ESTIMATION OF THE PARAMETERS IN THE BALANCED INCOMPLETE BLOCK DESIGN

A Thesis Submitted to the
Graduate School of Natural and Applied Sciences of
Dokuz Eylül University
in partial Fulfillment of the Requirements for
the Degree of Master of Science in Statistics

by
Selma ERDOĞAN

December, 2001
İZMİR

Ms.Sc. THESIS EXAMINATION RESULT FORM

We certify that we have read the thesis, entitled **"ESTIMATION OF THE PARAMETERS IN THE BALANCED INCOMPLETE BLOCK DESIGN"** completed by Selma ERDOĞAN under supervision of Prof. Dr. Serdar KURT and that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Prof. Dr. Serdar KURT

Supervisor

Prof. Dr. Şevkinaz GÜMÜŞOĞLU

Committee Member

Prof. Dr. Nilgün MORALI

Committee Member

Approved by the
Graduate School of Natural and Applied Sciences

Prof.Dr.Cahit Helvacı

Director

ACKNOWLEDGMENTS

I wish to express my sincere gratitude to my supervisor Prof. Dr. Serdar KURT for his close guidance and help throughout the course of this work.

I am also grateful to Alper VAHAPLAR and Songül ERCAN for all their encouragement and assistance.

Selma ERDOĞAN

ABSTRACT

In this study, the methods which are used for estimating the treatment effects in the balanced incomplete block design and the estimated values of the treatment effects obtained by using these methods have been compared.

When it is impossible to make the required number of treatments, which are needed for each block in the randomized complete block design, the experiment is designed in the balanced incomplete block design. In the balanced incomplete block design, it is suggested that other methods should be used rather than the least squares estimators to estimate the treatment effects. Therefore, in order to estimate the treatment effects in the balanced incomplete block design, the intrablock, the interblock and the combined estimates methods are introduced in literature.

In this study, the simulation studies were made by using the programs written in statistical software Minitab. The intrablock, the interblock and the combined estimates for the balanced incomplete block design and the least squares estimates for the randomized complete block design of the treatment effects were calculated 2500 times by simulation. The results were compared and the most appropriate estimator of the treatment effects for the balanced incomplete block design was investigated. The means and the standard deviations of the estimates were considered as the criteria of this comparison.

After the evaluation of the results, it was observed that each of the three methods gave unbiased results when the block effects were insignificant for the balanced incomplete block design. When the block effects were significant, it was seen that the results of the means of the intrablock and the combined estimates were unbiased. The interblock estimates results were observed as inappropriate. In both of the situations in which the block effects were significant and insignificant, the standard

deviation of the intrablock estimates is lower than the standard deviations of the interblock and the combined estimates.

ÖZET

Bu çalışmada, dengeli tamamlanmamış bloklar düzeninde deneme etkilerini tahmin etmek için kullanılabilen yöntemler tanıtılmış ve bu yöntemler kullanılarak elde edilen deneme etkilerinin tahmin değerleri karşılaştırılmıştır.

Tamamlanmış rasgele bloklar düzeninde bir blokta, denenmesi gereken deneme sayısı kadar deneme yapma olanağı olmadığı zaman, deneyi dengeli tamamlanmamış bloklar düzeni olarak düzenlemek yoluna gidilmektedir. Dengeli tamamlanmamış bloklar düzeninde ise, deneme etkilerinin bilinen en küçük kareler tahmin edicileri ile tahmin etmek yerine başka yöntemlerin kullanılması önerilmektedir. Bunun için literatürde dengeli tamamlanmamış blok düzeninde deneme etkilerinin tahmini için bloklar içi (interblock), bloklar arası (intrablock), ve bileşik (combined) tahmin edicileri olarak isimlendirilen yöntemler yer almaktadır.

Bu çalışmada, Minitab istatistiksel paket programı kullanılarak hazırlanan programlarla 2500 tekrarlı benzetim çalışmaları yapılmıştır. Çalışmada dengeli tamamlanmamış blok düzeninde deneme etkilerinin bloklar içi, bloklar arası, ve bileşik tahmin değerleri ile rasgele tamamlanmış bloklar düzeninde deneme etkilerinin en küçük kareler tahmin değerleri hesaplanmıştır. Elde edilen sonuçlar karşılaştırarak, dengeli tamamlanmamış bloklar düzeni için uygun deneme etkisi tahmin edicisi araştırılmıştır. Karşılaştırmalarda ölçüt olarak tahminlerin beklenen değerleri ile standart sapmaları göz önüne alınmıştır.

Yapılan değerlendirmeler sonucunda dengeli tamamlanmamış bloklar düzeni için blok etkisinin önemsiz olduğu durumda her üç yönteminde yansız sonuçlar verdiği gözlenmiştir. Blok etkisi önemli olduğunda, bloklar içi ve bileşik tahmin yöntemleri sonucu elde edilen deneme etkileri tahmin değerlerinin yansız sonuçlar verdiği

görülmüştür. Bloklar arası tahmin yöntemi sonucunda elde edilen deneme etkisi tahmin değerlerinin ise uygun olmadığı gözlenmiştir. Blok etkisinin hem önemli hem de önemsiz olduğu durumlarda ise bloklar içi tahmin değerlerinin standart sapmalarının, bloklar arası ve bileşik tahmin değerlerinin standart sapmalarına göre daha düşük olduğu gözlenmiştir.

CONTENTS

Chapter Three

SIMULATION RESULTS FOR TREATMENT EFFECTS ESTIMATION

Chapter Four

CONCLUSIONS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE
INTRODUCTION

1.1 Introduction

Much of research in engineering, science, and industry is empirical and makes extensive use of experimentation. Statistical methods can greatly increase the efficiency of these experiments and often strengthens the conclusions so obtained.

Design of experiments can be defined as a technique which includes the application of the treatments whose effect will be measured under certain rules and conditions and the measuring of the response given to these treatments by the experimental units and coming to a conclusion by comparing the obtained results. The word "treatment" in this description corresponds to the variable which effects to measure and to compare with other treatments.

The three basic principles of experimental design are replication, randomization, and blocking. By replication we mean a repetition of the basic experiment. Randomization is the cornerstone underlying the use of statistical methods in experimental design. By randomization we mean that both the allocation of the experimental material and the order in which the individual runs or trials of the experiment to be performed are randomly determined. Blocking is a technique used to increase the precision of an experiment. A block is a portion of the experimental material that should be more homogenous than the entire set of material. Blocking involves making comparisons among the conditions of interest in the experiment within each block.

An experimental unit or experimental plot is the unit of material to which one application of a treatment is applied. The treatment is the procedure whose effect is to be measured and compared with other treatments. The basic objective of designing experiments is to test whether the difference between different treatments applied on experimental units is meaningful or not. The treatments are applied to experimantal units according to the design type to be formed. Sometimes, in necessity of design, blocking is used and treatments are applied to experimental units in blocks. In this type of designs known as randomized complete block designs, all treatments are applied on the experimental units in each block randomly. Because of shortages of experimental apparatus or facilities or the pyhsical size of the block, we may not to be able to run all the treatment combinations in each block. Therefore, it is not possible to use randomized complete block design. The designs in which blocks do not contain all the treatments are known as incomplete block designs. The most widely used incomplete block design is balanced incomplete block design.

The purpose of this study is to examine the intrablock, the interblock and the combined estimates of the treatment effects in the balanced incomplete block design and to determine the most appropriate estimator by simulation method.

The study contains four chapters. The general information about the overall research is in the first chapter. The second chapter examines the randomized complete block design and the estimation of the treatment effects. In addition, this chapter gives general information about the balanced incomplete block design and deals with the intrablock, the interblock and the combined estimates of the treatment effects for this kind of designs and searches for the appropriate treatment effects estimator.

The third chapter aims to find the treatment effect estimator by using simulation method on statistical software program Minitab and to choose the most appropriate treatment effects estimator. The fourth chapter contains the results obtained in this research.

CHAPTER TWO

RANDOMIZED COMPLETE BLOCK AND BALANCED INCOMPLETE BLOCK DESIGNS

2.1 Introduction

In a randomized block design each block contains every level of the "treatment" factor. For this reason the design is sometimes called a *complete randomized block design.* In certain experiments using randomized block designs, we may not be able to run all the treatment combinations in each block. Situations like this occur because of shortages of experimental apparatus or facilities or the physical size of the block. So for this type of problem it is possible to use randomized block designs in which every treatment is not present in every block. These designs are known as *randomized incomplete block designs*, and they are the subject of this chapter. We shall also discuss such topics as the recovery of interblock information.

2.2 Randomized Complete Blocks Designs

A randomized block design is a restricted randomized design in which the experimental units are first sorted into homogeneous groups, which are called blocks, and the treatments are then assigned at random within the blocks.

Throughout this chapter it is used the term "experimental unit" to denote the unit that is allocated a treatment independently of the other units. The experimental unit can contain several observational units; for instance, a class of students that receive a certain method of teaching in common can be an experimental unit, while the individual students are observational units. The distinction is, as it is seen, very important, because, from the point of view of inference on the effects of treatments, the experimental unit must be considered as a whole, and the variation between the observational units within an experimental unit is usually of little value in assessing the errors of estimates of treatment effects.

The key objective in blocking the experimental units is to make them as homogeneous as possible within blocks with respect to the response variable under study, and to make the different blocks as heterogeneous as possible with respect to the response variable. The design in which each treatment is included once in each block is called a randomized complete block design.

The randomization procedure for a randomized block design is straightforward. Within each block a random permutation is used to assign treatments to experimental units. Independent permutations are selected for a several blocks.

In fact, a randomized complete block design may be viewed as corresponding to a two-factor study (blocks and treatments are the factors), with one observation in each cell. It is noted that the assumption of no interactions between the two factors permits an analysis of factor effects when there is only one observation in each cell and the factors have fixed effects.

The model for a randomized complete block design containing the assumption of no interaction effects, when both the block and treatment effects are fixed and there are b blocks (replications) and a treatments, is as follows:

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \quad i=1,2,\ldots,a\,;\, j=1,2,\ldots,b \tag{2.1}$$

where μ is an overall mean, τ_i is the effect of the ith treatment, β_j is the effect in the jth block, and ε_{ij} is the usual $\mathrm{NID}(0,\sigma^2)$ random error term. Treatments and blocks are considered initially as fixed factors. It is assumed that $\sum_{i=1}^{a}\tau_i=0$ and $\sum_{j=1}^{b}\beta_j=0$.

If the treatment effects are random, the only changes in model (2.1) are that the τ_i now represent independent normal variables with expectation zero and variance σ_τ^2 and that the τ_i are independent of the ε_{ij}.

The least squares estimators of the parameters in the randomized complete block model are obtained.

Table 2.1 Least Squares Estimators of the Parameters in the Randomized Complete Block Design

Parameter	Estimator
μ	$\hat{\mu}=\bar{y}_{..}$
τ_i	$\hat{\tau}_i=\bar{y}_{i.}-\bar{y}_{..}$
β_j	$\hat{\beta}_j=\bar{y}_{.j}-\bar{y}_{..}$

2.3 Balanced Incomplete Block Designs

In the randomized blocks design the size of the block of experimental units must be equal to the number of treatments to be compared. It is sometimes desirable or necessary to have the block size smaller than the number of treatments. These designs were introduced by Yates (1936).

When all treatment comparisons are equally important, the treatment combinations used in each block should be selected in a balanced manner, that is, so that any pair of treatments occur together the same number of times as any other *pair. Thus, a balanced incomplete block design is an incomplete block design in which any two treatments appear together an equal number of times.* (Montgomery, 1991, p.176)

It can very well happen that the available blocks are not big enough to accommodate a complete replication. The factor represented by "blocks" is often a natural one, such as days, positions in a field, factory, and so on. If for example, only 3 observations can be taken in one day, while the "treatments" factor has 4 levels, A, B, C, D, then a complete replicate cannot be observed in one day. However, a "balanced" design, in which there are 3 replicates arranged in 4 blocks, can be specified as in Table 2.2. The order in which the treatments are placed in each block has been randomized.

Table 2.2 A Balanced Incomplete Blocks for Four Treatments

BLOCK			
I	II	III	IV
B	A	C	B
A	B	A	D
C	D	D	C

This is called an "incomplete" randomized block design. It is described by the adjective "balanced," giving the name "balanced incomplete block". The word "balanced" does not simply mean that the blocks are of the same size, and each treatment level appears the same number of times. In a balanced incomplete block design it must also be true that each pair of treatment levels appears together in the same block the same number of times (twice in Table 2.3).

Table 2.3 Four Treatments in Four Blocks of Three Plots

TREATMENTS	BLOCKS			
	I	II	III	IV
A	x	x	x	
B	x	x		x
C	x		x	x
D		x	x	x

Suppose that there are *a* treatments and *b* blocks. Each block can hold exactly *k* treatments. The balanced incomplete block design is frequently used when $k<a$. In this design, each treatment is replicated *r* times. Since each treatment is replicated *r* times, and there are *k* plots in each block, there are $N = ar = bk$ total observations. If $a=b$, the design is said to be *symmetric*. Furthermore, the number of times each pair of treatment appears in the same block is as given below.

$$\lambda = \frac{r(k-1)}{a-1} \qquad (2.2)$$

The parameter λ must be an integer. To derive the relationship for λ, consider any treatment, say treatment 1. Since treatment 1 appears in *r* blocks and there are *k*-1 other treatments in each of those blocks, there are $r(k-1)$ observations in a block containing treatment 1. These $r(k-1)$ observations also have to represent the remaining $a-1$ treatments λ times. Therefore $\lambda(a-1)=r(k-1)$.

A large number of balanced designs has been worked out and tabulated for $r \leq 10$ and for *k* of various sizes. The set of (k,a) pairs for which designs have been shown to exist with block size not bigger than ten, and total number of units not exceeding 100, are listed in Table 2.4. (Mead, 1988, p.152)

Table 2.4 Pairs of Values of (*k*,*a*) For Which A Balanced Incomplete Block Design is Known to Exist For $k \leq 10$ and $bk < 100$.

k	*a*	*b*	*r*	λ	*k*	*a*	*b*	*r*	λ
3	4	4	3	2	4	13	13	4	1
3	5	10	6	3	4	16	20	5	1
3	6	10	5	2	5	6	6	5	4
3	6	20	10	4	5	9	18	10	5
3	6	10	5	2	5	10	18	9	4
3	7	7	3	1	5	11	11	5	2
3	9	12	4	1	6	7	7	6	5
3	10	30	9	2	6	9	12	8	5
3	13	26	6	1	6	10	15	9	5
4	5	5	4	3	6	11	11	6	3
4	6	15	10	6	6	16	16	6	2
4	7	7	4	2	7	8	8	7	6
4	8	14	7	3	8	9	9	8	7
4	9	18	8	3	9	10	10	9	8
4	10	15	6	2					

The randomization procedure for balanced incomplete block design follows:

(i) *Allot the entry numbers to the treatments at random unless there is a specific reason for not doing so.*

(ii) *Allot the groups to the b blocks at random.*

(iii) *Randomly allot the treatments to the k experimental units within each block.* (Federer, 1955, p.415)

2.3.1 Balanced Incomplete Block Designs with Intrablock Analysis

"We consider now the usual method of analyzing balanced incomplete block designs. This is called the *intrablock analysis* because the block differences are eliminated and the estimates of all contrasts in the treatment effects can be expressed in terms of comparisons between plots in the same block." (John, 1971, p.223)

The experimental model for this design is

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}, \quad i=1,2,...,a; j=1,2,...,b \tag{2.3}$$

where y_{ij} is the i th observation in the j th block, μ is the overall mean, τ_i is the effect of the i th treatment, β_j is the effect of the j th block, imposing $\sum_{i=1}^{a}\tau_i$ =0, $\sum_{j=1}^{b}\beta_j$ =0 and ε_{ij} is the NID$(0,\sigma^2)$ random error component. The mean μ and the treatment effects τ_i are unknown constants. The block effects may be fixed or random; it is treated as if they are fixed. It is important to note that the model assumes that there is no interaction between treatments and blocks.

For comparing the treatment effects in the balanced incomplete block design, firstly the hypothesis must be set.

$H_0 : \tau_i = 0 \quad$ i=1,2,...,a

$H_a : \tau_i \neq 0 \quad$ for at least one i

The model in (2.3) represents this design. For the least squares method, the normal equations are obtained to minimize the equation below;

$$L = \sum_{ij} n_{ij}(y_{ij} - \mu - \tau_i - \beta_j)^2 \tag{2.4}$$

where n_{ij}=1 if the i th treatment appears in the j th block, and zero otherwise.

Note that since there are r blocks that contain i, there are r-values of j for which $n_{ij} = 1$, and n_{ij}=0 for the other b-r values of j. As a result the number of blocks containing the ith treatment is $\sum_{j=1}^{b} n_{ij} = r$. Since there are k treatments contained in the block j, there are k values of i for which n_{ij}=1, and n_{ij}=0 for the other a-k values of i. As a result the number of treatments appearing in the j th block is $\sum_{i=1}^{a} n_{ij} = k$.

From the minimization of (2.4), the estimating equations are

$$\mu : N\hat{\mu} + r\sum_{i=1}^{a}\hat{\tau}_i + k\sum_{j=1}^{b}\hat{\beta}_j = y_{..} \qquad (2.5)$$

$$\tau_i : r\hat{\mu} + r\hat{\tau}_i + \sum_{j=1}^{b} n_{ij}\hat{\beta}_j = y_{i.} \qquad (2.6)$$

$$\beta_j : k\hat{\mu} + \sum_{i=1}^{a} n_{ij}\hat{\tau}_i + k\hat{\beta}_j = y_{.j} \qquad (2.7)$$

There are b+a+1 equations, and only b+a-1 of these equations is linearly independent. The estimators for the block and treatment effects can be obtained with the constraints $\sum_{i=1}^{a}\tau_i$ =0, $\sum_{j=1}^{b}\beta_j$ =0.

The first step in solving for the treatment effects is to adjust the treatment effects for the block effects by removing the block constants from the treatment equations. This accomplished by multiplying the β_j equation by $\sum_{j=1}^{b} n_{ij} / k$ and subtracting the sum of all of these altered β_j equations from the τ_i equation.

The resultant equation is

$$r\hat{\tau}_i - \frac{1}{k}\sum_{j=1}^{b}\sum_{p=1}^{a} n_{ij}n_{pj}\hat{\tau}_p = y_{i.} - \frac{1}{k}\sum_{j=1}^{b} y_{.j} \equiv Q_i \qquad (2.8)$$

where Q_i is called an *adjusted treatment total* (adjusted for block effects) and $\sum_{i=1}^{a} Q_i = 0$.

Suppose the ith treatment appears λ times in the same block with the pth treatment. Then it is used the following equations (Bose&Manvel, 1984, p.160).

$$\sum_{j=1}^{b} n_{ij} n_{pj} = r \quad \text{if } i=p$$

$$\sum_{j=1}^{b} n_{ij} n_{pj} = \lambda \quad \text{if } i \neq p \tag{2.9}$$

If $i=p$, than $\sum_{j=1}^{b} n_{ij} n_{pj}$ becomes $\sum_{j=1}^{b} n_{ij}^2 = r$. But $n_{ij} = n_{ij}^2$ since $n_{ij} = 0$ or 1.

Now note that, when $i \neq p$, $n_{ij} n_{pj} = 1$ if and only if both n_{ij} and n_{pj} are unity, that is, both the treatments i and p occur in the block j. Now there are exactly λ values for j for which the pair of treatments i and p occur together in the block j. Hence $n_{ij} n_{pj} = 1$ for exactly λ values of j for any fixed i and p, $i \neq p$, and is zero for the other values.

The Equation (2.8) may be rewritten as Equation (2.10). Hence the obtained equation and next step are given in below.

$$r\hat{\tau}_i - \frac{1}{k}\left[\sum_{j=1}^{b}\sum_{\substack{p=1 \\ p\neq i}}^{a} n_{ij} n_{pj} \hat{\tau}_p + \sum_{j=1}^{b}\sum_{\substack{p=1 \\ p=i}}^{a} n_{ij} n_{pj} \hat{\tau}_p\right] = Q_i \tag{2.10}$$

$$kr\hat{\tau}_i - \left[\lambda \sum_{\substack{p=1 \\ p\neq i}}^{a} \hat{\tau}_p + r \sum_{\substack{p=1 \\ p=i}}^{a} \hat{\tau}_p\right] = kQ_i \tag{2.11}$$

The known conditions $\sum_{i=1}^{a}\hat{\tau}_i = 0$, $\sum_{\substack{p=1 \\ p\neq i}}^{a}\hat{\tau}_p = -\hat{\tau}_i$ and $\lambda = \dfrac{r(k-1)}{(a-1)}$ are used.

$$kr\hat{\tau}_i - r\hat{\tau}_i + \lambda\hat{\tau}_i = kQ_i$$

$$r(k-1)\hat{\tau}_i + \lambda\hat{\tau}_i = kQ_i \qquad (2.12)$$

And as a result estimation of the treatment effect is

$$\hat{\tau}_i = \frac{kQ_i}{\lambda a} \qquad i=1,2,\ldots,a \qquad (2.13)$$

$\hat{\tau}_i$ is called intrablock estimator of the treatment effect.

The variance of $\hat{\tau}_i$ and of the estimate of the difference between two treatment effects, $\hat{\tau}_i - \hat{\tau}_{i'}$, may be obtained by consideration of the variance and covariance of Q_i:

$$V(Q_i) = \frac{(k-1)r}{k}\sigma^2$$

$$Cov(Q_i, Q_{i'}) = -\frac{\lambda}{k}\sigma^2 \qquad (2.14)$$

And using equations (2.13) and (2.14), $V(\hat{\tau}_i)$ can be estimated.

$$V(\hat{\tau}_i) = \frac{k(a-1)}{\lambda a^2}\sigma^2 \qquad (2.15)$$

$$V(\hat{\tau}_i - \hat{\tau}_{i'}) = \left(\frac{k}{\lambda a}\right)^2 V(Q_i - Q_{i'}) = \frac{2k}{\lambda a}\sigma^2 \qquad (2.16)$$

In assessing incomplete block designs, the efficiency of the design is measured by comparing the variance of the balanced incomplete block design with the variance for a complete block design, with the same replication per treatment and assuming, which is most realistic, the same σ^2. For a randomized complete block design with r observations on each treatment,

$$V(\hat{\tau}_i - \hat{\tau}_{i'}) = \frac{2\sigma^2}{r}$$

The *efficiency factor* (*Eff*) of any balanced incomplete block design is a lower limit to the efficiency of balanced incomplete blocks relative to randomized blocks. From the variance for $(\hat{\tau}_i - \hat{\tau}_{i'})$ the efficiency of the balanced incomplete block design is given as equation (2.17).

$$Eff = \frac{\frac{2\sigma^2}{r}}{\frac{2k\sigma^2}{\lambda a}} = \frac{\lambda a}{rk} \qquad (2.17)$$

It can be rewritten as below.

$$Eff = \frac{1 - \frac{1}{k}}{1 - \frac{1}{a}}$$

Hence it is clearly less than 1 since $k<a$ (Mead, 1988, p.166). As *Eff* takes values closer to 1 the efficiency of balanced incomplete block design increases.

Note that this efficiency depends critically on k as is shown in numerical values of *Eff* given in Table 2.5 (Mead, 1988, p.166).

Table 2.5 Values of *Eff* For *k*=2,3,4,6,8 and *a*=4,6,8,12,16,24,32.

	k				
a	2	3	4	6	8
4	0.67	0.88			
6	0.60	0.80	0.90		
8	0.57	0.76	0.86	0.95	
12	0.55	0.73	0.82	0.91	0.96
16	0.53	0.71	0.80	0.89	0.93
24	0.52	0.70	0.78	0.87	0.91
32	0.52	0.69	0.77	0.86	0.90

The efficiency factor tells us the loss caused by using an incomplete design. For balanced incomplete block designs the efficiency factor is smallest when the number of units per block is small and the number of treatments large. Thus with two units per block and a large number of treatments, the efficiency factor is 1/2. The efficiency factor is near one when the number of treatments does not greatly exceed the number of units per block is large.

'If the efficiency factor is less than about 0.85 and if the number of blocks exceeds about 10, a more complicated method of analysis may be used that makes any possible loss of precision as compared with a randomized block design small.' (Cox, 1992, p.230)

This method that is mentioned in the previous paragraph is based on the idea that if the variation between blocks is not too great, the block totals contain a certain amount of information about the treatment effects. The method, which is called the recovery of interblock information, will described in the following section.

2.3.2 Balanced Incomplete Block Designs with the Recovery of Interblock Analysis

The analysis of the balanced incomplete block design given in section 2.3.1 is called the intrablock or within block analysis. However, in that analysis only the information about treatments from comparisons within blocks was considered. The differences between block totals also contain information about treatment differences.

Yates (1940) introduced a new computing technique for balanced incomplete block design when the blocks selected randomly. "Yates noted that, if the block effects are uncorrelated random variables with zero means and variance σ_β^2, than one may obtain additional information about the treatment effects τ_i. Yates called the method of obtaining this additional information the *interblock analysis*." (Montgomery, 1976, p.184)

It will be assumed that they are a random sample of b from an infinite population that the β_j are independently and identically distributed. This assumption is appropriate if the blocks can be regarded as a random sample from a large finite population. This might be the case for an example if the cars are known as block were all of the same made and model.

In the balanced incomplete block model with blocks random, each treatment appears r times, and there are k plots in each block; every pair of treatments appears together in λ blocks. The model can be written

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \quad i=1,2,...,a; \quad j=1,2,...,b \tag{2.18}$$

The μ and τ_i are fixed unknown parameters, β_j is distributed $N(0,\sigma_\beta^2)$; e_{ij} is distributed $N(0,\sigma^2)$; and all random variables are independent.

The hypothesis in interblock analysis is constructed with the same as the intrablock analysis. That is

$H_0 : \tau_i = 0 \quad$ i=1,2,...,a

$H_a : \tau_i \neq 0 \quad$ for at least one i

The intrablock estimators for τ_i were given by the formula (2.13). When the blocks are random, alternative estimators for τ_i can be obtained from the block totals. Next we shall show that there exists another unbiased estimator of τ_i independent of the intrablock estimator $\hat{\tau}_i$. This estimator will be denoted by $\tilde{\tau}_i$ and will be a function of the block totals $y_{.j}$. The symbol B_i will be used to denote the total of all blocks that contain treatment i;

$$B_i = \sum_{j=1}^{b} n_{ij} y_{.j} \tag{2.19}$$

$$y_{.j} = k\mu + \sum_{i=1}^{a} n_{ij}\tau_i + (k\beta_j + \sum_{i=1}^{a} \varepsilon_{ij}) \tag{2.20}$$

where the term in parentheses may be regarded as error. The interblock estimators of μ and τ_i are found by minimizing the least squares function.

$$\mathrm{L} = \sum_{j=1}^{b} (y_{.j} - k\mu - \sum_{i=1}^{a} n_{ij}\tau_i)^2 \tag{2.21}$$

The following least squares normal equations

$$\mu : bk\tilde{\mu} + r\sum_{i=1}^{a} n_{ij}\tilde{\tau}_i = y_{..} \tag{2.22}$$

$$\tau_i : rk\tilde{\mu} + \sum_{j=1}^{b}\sum_{p=1}^{a} n_{ij} n_{pj} \tilde{\tau}_p = \sum_{j=1}^{b} n_{ij} y_{.j} \tag{2.23}$$

With $\sum_{i=1}^{a} \tilde{\tau}_i = 0$, the equation (2.22) can be rewrite as follows.

$$\tilde{\mu} = y_{..} / bk = \bar{y} \tag{2.24}$$

To obtain $\tilde{\tau}_i$, we use the following assumptions given in the previous section.

$$\sum_{j=1}^{b} n_{ij} n_{pj} = r \quad \text{if } i=p$$

$$\sum_{j=1}^{b} n_{ij} n_{pj} = \lambda \quad \text{if } i \neq p$$

Than the equation (2.23) becomes following equation.

$$rk\tilde{\mu} + \sum_{j=1}^{b}\sum_{\substack{p=1 \\ p\neq i}}^{a} n_{ij} n_{pj} \tilde{\tau}_p + \sum_{j=1}^{b}\sum_{\substack{p=1 \\ p=i}}^{a} n_{ij} n_{pj} \tilde{\tau}_p = \sum_{j=1}^{b} n_{ij} y_{.j} \tag{2.25}$$

$$rk\tilde{\mu} + \left[\lambda \sum_{\substack{p=1 \\ p\neq i}}^{a} \tilde{\tau}_p + r \sum_{\substack{p=1 \\ p=i}}^{a} \tilde{\tau}_p \right] = \sum_{j=1}^{b} n_{ij} y_{.j} \tag{2.26}$$

Using these equations $\sum_{i=1}^{a} \tilde{\tau}_i = 0$, $\sum_{\substack{p=1 \\ p\neq i}}^{a} \tilde{\tau}_p = -\tilde{\tau}_i$, $\sum_{p=1}^{a} \tilde{\tau}_p = 0$, the estimation of τ_i is obtained.

$$kr\tilde{\mu} + (r-\lambda)\tilde{\tau}_i = \sum_{j=1}^{b} n_{ij} y_{.j} \tag{2.27}$$

$$\tilde{\tau}_i = \frac{\sum_{j=1}^{b} n_{ij} y_{.j} - kr\bar{y}}{r - \lambda} \tag{2.28}$$

Now it was determined the two estimators for the treatment effects. So it can be said that the estimates $\hat{\tau}_i$ and $\tilde{\tau}_i$ are uncorrelated.

The variance of $\tilde{\tau}_i$ is given in equation (2.29). (Graybill, 1961, p.408)

$$V(\tilde{\tau}_i) = \frac{k(a-1)}{a(r-\lambda)}(\sigma^2 + k\sigma_\beta^2) \tag{2.29}$$

Now it is considered that it always necessary or useful to calculate the full analysis of variance with both sets of estimates or not. In practice often only the intrablock analysis is calculated and it's proper to consider when we should additionally attempt to use the interblock information.

'*The principle of using incomplete blocks stems from a recognition that* σ^2 *should be reduced by using smaller blocks. As the selection of blocks becomes more successful the ratio* $\sigma^2/(\sigma^2 + k\sigma_\beta^2)$ *will become smaller and consequently the contribution of the interblock information will become less. Thus the more successful we are in making the within block analysis efficient the less benefit will accrue from using the interblock information.*' (Mead, 1988, p.171)

That is by choosing the blocks efficiently the precision of that intrablock proportion of information will be made as much greater than the precision of the interblock information as is possible.

That is we should expect to use the interblock information when the intrablock proportion of information, which is measured by *Eff*, is not very large, or the gain from using small blocks is not very substantial. Even when *Eff* is not large, it will be appropriate to use the interblock information only when the ratio $(\sigma^2 + k\sigma_\beta^2)/\sigma^2$ is not large.

By applying interblock analysis, treatment effect estimates can be calculated and by using intrablock and interblock estimation, combined estimates can be determined. In this case in determining treatment effect estimates, combined estimates are also used.

2.4 Combining the Intrablock and Interblock Estimates

The intrablock and interblock analyses provide two independent sets of estimates of treatment effects. There must, inevitably, be combined estimate, calculated from the two separate estimates of each treatment effect, which will be the best estimate. Since both sets of estimates are unbiased by the general linear model theory, any linear combination of the estimates will also be unbiased and the linear combination with the smallest variance will be the best linear unbiased estimate. This minimum variance combined estimate can be shown to require weighting each estimate inversely by its variance.

Now we wish to combine the interblock and intrablock estimators to obtain a single, unbiased, minimum variance estimate of each τ_i. We know that $\hat{\tau}_i$ and $\tilde{\tau}_i$ are unbiased and also that

$$V(\hat{\tau}_i) = \frac{k(a-1)}{\lambda a^2}\sigma^2 \qquad \text{(intrablock)}$$

and

$$V(\tilde{\tau}_i) = \frac{k(a-1)}{a(r-\lambda)}(\sigma^2 + k\sigma_\beta^2) \qquad \text{(interblock)}$$

We use a linear combination of the two estimators. By a linear estimator of τ_i we mean a linear combination of $\hat{\tau}_i$ and $\tilde{\tau}_i$; that is, $\tau_i^* = \alpha_1\hat{\tau}_i + \alpha_2\tilde{\tau}_i$. The problem is to determine α_1 and α_2 such that τ_i^* is unbiased $\left[E(\tau_i^*)=\tau_i\right]$ and such that $V(\tau_i^*)$ is a minimum. (Graybill, 1961, p409) The expectation of τ_i^* is equated as follows.

$$\tau_i = E(\tau_i^*) = E(\alpha_1\hat{\tau}_i + \alpha_2\tilde{\tau}_i) = \alpha_1\tau_i + \alpha_2\tau_i = (\alpha_1 + \alpha_2)\tau_i$$

It is seen from the above that:

$$\alpha_1 + \alpha_2 = 1 \text{ or } \alpha_1 = 1 - \alpha_2$$

Now the variance of τ_i^* is equated as follows.

$$V(\tau_i^*) = V(\alpha_1\hat{\tau}_i + \alpha_2\tilde{\tau}_i)$$
$$V(\tau_i^*) = \alpha_1^2 V(\hat{\tau}_i) + \alpha_2^2 V(\tilde{\tau}_i)$$
$$V(\tau_i^*) = \alpha_1^2 V(\hat{\tau}_i) + (1-\alpha_1)^2 V(\tilde{\tau}_i)$$

The only unknown in $V(\tau_i^*)$ is α_1. The value of α_1 that minimizes $V(\tau_i^*)$ is found by

$$\frac{\partial V(\tau_i^*)}{\partial \alpha_1} = 2\alpha_1 V(\hat{\tau}_i) - 2(1-\alpha_1)V(\tilde{\tau}_i) = 0$$

$$\alpha_1 = V(\tilde{\tau}_i)\left(\frac{1}{(V(\hat{\tau}_i) + V(\tilde{\tau}_i))}\right)$$

and

$$\alpha_2 = V(\hat{\tau}_i)\left(\frac{1}{(V(\hat{\tau}_i) + V(\tilde{\tau}_i))}\right)$$

So

$$\tau_i^* = \left(V(\tilde{\tau}_i)\hat{\tau}_i + V(\hat{\tau}_i)\tilde{\tau}_i\right)\left(\frac{1}{(V(\hat{\tau}_i) + V(\tilde{\tau}_i))}\right)$$

This implies that the best combined estimator is

$$\tau_i^* = \frac{\frac{k(a-1)}{a(r-\lambda)}(\sigma^2 + k\sigma_\beta^2)\hat{\tau}_i + \frac{k(a-1)}{\lambda a^2}\sigma^2\tilde{\tau}_i}{\frac{k(a-1)}{\lambda a^2}\sigma^2 + \frac{k(a-1)}{a(r-\lambda)}(\sigma^2 + k\sigma_\beta^2)} \qquad i=1,2,\ldots,a \qquad (2.30)$$

which can be simplified to equation (2.31).

$$\tau_i^* = \frac{kQ_i(\sigma^2 + k\sigma_\beta^2) + (\sum_{j=1}^{b} n_{ij} y_{.j} - kr\bar{y})\sigma^2}{(r-\lambda)\sigma^2 + \lambda a(\sigma^2 + k\sigma_\beta^2)} \qquad i=1,2,\ldots,a \qquad (2.31)$$

Unfortunately, equation (2.31) cannot be used to estimate the τ_i because the variances σ^2 and σ_β^2 are unknown. The usual approach is to estimate σ^2 and σ_β^2 from the data and replaces these parameters in equation (2.31) by the estimates. The estimate usually taken for σ^2 is the error mean square from the intrablock error. Thus,

$$\sigma^2 = MS_E \qquad (2.32)$$

To obtain an analysis of variance, the model will be considered with blocks fixed. Then, when an analysis of variance table is obtained, the model will be considered with blocks random, and the appropriate mean squares and expected mean squares will be equated, so as to give rise to estimators of σ^2 and σ_β^2.

The total variation in the data is expressed by the total corrected sum of squares:

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b} y^2_{ij} - \frac{y_{..}^2}{N} \qquad (2.33)$$

Total variability may be partitioned into

$$SS_T = SS_{Tr.(adjusted)} + SS_{Blocks} + SS_E \qquad (2.34)$$

where the sum of squares for treatments is adjusted to separate the treatment and the block effects. This is often called the *method of treatments eliminating (adjusted for) blocks*. This adjustment is necessary because each treatment is represented in a different set of r blocks. This method is called as *Method A*. Its analysis of variance table is given in Table 2.6. (Graybill, 1961, p.412)

Table 2.6 Analysis of Variance For Incomplete Block Model

Method A		
Source of Variation	d.f.	Sum of Squares
Total	bk-1	$A=\sum_{i=1}^{a}\sum_{j=1}^{b}y_{ij}^2-\frac{y_{..}^2}{bk}$
Treatment Adjusted	a-1	$A_1=\frac{k}{\lambda a}\sum_{i=1}^{a}Q_i^2$
Blocks Unadj.	b-1	$A_2=\sum_{j=1}^{b}\frac{y_{.j}^2}{k}-\frac{y_{..}^2}{bk}$
Intrablock Error	bk-b-a+1	$A_3=A-A_1-A_2$

Another way of partitioning the total variability in the data is the *method of blocks eliminating (adjusted for) treatments*. The total variability may be partitioned into

$$SS_T = SS_{Tr} + SS_{Blocks.(adjusted)} + SS_E \quad (2.35)$$

This method is called as *Method B*. Its analysis of variance table is given in Table 2.7. (Graybill, 1961, p.412)

Table 2.7 Analysis of Variance For Incomplete Block Model

Method B		
Source of Variation	d.f.	Sum of Squares
Total	bk-1	$B=\sum_{i=1}^{a}\sum_{j=1}^{b}y_{ij}^2-\frac{y_{..}^2}{bk}$
Treatment Unadj.	a-1	$B_1=\sum_{i=1}^{a}\frac{y_{i.}^2}{r}-\frac{y_{..}^2}{bk}$
Blocks Adjusted	b-1	$B_2=\frac{k}{\lambda a}\sum_{i=1}^{a}Q_i^2+\sum_{j=1}^{b}\frac{y_{.j}^2}{k}-\sum_{i=1}^{a}\frac{y_{i.}^2}{r}$
Intrablock Error	bk-b-a+1	$B_3=B-B_1-B_2$

As it is seen in the Table 2.6 and Table 2.7, there are two methods of analysis that could be used.

From the above it follows that

$$SS_{Tt(adjusted)} + SS_{Blocks} = SS_{Tr} + SS_{Blocks.(adjusted)} \tag{2.36}$$

or, in words, the sum of squares due to blocks(adj.) plus the sum of squares due to treatments(unadj.) equals the sum of squares due to treatments(adj.) plus the sum of squares due to blocks(unadj.). The expected mean square for intrablock error is as shown because A_3=B_3 and A_3 has the same expectation whether the model has blocks fixed or blocks random.

The estimate of σ_β^2 is found from the mean square for blocks adjusted for treatments. This mean square is given as equation (2.37).

$$MS_{blocks(adj.)} = \left\{ \frac{k\sum_{i=1}^{a} Q_i^2}{\lambda a} + \frac{\sum_{j=1}^{b} y_{.j}^2}{k} - \frac{\sum_{i=1}^{a} y_{i.}^2}{r} \right\} /(b-1) \tag{2.37}$$

The result equation of expectation of $MS_{blocks(adj.)}$ is as follows (Graybill, 1961, p412).

$$E(MS_{blocks(adj.)}) = \sigma^2 + \frac{a(r-1)}{b-1}\sigma_\beta^2 \tag{2.38}$$

Thus, if $MS_{blocks(adj.)} > MS_E$, the estimate of $\hat{\sigma}_\beta^2$ is given as equation (2.39). If $MS_{blocks(adj.)} < MS_E$, we set $\hat{\sigma}_\beta^2$=0.

$$\hat{\sigma}_\beta^2 = \frac{\left(MS_{Blocks(adj.)} - MS_E\right)(b-1)}{a(r-1)} \tag{2.39}$$

In order to obtain a single, unbiased, minimum variance estimates of each τ_i, $\hat{\sigma}_\beta^2$ and σ^2 are replaced in equation (2.40). As a result, the equation (2.41) is obtained.

$$\tau_i^* = \frac{kQ_i(\sigma^2 + k\sigma_\beta^2) + (\sum_{j=1}^{b} n_{ij} y_{.j} - kr\bar{y})\sigma^2}{(r-\lambda)\sigma^2 + \lambda a(\sigma^2 + k\sigma_\beta^2)} \qquad i=1,2,\ldots,a \tag{2.40}$$

$$\hat{\tau}_i^* = \begin{cases} \dfrac{kQ_i(\hat{\sigma}^2 + k\hat{\sigma}_\beta^2) + (\sum_{j=1}^{b} n_{ij} y_{.j} - kr\bar{y})\hat{\sigma}^2}{(r-\lambda)\hat{\sigma}^2 + \lambda a(\hat{\sigma}^2 + k\hat{\sigma}_\beta^2)} & if\ \hat{\sigma}_\beta^2 \rangle 0 \\ \dfrac{y_{i.} - (1/a)y_{..}}{r} & if\ \hat{\sigma}_\beta^2 = 0 \end{cases} \tag{2.41}$$

That is in determining of combined estimator of the treatment effects, it is used the equation (2.41) according to the values of $\hat{\sigma}_\beta^2$.

The problem of estimating a common mean of two normal distributions and the related problem of recovery of interblock information has been studied in several papers.

Yates (1940) was apparently the first to suggest that information could be recovered in balanced incomplete block designs. Graybill and Deal (1959) initiated a close study of the properties of combined interblock and intrablock estimates of treatment differences. Further results were obtained by Seshadri (1963). Seshadri has constructed an estimator by combining the intrablock and interblock estimates, which is, in the sense of minimum variance, uniformly better than either of the components whenever the balanced incomplete block design has more than 8 treatments. Another striking feature of this theorem is that it advocates the use of interblock information in the balanced incomplete block design whenever $a \geq 9$, conforming to the conventional theory which does not favor recovery of interblock information unless the experiment is *large*. Shah (1964) improved this result, and showed interblock information should be recovered it $a \geq 6$, and Stein (1966) showed recovery should be made, provided $a \geq 4$.

Weerakkody (1992) shows that the recovery of interblock information through the estimated generalized least squares estimator may improve the estimation, provided $(b-a)$>4. This suggests that the recovery of information significantly improves the estimation when blocking effect is not significant.

CHAPTER THREE
SIMULATION RESULTS FOR TREATMENT EFFECTS ESTIMATION

3.1 Introduction

The explanations about the estimation of treatment effects of randomized complete block design and balanced incomplete block design and the variances of these estimates have been explained in the previous chapter. In this chapter, the intrablock, the interblock and the combined estimates of the balanced incomplete block design will be discovered, and the subject of the most appropriate estimator for treatment effects will be investigated. As mentioned in the previous chapter, the significance of the block effects is effective on estimates of the treatment effects for balanced incomplete block design. For this reason, searching for the most appropriate estimates for treatment effects will be conducted separately for both the significance and the insignificance of the block effects. In this study, a 4x4 randomized complete block design in which each treatment was repeated only once and treatment levels were specifically selected and a balanced incomplete block design formed from this randomized complete block design were used. To achieve this, the steps below were followed:

1- For a randomized complete block design a normally distributed population was generated. In this population the treatment effects are significant. There will be two cases for the block effects: they are insignificant and significant. In the first case the block effects will be considered as insignificant.

2- A random 4x4 randomized complete block design was selected from this population.
3- The least squares estimates of the treatment effects for randomized complete block design were found.
4- Balanced incomplete block design was obtained by taking out the defined treatments from each block in randomized complete block design.
5- Intrablock, interblock and combined estimates were determined for this balanced incomplete block design.
6- By returning to Step 2, this process was repeated for 2500 times.
7- The simulation values obtained for the least squares estimates for randomized complete block design and the intrablock, the interblock and the combined estimates for balanced incomplete block design were compared and appropriate method for treatment effect estimates were figured out.
8- This process was repeated for the condition in which the treatment effects and the block effects are both significant.

3.2 Generating the Population

Independent populations which were normally distributed with fixed variance and different averages were constructed for the 16 independent units in 4x4 randomized complete block design. So, the assumption of NID$(0,\sigma^2)$ for the ε_{ij} error terms in the model equation of randomized complete block design which was designed as 4 treatment 4 blocks and 1 observation in each unit was provided.

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \qquad i=1,2,3,4 \quad j=1,2,3,4 \qquad (3.1)$$

Here, σ^2 was equal to 25. To generate the population, the following Minitab program was used:

SET0.MTB

```
SET C199
4(50:80/10)
END
SET C91
4(1:4)
END
SET C92
(1:4)4
END
LET K1=0
LET K3=100
NAME C199='MEANS'
NAME C101='POP11' C102='POP21' C103='POP31' C104='POP41'
NAME C105='POP12' C106='POP22' C107='POP32' C108='POP42'
NAME C109='POP13' C110='POP23' C111='POP33' C112='POP43'
NAME C113='POP14' C114='POP24' C115='POP34' C116='POP44'
EXEC 'SET1' 16
STACK C101-C104 C151
STACK C105-C108 C152
STACK C109-C112 C153
STACK C113-C116 C154
STACK C101 C105 C109 C113 C161
STACK C102 C106 C110 C114 C162
STACK C103 C107 C111 C115 C163
STACK C104 C108 C112 C116 C164
STACK C101-C116 C150
DESC C150
DESC C101-C116
DESC C151-C154
DESC C161-C164
```

SET1.MTB

```
LET K1=K1+1
LET K2=C199(K1)
LET K3=K3+1
LET K4=SQRT(25)
RANDOM 1000 CK3;
NORMAL K2 K4.
LET CK3=(CK3-MEAN(CK3))/STDEV(CK3)
LET CK3=CK3*K4+K2
```

The parameters related with these populations, which were constructed by using this program, were given in Table 3.1.

Table 3.1 Generated Population Parameters in Which $\sigma^2 = 25$ and the Block Effects are not Significant

		BLOCKS				MEAN ST.DEV.
		1	2	3	4	
TREATMENTS	1	50 5	50 5	50 5	50 5	50 5
	2	60 5	60 5	60 5	60 5	60 5
	3	70 5	70 5	70 5	70 5	70 5
	4	80 5	80 5	80 5	80 5	80 5
MEAN ST.DEV.		65 12.248	65 12.248	65 12.248	65 12.248	μ=65 12.248

Although each treatment has different average in the population shown in Table 3.1, the block averages are the same. The treatment effects and the block effects for the generated population are given in the Table 3.2.

Table 3.2 The True Treatment and the Block Effects For the Generated Population

	Treatments	Blocks
	τ_i	β_j
1	-15	0
2	-5	0
3	+5	0
4	+15	0

The following graphics show that the populations called POP1.1, POP1.2,..., POP1.16 are distributed normally.

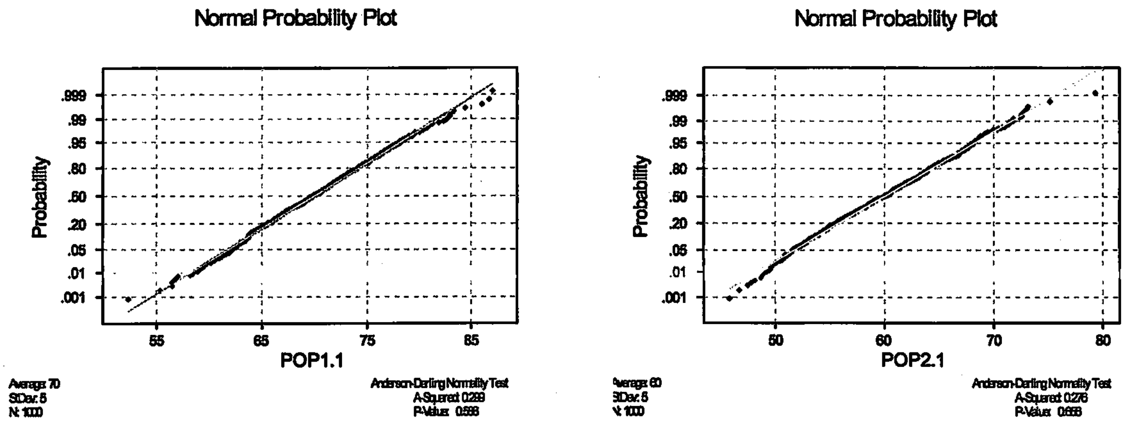

Figure 3.1 Normal Probability Plot of Generated Populations

An Anderson-Darling test for normality is performed and the numerical results are displayed with the graphs.

In these graphs, a straight (or close to straight) line indicates normality. A lot of curvature indicates non-normal data. The null hypothesis is that the data are normal; the alternative hypothesis is that the data are not normal. A p-value greater than the cut-value of our choice (0.05), says not to reject the null hypothesis, that is, not to reject the hypothesis that the population data are normal.

3.3 Forming the Randomized Complete Block Design and the Balanced Incomplete Block Design From the Generated Population

Random samples were taken from each population in order to construct the 4x4 randomized complete block design with single observation. Sampling was done without replacement.

After calculating the treatment effects for randomized complete block design by using the simulation program, appropriate balanced incomplete block design was constructed from this design by deleting some units. Regarding the condition of being balanced, the observation values in the 2nd, 8th, 9th and 15th units were erased.

The Minitab program was used to calculate the treatment effects estimates in randomized complete block and balanced incomplete block design. The Minitab program that calculates the treatment effects estimates in randomized complete block design and balanced incomplete block design is as follows:

SET2.MTB

```
LET K1=100
LET K2=150
EXEC 'SET3' 16
STACK C151-C166 C90
GLM C90=C91 C92;
COEF C93.
ERASE C83
LET C83(1)=C93(2)
LET C83(2)=C93(3)
LET C83(3)=C93(4)
LET K20=SUM(C83)
LET K21=(-1)*K20
LET C83(4)=K21
DELETE 2 8 9 15 C90
COPY C90 C1
LET K1=4
LET K2=3
LET K3=2
LET K4=3
LET K5=4
STAT C1;
BY C2;
SUMS C5.
STAT C1;
BY C3;
SUMS C6.
LET C54=C50*C6
LET C55=C51*C6
LET C56=C52*C6
LET C57=C53*C6
LET C58=C46*C5
```

```
LET C59=C47*C5
LET C60=C48*C5
LET C61=C49*C5
LET C8(1)=SUM(C54)
LET C8(2)=SUM(C55)
LET C8(3)=SUM(C56)
LET C8(4)=SUM(C57)
LET C9(1)=SUM(C58)
LET C9(2)=SUM(C59)
LET C9(3)=SUM(C60)
LET C9(4)=SUM(C61)
LET C7(1)=C5(1)-(1/K2)*C8(1)
LET C7(2)=C5(2)-(1/K2)*C8(2)
LET C7(3)=C5(3)-(1/K2)*C8(3)
LET C7(4)=C5(4)-(1/K2)*C8(4)
LET C10(1)=C6(1)-(1/K2)*C9(1)
LET C10(2)=C6(2)-(1/K2)*C9(2)
LET C10(3)=C6(3)-(1/K2)*C9(3)
LET C10(4)=C6(4)-(1/K2)*C9(4)
LET C80=(K2/(K1*K3))*C7
LET K10=MEAN(C1)
GLM C1=C2 C3;
RESIDUALS C99.
LET C100=C99*C99
LET K13=SUM(C100)/(K1*K4-K1-K5+1)
LET C99=C10*C10
LET K14=(K4*SUM(C99)/(K3*K5))/(K5-1)
LET K15=((K14-K13)*(K5-1))/(K1*(K4-1))
LET C99=K2*C7*(K13+K2*K15)+(C8-K2*K4*K10)*K13
LET K21=(K4-K3)*K13+(K3*K1)*(K13+K2*K15)
LET C82=C99/K21
LET C81=(C8-K2*K4*K10)/(K4-K3)
LET K51=K51+1
LET C11(K51)=C80(1)
LET C12(K51)=C80(2)
LET C13(K51)=C80(3)
LET C14(K51)=C80(4)
LET C15(K51)=C81(1)
LET C16(K51)=C81(2)
LET C17(K51)=C81(3)
LET C18(K51)=C81(4)
LET C19(K51)=C82(1)
LET C20(K51)=C82(2)
LET C21(K51)=C82(3)
LET C22(K51)=C82(4)
```

```
LET C23(K51)=C83(1)
LET C24(K51)=C83(2)
LET C25(K51)=C83(3)
LET C26(K51)=C83(4)
PRINT K51
```

SET3.MTB

```
LET K1=K1+1
LET K2=K2+1
SAMPLE 1 CK1 CK2
```

So, the balanced incomplete block design was obtained with the parameters given below.

a=4, b=4 r=3, k=3 ,λ =2 and N=12

By the help of the program, the necessary calculations for intrablock, interblock and combined estimates were obtained. The equations used for these estimates are given below in order:

$$\hat{\tau}_i = \frac{kQ_i}{\lambda a} \qquad i=1,2,\ldots,a \tag{3.4}$$

$$\tilde{\tau}_i = \frac{\sum_{j=1}^{b} n_{ij} y_{.j} - kr\bar{y}}{r-\lambda} \tag{3.5}$$

$$\hat{\tau}_i^* = \begin{cases} \dfrac{kQ_i(\hat{\sigma}^2 + k\hat{\sigma}_\beta^2) + (\sum_{j=1}^{b} n_{ij} y_{.j} - kr\bar{y})\hat{\sigma}^2}{(r-\lambda)\hat{\sigma}^2 + \lambda a(\hat{\sigma}^2 + k\hat{\sigma}_\beta^2)} & if \ \hat{\sigma}_\beta^2 \rangle 0 \\ \dfrac{y_{i.} - (1/a)y_{..}}{r} & if \ \hat{\sigma}_\beta^2 = 0 \end{cases} \tag{3.6}$$

After executing the program SET2.MTB for 2500 times, the intrablock, the interblock and the combined estimates of the treatment effects were obtained.

3.4 A Sample Application

After generating the population to be used in this study, the randomized complete block design and balanced incomplete block design were obtained from this population. The estimates of the treatment effects were calculated by using the program written in Minitab software package.

The results obtained by running SET2.MTB program only once are given in the tables below:

Table 3.3 An Example For Randomized Complete Block Design Obtained by Simulation

		BLOCKS			
		1	2	3	4
TREATMENTS	1	52.6429	50.7093	49.7790	44.2755
	2	66.8400	57.7852	61.6684	59.6934
	3	78.0609	70.9539	61.2230	79.7252
	4	90.0230	81.3956	72.0152	84.7867

Table 3.3 shows the randomized complete block design obtained by taking samples from the population. The analysis of variance for the randomized complete block design is given in the Minitab printout.

```
Analysis of Variance (Balanced Designs)

Factor      Type Levels Values
Treatments          fixed       4     1     2     3     4
Blocks              fixed       4     1     2     3     4

Analysis of Variance for RESPONSE

Source        DF         SS          MS        F       P
T              3    2387.43      795.81    30.95   0.000
B              3     237.67       79.22     3.08   0.083
Error          9     231.43       25.71
Total         15    2856.53
```

The p-value was compared with α=0.05 and it was observed that there were differences between the treatment averages.

The treatment effects of the population and the randomized complete block design treatment effects obtained by SET2.MTB program can be seen in Table 3.4. Estimated values and true values are close to each other.

Table 3.4 Treatment Effects For the Population and For Randomized Complete Block Design

Treatments Effects of the Population	Treatments Effects Estimates of the Randomized Complete Block Design
$\tau_1 = -15$	$\hat{\tau}_1$=-16.9969
$\tau_2 = -5$	$\hat{\tau}_2$=-4.85184
$\tau_3 = +5$	$\hat{\tau}_3$=6.14217
$\tau_4 = +15$	$\hat{\tau}_4$=15.7066

Balanced incomplete block design obtained from the randomized complete block design is shown in Table 3.5.

Table 3.5 Formed Balanced Incomplete Block Design

		BLOCKS			
		1	2	3	4
TREATMENTS	1	52.6429	50.7093	-	44.2755
	2	-	57.7852	61.6684	59.6934
	3	78.0609	70.9539	61.2230	-
	4	90.0230	-	72.0152	84.7867

Analysis of variance for balanced incomplete block design is given in Minitab printout. Here the "GLM" command was used.

```
Analysis of Variance for response

Source        DF    Seq. SS     Adj. SS     Adj. MS       F        P
Treatments     3    1803.26    1740.74     580.25     26.39    0.002
Blocks         3     250.49     250.49      83.50      3.80    0.092
Error          5     109.94     109.94      21.99
Total         11    2163.69
```

The intrablock, the interblock and the combined estimates of the treatment effects in balanced incomplete block design obtained by running program SET2.MTB and the estimates of treatment effects of randomized complete block design are given in Table 3.6.

Table 3.6 Estimates of the Treatment Effects

Treatment	True Parameter Values	$\hat{\tau}_i$ For Randomized Complete Block Design	Intrablock Estimates $\hat{\tau}_i$	Interblock Estimates $\tilde{\tau}_i$	Combined Estimates $\hat{\tau}_i^*$
1	-15.0000	-16.9969	-18.2560	1.0527	-17.6910
2	-5.0000	-4.8518	-3.2087	-24.7675	-3.8394
3	5.0000	6.1421	4.4539	7.2037	4.5343
4	15.0000	15.7066	17.0107	16.5110	16.9961

3.5 Treatment Effects Estimates in the Balanced Incomplete Block Design and the Randomized Complete Block Design by Simulation When the Block Effects are Insignificant

In this section of the study, the estimates of the treatment effects for balanced incomplete block design and randomized complete block design were obtained for the situation in which the variance of the population was 25, 50 and 100. The population parameters are shown in Table 3.7.

Table 3.7 The Generated Population in which $\sigma^2 = 25$ and the Treatment Effects are Significant and the Block Effects are Insignificant

		BLOCKS				MEAN ST.DEV
		1	2	3	4	
TREATMENTS	1	50 5	50 5	50 5	50 5	50 5
	2	60 5	60 5	60 5	60 5	60 5
	3	70 5	70 5	70 5	70 5	70 5
	4	80 5	80 5	80 5	80 5	80 5
MEAN ST.DEV.		65 12.248	65 12.248	65 12.248	65 12.248	μ=65 12.248

As seen in Table 3.7, the treatments have different averages in the population for randomized complete block design. While estimating the treatment effects, the block effects are considered as insignificant. The true treatment effects and the block effects of the population are given in Table 3.8.

Table 3.8 The True Treatment and the Block Effects For the Generated Population

	Treatments	Blocks
	τ_i	β_j
1	-15	0
2	-5	0
3	+5	0
4	+15	0

In order to determine the best estimates of the treatment effects, the program SET2.MTB was run for 2500 times and the intrablock($\hat{\tau}_i$), the interblock($\tilde{\tau}_i$) and the combined($\hat{\tau}_i^*$) estimates were obtained. These estimates related to the treatment effects are independent from each other and they are normally distributed. Figure 3.2 shows that the intrablock, the interblock and the randomized complete block design treatment effects estimates have the normal distribution. When the block effects are insignificant, the combined estimates values are not normally distributed according to the simulation results.

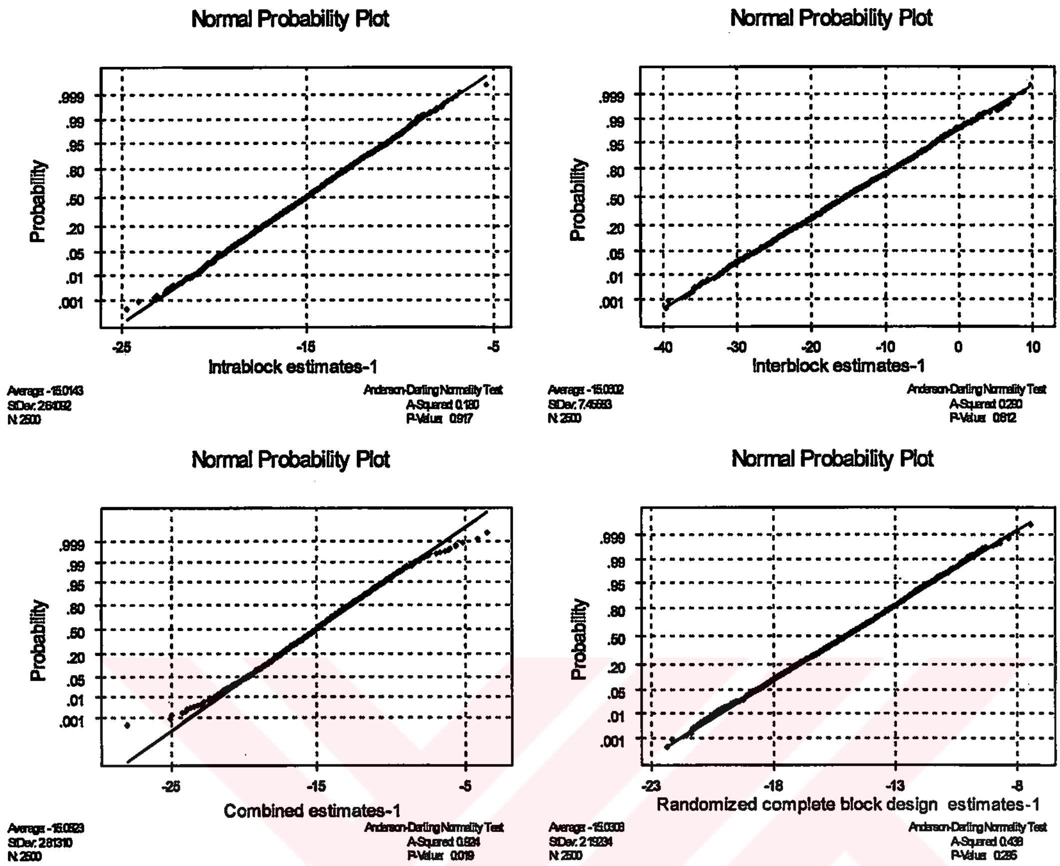

Figure 3.2 Normal Probability Plot of the Treatment Effects Estimates

The correlation of the treatment effects estimates for the intrablock and for the randomized complete block design are given in Table 3.9.

Table 3.9 Correlation Between the Intrablock Estimates and the Randomized Complete Block Design Estimates

	Intra-1	Intra-2	Intra-3	Intra-4	Comp-1	Comp-2	Comp-3
Intra-2	-0.345						
Intra-3	-0.332	-0.321					
Intra-4	-0.316	-0.361	-0.325				
Comp-1	0.814	-0.296	-0.278	-0.234			
Comp-2	-0.307	0.824	-0.249	-0.290	-0.375		
Comp-3	-0.262	-0.259	0.816	-0.277	-0.328	-0.309	
Comp-4	-0.261	-0.274	-0.277	0.811	-0.317	-0.321	-.349

As seen in Table 3.9, the correlations between the values of the intrablock estimates and the values of the treatment effects estimates of the randomized complete block design are high.

Table 3.10 Correlation Between the Interblock Estimates and the Randomized Complete Block Design Estimates

	Inter-1	Inter-2	Inter-3	Inter-4	Comp-1	Comp-2	Comp-3
Inter-2	-0.330						
Inter-3	-0.333	-0.356					
Inter-4	-0.336	-0.317	-0.328				
Comp-1	0.302	-0.103	-0.105	-0.094			
Comp-2	-0.110	0.286	-0.099	-0.078	-0.375		
Comp-3	-0.107	-0.088	0.277	-0.087	-0.328	-0.309	
Comp-4	-0.092	-0.097	-0.069	0.263	-0.317	-0.321	-0.349

In Table 3.10, the correlations between the interblock estimates and the treatment effects estimates for randomized complete block design are low. Although the interblock estimates are close to the values of the treatment effects estimates for randomized complete block design, the correlation between them is low.

Table 3.11 Correlation Between the Combined Estimates and the Randomized Complete Block Design Estimates

	Comb-1	Comb-2	Comb-3	Comb-4	Comp-1	Comp-2	Comp-3
Comb-2	-0.337						
Comb-3	-0.324	-0.330					
Comb-4	-0.311	-0.346	-0.352				
Comp-1	0.761	-0.269	-0.234	-0.237			
Comp-2	-0.295	0.761	-0.221	-0.255	-0.375		
Comp-3	-0.243	-0.242	0.733	-0.252	-0.328	-0.309	
Comp-4	-0.238	-0.255	-0.268	0.754	-0.317	-0.321	-0.349

When Table 3.11 is inspected, it can be seen that the correlations between the combined estimates and the treatment effects estimates for the randomized complete block design are high. The correlations between the intrablock estimates and the randomized complete block design estimates are higher than the correlations between the combined estimates and the randomized complete block design estimates.

Table 3.12 gives the $\hat{\tau}_i$, $\tilde{\tau}_i$, $\hat{\tau}_i^*$ estimates of the treatment effects in the balanced incomplete block design.

Table 3.12 Estimates of the Treatment Effects Obtained by Simulation For $\sigma^2 = 25$

	Intrablock Estimates $\hat{\tau}_i$		Interblock Estimates $\tilde{\tau}_i$		Combined Estimates $\hat{\tau}_i^*$		$\hat{\tau}_i$ for Randomized Complete Block Design	
	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
1	-15.014	2.641	-15.060	7.456	-15.052	2.813	-15.031	2.192
2	-4.965	2.708	-4.848	7.482	-4.912	2.901	-4.983	2.168
3	5.051	2.610	4.930	7.559	5.022	2.886	5.077	2.136
4	14.928	2.660	14.977	7.353	14.942	2.894	14.937	2.138

It can be seen in Table 3.12 that the treatment effects estimates of the balanced incomplete block design are close to the treatment effects estimates in the randomized complete block design. Examining the standard deviations of the treatment effects estimates, it is observed that the standard deviations of the interblock estimates are high. Among the estimation methods in the balanced incomplete block design, the lowest standard deviation belongs to the intrablock estimates values.

In order to examine the effect of the population variance on the treatment effects estimates, the population variance was increased and the treatment effects estimates were repeated for $\sigma^2 = 50$. These estimates values are given in Table 3.13.

Table 3.13 Estimates of the Treatment Effects Obtained by Simulation For $\sigma^2 = 50$

	Intrablock Estimates $\hat{\tau}_i$		Interblock Estimates $\tilde{\tau}_i$		Combined Estimates $\hat{\tau}_i^*$		$\hat{\tau}_i$ for Randomized Complete Block Design	
	Mean	St.Dev	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
1	-14.977	3.612	-14.869	10.592	-15.023	3.997	-14.995	3.068
2	-4.979	3.793	-4.687	10.785	-4.952	4.175	-4.944	3.075
3	4.942	3.632	4.684	10.286	4.903	3.987	4.952	2.981
4	15.015	3.674	14.872	10.584	15.071	4.090	14.988	3.063

The standard deviation of the treatment effects estimates increased because the variance of the population was increased to 50. The standard deviation of interblock estimates increased more and therefore the reliability of these estimates decreased. When the interblock estimates were compared with the treatment effects estimates in the randomized complete block design, it was seen that, the results of the interblock estimates were more biased than the intrablock and the combined estimates. Even though the standard deviation of the intrablock and the combined estimates increased because of increasing the population variance to 50, these estimates were better than the interblock estimates.

When the population variance was increased up to 100, the standard deviation of the estimates increased more, as seen in the Table 3.14. The higher the population variance is, the higher the variance of estimates gets. The minimum standard deviations are obtained from the intrablock estimates.

Table 3.14 Estimates of the Treatment Effects Obtained by Simulation For $\sigma^2 = 100$

	Intrablock Estimates $\hat{\tau}_i$		Interblock Estimates $\tilde{\tau}_i$		Combined Estimates $\hat{\tau}_i^*$		$\hat{\tau}_i$ for Randomized Complete Block Design	
	Mean	St.Dev	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
1	-15.027	5.239	-15.467	15.147	-15.002	5.928	-15.136	4.360
2	-4.919	5.358	-4.632	14.862	-4.968	6.183	-4.979	4.361
3	4.891	5.266	4.683	14.861	4.880	5.913	4.960	4.329
4	15.055	5.131	15.415	14.966	15.091	6.007	15.155	4.286

Table 3.15 is constructed to examine the variation between the mean and the standard deviation of the intrablock estimates according to the population variance.

Table 3.15 Intrablock Estimates For the Population Variance σ^2=25, 50, 100

	σ^2=25		σ^2=50		σ^2=100	
	Mean	St.Dev	Mean	St.Dev.	Mean	St.Dev.
1	-15.014	2.641	-14.977	3.612	-15.027	5.239
2	-4.965	2.708	-4.979	3.793	-4.919	5.358
3	5.051	2.610	4.942	3.632	4.891	5.266
4	14.928	2.660	15.015	3.674	15.055	5.131

It is seen that, the intrablock estimates are unbiased but the standard deviations of the estimates are increasing when the population variance is increasing.

3.6 Treatment Effects Estimates in the Balanced Incomplete Block Design and the Randomized Complete Block Design by Simulation When the Block Effects are Significant

In this section, when the block effects are significant, all studies in the section 3.5 were repeated for the population variance is 25 and 50. Generated population parameters are described in Table 3.16.

Table 3.16 Generated Population Parameters in which $\sigma^2 = 25$ and the Block Effects are Significant

		BLOCKS				MEAN ST.DEV
		1	2	3	4	
TREATMENTS	1	50 5	60 5	70 5	80 5	65 12.248
	2	60 5	70 5	80 5	90 5	75 12.248
	3	70 5	80 5	90 5	100 5	85 12.248
	4	80 5	90 5	100 5	110 5	95 12.248
MEAN ST.DEV.		65 12.248	75 12.248	85 12.248	95 12.248	μ=80 16.583

As seen in Table 3.16, the treatments and the blocks have different averages in the generated population for the randomized complete block design. The true treatment and block effects for the population are given in Table 3.17.

Table 3.17 The True Treatment and the Block Effects For the Generated Population

	Treatments	Blocks
	τ_i	β_j
1	-15	-15
2	-5	-5
3	+5	+5
4	+15	+15

In order to decide on the best estimator in the balanced incomplete block design which was obtained from the population with the block effects considered as significant, SET2.MTB program was run for 2500 times and the intrablock($\hat{\tau}_i$), the interblock($\tilde{\tau}_i$) and the combined($\hat{\tau}_i^*$) estimates were obtained.

Figure 3.3 shows that the intrablock, the interblock, the combined and the randomized complete block design estimates values have the normal distribution.

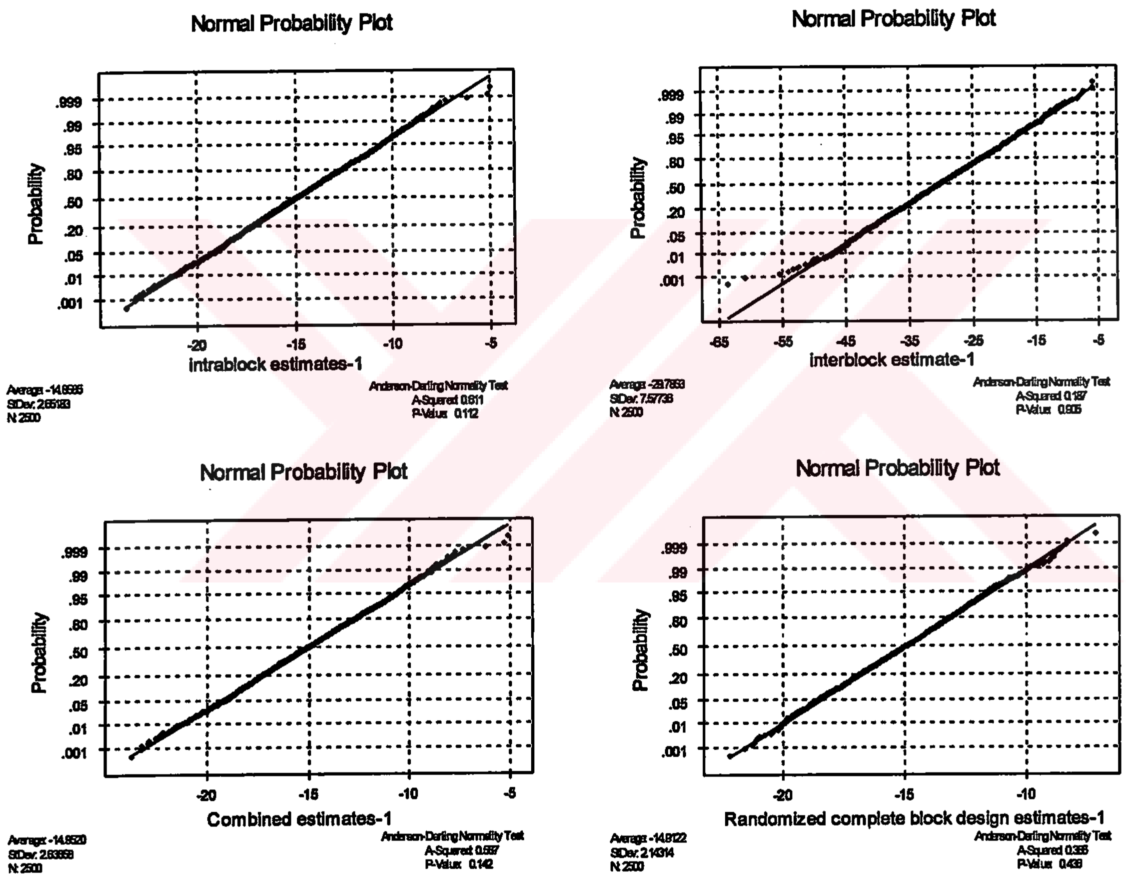

Figure 3.3 Normal Probability Plot of the Treatment Effects Estimates

Table 3.18 gives the $\hat{\tau}_i$, $\tilde{\tau}_i$, $\hat{\tau}_i^*$ estimates of the treatment effects in the balanced incomplete block design.

Table 3.18 Estimates of the Treatment Effects Obtained by Simulation For $\sigma^2 = 25$

	Intrablock Estimates $\hat{\tau}_i$		Interblock Estimates $\tilde{\tau}_i$		Combined Estimates $\hat{\tau}_i^*$		$\hat{\tau}_i$ for Randomized Complete Block Design	
	Mean	St.Dev	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
1	-14.858	2.652	-29.785	7.577	-14.952	2.639	-14.912	2.143
2	-5.079	2.632	39.719	7.453	-4.799	2.639	-5.083	2.129
3	5.034	2.624	-39.974	7.494	4.753	2.628	5.044	2.166
4	14.903	2.610	30.040	7.342	14.998	2.596	14.951	2.150

Examining the estimates obtained by executing SET2.MTB program 2500 times, it is seen that the interblock estimates differ from the true values of the treatment effects. In this case, the values of the combined estimates which are the combined estimates of the interblock and the intrablock estimates are closer to the treatment effect estimates of randomized complete block design. The standard deviations of the interblock estimates are higher than the standard deviations of the other estimates.

Table 3.19 shows the estimates of the treatment effects when the population variance is 50.

Table 3.19 Estimates of the Treatment Effects Obtained by Simulation For $\sigma^2 = 50$

	Intrablock Estimates $\hat{\tau}_i$		Interblock Estimates $\tilde{\tau}_i$		Combined Estimates $\hat{\tau}_i^*$		$\hat{\tau}_i$ for Randomized Complete Block Design	
	Mean	St.Dev	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
1	-15.130	3.807	-30.068	10.435	-15.304	3.775	-15.085	3.113
2	-5.005	3.721	39.985	10.579	-4.470	3.750	-4.969	3.085
3	5.109	3.707	-39.810	10.488	4.576	3.740	5.029	3.016
4	15.026	3.837	29.893	10.749	15.199	3.804	15.024	3.078

It is seen that the interblock estimates are different than the other estimates. When the standard deviations of the estimates are examined, it is seen that the combined estimates and the intrablock estimates are better estimates of the treatment effects when the block effects are significant.

CHAPTER FOUR
CONCLUSIONS

4.1 Conclusions

In this thesis, the treatment effects estimates obtained by using three methods in the balanced incomplete block design have been compared with the treatment effects estimates in the randomized complete block design. The intrablock, the interblock and the combined estimates for the balanced incomplete block design and the least squares estimates for the randomized complete block design of the treatment effects were calculated 2500 times by simulation. In this study as the statistical comparison criteria, mean of 2500 treatment effects estimates and standard deviation of those estimates were used.

The populations were generated by using a macro program written in the Minitab statistical software. Searching for the most appropriate estimator for the treatment effects, the simulation study was conducted separately for both the insignificance and the significance of the block effects in the generated population.

After the evaluation of the results, it was observed that each of the three methods gave unbiased results for the balanced incomplete block design when the block effects were insignificant. When evaluating the correlations between the treatment effects estimates in the randomized complete block design and the treatment effects estimates using the suggested methods in the balanced incomplete block design, it was observed that the correlations between the intrablock estimates and the randomized complete block design estimates are higher than the others. As for the standard deviations of the treatment effects estimates, the values of the intrablock estimates are lowest among the others.

When the block effects were significant in the generated population, the intrablock and the combined estimates give unbiased results but the interblock estimates are different from the randomized complete block design estimates because of the great variation between blocks. The standard deviations of the intrablock estimates and the combined estimates are better estimates of the treatment effects when the block effects were significant. Although the intrablock estimates and the combined estimates give approximately similar results, in practice only using intrablock estimation method is more convenient because of the computational difficulties of the combined estimation method.

REFERENCES

Anderson, R.L.&Bancroft, T.A. (1952). Statistical Theory in Research. USA: McGraw-Hill.

Bose, R.C. & Manvel, B. (1984). Introduction to Combinatorial Theory. New York: John Wiley & Sons.

Cochran, W.G. & Cox, G.M. (1957). Experimental Designs. New York: John Wiley & Sons.

Cox, D.R. (1992). Planning of Experiments. New York: John Wiley & Sons.

Federer, W.T. (1955). Experimental Design. New York: Macmillan Company.

Graybill, F.A. (1961). An Introduction to Linear Statistical Models. USA: McGraw-Hill.

Graybill, F.A. (1976). Theory and Application of the Linear Model. USA: Duxbury Press.

Graybill, F.A.&Deal, R.B. (1959). Combining Interblock and Intrablock Information in Balanced Incomplete Blocks. Ann. Math. Statist. 30, 799-805

John, P.W.M. (1971). Statistical Design and Analysis of Experiments. New York: Macmillan Company.

Johnson, N.L. & Leone, F.C. (1964). Statistics and Experimental Design. New York: John Wiley&Sons.

Kempthorne, O. (1952). The Design and Analysis of Experiments. New York: John Wiley&Sons.

Kurt, S. (1991). Deney Düzenleme: Lecture Notes. İzmir.

Mead, R. (1988). The Design of Experiments. New York: Cambridge University Press.

Meyer, R.&Krueger, D. (1998). A Minitab Guide to Statistics. USA: Prentice-Hall.

Monthgomery, D.C. (1992). Design and Analysis of Experiments. New York: John Wiley & Sons.

Rao, P.S.R.S. (1997). Variance Components Estimation. London: Chapman-Hall.

Seshadri, V. (1963). Constructing Uniformly Better Estimators. Journal of the American Association. 58, 172-175

Shah, K.R. (1964). Use of Interblock Information to Obtain Uniformly Better Estimators. Ann. Math. Statist. 35, 1064-1078

Weerakkody, G.J. (1992). A Note on the Recovery of Interblock Information in Balanced Incomplete Block Designs. Commun. Statist.-Theory Meth. 21(4), 1125-1136