be written as

$$bt_m - \frac{b}{k+p}t_m - \frac{b}{k+p}\sum_{m'\neq m}^{v+p} t_{m'} - \frac{r}{k+p}\sum_{i=1}^{v} t_i = Q_m \qquad 2.19$$

But

$$\frac{b}{k+p}t_m + \frac{b}{k+p}\sum_{m'\neq m}^{v+p} t_{m'} = \frac{b}{k+p}\sum_{m=v+1}^{v+p} t_m$$

Thus

$$bt_m - \frac{b}{k+p}\sum_{m=v+1}^{v+p} t_m - \frac{r}{k+p}\sum_{i=1}^{v} t_i = Q_m$$

$$bt_m - \frac{1}{k+p}\left(b\sum_{m=v+1}^{v+p} t_m + r\sum_{i=1}^{v} t_i\right) = Q_m \qquad 2.20$$

Taking the restriction that $b\sum t_m + r\sum t_i = 0$, the equation 2.20 becomes

$$bt_m = Q_m \qquad 2.21$$

That is,

$$t_m = \frac{Q_m}{b} \quad (m = v+1, v+2, \cdots, v+p) \text{ and } \sum_{m'} t_{m'} = \frac{\sum Q_{m'}}{b}$$

Since

$$\sum_{i=1}^{v} t_i = \sum_{i'\neq i}^{v} t_{i'} + t_i$$

Then equation 2.14 can be written as

$$\left(r-\frac{r}{k+p}+\frac{\lambda}{k+p}\right)t_i-\frac{r}{k+p}\sum t_{m'}-\frac{\lambda}{k+p}\sum_{i=1}^{v}t_i=Q_i \qquad 2.22$$

Writing $\sum t_i=-\frac{b}{r}\sum t_m$ and $\sum t_m=\frac{\sum Q_m}{b}$, these become

$$\left(r-\frac{r}{k+p}+\frac{\lambda}{k+p}\right)t_i=Q_i+\frac{\sum Q_m}{b(k+p)}\left(r-\frac{b\lambda}{r}\right)$$

That is,

$$t_i=\frac{1}{(k+p)r-r+\lambda}\left\{(k+p)Q_i+\frac{\sum Q_m}{b}\left(r-\frac{b\lambda}{r}\right)\right\}(i=1,2,\cdots v) \qquad 2.23$$

Thus, the estimates of all the treatments effects have been obtained. The adjusted treatment sum of squares can now be obtained from

$$\sum t_iQ_i+\sum t_mQ_m$$

Error S.S. = Total S.S. - $\left(\sum\frac{B_j^2}{k+p}-C.F\right)$ - Adjusted tr. S.S.

where B_j is the j^{th} block total.

The estimate of the treatment differences can be obtained from equations 2.18, 2.20 and 2.22 and are given below:

1). $t_m-t_{m'}=\frac{Q_m-Q_{m'}}{b}$

2). $t_i-t_{i'}=\frac{k+p}{(k+p)r-r+\lambda}(Q_i-Q_{i'})$

3). $t_m-t_i=\frac{Q_m}{b}-\frac{1}{(k+p)r-r+\lambda}\left\{(k+p)Q_i+\frac{\sum Q_m}{b}\left(r-\frac{b\lambda}{r}\right)\right\}$

2.5 AUGMENTED BLOCK DESIGN

In augmented block designs a randomized complete block design is selected for the control treatments and then the blocks are enlarged to accommodate the new treatments. Augmented Block Designs (ABDs) were introduced by Federer (1956, 1961a, 1961b) as an alternative to the systematically arranged check and new treatments. ABDs have several advantages over the systematic arrangement. They are useful for screening new treatments such as genotypes, insecticides, herbicides, drugs, among others.

2.5.1 The General Theory

Suppose that w test varieties are to be tested and that sufficient seeds or plants are available to plant only single replicates of each variety. Furthermore, suppose that u other varieties, called check or control varieties are available in such quantities that r replications of each variety may be planted. The $v = w + u$ varieties included in a particular experiment are laid out in appropriate experimental design for controlling the heterogeneity effects in the experimental area. Sufficient replications of the control varieties are included to have sufficient degrees of freedom for estimating the experimental error variance and for estimating effects of the blocking used to control the heterogeneity.

The statistical analysis for experimental designs in which u control varieties have been replicated r times and where w test varieties have been replicated once, may be carried out in the following two ways:

a. The trial on v varieties may be analyzed using standard methods for disproportionate numbers in the subclasses; then, contrasts among the control varieties, among the test varieties, and among the control and the new varieties may be made

b. A statistical analysis is performed on the control variety yield only, and blocking effects, and the control variety effects are estimated; an estimate of the experimental error variance is obtained, Then, the estimated test variety means and effects are obtained and the varietal contrasts are made as in (a).

Though methods (a) and (b) might appear to result in different estimators for the effects and experimental error variance, it can be shown that this is not the case. Let y be an $n \times 1$ observational vector with

$$\left.\begin{aligned} E(y) &= \boldsymbol{X}_{11}\boldsymbol{\beta} \\ V(y) &= \sigma^2 \boldsymbol{I}_n \end{aligned}\right\} \qquad 2.24$$

Where $E(.)$ and $V(.)$ denote the expected value and the dispersion matrix, respectively, of the quantity inside the parentheses; $\boldsymbol{I}_n$ is an identity matrix of order n, β is a $p \times 1$ column vector of unknown parameters, σ^2 is unknown scalar, and X_{11} is a $n \times p$ matrix with unknown coefficients. Let $\boldsymbol{z}$ be another $m \times 1$ observational vector with

$$\left.\begin{aligned} E(z) &= \boldsymbol{X}_{21}\boldsymbol{\beta} + \boldsymbol{X}_{22}\boldsymbol{\gamma} \\ V(z) &= \sigma^2 \boldsymbol{I}_m \end{aligned}\right\} \qquad 2.25$$

Where γ is a $q \times 1$ column vector of another set of parameters and $\boldsymbol{X}_{21}$ and $\boldsymbol{X}_{22}$ are $m \times p$ and $m \times q$ matrices respectively, of known coefficients. We assume that $m = q$ and that the rank of $\boldsymbol{X}_{22}$ is equal to q.

Then γ can be estimated either by estimating β form equation 2.24 and substituting in equation 2.25, which is method (b), or by using the combined set of equations in 2.24 and 2.25 to estimate the varietal effects, which is method (a). By the method of least squares, the estimated vector $\hat{\beta}^{(1)}$ of β from equation 2.24 is:

$$\hat{\beta}^{(1)} = (X_{11}'X_{11})^{-1}X_{11}'y \qquad 2.26$$

Where $(X_{11}'X_{11})^{-1}$ denotes a generalized inverse of $X_{11}'X_{11}$. Substituting the value $\hat{\beta}^{(1)}$ for β in (4.2), we obtain the estimate $\hat{\gamma}^{(1)}$ of γ as follows:

$$\hat{\gamma}^{(1)} = X_{22}^{-1}\left[z - X_{21}(X_{11}'X_{11})^{-1}X_{11}'y\right] \qquad 2.27$$

Alternatively, from the combined set of equations 2.24 and 2.25 that is,

$$E\begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} X_{11} & 0_{n,p} \\ X_{21} & X_{22} \end{bmatrix}\begin{bmatrix} \beta \\ \gamma \end{bmatrix}, \qquad 2.28$$

Where $0_{n,p}$ is the $n \times q$ null matrix, the estimate $\left(\hat{\beta}^{(2)'}\hat{\gamma}^{(2)'}\right)'$ of $(\beta'\gamma')'$ is obtained by the method of least squares described below. Now,

$$\begin{bmatrix} X_{11}'X_{11} + X_{21}'X_{21} & X_{21}'X_{22} \\ X_{22}'X_{21} & X_{22}'X_{22} \end{bmatrix} \cdot \begin{bmatrix} \hat{\beta}^{(2)} \\ \hat{\gamma}^{(2)} \end{bmatrix} = \begin{bmatrix} X_{11}'y + X_{21}'z \\ X_{22}'z \end{bmatrix} \qquad 2.29$$

Since the rank of X_{22} is q, there exists a $q \times p$ matrix L such that

$$X_{21} = X_{22}L \qquad 2.30$$

After substituting this value in equation 2.29 and eliminating $\hat{\beta}^{(2)}$, we obtain $\hat{\gamma}^{(2)} = \hat{\gamma}^{(1)}$. Further, $\hat{\beta}^{(2)} = \hat{\beta}^{(1)}$. Thus, method (a) and method (b), as described

previously, lead to the same estimates of the varietal effects for both control and test varieties.

The error mean sum squares for both methods can also be easily verified to be the same. The inferences drawn from both methods are thus identical whenever the test treatments are replicated exactly once. It is recommended that method (b) be used for the statistical analyses as it minimizes the algebra and the computations. Use of experimental designs with known statistical analysis further minimizes the algebraic and numerical computation.

2.5.2 Augmented Randomized Complete Block Design

Let us consider the experimental situation where w test treatments are to be compared with u control treatments using n experimental units arranged in b blocks such that j^{th} block is of size $k_j (> u)$. For an augmented randomized complete block design which we will call augmented block design (ABD), each of the control treatments is replicated r times and occurs once in every block and test treatments occur once in one of the blocks. Therefore, it can easily be seen that in the j^{th} block there are $k_j - u = n_j$ test treatments. The randomization procedure is given as follows:

1. Randomly allot u controls to u of the k_j experimental units in each block.
2. Randomly allot the w test treatments to the remaining experimental units.
3. If a new treatment appears more than once, assign the different entries of the treatment to a complete block at random with the provision that no treatment

appears more than once in a complete block until that treatment occurs once in each of the complete blocks.

For augmented randomized complete block design standard errors for comparing mean differences are generated from the mean square error of the original RCBD calculated as in chapter 2.

Standard Errors

1. Between two control treatment means

$$SE(1) = \sqrt{\frac{2MSE}{b}}$$

2. Between two test treatments in the same block

$$SE(2) = \sqrt{2MSE}$$

3. Between two test treatments not in the same block

$$SE(3) = \sqrt{2MSE(1 + 1/u)}$$

4. Between a test treatment and a control treatment

$$SE(3) = \sqrt{2MSE(1 + 1/b + 1/u - 1/ub)}$$

The analysis of variance of the data generated from an augmented block design with $v = u + w$ treatments comprising of w tests and u controls arranged in b blocks having k_1 plots in block 1, k_2 plots in block 2, and so on, and k_b plots in block b, such that $k_1 + k_2 + \cdots + k_b = n$, the total number of plots in the design, is shown in Table 6

Table 6: Analysis of Variance (ANOVA) table for augmented block design

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Sum of squares
Blocks (eliminating treatments)	b -1	ASSB	MSSB
Treatments (eliminating blocks)	v -1	ASST	
Among Tests	w-1	SST	MSST
Among Controls	u-1	SSC	MSSC
Tests vs Controls	1	SSTC	MSSTC
Error	n–v–b+1	SSE	MSE
Total	n -1	TSS	

3 CHAPTER THREE - METHODOLOGY

3.1 PROBLEM FORMULATION

We have w treatments divided into two disjoint sets, H and G of cardinality respectively v and p, $v + p = w$, $G \cap H = \phi$. The set H has v test treatments labelled as $1,2,\cdots,v$ while the set G has p control treatments labeled as $v+1, v+2, \cdots, v+p$. The problem here is to design an experiment to compare test treatments belonging to the set H with control treatments belonging to the set G with as high a precision as possible. Suppose that n experimental units are available for experimentation and these n experimental units can be arranged in b blocks of sizes $k_1, k_2, \cdots, k_b$, respectively, $k_1 + k_2 + \cdots + k_b = n$.

As described above, the contrasts of interest in the experiment are $t_g - t_h$, $g \in G, h \in H$. Comparisons of treatments within G and within H are of secondary importance, though in many practical situations, the comparisons among treatments within the groups are also important, though not as important as the between group comparisons. The contrasts of interest then are also $t_g - t_{g'}, \mathrm{g} \neq \mathrm{g}' \in \mathrm{G}$ and $t_h - t_{h'}, \mathrm{h} \neq \mathrm{h}' \in \mathrm{H}$.

As such we are interested in a design that permits the estimation of all the $v - 1$ linearly independent treatment contrasts called a connected design with Rank (**C**) = v - *1*. Further, it would also be desirable that the comparisons of interest are estimated through the design with the same variance. The precision of other comparisons is of

no consequence to us although these are also estimable through a connected design. A design is said to be variance balanced for the estimation of test treatments versus control treatments comparisons if it permits the estimation of these comparisons with the same variance and the covariance between any two estimated test treatments versus control treatments comparisons is also the same.

3.2 MODEL AND ANALYSIS OF VARIANCE

A statistical model is actually a linear relation of the effects of different levels of a number of factors involved in an experiment along with one or more terms representing error effects. The effects of any factor can be either fixed or random but the error effects are always random and may belong either to a finite or infinite population. A model in which each of the factors has fixed effects and only the error effects are random is called a fixed model. Models in which some factors have fixed effects and some random effects are called mixed models. In the same context, models where all the factors have random effects are called random models.

In fixed effects models, the main objectives are to estimate the effects, find a measure of variability among the effects of each of the factors and finally find the variability among the error effects. In this study we shall restrict ourselves to fixed effects model since it is enough to achieve our objectives.

If the treatment t is assigned to the i^{th} plot of the j^{th} block $(1 \le t \le v;\ 1 \le i \le n_{tj};$ $1 \le j \le b)$, we shall have y_{tij} denote the corresponding random variable: we assume the usual fixed effects, additive, and heteroscedastic linear model without interaction.

$$y_{ijk} = \mu + t_i + b_j + \varepsilon_{ijk}$$

Where

μ is the grand mean

t_i is the effect of treatment i w.r.t. grand mean

b_j is the effect of block j w.r.t. grand mean

ε_{ijk} is error and is assumed to be N(0, $\sigma^2 k_j^{\alpha}$)

Here $\alpha \ge 0$ is a scalar constant, generally unknown. This model is in fact a generalization of Fairfield Smith's Variance Law. The value of α may be estimated from the uniformity trial data. Lee and Jacroux (1987), Gupta, Das and Dey (1991), Das, Gupta and Das (1992), Parsad and Gupta (1994a, 1994b), Parsad, Gupta and Singh (1996) and Srivastava, Gupta and Parsad (2000) have earlier studied this model. For $\alpha = 0$ we get the usual homoscedastic model.

The assumptions on the behavior of ε_{ijk} are necessary for drawing inference by adopting known statistical methodology. The methodology that is adopted is the analysis of variance technique where inference is drawn by using F test. For the F test it is necessary that the observations, that is, the error components should be

normally and independently distributed with a common variance. Further assumption that has to be made in the model is that the effects are additive.

3.3 MONTE-CARLO SIMULATION

Monte Carlo Methods are a class of computational algorithms that rely on repeated random sampling to compute their results, Fishman (1995). Monte Carlo methods are often used when simulating physical and mathematical systems. Because of their reliance on repeated computation and random or pseudo-random numbers, Monte Carlo methods are most suited to calculation by a computer. Monte Carlo methods tend to be used when it is infeasible or impossible to compute an exact result with a deterministic algorithm.

To compare the efficiencies of ABD and RBIBD we needed to simulate data from the standard normal distribution and uniform distribution. We used Genstat's Release 12 *GRANDOM* procedure to generate pseudo-random numbers from the Normal and uniform distributions. The *RESTRICT* directive was used to restrict the value of yield to greater than zero and less that 200tch since we don't expect to have negative yield or yield greater than 200 tonnes of cane per hectare.

3.4 TREATMENTS DISTRIBUTION CHARACTERISTICS

We used five hypothetical families and two controls whose desired characteristic yield, follows a normal distribution with distinct means and standard deviations as shown:

Family 1 ~ $N(80, 10^2)$

Family 2 ~ $N(60, 5^2)$

Family 3 ~ $N(50, 4^2)$

Family 4 ~ $N(100, 4^2)$

Family 5 ~ $N(30, 2^2)$

Control 1 ~ $N(60,1.5^2)$

Control 1 ~ $N(62,2^2)$

Where Control 1 and Control 2 are the standard commercial sugarcane varieties

Since we know all the parameters of the distribution for each treatment, we will generate normal random numbers using Genstat Release 12 GRANDOM procedure to fit the two experimental designs. Each block will be given a distinct environmental effect that either suppresses or enhances the treatment mean. From the generated figures we will carry out analysis of variance to test for the difference of treatment effects. The results of the ANOVA will be compared to the expected results through a paired sample t-test.

3.5 REINFORCED BALANCED INCOMPLETE BLOCK DESIGN

A Balanced Incomplete Block design with parameters $v = 5$, $b = 5$ and $k = 4$ was generated using IASRI design resource server in accordance to Rathore et al (2006) and the output is given below.

"OUTPUT GENERATED BY IASRI DESIGN RESOURCE SERVER"

V=5 B=5 K=4 A-Efficiency=1.0 D-Efficiency=1.0

$\lambda = 4$

r1=4 r2=4 r3=4 r4=4 r5=4

BLOCK DIAGRAM

Block 1 (	1	2	3	5	)
Block 2 (	2	3	4	5	)
Block 3 (	1	2	3	4	)
Block 4 (	1	3	4	5	)
Block 5 (	1	2	4	5	)

CONCURRENCE MATRIX

	T2	T3	T4	T5
T1	3	3	3	3
T2		3	3	3
T3			3	3
T4				3

The generated design was reinforced by adding two control varieties (i.e. Control 1 and Control 2) in each of the 5 blocks making the design a Reinforced Balanced Incomplete Block Design (RBIBD). The treatments were randomized in each block.

BLOCKS						
1	Plot 1 Family 1	Plot 2 Family 2	Plot 3 Family 3	Plot 4 Family 5	Plot 5 Control 1	Plot 6 Control 2
2	Plot 12 Family 2	Plot 11 Family3	Plot 10 Family4	Plot 9 Family 5	Plot 8 Control 1	Plot 7 Control 2
3	Plot 17 Family 1	Plot 18 Family2	Plot 16 Family 3	Plot 13 Family 4	Plot 15 Control 1	Plot 14 Control 2
4	Plot 24 Family 1	Plot 23 Family 3	Plot 22 Family 4	Plot 20 Family 5	Plot 19 Control 1	Plot 21 Control 2
5	Plot 28 Family 1	Plot 26 Family 2	Plot 27 Family 4	Plot 30 Family 5	Plot 29 Control 1	Plot 25 Control 2

Figure 3: Plots layout for Reinforced Incomplete Block Design (RBIBD)

Table 7: Frequencies for different treatments in the blocks of RBIBD

	Treatments							
Block	Family 1	Family 2	Family 3	Family 4	Family 5	Control 1	Control 2	Block Size
1	1	1	1	0	1	1	1	6
2	0	1	1	1	1	1	1	6
3	1	1	1	1	0	1	1	6
4	1	0	1	1	1	1	1	6
5	1	1	0	1	1	1	1	6
Replication	4	4	4	4	4	5	5	

Data was simulated using Monte Carlo simulation as described in Chapter 1. Each treatment had unique properties as shown in Figure 5. Family 1 had the largest dispersion due to large variance while Control 1 had the least dispersion due to small variance

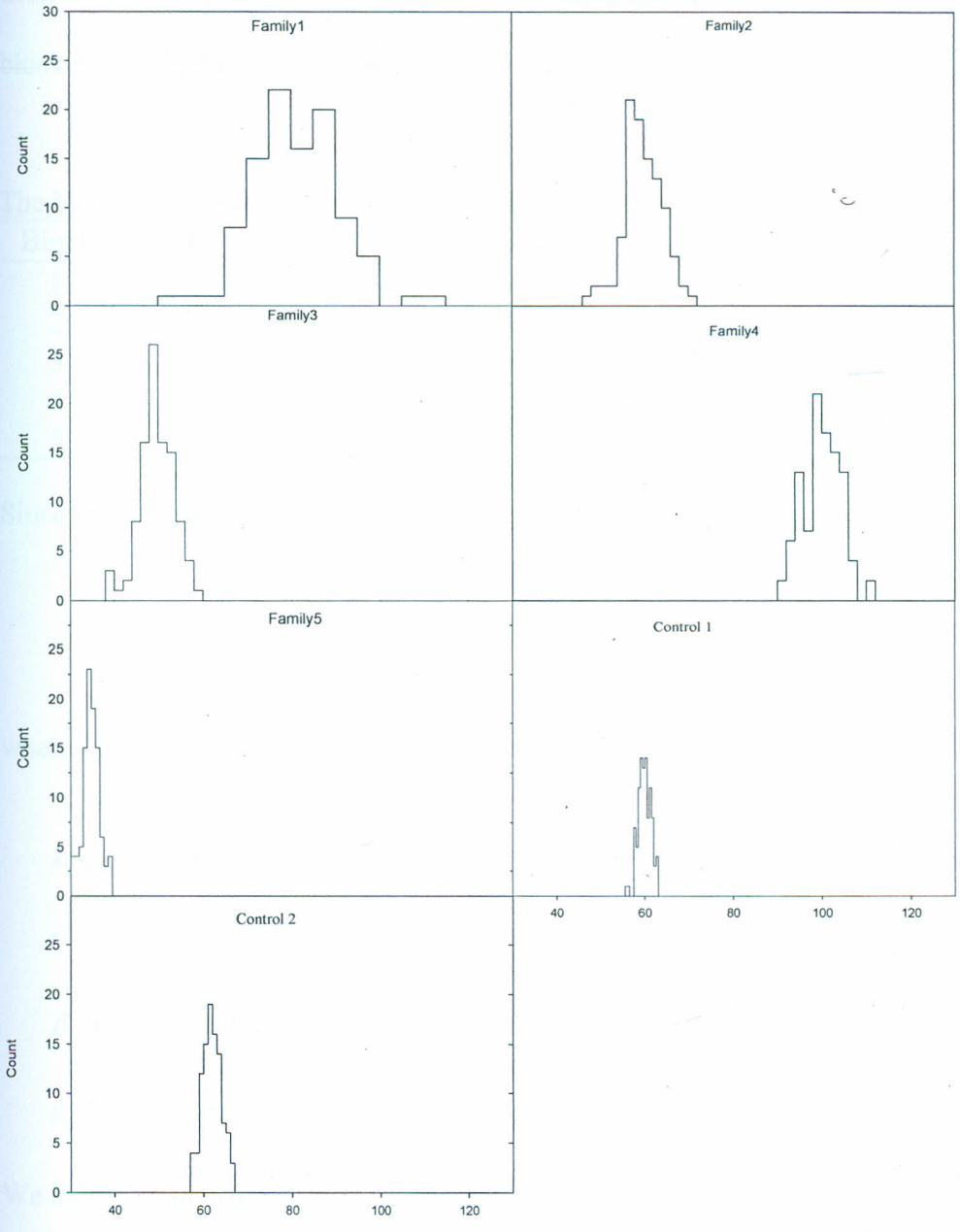

Figure 4: Histogram of data generated for various treatments (Sample size 100)

The data generated was inserted in the design and adjusted for block effects. For ease of computation, block effects were assumed to be uniformly distributed in the range [-5, 5]. The block effects were again generated through Monte Carlo simulation as

described in chapter 1. From the generated data, the reinforced balanced incomplete block design had the following block effects.

The blocks were assumed to have the following properties:

Block	Random No	Block Effect
1	3	Enhance the mean by 3 units
2	-3	Depress the mean by 3 units
3	1	Enhance the mean by 1 units
4	4	Enhance the mean by 4 units
5	-5	Depress the mean by 5 units

Since in this analysis we assume additive and homoscedastic linear model

$$Y_{ij} = \mu + t_i + B_j + \varepsilon_{ij}$$

where,

μ is the grand mean

t_i is the effect of treatment i

B_j is the effect of block j

ε_{ij} is error and is assumed to be $N(0,\sigma^2)$

We included an error term in the generated data. This error term was distributed as N(0,25) and was added to each data point (plot).

3.6 AUGMENTED BLOCK DESIGN

An Augmented Block Design with parameters $v = 7$, $w = 5$, $u = 2$, and $b = 3$ was generated using IASRI design resource server. The two control treatments were laid in a randomized complete block design with three replicates. In each block the

treatments were augmented with inclusion of two test families randomly allotted to block 1 and 2 and one test family randomly allotted to block 3. This was done in conformity to the optimum replication number r of each check varieties in every block of the augmented block design that was given by Gupta and Parsad (2001) as shown below:

$$r = (c+b-1)^{1/2} * \left(\frac{t}{c*b}\right)^{1/2}, \text{ for } t > (c+b-1) \;\&\; c > 1$$

Else $$r = \left(\frac{t}{b}\right)^{1/2}$$

Where: c = *no. of check varieties*
b = *no. of blocks*
t = *no. of test varieties*

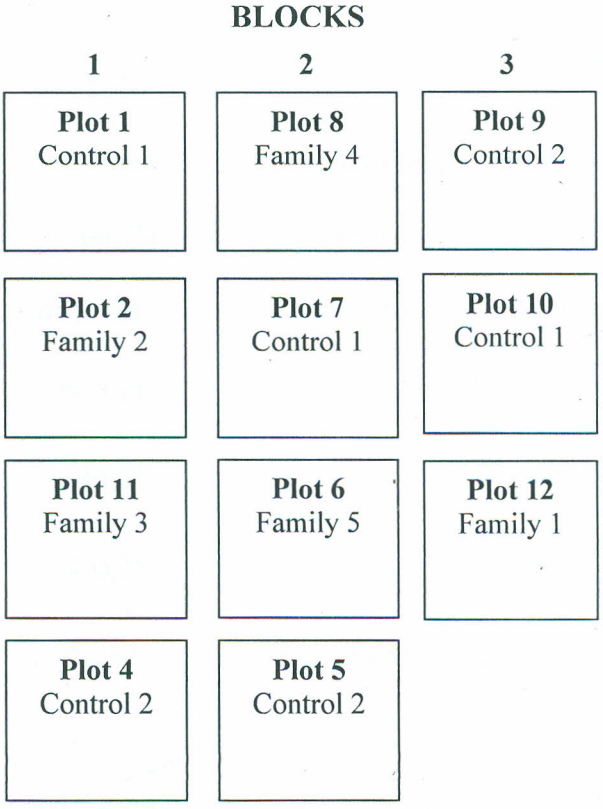

Figure 5: Augmented Block Design (ABD)

For uniformity and so as to have a basis for comparison of RBIBD and ABD, data which was simulated using Monte Carlo simulation in chapter 2 was used for each treatment.

The data generated was inserted in the design and adjusted for block effects. For ease of computation, block effects were assumed to be uniformly distributed in the range [-5, 5]. The block effects were again generated through Monte Carlo simulation as described in chapter 1. From the generated data, the augmented block design had the following block effects.

Block	Random No	Block Effect
1	3	Enhance the mean by 3 units
2	-3	Depress the mean by 3 units
3	1	Enhance the mean by 1 units

From the experiment it was observed that the results of the augmented block designs are dependent on the experiment error term. Thus to explore different scenarios we added three different random errors terms to the simulated data, namely;

A. ε_{ij} ~N(0,1)

B. ε_{ij} ~N(0,5)

C. ε_{ij} ~N(0,25)

Data were analyzed for each level of error term

The generated data were analyzed using SAS PROC GLM as shown below

```
TITLE 'SIMULATED DATA ON AUGMENTED BLOCK DESIGN - N (0,1);
OPTIONS LS=75;
DATA YIELDABD;
INPUT Block Trtment$     Yield;
```

```
CARDS;
.
.
.
.
.
.
;
proc glm;
class Block Trtment;
model yield = Block Trtment/ss2;
lsmeans Trtment/stderr pdiff ;
contrast 'controls' Trtment 0 0 0 0 0 -1 1;
contrast 'tests'  Trtment 1 -1 0 0 0 0 0 ,
                          Trtment 1 1 -2 0 0 0 0 ,
                          Trtment     1 1 1 -3 0 0 0 ,
                          Trtment     1 1 1 1 -4 0 0 ,
contrast 'tests vs controls' Trtment 2 2 2 2 2 -5 -5;
means Trtment/ duncan;
run;
```

4 CHAPTER FOUR – RESULTS AND DISCUSSION

4.1 REINFORCED BALANCED INCOMPLETE BLOCK DESIGN

The means generated from RBIBD are shows in **Error! Not a valid bookmark self-reference.**. We note how unhelpful it is to compare treatment means from the table below. Treatments that fell on blocks that depress yield will have less mean as compared to those treatments that fell on blocks that enhance yield due to the effect of treatment-block interaction. The only exception are the controls (Control 1 and Control 2), this is because they appeared in all blocks thus comparing them will give the true performance. The treatment means have to be adjusted to remove the block effects.

Table 8: RBIBD treatments performance per block and their related means

	Treatments								
Block	Family 1	Family 2	Family 3	Family 4	Family 5	Control 1	Control 2	Block Total	Block Means
1	85.99	55.80	51.07	-	37.76	59.59	65.77	355.98	59.33
2	-	58.60	49.87	91.08	34.91	63.11	65.65	363.22	60.54
3	85.89	63.43	48.88	104.52	-	61.15	53.48	417.35	69.56
4	90.85	-	54.46	100.84	27.02	68.08	60.62	401.87	66.98
5	80.96	48.73	-	93.73	23.95	51.46	52.27	351.10	58.52
Total	343.69	226.56	204.28	390.17	123.64	303.39	297.79		
Means	85.92	56.64	51.07	97.54	30.91	60.68	59.56		

Computing

Taking $i=(1,2,\ldots 5)$ to represent family1 to family 5 and $m=(1\ and\ 2)$ to represent Control 1 and Control 2 respectively.

G = Grand total of observations

μ = Grand mean = G/n, where n=total number of observations

T_i = Sum of observations for treatment i, $(i = 1, 2, \ldots, 7)$

B_j = Sum of observations in block j, $(j = 1, 2, \ldots, 5)$

$CF = G^2/n$

Q_i = adjusted i^{th} treatment total

Thus

G = 1889.52, n=30, μ =63.19 and CF= $1889.52^2/30 = 119009.53$

Since k=4, p=2, r=4, λ=3 and b=5 we have

$$(k+p)r - r + \lambda = 23$$

$$\left(r - \frac{b\lambda}{r}\right) = 0.25$$

Table 9: Calculation of treatment effects, adjusted treatment total and adjusted treatment sum of squares

Block no	Treat no	T_i	B_j	Block No.'s in which treatment i occurs	Total of Blocks in which treatment i occurs	$\sum_{j(i)} B_j / k$	Q_i	t_i	$t_i Q_i$
1	1	343.69	355.98	1,3,4,5	1526.30	254.38	89.31	23.24	2075.05
2	2	226.56	363.22	1,2,3,5	1487.65	247.94	-21.38	-5.64	120.60
3	3	204.28	417.35	1,2,3,4	1538.42	256.40	-52.12	-13.66	711.99
4	4	390.17	401.87	2,3,4,5	1533.54	255.59	134.58	35.05	4716.43
5	5	123.64	351.10	1,2,4,5	1472.17	245.36	-121.72	-31.82	3872.67
		T_m				$\sum_{j(m)} B_j / b$	Q_m	t_m	$t_m Q_m$
	6	303.39		1,2,3,4,5	1889.52	314.92	-11.53	-2.31	26.59
	7	297.79		1,2,3,4,5	1889.52	314.92	-17.13	-3.43	58.69

Total SS (TSS) = $\sum_{ij} y_{ij}^2 - C.F = 12{,}547.04$

Treatment SS unadjusted (SST_u) = $\left(\sum_{i=1}^{5} T_i^2/r + \sum_{m=6}^{7} T_m^2/b\right) - C.F = 11{,}810.86$

Block SS unadjusted (SSB_u) = $\left(\sum_{j=1}^{5} B_j^2/k+p\right) - C.F = 590.85$

Treatment SS adjusted (SST_A) = $\sum_{i=1}^{5} \hat{t}_i Q_i + \sum_{m=6}^{7} \hat{t}_m Q_m = 11{,}582$

Error SS (ESS) = $TSS - SSB_U - SST_A = 374.19$

Block SS adjusted (SSB_A) = $SST_A + SSB_u - SST_u = 361.99$

Plugging in the figures we get the analysis of variance table below

Table 10: Analysis of Variance Table for Simulated RBIBD

Source of Variation	DF	Sum of Squares	Mean Sum of Squares	F Value	Pr > F
Blocks	4	361.99	90.50	4.6	0.0092
Treatments	6	11582.00	1930.33	98.02	<.0001
Among Tests	4	11455.66	2863.91	145.42	<.0001
Among Controls	1	3.14	3.14	0.16	0.6943
Tests vs Controls	1	123.21	132.21	6.26	0.0217
Error	19	374.19	19.69		
Corrected Total	29	12547.04			

From Table 10 it can be seen that difference among test families is statistically significant (p-value < 0.0001) but difference among controls is not (p-value 0.6943). The difference between control and tests families is also significant (p-value 0.0217). This is as we expected and we can conclude that in this case RBIBD is able to bring out the differences.

Table 11 gives the adjusted treatment means for all treatments. These are the observed treatment means adjusted for block effect using the formula:

Adjusted treatment means for treatment i = i^{th} treatment effect + grand mean

Table 11: Adjusted treatments means for RBIBD

Treatment	Grand Mean (μ)	Treatment effect (t_i)	Adjusted treatment mean ($\overline{T}_i$)
Family1	63.19	23.24	86.43
Family2	63.19	-5.64	57.55
Family3	63.19	-13.66	49.53
Family4	63.19	35.05	98.24
Family5	63.19	-31.82	31.37
Control 1	63.19	-2.31	60.88
Control 2	63.19	-3.43	59.76

4.1.1 Relative Efficiency

To get the relative efficiency of the reinforced BIBD we used the same simulated data to construct a randomized complete block design (RCBD) with 4 replications. The replications have properties similar to the blocks of the RBIBD with 28 experimental units as opposed to 30 experimental units in RBIBD.

The analysis of the data produced the following results:

RCBD Analysis of variance

Table 12: The ANOVA table for simulated RCBD

Source	**DF**	**Sum of Squares**	**Mean Sum of Squares**	**F - Statistic**	**Pr >F**
Block	3	721.28	240.43	6.47	0.0037
Treatments	6	11077.53	1846.25	49.67	<.0001
Error	18	669.09	37.17		
Total	**27**	**12467.90**			

Table 11 shows that the treatments are statistically different at p-value <0.0001 while blocking was also significant (p-value 0.0037).

Multiple Comparisons

Duncan's Multiple Range Test for Yield: RCBD

Duncan Grouping	Mean	N	Treatment
A	98.388	4	Family4
B	83.773	4	Family1
C	63.930	4	Control 2
C	60.163	4	Control 1
D C	58.183	4	Family2
D	50.480	4	Family3
E	32.943	4	Family5

Means with the same letter are not significantly different i.e. family4, Family1 and Family5 are each different from all the other treatments. Family2, Control 2 and C 1 are not statistically different while Family2 and Family3 are not statistically different.

Table 13: Comparison of the treatments means

Treatment	RBIBD Means (Tons Ha^{-1})	RCBD Means (Tons Ha^{-1})	Simulated Means (Tons Ha^{-1})	Assumed Means (Tons Ha^{-1})
Family1	86.43	83.77	81.19	80.00
Family2	57.55	58.18	59.86	60.00
Family3	49.53	50.48	49.65	50.00
Family4	98.24	98.38	100.20	100.00
Family5	31.37	32.94	29.84	30.00
Control 1	60.88	60.16	59.96	60.00
Control 2	59.76	63.93	61.91	62.00

To compare the efficiencies of the two designs we use the Fishers approach by calculating the 'amount of information' which the estimated difference between two

treatments means supplies about the true difference. Thus the relative efficiency of the RBIBD design to the RCBD design is estimated as:

$$\frac{(n_1+1)(n_2+3)s_2^2}{(n_2+1)(n_1+3)s_1^2}$$

Where

n_1 are the RBIBD error degrees of freedom

n_2 are the RCBD error degrees of freedom

s_1^2 is the RBIBD error mean sum of squares

s_2^2 is the RCBD error mean sum of squares

The required parameters are found in the ANOVA tables. For our scenario, $n_1 = 19$, $n_2 = 18$, $s_1 = 19.69$ and $s_2 = 37.17$. Thus RBIBD was about 3 times more efficient than RCBD. We note that this efficiency is dependent on the random numbers generated and as such will tend to vary.

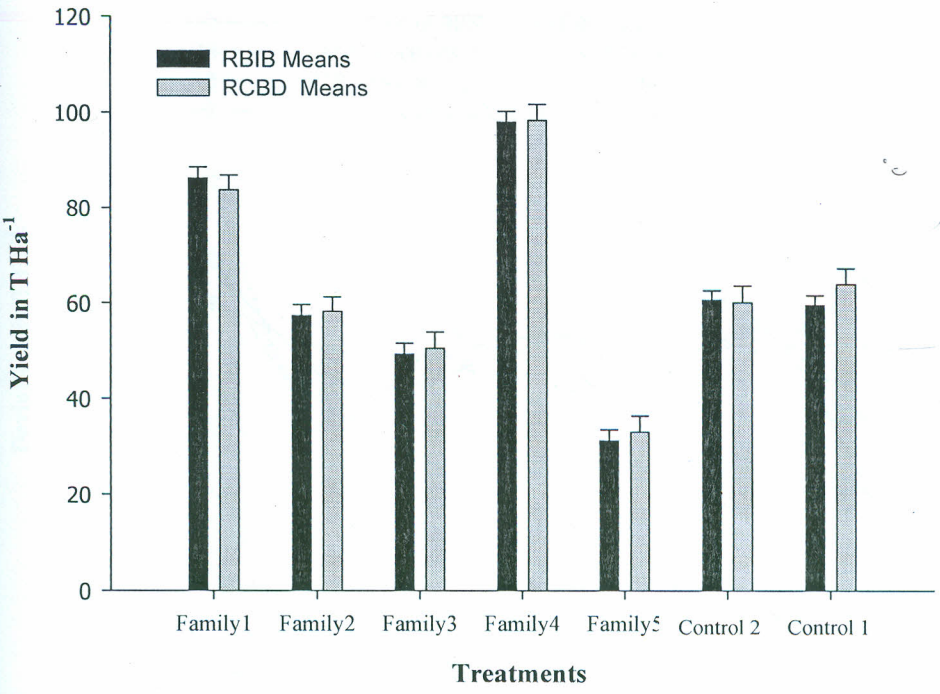

Figure 6: Comparison of RBIBD and RCBD generated means

From Figure 6 above it can be noted that the means generated by the two designs are not significantly different. The error bars also show that the precision of the RBIBD is higher than that of RCBD. There is also a marked difference from generated mean of Control 1 though the difference is still within the margin of error.

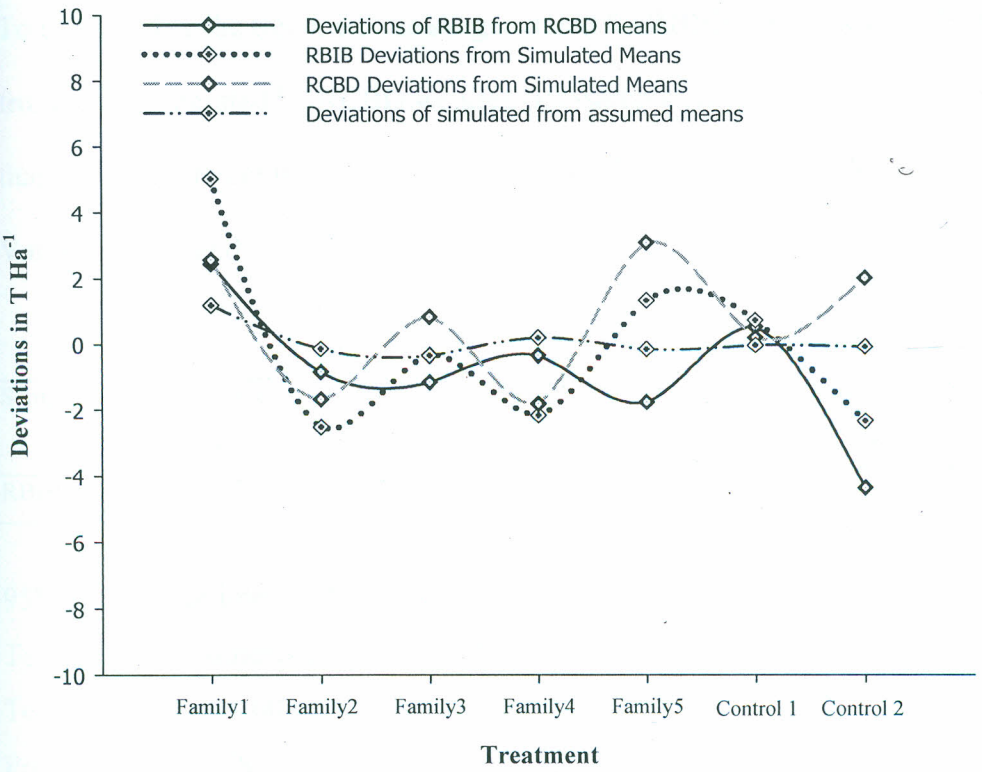

Figure 7: Comparison of mean deviations

From Figure 1 above we note that there is considerable difference across treatments for deviations of RBIBD means from RCBD means, RBIBD means from simulated means, and RCBD means from simulated means. The small difference across treatments for deviations of the simulated means from the assumed means can be attributed to the fact that the simulated means are an average of 100 simulations. This difference would have been minimized if the observation were averaged over many simulations.

4.1.2 Paired Sample T-Test

To test whether the treatment means generated by RBIBD are significantly different from those generated by RCBD we used one-sample paired t-test. The test showed that we cannot reject the null hypothesis (p=0.358). See results below.

Variate: Plot Yield

Sample	Size	Mean	Variance	Standard deviation	Standard error of mean
RBIBD –RCBD	7	-0.7886	4.387	2.095	0.7917

95% confidence interval for mean: (-2.726, 1.149)

Test of null hypothesis that mean of RBIBD –RCBD is equal to 0

Test statistic t = -1.00 on 6 d.f.

Probability = 0.358

4.2 AUGMENTED BLOCK DESIGN

4.2.1 Scenario One: ε_{ij} ~N(0,1)

When the error term is distributed as N(0,1)

Table 14: ANOVA table for augmented block design when ε_{ij} ~ N(0,1)

Source of Variation	DF	Sum of Squares	Mean Sum of Squares	F Value	Pr > F
Blocks	2	33.84	16.92	46.96	0.0209
Treatments	6	2889.52	481.59	1336.44	0.0007
Among Tests	4	2681.47	670.37	1860.32	0.0005
Among Controls	1	3.11	3.11	8.63	0.9900
Tests vs Controls	1	52.75	52.75	156.37	0.0068
Error	2	0.7207	0.3604		
Corrected Total	10	2924.09			

From Table 14 it can be seen that difference among test families is significant (p-value = 0.0005) but difference among control is not (p-value 0.9900). The controls were also statistically different from the test families (p-value 0.0068). This is as expected and we can conclude that in this case ABD is able to bring out the differences.

Table 15: The Least squares mean yield for ABD Yield: When ε_{ij} ~N(0,1)

Treatment	Lsmean yield	Standard Error	Pr>|t|
Family1	80.32	0.69316	>.0001
Family2	62.77	0.69316	0.0001
Family3	51.82	0.69316	0.0002
Family4	99.77	0.69316	>.0001
Family5	30.37	0.69316	0.0005
Control 1	59.81	0.34658	>.0001
Control 2	61.25	0.34658	>.0001

The same treatments and data scheme when applied to a randomized complete block design with three blocks gives the following results:

Table 16: ANOVA table for randomized complete block design when ε_{ij} ~N(0,1)

Source	**DF**	**Sum of Squares**	**Mean Sum of Squares**	**F – Statistic**	**Pr >F**
Treatments	6	9086.94	1514.49	1019.63	<.0001
Blocks	2	163.71	81.85	55.11	<.0001
Error	12	17.82	1.49		
Total	**20**	**9268.47**			

Table 16 shows that the treatments were statistically different at p-value <0.001

Table 17: The Least squares mean yield for RCBD Yield: When ε_{ij} ~N(0,1)

Treatment	Lsmean yield	Standard Error	Pr>|t|
Family1	81.59	0.70364	>.0001
Family2	60.30	0.70364	>.0001
Family3	49.70	0.70364	>.0001
Family4	100.59	0.70364	>.0001
Family5	30.20	0.70364	>.0001
Control 1	59.94	0.70364	>.0001
Control 2	61.52	0.70364	>.0001

When we use the RCBD the p-values for treatment and block are highly significant ($p < 0.001$, Table 10).

4.2.2 Scenario Two: ε_{ij} ~N(0,5)

When the error term is distributed as N(0,5)

Table 18: ANOVA table for augmented block design when ε_{ij} ~ N(0,5)

Source of Variation	DF	Sum of Squares	Mean Sum of Squares	F Value	Pr > F
Blocks	2	33.23	16.61	6.59	0.1318
Treatments	6	2883.43	480.57	190.62	0.0052
Among Tests	4	2679.53	669.88	265.72	0.0038
Among Controls	1	8.28	8.28	3.29	0.2116
Tests vs Controls	1	8.83	8.83	3.50	0.2021
Error	2	5.04	2.52		
Corrected Total	10	2921.70			

Table 19: The Least squares mean yield for ABD Yield: When ε_{ij} ~ N(0,5)

Treatment	Lsmean yield	Standard Error	Pr>|t|
Family1	80.71	1.83341	0.0005
Family2	59.56	1.83341	0.0009
Family3	49.74	1.83341	0.0014
Family4	97.63	1.83341	0.0004
Family5	28.96	1.83341	0.0040
Control 1	60.31	0.91671	0.0002
Control 2	62.66	0.91671	0.0002

When the variance of the error term increases from 1 to 5, the differences among test families remain significant while the difference for test vs control treatments increases move from being significant. The p-value for blocking effect increases from 0.02 to 0.13 implying that blocking becomes ineffective at a higher error variance. The structure of the Duncan's multiple range test remains unchanged though the treatments means are different from scenario one especially among the test families. The control treatments are stable. This can be attributed to lack of replication in the test families and thus the errors are not averaged out.

When we used the same treatments and data scheme as used in scenario 2 of the ABD and applied to a randomized complete block design with three blocks, the following results were observed:

Table 20: ANOVA table for randomized complete block design when ε_{ij} ~ N(0,5)

Source	DF	Sum of Squares	Mean Sum of Squares	F - Statistic	Pr >F
Treatments	6	8793.05	1465.51	334.94	<.0001
Blocks	2	174.12	87.06	19.90	0.0002
Error	12	52.50	4.37		
Total	**20**	**9019.67**			

In the RCBD the difference between treatments remains highly significant but as shown in Table 21 below the treatments standard error increases to 1.2.

Table 21: The Least squares mean yield for RCBD Yield: When ε_{ij} ~ N(0,5)

Treatment	Lsmean yield	Standard Error	Pr>|t|
Family1	80.45	1.20767	>.0001
Family2	62.60	1.20767	>.0001
Family3	50.72	1.20767	>.0001
Family4	100.66	1.20767	>.0001
Family5	30.56	1.20767	>.0001
Control 1	60.17	1.20767	>.0001
Control 2	61.68	1.20767	>.0001

It can be noted the results are very similar to scenario one RCBD results. This is because even though the errors have different variances, when averaged they give the nearly the same figure in the neighborhood of zero.

4.2.3 Scenario Three: ε_{ij} ~N(0,25)

When the error term is distributed as N(0,25)

Table 22: ANOVA table for augmented block design when ε_{ij} ~ N(0,25)

Source of Variation	DF	Sum of Squares	Mean Sum of Squares	F Value	Pr > F
Blocks	2	56.61	28.30	0.89	0.5285
Treatments	6	3315.80	552.63	17.42	0.0553
Among Tests	4	3113.29	778.32	24.53	0.0396
Among Controls	1	37.35	37.35	1.18	0.3913
Tests vs Controls	1	14.71	14.71	0.46	0.5662
Error	2	63.45	31.73		
Corrected Total	10	3435.87			

In this scenario the augmented block design is unable to bring out the difference among the test vs control treatments (p-value 0.56). The difference among test treatments is significance but blocking becomes more ineffective (p-value > 0.5).

Table 23: The Least squares mean yield for ABD Yield: When ε_{ij} ~ N(0,25)

Treatment	Lsmean yield	Standard Error	Pr>|t|
Family1	80.45	6.50407	0.0065
Family2	62.46	6.50407	0.0107
Family3	51.46	6.50407	0.0156
Family4	100.81	6.50407	0.0041
Family5	25.63	6.50407	0.0588
Control 1	59.3	3.25203	0.0030
Control 2	64.29	3.25203	0.0025

When we used the same treatments and data scheme as used in scenario 3 of the ABD and applied to a randomized complete block design with three blocks, the following results were observed:

Table 24: ANOVA table for randomized complete block design when ε_{ij} ~N(0,25)

Source	DF	Sum of Squares	Mean Sum of Squares	F - Statistic	Pr >F
Treatments	6	7986.96	1331.16	59.93	0.0008
Blocks	2	600.30	300.15	13.51	<.0001
Error	12	266.55	22.21		
Total	**20**	**8853.80**			

Table 25: The Least squares mean yield for RCBD Yield: When ε_{ij} ~N(0,25)

Treatment	Lsmean yield	Standard Error	Pr>|t|
Family1	83.22	2.72104	>.0001
Family2	62.02	2.72104	>.0001
Family3	52.14	2.72104	>.0001
Family4	100.42	2.72104	>.0001
Family5	35.39	2.72104	>.0001
Control 1	59.99	2.72104	>.0001
Control 2	64.14	2.72104	>.0001

For the RCBD, Table 24 shows that the difference in treatments effects remains significant for both N(0,1) and N(0,5). The least squares means results show that the RCBD is still robust in bringing out difference among treatments

4.2.4 Relative Efficiency

To compare relative efficiencies of RBIBD design to RCBD design we again use Fishers method given earlier as:

$$\frac{(n_1+1)(n_2+3)s_2^2}{(n_2+1)(n_1+3)s_1^2}$$

Where

n_1 are the ABD error degrees of freedom

n_2 are the RCBD error degrees of freedom

s_1^2 is the square of ABD error mean square

s_2^2 is the square of RCBD error mean square

The required parameters are found in the ANOVA table. Calculating the relative efficiency for our three scenarios we find that the ABD has the following efficiencies (Table 26)

Table 26: Relative efficiency of ABD against RCBD

SCENARIO	DESIGN	ERROR DF	ERROR MS	EFFICIENCY OF ABD AGAINST RCBD
N(0,1)	ABD	2	0.36	11.86
	RCBD	12	1.49	
N(0,5)	ABD	2	2.52	2.08
	RCBD	12	4.37	
N(0,25)	ABD	2	13.73	1.81
	RCBD	12	22.21	

From Table 26 it is evident that the relative efficiency of ABD against RCBD depended on the variance of the error term. ABD's are relatively more efficient than RCBD for standard normal error but this efficiency deteriorates rapidly as the variance of the error term increases. From the error variance of 5 to 25 the relative efficiency drops remains stable. We would expect the relative efficiency of ABD to be the same as the efficiency if RCBD as the variance of the error $\sigma^2 \to \infty$. As such it would be wise to use augmented block design even when variance of experimental error is unknown as is the case with agricultural experiments.

Table 27: Comparison of ABD and RCBD means and paired sample t-test result

ε_{ij} ~ N(0,1)		ε_{ij} ~ N(0,5)		ε_{ij} ~ N(0,25)	
ABD	RCBD	ABD	RCBD	ABD	RCBD
80.32	81.59	80.71	80.45	80.45	83.22
62.77	60.30	59.56	62.60	62.46	62.02
51.82	49.70	49.74	50.72	51.46	52.14
99.77	100.59	97.63	100.66	100.81	100.42
30.37	30.20	28.96	30.56	25.63	35.39
59.81	59.94	60.31	60.17	59.30	59.99
61.25	61.52	62.66	61.68	64.29	64.14
T-test: p=0.57		T-test: p=0.14		T-test: p=0.23	

From Table 26 above it can be seen that there is no significant difference between means generated by augmented block design and those generated by randomized complete block design. We can conclude that augmented block design and randomized complete block design are equally effective.

4.3 RBIBD and ABD

Reinforced incomplete block design (RBIBD) and augmented block design (ABD) are similar in construction given that in reinforced incomplete block design the control treatments are added to a design containing the test treatments whereas in the augmented block design the test treatments are added to a design containing the control treatments. The results produced by the two methods were not significantly different. See Table 28 below.

Table 28: Comparison of treatment means generated by both RBIBD and ABD

		Augmented Block Design Scenario		
Treatment	RBIBD	ABD-N(0,1)	ABD-N(0,5)	ABD-N(0,25)
Family1	86.22	80.32	80.71	80.45
Family2	57.34	62.77	59.56	62.46
Family3	49.32	51.82	49.74	51.46
Family4	98.03	99.77	97.63	100.81
Family5	31.17	30.37	28.96	25.63
Control 1	60.68	59.81	60.31	59.30
Control 2	59.56	61.25	62.66	64.29

The paired sample t-test generated the following results:

Sample	Size	Mean	Variance	Standard deviation	Std error of mean
RBIBD- ABD_N(0,1)	7	-0.5414	12.63	3.554	1.343

95% confidence interval for mean: (-3.828, 2.745)
H_0: RBIBD - ABD_N(0,1) = 0
Test statistic t = -0.40 on 6 d.f.
Probability = 0.701

Sample	Size	Mean	Variance	Standard deviation	Std error of mean
RBIBD- ABD_N(0,5)	7	0.3929	8.196	2.863	1.082

95% confidence interval for mean: (-2.255, 3.041)
H_0: RBIBD - ABD_N(0,5) = 0
Test statistic t = 0.36 on 6 d.f.
Probability = 0.729

Sample	Size	Mean	Variance	Standard deviation	Standard error of mean
RBIBD- ABD_N(0,25)	7	-0.2971	21.03	4.586	1.733

95% confidence interval for mean: (-4.538, 3.944)
H_0: RBIBD - ABD_N(0,25) = 0
Test statistic t = -0.17 on 6 d.f.
Probability = 0.870

Thus from the T-test results above, RBIBD and ABD are not significantly different and can be used interchangeably.

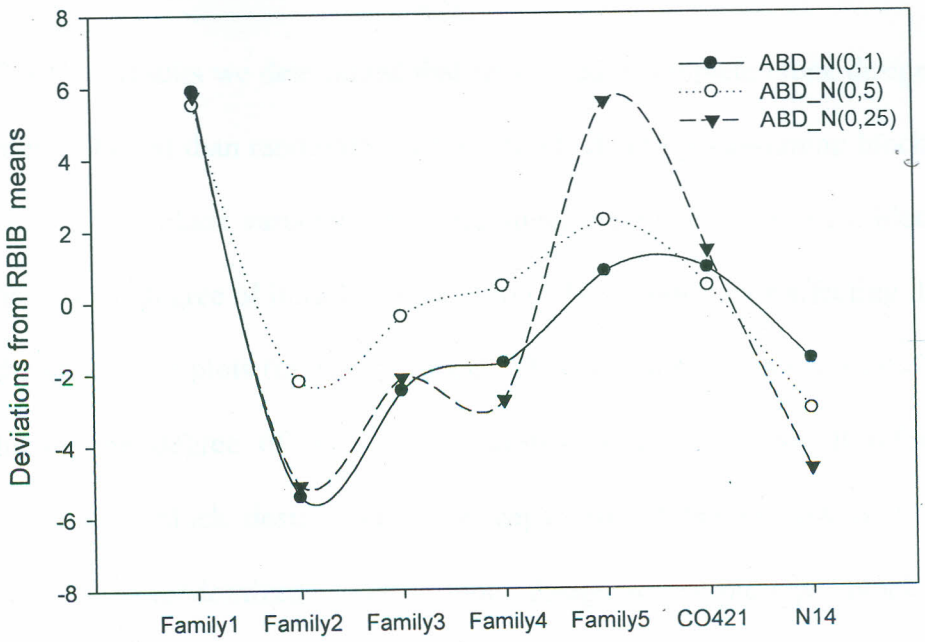

Figure 8: Deviations of various ABD means from RBIBD means

Figure 9 above shows that deviations were equally likely to occur in the both positive and negative side. RBIBD produced higher figures for family 1 though it is still within the family standard deviation.

5 CONCLUSION AND RECOMMENDATIONS

From the results we determined that reinforced incomplete block designs seems to be more efficient than randomized complete block designs assuming block homogeneity i.e. nil intra-block variation. In agricultural experiments setting a block will always have some degree of intra-block variation. The major factor affecting this variation is the number of plots (units) per block. The more the number of plots in a block: the higher the degree of intra-block variation and vice versa. Reinforced balanced incomplete block designs have the capability of having few plots per block as opposed to randomized complete block design whose plots per block is dictated by the number of treatments being evaluated. As such when evaluating large number of treatments (families), reinforced balanced incomplete block designs (RBIBD) will always be preferred to randomized complete block designs (RCBD).

We also determined that augmented block design seems to more efficient than randomized complete block design when error variance is small. In most cases under agricultural experiments the error variances are assumed small especially with blocking. The augmented block designs also have an advantage in that they require only one replication of the test treatments. When test materials are scarce or experimental units are limited the augmented block design can be used instead of randomized complete block design.

We recommend cautious use of the results given the difficulty of comparing experimental designs using simulated data. Concrete result in this study area will be very useful to plant breeders for efficient evaluation of genetic materials.

6 REFERENCES

Bose, R. C. (1939). "On construction of balanced incomplete block designs.". *Annal of Eugenics* , 353.

Box, G. E. (1958). A Note on the Generation of Random Normal Deviates. *Annals of Mathematical Statistics , 29* (2), 610-611.

Das, A. G. (1992). E-optimal block designs under heteroscedastic model. *Communication in Statistics – Theory & Methods , 21* (6), 1651-1666.

Das, M. N. (1979). *Design and Analysis of Experiments.* London: Wiley Eastern Ltd.

Das, M. N. (1958). On reinforced incomplete block designs. *Journal of Indian Society of Agricultural Statistics , 10*, 73-77.

Federer, W. T. (1956). Augmented (hoonuiaku) designs. *Hawaiian Planters' Record , 55*, 191-208.

Federer, W. T. (1961a). Augmented designs with one-way elimination of heterogeneity. *Biometrics* , 447- 473.

Federer, W. T. (1961b). Augmented designs with two-, three- and higher-way elimination of heterogeneity. *Biometrics , 3*, 166.

Fisher, R. A. (1940). "An examination of possible different solutions of a problem in incomplete blocks." . *Annals of Eugenics , 9*, 353-400.

Fisher, R. A. (1938). The mathematics of experimentation. *Nature , 142*, 442-443.

Fishman, G. (1995). *Monte Carlo: Concepts, Algorithms, and Applications.* New York: Springer Verlag.

Gupta, V. D. (1991). Universal optimality of block designs with unequal block sizes. *Statististical Probabilities Letters , 11*, 177-180.

Gupta, V. K. (2001). Block designs for comparing test treatments with control treatments – an overview. *Statistics and Applications* , 133-146.

ICAR 110 012, I. (2008). *Design Resources Server.* New Delhi: Indian Agricultural Statistics Research Institute.

Jacroux, M. (2002). On the determination and construction of A- and MV-optimal block designs for comparing a set of test treatments to set of standard treatments. *Journal of Statistical Planning and Inference , 106*, 191-204.

Jacroux, M. (1984). On the Optimality and Usage of Reinforced Block Designs for Comparing Test Treatments with a Standard Treatment. *Journal of the Royal Statistical Society- Series B , 46* (2), 316-322.

Jacroux, M. (2000). Some optimal orthogonal and nearly orthogonal block designs for comparing a set of test treatments to a set of standard treatments. *Sankhya Series B , 62*, 276-289.

Jaggi, S. G. (1996). A-efficient block designs for comparing two disjoint sets of treatments. *Communications-in-Statistics: Theory & Methods , 25* (5), 967-983.

Lee, K. a. (1987). On E-optimality of block designs having unequal block sizes. *Sankhya Series B , 49*, 126-136.

Majumdar, D. (1996). Optimal and efficient treatment control designs. In e. Ghosh and C.R. Rao, *Handbook of Statistics* (pp. 1007-1053). Amsterdam: Elsevier Science.

Majumdar, D. (1986). Optimal designs for comparisons between two sets of treatments. *Journal of Statistical Planning and Inference , 14*, 359-372.

Parsad, R. a. (2000). A Note on augmented designs. *Indian Journal of Plant Genetic Resources , 13* (1), 53-58.

Parsad, R. a. (1994b). A-optimality of group divisible designs with unequal block sizes for making test treatments - control comparisons under a heteroscedastic model. *International Journal of Mathematics and Statististical Science* , 5-31.

Parsad, R. a. (1994a). Optimal block designs with unequal block sizes for making test treatments control comparisons under a heteroscedastic model. *Sankhya Series B , 56* (3), 449-461.

Parsad, R. G. (1996). Trace optimal designs for comparing two disjoint sets of treatments. *Sankhya Series B , 58* (3), 414-426.

Rao, C. R. (1961). "Study of B. I. B. designs with replications 11 to 15.". *Sankhya* , 117-127.

Rathore, A. P. (2006). Computer Aided Search of Efficient Block Designs for Making All Possible Pairwise Treatment Comparisons. *Journal of Statistics & Applications* , 15-34.

Sharma, V. K. (2000). *Balanced Incomplete Block Designs.* New Delhi: IASRI.

Sprott, D. A. (1962). "Listing of B. I. B. designs from r=16 to 20.". *Sankhya* , 203-204.

Srivastava, R. G. (2000). *Studies on optimality of block designs for making test treatments-control comparisons.* New Delhi: I.A.S.R.I.

Yates, F. (1936). "A New Method of Arranging Variety Trials Involving a Large Number of Varieties" . *Journal of Agricultural Science* , 424-455.

Yong-Bi, F. A. (1999). Incomplete block designs for genetic testing: some practical considerations. *Canadian Journal of Forest Research* , 1871–1878 .