ASYMMETRICAL RESPONSE SURFACE DESIGNS

BY

P.R. RAMACHANDER

Dissertation submitted in fulfilment of the requirements for the award of Diploma in Agricultural and Animal Husbandry Statistics of the Institute of Agricultural Research Statistics (I.C.A.R.), NEW DELHI.

1963.

ACKNOWLEDGEMENT

I have great pleasure in expressing my deep sense of gratitude to Prof. M.N.Das, Professor of Statistics, Institute of Agricultural Research Statistics, for his valuable guidance, keen interest, valuable criticisms and constant encouragement throughout the course of the investigation and immense help in writing this thesis.

I am grateful to Dr. V. G. Panse, Statistical Adviser, Institute of Agricultural Research Statistics, for providing necessary facilities to carry out this work.

(P. R. Ramachander)

CONTENTS

INTRODUCTION

Of late considerable emphasis has been laid on evolving designs suitable for fitting response surfaces. These designs are used to solve mainly industrial problems where a knowledge of the functional relationship between a quantitative response such as yield, cost etc. and a set of controlled variables such as temperature, preassure etc. is necessary. When this functional relationship is not known we have to lay down experiments with a view to approximating it by a polynomial. With this end in view Box and Hunter (1957) introduced a series of response surface designs called rotatable designs. They presented several designs using geometrical confligurations. By following a similar method Gardiner and others (1959) obtained some third order designs. Bose and Draper (1959) obtained these designs by using a different method. Many other designs were given by Draper (1960), Box and Behnken (1960) and Thaker (1960).

A radically different approach for the construction of the above type of designs was given by Das (1962), Das and Narasimham (1962) and Das (1963) by using factorial experiments and incomplete block designs.

Box and Hunter (1957) gave the criteria for blocking in these designs. Box and Behnken (1957), Das (1962), Das and Narasimham (1962) and Das (1963) gave some rotatable designs with blocking.

The problem of estimation of missing value in rotatable designs was considered by Draper (1962).

De Baun (1959), Box and Draper (1959) and Box and Behnken(1960) considered some other response surface designs which are not rotatable. De Baun (1959) has compared the rotatable designs with some of these designs.

In this thesis we have attempted to present some new series of response surface designs which are of asymmetrical type and can be used for fitting second degree response surfaces with advantage. Some of these designs enjoy some advantages over the corresponding rotatable designs. A number of designs split into blocks have also been presented in the thesis. The problem of estimation of missing value in these designs has also been considered

ASYMMETRICAL RESPONSE SURFACE DESIGNS

One most important series of response surface designs is the rotatable designs introduced by Box and Hunter (1957). De Baun (1959) and Box and Behnken (1960) gave some other designs suitable for fitting response surfaces but they are not always rotatable. Two important criteria with which such designs have been evolved are that, the parameters defining the surface should be estimable without much involvement in the solution of their normal equations when the surface is fitted through least squre technique and that it should be rotatable i.e. the variance of an estimabled response at a certain point defined by the levels of the factors on suitably chosen original scale should be a function of the distance of the point from the origin. It is seen that the second criterian is not of the kind that necessarily must be satisfied, but rather it gives an objective which might be aimed at, if it can be satisfied without creating other difficulties.

All the response surface designs obtained so far are of the symmetric type that is each of the factors under study have got the same number of levels. It is well known that there are situations in agronomic and other trials where it becomes desirable to include different numbers of levels for the different factors. This consideration necessitated evolving designs of the asymmetrical type. The same considerations apply in the case of response surface designs also. In the present thesis we have therefore made an attempt to evolve suitable asymmetrical designs for the study of second degree response surfaces. Two series of such designs have been obtained together with expressions for the estimates of the various regression coefficients and an expression for the variance of estimated response at given points. Some other designs for various numbers of factors have also been obtained.

In this thesis we have confined our attention to designs which have two types of levels of factors viz. one set of factors has k levels and the other set has l levels each. Let there be n factors of which n_1 factors have k levels each and n_2 factors have l levels each ($n_1 + n_2 = n$). Let x_{ijk} denote the level of the ith factor in the jth group, in the kth of a total of N combinations of the levels. Such combinations will also be called the design points. The total number of design points in a design will be denoted by N.

The N design points chosen so as to satisfy the following relations will ensure simplicity in the solution of normal equations obtained through the least square technique for estimating parameters (regression coeeficients) in the equation of the response surface. This becomes evident from a glance at the normal equations for estimating the regression coefficients of the surface.

In all the following relations the summation is over N design points:

$$\sum x_{ij_1}^{k_1} \, x_{kj_2}^{k_2} \, x_{lj_3}^{k_3} \, x_{mj_4}^{k_4} = 0 \qquad \text{(I)}$$

whenever atleast one of the k_i's is odd and

$$k_1 + k_2 + k_3 + k_4 \leqslant 4$$

$$\left.\begin{aligned} \sum x_{i1}^{2} &= \text{constant} = \lambda_1 \\ \sum x_{i2}^{2} &= \text{constant} = \lambda_2 \end{aligned}\right\} \qquad \text{(II)}$$

$$\sum x_{i1}^{4} - (\sum x_{i1}^{2})^2 / N = \text{constant} = A$$

$$\sum x_{i2}^{4} - (\sum x_{i2}^{2})^2 / N = \text{constant} = C$$

$$\sum x_{i1}^{2} \, x_{j1}^{2} - (\sum x_{i1}^{2})^2/N = \text{constant} = B \qquad \text{(III)}$$

$$\sum x_{i2}^{2} \, x_{j2}^{2} - (\sum x_{i2}^{2})^2/N = \text{constant} = D$$

$$\sum x_{i1}^{2} \, x_{i2}^{2} - (\sum x_{i1}^{2})(\sum x_{i2}^{2})/N = \text{constant} = E$$

When these conditions are satisfied the normal equations for fitting the second degree response surface viz.

$$y = B_0 + \sum_{i_1} B_{i1} \, x_{i1} + \sum_{i_1} B_{i1,i1} \, x_{i1}^2 + \sum_{i_2} B_{i2} \, x_{i2} + \sum_{i_2} B_{i2,i2} \, x_{i2}^2$$

$$+ \sum_{\substack{i_1, j_1 \\ i_1 \neq j_1}} B_{i1,j1} \, x_{i1} x_{j1} + \sum_{\substack{i_2, j_2 \\ i_2 \neq j_2}} B_{i2,j2} \, x_{i2} x_{j2} + \sum_{i_1, i_2} B_{i1,i2} \, x_{i1} \, x_{i2}.$$

(where $B_{ij,ik}$ denotes coefficient of $x_{ij}\,x_{ik}$, B_{ij} denotes coefficients of x_{ij} and B_0 is the constant term) of the n independent factors defined earlier on the response y produced by their combinations are much simplified. Denoting by b's the corresponding estimates of B's we have the various estimates as follows:

$$b_{i1} = \frac{1}{\lambda_1}\sum x_{i1}y \qquad b_{i2} = \frac{1}{\lambda_2}\sum x_{i2}y$$

$$b_{i1,j1} = \frac{1}{E+\lambda_1^2/N}\sum x_{i1}x_{j1}y$$

$$b_{i1,i2} = \frac{1}{B+\lambda_1\lambda_2/N}\sum x_{i1}x_{i2}y$$

$$b_{i2,j2} = \frac{1}{D+\lambda_2^2/N}\sum x_{i2}x_{j2}y$$

$$n_1 b_{i1,i1} = \frac{\sum x_{i1}^2 y}{(A-E)\,K}\left[\begin{array}{l}(C+D(n_2-1))(n_1A+E((n_1-1)^2-1)) \\ -n_1n_2(n_1-1)B^2\end{array}\right]$$

$$-\frac{\sum\limits_{j\neq i}\sum x_{j1}^2 y}{(A-E)\,K}\left\{n_1n_2B^2-n_1E\left[C+D(n_2-1)\right]\right\}$$

$$-\frac{G}{NK(A-E)}\left[\begin{array}{l}\lambda_1\left[\begin{array}{l}[C+D(n_2-1)][n_1A-E((n_1-1)^2-1)] \\ -n_1n_2(n_1-2)B^2-n_1B[C+D(n_2-1)]\end{array}\right] \\ -\lambda_2 n_1 B(A-E)\end{array}\right]$$

$$-\frac{n_1B}{K}\sum_j x_{j2}^2 y.$$

where $K = \left[A+E(n_1-1)\right]\left[C+D(n_2-1)\right] - n_1n_2B^2$.

$$n_2 b_{12,12} = \frac{\sum x_{i2}^2 y}{(C-D)K}\left[\left[A+E(n_1-1)\right]\left[n_2 B+D\left[(n_2-1)^2-1\right]\right] - n_1 n_2 (n_2-1) B^2\right]$$

$$+ \frac{\sum_{j\neq i}\sum x_{i2}^2 y}{(C-D)K}\left\{n_1 n_2 B^2 - n_2 D\left[A-E(n_1-1)\right]\right\}$$

$$- \frac{G}{NK(C-D)}\left[\lambda_1\left[\left[A+E(n_1-1)\right]\left[n_2 B+D\left[(n_2-1)^2-1\right]\right] - n_1 n_2 (n_2-2) B^2 - n_1 D\left[A+E(n_1-1)\right]\right] - \lambda_1 n_2 B(A-E)\right]$$

$$- \frac{n_2 B}{K} \sum_j x_{j1}^2 y$$

$$N b_0 = G - \frac{1}{K}\left[\sum_i \sum x_{i1}^2 y \left[\lambda_1\left[C+D(n_2-1)\right] - \lambda_2 n_2 B\right] + \sum_i \sum x_{i2}^2 y \left[\lambda_2\left[A-E(n_1-1)\right] - \lambda_1 n_1 B\right]\right]$$

$$- \frac{G}{NK}\left[n_1 \lambda_1^2\left[C+D(n_2-1) - n_1 n_2 \lambda_2 B\right] + n_2 \lambda_2^2\left[A-E(n_1-1) - n_1 n_2 \lambda_1 B\right]\right]$$

From these solutions it is evident that if finite solutions are to exist for the normal equations of the above type of designs then the constants must satisfy

(1) $C \neq D$ and $A \neq E$

(2) $K \neq 0$

i.e. $\left[A+E(n_1-1)\right]\left[C+D(n_2-1)\right] \neq n_1 n_2 B^2$

In most cases condition 1 is satisfied. The second condition can be satisfied, in most of the cases by adding the cent-ral points (0,....,0). In other cases it can be satisfied by adding points of the type (0,......,0),(- ,0,.....,0), (0, 0,....,0) etc. which means that the number of levels in one group of factors is increased.

The variances and covariances of the various coefficients have been obtained from the following considerations:

$V(b_{ij})$ = Coefficient of $\sum x_{ij}y$ in solution of b_{ij}

$V(b_{ij,kj})$ = Coefficient of $\sum x_{ij}x_{kj}y$ in solution of $b_{ij,kj}$

$Cov(b_{ij,kj}, b_{lj,mj})$ = 2(Coefficient of $\sum x_{lj}x_{mj}y$ in solution of $b_{ij,kj}$)

$Cov(b_0, b_{ij,kj})$ = 2(Coefficient of $\sum x_{ij}x_{kj}y$ in solution of b_0)

These results follow from the theory of least squares (Rao, C.R., 1952) and the variances and covariances are as follows, taking σ^2 as the error variance.

$$\frac{Var(b_{i1})}{\sigma^2} = \frac{1}{\lambda_1} \qquad \frac{Var(b_{i2})}{\sigma^2} = \frac{1}{\lambda_2}$$

$$\frac{Var(b_{i1,j1})}{\sigma^2} = \frac{N}{NE + \lambda_1^2}$$

$$\frac{Var(b_{i2,j2})}{\sigma^2} = \frac{N}{ND + \lambda_2^2}$$

$$\frac{Var(b_{i1,i2})}{\sigma^2} = \frac{N}{NB + \lambda_1\lambda_2}$$

$$\frac{Var(b_0)}{\sigma^2} = 1/N + 1/N^2K \left[\begin{array}{c} \lambda_1^2 n_1 [C+D(n_2-1)] + \lambda_2^2 n_2 [A+E(n_1-1)] \\ -2 \lambda_1 \lambda_2 n_1 n_2 B^2 \end{array} \right]$$

$$\frac{Cov(b_0, b_{11,11})}{\sigma^2} = -\frac{2}{NK} \left[\lambda_1 [C+D(n_2-1)] - \lambda_2 n_2 B \right]$$

$$\frac{Cov(b_0, b_{12,12})}{\sigma^2} = -2/NK \left[\lambda_2 [A+E(n_1-1)] - \lambda_1 n_1 B \right]$$

$$\frac{Var(b_{11,11})}{\sigma^2} = \frac{1}{n_1 K(A-E)} \left[\begin{array}{c} [C+D(n_2-1)][n_1 A+E((n_1-1)^2-1)] \\ -n_1 n_2 (n_1-1) B^2 \end{array} \right]$$

$$\frac{Var(b_{12,12})}{\sigma^2} = \frac{1}{n_2 K(C-D)} \left[\begin{array}{c} [A+E(n_1-1)][n_2 C+D((n_2-1)^2-1)] \\ -n_1 n_2 (n_2-1) B^2 \end{array} \right]$$

$$\frac{Cov(b_{11,11}, b_{j1,j1})}{\sigma^2} = \frac{2}{K(A-E)} \left[n_2 B^2 - E[C+D(n_2-1)] \right]$$

$$\frac{Cov(b_{12,12}, b_{j2,j2})}{\sigma^2} = \frac{2}{K(C-D)} \left[n_1 B^2 - D[A+E(n_1-1)] \right]$$

$$\frac{Cov(b_{11,11}, b_{j2,j2})}{\sigma^2} = \frac{-2B}{K}$$

We have evolved some series of designs of the above type. We confine ourselves to the case where the levels of the various factors are three or five. No factor with two levels has been included because if we include a factor with two levels some constant in the second degree surface cannot be estimated. Also all the designs which we are going to consider have either one or two factors which cause asymmetry though more general cases can be obtained.

Before we present the actual designs in terms of actual values of x_{ijk} we shall explain the following method of their construction which we have adopted.

A design of the above type in n factors can be constructed by taking one or more combinations of some unknown levels a,b,c.......etc. together with the zero level. Such combinations for four factors, say, will be like 1) a a a a, 2) a b b b, 3) b a a b and so on.

Next we shall have another design in n factors of the form 2^n where the two levels of each of the factors are + 1 and - 1. We can now get one more set of combinations when any combination of the first design is associated with any combination of the second design, by multiplying the corresponding entries that is the levels of the same factor in the two combinations and writing the products in the same order. This method of association of any two combinations of the two designs will hereafter be called multiplication. By 'multiplying' any combination of the first design, with all the combinations of the second design, we shall get 2^m distinct combinations where m denotes, the number of non-zero unknown levels in the combination considered of the first design.

Thus we have 3types of combinations :

(i) Factorial combinations of the unknown levels a,b,....etc. together with zero.

(ii) Factorial combinations of the levels + 1 and - 1

(iii) combinations when each O, a, b....., is associated with + 1 and - 1 through multiplication.

The first type of factorial combinations will be called combination of unknown levels, the second will be called associate combinations, and the third which actually constitute the design are our design points.

It will be seen easily that if a design be formed by including all the distinct points which are got by multiplying any combination of the unknown levels with all the associate combinations, these points will always satisfy relation - I on page 5. When $m > 4$ a suitable fraction of the 2^m associated combinations, where no interaction with less than five factors is included in its identity group will again satisfy relation I.

For satisfying relations II and III we have to choose one or more combinations of unknown levels.

In what follows we shall always give the factorial combinations involving unknown levels which will be denoting the design points that can be generated from them.

For constructing an asymmetric design in n factors we proceed as follows:

Consider first a central composite design in (n-1) factors which can be obtained from the following sets through multiplication with the appropriate associate combinations.

Set I: (p p .... p)

Set II: (q 0 .... 0)

....

(0 0 .... q)

Our aim is to add another factor with three levels to this design so that conditions I, II, and III on page 5 are satisfied. The obvious method is to add an unknown level 'a' to each of the sets together with '0' to each of the sets. B ut this makes the number of points two large we have however obtained below two series of designs as follows:

(1) adding '0' to set I and unknown 'a' to set II

(2) adding an unknown 'a' to set I above and '0' to set II.

SERIES -1

The following sets of unknown levels when multiplied by the appropriate associate combination give an asymmetrical design with the first factor at 3 levels and each of the rest at 5 levels.

Set I: (0 p p ...... p)

Set II: (a q 0 ...... 0)

(a 0 q ...... 0)

....

(a 0 ...... q)

Set III: (0 0 .... 0)

The central point (0 0 .... 0) is to be added to satisfy condition No.2 required for the existance of the solution.

Here

N = Total number of points

$= 2^{s-1} + 4n-3$, where 2^{s-1} points are got from set I

$n_2 = n-1$, $n_1 = 1$

For $n \geqslant 6$, a fraction of set I points is to be taken as explained earlier.

Here we shall denote the levels of the first factor by x_1 and those of the kth factor in the second group by x_k, $k = 2 \ldots\ldots n$.

So far we have not attempted to impose any other restriction on the summation of powers or products of powers of x_i's excepting those required to achieve simplicity of solutions of normal equations. By imposing certain other restrictions it is possible to make such a design rotatable in respect of certain of the factors. For example by imposing the restriction

$$\sum x_i^4 = 3\sum x_i^2 x_j^2, \quad i,j = 2,\ldots\ldots n, \; i \neq j$$

the design will be rotatable in so far as the (n-1) factors in the second group are concerned, that is when the levels of the first set of factors remains the same.

Further by putting $\sum x_i^2 = N$, $i = 1 \ldots\ldots n$, the variability of each variable x_{ik} can be kept constant. We note that with the above restrictions all the unknowns in the above series are determined.

From the rotatability condition for the second group the following equation is obtained $q^4 = 2^{s-2}p^4$
Also from the condition $\sum x_i^2 = N \quad i = 1,\ldots,n$
we have $4(n-1)a^2 = N$ and $2^s p^2 + 4q^2 = N$.

The solutions of the normal equations obtained through the least square technique for estimating the regression coefficients of a second degree response surface obtained through the design come out as follows:

$$b_1 = \frac{\sum x_1 y}{4(n-1)a^2} \qquad b_i = \frac{\sum x_i y}{2^{s-1}p^2 + 4q^2}$$

$$b_{1i} = \frac{\sum x_1 x_i y}{4a^2 q^2} \qquad b_{ij} = \frac{\sum x_i x_j y}{2^{s-1}p^4}$$

$$b_0 = G - \frac{(n-1)p^2 - q^2}{(n-1)p^2 a^2} \sum x_1^2 y - \frac{1}{(n-1)p^2}\left(\sum x_2^2 y + \ldots \sum x_n^2 y\right)$$

$$b_{11} = -\frac{(n-1)p^2 - q^2}{(n-1)p^2 a^2} G + \frac{2^{s-1}(4n-3)(n-1)p^4 + (2^{s+1}+4)q^4 - 2^{s+2}p^2 q^2 (n-1)}{2^{s-1}(n-1)^2 p^4 a^4} \sum x_1^2 y$$

$$- \frac{4a^2\left[(2^{s-1}+1)q^2 - 2^{s-1}(n-1)p^2\right]}{2^{s-1}(n-1)^2 p^4 a^4}\left[\sum x_2^2 y + \ldots \sum x_n^2 y\right]$$

$$b_{ii} = -\frac{1}{(n-1)p^2} G - \frac{4a^2\left[(2^{s-1}+1)q^2 - 2^{s-1}(n-1)p^2\right]}{2^{s-1}(n-1)^2 p^4 a^4} \sum x_1^2 y$$

$$+ \left[\frac{n-2}{4(n-1)q^4} + \frac{2^{s-1}+1}{(n-1)2^{s-1}p^4}\right] \sum x_i^2 y$$

$$- \frac{1}{2q^4} \sum_{j \neq i} \sum x_j^2 y$$

$i, j = 2, 3 \cdots n \quad i \neq j.$

From the relationship of the normal equations we get variances and covariances of the different regression coefficients as follows:

$$\frac{Var(b_1)}{\sigma^2} = \frac{1}{4(n-1)a^2} \qquad \frac{Var(b_i)}{\sigma^2} = \frac{1}{2^{s}p^2 + 4q^2}$$

$$\frac{Var(b_{1i})}{\sigma^2} = \frac{1}{4a^2q^2} \qquad \frac{Var(b_{ii})}{\sigma^2} = \frac{1}{2^{s-1}p^4}$$

$$Var(b_0) = \sigma^2, \quad \frac{Cov(b_0, b_{11})}{\sigma^2} = \frac{-2\left[(n-1)p^2 - q^2\right]}{(n-1)p^2a^2}$$

$$\frac{Cov(b_{ii}, b_{jj})}{\sigma^2} = \frac{2^{s-1}-1}{a^4p^4 2^{s-2}(n-1)^2} - \frac{1}{2(n-1)q^4}$$

$$\frac{Cov(b_0, b_{ii})}{\sigma^2} = \frac{-2}{(n-1)p^2}$$

$$\frac{Cov(b_{11}, b_{ii})}{\sigma^2} = \frac{-8a^2\left[(2^{s-1}+1)q^2 - 2^{s-1}(n-1)p^2\right]}{2^{s-1}(n-1)^2p^4a^4}$$

$$\frac{Var(b_{11})}{\sigma^2} = \frac{p^4 2^{s-1}(4n-3)(n-1) + q^4(2^{s+1}+4) - 2^{s-2}p^2q^2(n-1)}{2^{s-1}(n-1)^2p^4a^4}$$

$$\frac{Var(b_{ii})}{\sigma^2} = \frac{n-2}{4(n-1)q^4} + \frac{2^{s-1}+1}{2^{s-1}p^4(n-1)^2}$$

where

$i, j = 2, \ldots, n, \; i \quad j$

Let $(x_{10}, \ldots x_{n0})$ be any point. Let Y be the estimated response at that point.

Putting $d^2 = x_{20}^2 + \ldots\ldots x_{n0}^2$ we have :

$$\frac{Var(Y)}{\sigma^2} = 1 + x_{10}^2\left[\frac{1}{4(n-1)a^2} - \frac{2(n-1)p^2 - q^2}{(n-1)p^2a^2}\right] + d^2\left[\frac{1}{2^{s-1}p^2 + 4q^2} - \frac{2}{(n-1)p^2}\right]$$

$$+ \frac{x_{10}^4\left[2^{s-1}(4n-3)(n-1)p^4 + (2^{s+1}+4)q^4 - 2^{s+2}p^2q^2(n-1)\right]}{2^{s+1}(n-1)^2p^4a^4}$$

$$+ d^4\left[\frac{n-2}{4(n-1)q^4} + \frac{2^{s-1}-1}{(n-1)2^{s-1}p^4}\right]$$

$$+ x_{10}^2 d^2\left[\frac{-8a^2\left[(2^{s-1}+1)q^2 - 2^{s-1}(n-1)p^2\right]}{2^{s+1}(n-1)^2p^4a^4} + \frac{1}{4a^2q^2}\right]$$

It will be seen that when x is held constant, the variance becomes a function of d^2.

SERIES 2

The following points can be taken to constitute another series of asymmetrical design.

Set I: $(a\ p\ p\ \ldots\ldots\ p)$

Set II: $(0\ q\ 0\ \ldots\ldots\ 0)$

$\ldots\ldots\ldots\ldots$

$(0\ 0\ 0\ \ldots\ldots\ q)$

Set III: $(0\ \ \ldots\ldots\ 0)$

Here

N = Total number of points

= 2^s+2n-1 where 2^s points are got from Set I

$n_1 = 1$, $n_2 = n-1$

For $n \geq 5$ a fraction of Set I points is to be taken as explained earlier.

Rotatability for second group of factors gives $q^4 = 2^s p^4$ and $\sum x_i^2 = N$ gives $2^s a^2 = N$ and $2^s p^2 + 2q^2 = N$.

The solution for the normal equations for this design are:

$$b_1 = \frac{1}{2^s a^4} \sum x_1 y \qquad b_i = \frac{1}{2^s p^2 + 2q^2} \sum x_i y$$

$$b_{1i} = \frac{1}{2^s a^2 p^2} \sum x_1 x_i y \quad b_{ij} = \frac{1}{2^s p^4} \sum x_i x_j y$$

$$b_0 = 0 + \frac{(n-1)p^2 - q^2}{a^2 q^2} \sum x_1^2 y - \frac{1}{q^2}\left[\sum x_2^2 y + \ldots \sum x_n^2 y\right]$$

$$b_{11} = \frac{(n-1)p^2 - q^2}{a^2 q^2}\, 0 + \frac{2^s(2n-1)(n-1)p^2 + (2^s+8)q^4 - 2^{s+2}(n-1)p^2 q^2}{2^{s+1} a^4 q^4} \sum x_1^2 y$$

$$+ \frac{q^2 - (n-1)p^2}{a^2 q^4}\left[\sum x_2^2 y + \ldots\ldots \sum x_n^2 y\right]$$

$$b_{ii} = -\frac{1}{q^2}\, 0 + \frac{q^2 - (n-1)p^2}{a^2 q^4}\, x_1^2 y + \frac{1}{2q^4} \sum x_i^2 y$$

$$+ \frac{1}{q^4} \sum_{j \neq i} \sum x_j^2 y$$

Where $\quad i, j = 2, \ldots, n, i \; j$

Variances and covariances of the different regression coefficients are as follows:

$$\frac{Var(b\ )_1}{\sigma^2} = \frac{1}{2^s a^4} \qquad \frac{Var(b_i)}{\sigma^2} = \frac{1}{2^s p^2 + 2q^2}$$

$$\frac{Var(b_{1i})}{\sigma^2} = \frac{1}{2^s a^2 p^2} \qquad \frac{Var(b_{ii})}{\sigma^2} = \frac{1}{2^s p^4}$$

$$Var(b_0) = \sigma^2$$

$$\frac{Var(b_0, b_{11})}{\sigma^2} = \frac{2(n-1)p^2 - 2q^2}{a^2 q^2}$$

$$\frac{Cov(b_{ii}, b_{jj})}{\sigma^2} = \frac{2}{q^4} \qquad \frac{Cov(b_0, b_{ii})}{\sigma^2} = -\frac{2}{q^2}$$

$$\frac{Var(b_{11})}{\sigma^2} = \frac{2^s(2n-1)(n-1)p^4 + (2^{s+1}+2)q^4 - 2^{s-2}(n-1)p^2q^2}{2^{s+1} a^4 q^4}$$

$$\frac{Cov(b_{11} b_{ii})}{\sigma^2} = \frac{2q^2 - 2(n-1)p^2}{a^2 q^4}$$

$$\frac{Var(b_{ii})}{\sigma^2} = 3/2q^4$$

where $i, j = 2, \ldots\ldots n \quad i \neq j.$

If Y is the estimated response at the point $(x_{10}, x_{20} \ldots\ldots x_{n0})$ then

$$\frac{Var(Y)}{\sigma^2} = 1 + x_{10}^2\left[\frac{1}{2^s a^2} - \frac{2(n-1)p^2 - 2q^2}{a^2 q^2}\right] + d^2\left[\frac{1}{2^s p^2 + 2q^2} - \frac{2}{q^2}\right] + \frac{x_{10}^4\left[2^s(2n-1)(n-1)p^4 + (2^{s+1}+2)q^4 - 2^{s-2}(n-1)p^2q^2\right]}{2^{s+1} a^4 q^4}$$

$$+ \frac{3d^4}{2q^4} + x_{10}^2 d^2 \left[\frac{1}{2^s a^2 p^2} + \frac{2q^2 - 2(n-1)p^2}{a^2 q^4} \right]$$

Analysis of the above designs:

If there are N design points and if there are V constants involved in fitting the surface we have the following analysis of variance table in case none of the observation is repeated:

Due to	d.f.
Filled constants	V-1
Lack of fit	N-V
Total	N-1

Here we cannot test the goodness of fit of the surface. If we have previous knowledge that the second degree surface is a good fit, an estimate of σ^2 can be obtained from the lack of fit component of the sum of squares. On the other hand if some of the design points are repeated, we can get an estimate of error from those points and the lack of fit sum of squares can be tested against this error sum of squares. Comparison of the above two series of designs with rotatable designs:

We have now got 3 similar series of response surface designs viz. two fresh series presented above in this thesis and another series called central composite rotatable designs introduced by Box and Hunter (1957). It will be interesting to make a comparison among these series of designs in respect of the number of points in them and also the variances of response at specified points estimated from these surfaces. With this end in view the following investigation was made.

First we have presented in the following table the number of points required for the design with different number of factors.

TABLE

Numbers of points in the different designs.

Designs \ Factors	3	4	5	6	7
Series I	13	21	33	37	57
Series II	13	23	25	43	77
Rotatable design	14	24	26	44	78

From this we observe that the designs in series II have always one point less than that in rotatable designs.

For comparison in respect of variance of estimated response at specified points we have taken two particular cases, that is when the number of factors is three or four.

Case I: Number of factors 3.

When the response at the point x_{10}, x_{20}, x_{30} is estimated from the surface fitted with the help of the design presented in series 1 the variance of the estimated response come out as

$$\frac{V(Y_0)}{\sigma^2} = 1 - .2835x_{10}^2 - .6660d^2 + .1271x_{10}^4 + .3794d^4 + .0065x_{10}^2d^2$$

$$= V_1 \text{ (say)}$$

Similarly series 2 design will give more precise estimate than rotatable design if $V_2 - V_4 < 0$

i.e. $-.7054\, x_{10}^2 - 1.0398 d^2 + .0486\, x_{10}^4 + .1283 d^4 + .0731 x_{10}^2 d^2 < 0$

In particular if we consider only the factors in the second group keeping the level of the factor in the first group at zero we get

1) when $x_{10} = 0$

Series 1 gives more precise estimate than rotatable design if $d^2 < 1.4142$

Series 2 gives more precise estimate than rotatable design if $d^2 < 5.8170$

Similarly keeping the level of the second group the same

2) When $d^2 = 0$

Series 1 gives more precise estimate than rotatable design if $x_{10}^2 < .88$

Series 2 given more precise estimate than rotatable design if $x_{10}^2 < 8.75$

From the above it is clear that series 2 is better than series 1 in certain region as far as the comparison in respect of variances is concerned.

Case II: Designs with four factors:

Series: I

$$\underline{V(Y_0)} = 1 - 7/21\, x_{10}^2 - 29/63\, d^2 + 48/21^2\, x_{10}^4 + 320/3 \times 21^2\, d^4 + 248/21^2\, x_{10}^2 d^4$$

$$= V_1 \text{ (say), where } d^2 = x_{20}^2 + x_{30}^2 + x_{40}^2$$

Series 2:

$$\frac{V(Y_0)}{\sigma^2} = 1-7/23\, x_{10}^2 - 11/23\, d^2 + 116/23^2\, x_{10}^4 + 54/23^2 d^4 + 72/23^2\, x_{10}^2 d^2$$

$$= V_2 \text{ (say)}$$

For central composite design

$$\frac{V(Y_0)}{\sigma^2} = 1-1/5\, x_{10}^2 - 1/5\, d^2 + 51/25^2 x_{10}^4 + 51/25^2 d^4 + 102/25^2\, x_{10}^2 d^2$$

$$= V_3 \text{(say)}$$

In this case an addition of one more central point does not reduce the variance appreciably.

Here we find that

Series 1 gives a more precise estimate than rotatable design if

$$-.1333\, x_{10}^2 - .2603\, d^2 + .0272\, x_{10}^4 + .1603 d^4 + .3992 x_{10}^2 d^2 < 0$$

Series 2 gives a more precise estimate than rotatable design if

$$-.1043\, x_{10}^2 - .2783 d^2 + .1377 x_{10}^4 + .0205 d^4 - .0271 x_{10}^2 d^2 < 0$$

As before when

1) $x_{10} = 0$

Series 1 gives more precise estimate than rotatable design if $d^2 < 1.6238$

Series 2 gives more precise estimate than rotatable design if $d^2 < 13.5756$

2) $d^2 = 0$

Series 1 gives more precise estimate than rotatable design if $x_{10}^2 < 4.9$

Series 2 gives more precise estimate than rotatable design if $x^2_{10} < .75$

We see from above that within the region of interest series 2 designs can replace profitably rotatable designs. When we have some idea about the optimum combination by taking this combination as the central point, we see that the asymmetric designs are equally efficient as rotatable designs for estimating responses at the central region.

A design in 4 factors

So far we have described designs which satisfy the condition of rotatability only for one of the two sets of factors. It was, however, possible to obtain a design in four factors such that there are two factors each with 7 levels and another two factors with five levels each and that the design is rotatable in respect of each pair of the factors having either 5 or 7 levels. This design with 37 points is presented below using the usual notation

Set I:	(a b p p)
Set II:	(b a q 0)
Set III:	(b a 0 q)
Set IV:	(c 0 0 0)
Set V:	(0 c 0 0)
Set VI:	(0 0 0 0)

For estimating the unknown levels a, b, p, q and c we have the equations

$$\Sigma x_1^2 = 16a^2+16b^2+2c^2 = 16p^2+8q^2 = 37$$

$$\Sigma x_1^4 = \Sigma x_2^4 = 16a^4+16b^4+2c^4 = 96a^2b^2 = 3\Sigma x_1^2 x_2^2$$

$$\Sigma x_3^4 = \Sigma x_4^4 = 16p^4+8q^4 = 48p^4 = 3\Sigma x_3^2 x_4^2$$

Our ultimate aim in the construction of the above design is to make the variance of an estimated response, a function of d_1^2 and d_2^2 where $d_1^2 = x_{10}^2+x_{20}^2$ and $d_2^2 = x_{30}^2+x_{40}^2$. To achieve this we have to ensure that

$$\Sigma x_1^2 x_2^2 = \Sigma x_2^2 x_3^2 = \Sigma x_1^2 x_4^2 = \Sigma x_2^2 x_4^2 \qquad (A)$$

This condition actually follows from the rotatability conditions. Had they not followed from the rotatability condition, no design could be possible, as these conditions give three equations while we have only one free parameter.

As we have only 3 equations with four unknowns we can fix any one of the constants arbitrarily thereby getting an infinite series of designs.

Here

$$a = 1/64\sqrt{3737 - 276c^2 - 36c^4} + 1/64\sqrt{999 - 20c^2 - 36c^4}$$

$$\text{and } b = 1/64\sqrt{3737 - 276c^2 - 36c^4} + 1/64\sqrt{999 - 20c^2 - 36c^4}$$

We note that $c^2 > 0$ and $c^2 \leq 5.553$ since for $c^2 > 5.553$ a and b become imaginary.

The solution of normal equations of the regression coefficients from this design are:

$$b_1 = \frac{\sum x_1 y}{16(a^2+b^2)+2c^2}; \qquad b_2 = \frac{\sum x_2 y}{16(a^2+b^2)+2c^2}$$

$$b_3 = \frac{\sum x_3 y}{16q^2}; \qquad b_4 = \frac{\sum x_4 y}{16q^2}$$

$$b_{12} = \frac{\sum x_1 x_2 y}{32a^2b^2}; \qquad b_{13} = \frac{\sum x_1 x_3 y}{8q^2(a^2+b^2)}$$

$$b_{14} = \frac{\sum x_1 x_4 y}{8q^2(a^2+b^2)}; \qquad b_{23} = \frac{\sum x_2 x_3 y}{8q^2(a^2+b^2)}$$

$$b_{24} = \frac{\sum x_2 x_4 y}{8q^2(a^2+b^2)}; \qquad b_{34} = \frac{\sum x_3 x_4 y}{4q^4}$$

$$b_0 = -\frac{1}{c^2}\sum\left[x_1^2+x_2^2-(a^2+b^2)\right]y + \frac{a^2+b^2-c^2}{c^2q^2}\sum(x_3^2+x_4^2-q^2)y$$

$$2b_{11} = \frac{5}{2c^4}\sum\left[x_1^2+x_2^2-(a^2+b^2)\right]y - \frac{\left[5(a^2+b^2)-4c^2\right]}{2c^4q^2}\sum(x_3^2+x_4^2-q^2)y$$

$$+ \frac{1}{16(a^2-b^2)^2+2c^4}\sum(x_1^2-x_2^2)y$$

$$2b_{22} = \frac{5}{2c^4}\sum\left[x_1^2+x_2^2-(a^2+b^2)\right]y - \frac{\left[5(a^2+b^2)-4c^2\right]}{2c^4q^2}\sum(x_3^2+x_4^2-q^2)y$$

$$- \frac{1}{16(a^2-b^2)^2+2c^4}\sum(x_1^2-x_2^2)y$$

$$2b_{33} = \frac{q}{16q^2} + \frac{1}{16q^2}\left[\frac{37}{c^2} - \frac{5\left[16(a^2+b^2)-2c^2\right]}{2c^4}\right]$$

$$\times\sum\left[x_1^2+x_2^2-(a^2+b^2)\right]y$$

$$- \frac{1}{16q^2}\left[\frac{37(a^2+b^2-c^2)}{c^2q^2} - \left[16(a^2+b^2)-2c^2\right] \times \frac{\left[5(a^2+b^2)-4c^2\right]}{2c^4q^2}\right]\sum(x_3^2+x_4^2-q^2)y$$

$$+ \frac{1}{8q^4}\sum(x_3^2-x_4^2)y$$

$$2b_{44} = \frac{C}{16q^2} + \frac{1}{16q^2}\left[\frac{37}{c^2} - \frac{5[16(a^2+b^2)+2c^2]}{2c^4}\right]\sum\{x_1^2+x_2^2-(a^2+b^2)\}y$$

$$- \frac{1}{16q^2}\left[\frac{37(a^2+b^2-c^2)}{c^2q^2} - \frac{[16(a^2+b^2)-2c^2] \times [5(a^2+b^2)-4c^2]}{2c^4q^2}\right]\sum(x_3^2+x_4^2-q^2)y$$

$$- \frac{1}{8q^4}\sum(x_3^2+x_4^2)y$$

The variance of estimated response at the point $(x_{10}, x_{20}, x_{30}, x_{40})$ is

$$\frac{Var(Y_0)}{\sigma^2} = 1 + \frac{d_1^2}{16(a^2-b^2)+2c^2} - \frac{2d_1^2}{c^2}$$

$$+ d_2^2\left[\frac{1}{16q^2} - \frac{2(a^2+b^2-c^2)}{c^2}\right]$$

$$- d_1^4\left[\frac{5}{4c^4} - \frac{1}{2[16(a^2+b^2)^2+2c^4]}\right]$$

$$- d_2^4\left[-\frac{1}{32}\left[\frac{37(a^2+b^2-c^2)}{c^2q^2} - \frac{(16(a^2+b^2)+2c^2) \times (5(a^2+b^2)-4c^2)}{2c^4q^2}\right] + \frac{1}{16q^2}\right]$$

$$- d_1^2d_2^2\left[\frac{-[5(a^2+b^2)-4c^2]}{2c^4q^2} - \frac{1}{8q^2(a^2+b^2)}\right]$$

where

$$d_1^2 = x_{10}^2 + x_{20}^2\;;\quad d_2^2 = x_{30}^2 + x_{40}^2$$

As we have mentioned above $0 < c^2 < 5.553$. We have compared the variances at the estimated response of this design at $c^2 = 1$ and $c^2 = 5$ with the central composite design with four factors:

Case 1:

$$c^2 = 1,\ a^2 = 2.1,\ b^2 = .2,\ q^2 = 2.3$$

$$\frac{V(Y_0)}{\sigma^{-2}} = 1-19.970\, d_1^2 -43.9370\, d_2^2 +125.0086\, d_1^4 + 3.3097\, d_2^4 -115.6285\, d_1^2\, d_2^2 = V_1' \text{ (say)}$$

Case 2:

$$c^2 = 5,\ a^2 = 1.24,\ b^2 = .45,\ q^2 = 2.3$$

$$\frac{V(Y_0)}{\sigma^{-2}} = 1-.1730\, d_1^2 +1.3510\, d_2^2 +.0583\, d_1^4 +.3251\, d_2^4 -.0030\, d_1^2\, d_2^2 = V_2' \text{ (say)}$$

Case 3: Central composite design:

$$\frac{V(Y_0)}{\sigma^{-2}} = 1-.2d_1^2 - .2d_2^2 +.8161\, d_1^4 +.8161 d_2^4 +1.632 d_1^2 d_2^2$$

$$= V_{rot} \text{ (say)}$$

As before we make comparisons keeping the levels of factor in one group fixed:-

We have

(1) when $d_1^2 = 0$

$$V_1' < V_{rot} \text{ if } d_2^2 < 17.53$$

$$V_2' < V_{rot} \text{ if } d_2^2 < .37$$

(2) when $d_2^2 = 0$

$$V_1' < V_{rot} \text{ if } d_1^2 < .1592$$

$$V_2' < V_{rot} \text{ if } d_1^2 < .0356.$$

BLOCKING OF RESPONSE SURFACE DESIGNS

In the previous chapter we have presented two series of asymmetrical designs by means of which response surfaces can be fitted with some advantage over the existing rotatable designs. Jut like most of the response surface designs these designs also are not split into blocks. But for experimentation, particularly in the field of agriculture, designs with blocks of equal but small size are necessary. Hence we tried to split some of these designs obtained from the above two series into blocks.

As pointed out by Box and Hunter (1957) the design points should satisfy certain further relations for blocking, so that the regression coefficients can be estimated orthogonal to block effect. These relations in case of equal block sizes are:

(1) $\sum x_{ij}^2 =$ constant within each block

(2) $\sum x_{ij} x_{kl} = 0$ within each block

(3) $\sum_d x_{ij}^2 = \sum_l x_{ij}^2$ for all d and l where $\sum_d$ is summation over treatments in the dth block.

The method followed for splitting into blocks in the above two series is as follows. Here the points obtainable from the set I of unknown levels (a p ....p) and (0 p .....p) in the two series respectively can each be split into two or more groups when m 3 by confounding an interaction of order not less than 2. The points from the set (q 0.....0) (0 q 0 ....0) etc. are to be alloted to the blocks in each case, through a trial procedure in such a way that the above requirements for blocking are satisfied. Though for the general case it is possible to make into blocks each of the two series of designs, we have presented below only some particular designs which are likely to be more useful in actual situations.

Design No.1:

Block 1			Block 2		
x_1	x_2	x_3	x_1	x_2	x_3
a	q	0	-a	q	0
a	-q	0	-a	-q	0
-a	0	q	a	0	q
-a	0	-q	a	0	-q
0	p	p	0	p	p
0	p	-p	0	p	-p
0	-p	p	0	-p	p
0	-p	-p	0	-p	-p
0	0	0	0	0	0

From the relations in the design so as to make the solutions of normal equations easy and also to make it rotatable with respect to the last two factors we get,

$$\Sigma x_1^2 = \Sigma x_2^2 = \Sigma x_3^2 = 8a^2 = N = 18 = 8p^2+4q^2$$

$$\Sigma x_2^4 = \Sigma x_3^4 = 8p^4 + 4q^4 = 24p^4 = 3\Sigma x_2^2 x_3^2$$

That is $a^2 = 2.25$, $p^2 = 1.125$, $q^2 = 2.25$.

Though the designs are given interms of coded doses such that the origin is at the mean dose and the scale corresponds to second moment of the doses as unit, we can transform doses into actual doses from the following relation, coded doses:

Coded doses:	$-q$	$-p$	0	p	q
Actual doses:	0	$\frac{M}{2}(1-\frac{p}{q})$	$\frac{M}{2}$	$\frac{M}{2}(1+\frac{p}{q})$	M

when M is the maximum dose.

For the ultimately layout of the design we require the actual doses, while for analysis the coded doses are taken.

The estimates of the variances regression coefficients are

$$b_1 = .0556 \Sigma x_1 y; \qquad b_2 = .0556 \Sigma x_2 y$$

$$b_3 = .0556 \Sigma x_3 y; \qquad b_{12} = .0494 \Sigma x_1 x_2 y$$

$$b_{13} = .0494 \Sigma x_1 x_3 y; \qquad b_{23} = .0990 \Sigma x_2 x_3 y$$

$$b_0 = .5\ 0 - .2222 [\Sigma x_2^2 y + \Sigma x_3^2 y]$$

$$b_{11} = .0494 \ x_1^2 y - .0247 [\Sigma x_2^2 y + \Sigma x_3^2 y]$$

$$b_{22} = -.22220 + .1482 \ x_2^2 y + .0988 \ x_3^2 y - .0247 \ x_1^2 y$$

$$b_{33} = -.22220 + .0988 \ x_2^2 y + .1482 \ x_3^2 y - .0247 \ x_1^2 y.$$

The variance of the estimated response Y_0 at the point (x_{10}, x_{20}, x_{30}) is

$$\frac{V(Y_0)}{\sigma^2} = .5 + .0556 x_{10}^2 - .3888 d^2 + .0494 x_{10}^4 + .1482 d^4$$

For obtaining the error variance as also the sum of squares due to lack of fit we may apply the analysis of variance technique as given below:

Due to Blocks	1
Due to filled constants	9
Due to lack of fit	3
Due to error	4
Total	17

One of our aims being to test the significance of lack of fit i.e. whether the second degree surface is an adequate fit for the observed responses, the degrees of freedom for error is too low in this case. A method of getting more degrees of freedom for error is to repeat the whole design once more, that is to have 2 replications. In that the analysis is of variance partitioning will be:

	d.f.
Due to Replication	1
Due to Blocks within replication	2
Due to filled constants	9
Due to lack of fit	3
Due to error	20
	35

The various sums of squares in the analysis of variance of table where there is only one replication are obtained as follows:

To get the sum of squares due to fitted constants we use the following expression:

$$b_0 G + b_1 \Sigma x_1 y + b_2 \Sigma x_2 y + b_{12} \Sigma x_1 x_2 y$$

$$+ b_{13} \Sigma x_1 x_3 y + b_{23} \Sigma x_2 x_3 y + b_{11} \Sigma x_1^2 y$$

$$+ b_{22} \Sigma x_2^2 y + b_{33} \Sigma x_3^2 y - G^2/N$$

It will be seen that the 5 points

0	p	p
0	p	-p
0	-p	p
0	-p	-p
0	0	0

are repeated in each of the 2 blocks. Hence by considering the observations from this combination only, an analysis of variance like

	d.f
Between blocks	1
Between treatments	4
Error	4
	9

is possible. The error obtained from this analysis of variance is also error in the analysis is of variance of one replication of the design.

The lack of fit is obtained by subtraction. In case of two replications the expression for sum of squares due to fitted constants remains the same as above.

In this case the 5 points

$$\begin{array}{ccc} 0 & p & p \\ 0 & p & -p \\ 0 & -p & p \\ 0 & -p & -p \\ 0 & 0 & 0 \end{array}$$

are repeated in each of the four blocks. Hence by considering the observations from this combination only an analysis of variance like

Between blocks:	3	
Between treatments:	4	(I)
Error:	12	
	19	

is possible:

Also the points

$$\begin{array}{ccc} a & q & 0 \\ a & -q & 0 \\ -a & 0 & q \\ -a & 0 & -q \end{array}$$

are repeated in two blocks. Hence we have considering these observations only, an analysis of variance

Between Blocks	1	
Between treatments	3	(II)
Error	3	
	7	

and because the points

$$\begin{array}{ccc} -a & q & 0 \\ -a & -q & 0 \\ a & 0 & q \\ a & 0 & -q \end{array}$$

are again repeated in two blocks, we again have an analysis of variance table

Between blocks	1	
Between treatments	3	(III)
Error	3	
	7	

Now consider the observations in block 1. We get a contrast between the observation for (0 0 0) and the mean of the observations from the other 8 points in it. We can get the same contrast from block 1 of replication 2, which also contains the same points. The sum of squares due to the contrast between the two contrasts gives a sum of squares with 1 degree of freedom, which also belongs to error. Similarly we can get sum of squares with 1 degree of freedom from block 2 in the two replications.

The error obtained by the addition of the error sums of squares in the analysis of variance tables I, II and III and the sum of squares for 2 degrees of freedom obtained as mentioned above gives the error sum of squares in the analysis of variance for two replications.

The lack of fit sum of squares is got by subtraction.

It will be seen that we have not attempted to present the classifical analysis of the data in form of presentation of main effects and interaction. As we feel that the classifical approach of analysis is some what inadequate in so far as it does not throw any light as to the optimum combination of the level of the factor, the emphasis has been laid to obtain data through which response surfaces can be fitted with advantage and the optimum combination can also be obtained from it. But it may sometimes be necessary to have some information regarding the main effect and two factor interactions. For this purpose the regression coefficient can be used. Thus the estimate of b_i, the coefficient of x_i in the equation of the surface gives the linear contrast of the main effect of the ith factor. Again, b_{ij} the coefficient of $x_i x_j$ gives the linear x linear component contrast of the two factor interactions involving the ith and jth factors.

For all the designs which are mentioned below a similar method can be employed to get the sum of squares.

Design No.2

Block 1			Block 2		
x_1	x_2	x_3	x_1	x_2	x_3
a	-p	-p	-a	-p	-p
-a	p	-p	a	p	-p
-a	-p	p	a	-p	p
a	p	p	-a	p	p
0	q	0	0	q	0
0	-q	0	0	-q	0
0	0	q	0	0	q
0	0	-q	0	0	-q
0	0	0	0	0	0

Here the relationship in the design viz.

$$\Sigma x_1^2 = \Sigma x_2^2 = \Sigma x_3^2 = 8a^2 = 18 = N = 8p^2+4q^2$$

$$\Sigma x_2^4 = \Sigma x_3^4 = 8p^4+4q^4 = 24p^4 = 3\Sigma x_2^2 x_3^2$$

are the same as in design No. 1 and we hence have $a^2 = 2.25$, $p^2 = 1.125$, $q^2 = 2.25$. On substituting these values for a, p and q the normal equations for estimation of parameters of this design becomes the same as that for design No.1. Hence the same expressions for solutions and variance at estimated response as for design No.1 holds good here also.

The method of analysis also remains the same as in design No.1.

Design No.3

Instead of taking two replications of either design No.1 or design No.2, we can use a design obtainable by taking the above two designs together.

The design is this:

Block 1			Block 2			Block 3			Block 4		
x_1	x_2	x_2	x_1	x_2	x_3	x_1	x_2	x_3	x_1	x_2	x_3
a	q	0	-a	q	0	a	-p	-p	-a	p	p
a	-q	0	-a	-q	0	-a	p	-p	-a	p	-p
-a	0	q	a	0	q	-a	-p	p	a	-p	p
-a	0	-q	a	0	-q	a	p	p	-a	p	p
0	p	p	0	p	p	0	q	0	0	q	0
0	p	-p	0	p	-p	0	-q	0	0	-q	0
0	-p	p	0	-p	p	0	0	q	0	0	q
0	-p	-p	0	-p	-p	0	0	-q	0	0	-q
0	0	0	0	0	0	0	0	0	0	0	0

Here from the conditions of the design we get

$a^2 = 2.25,\ p^2 = 1.125,\ q^2 = 2.25$

The solution for the regress coefficients are:-

$b_1 = .0278 \sum x_1 y, \quad b_2 = .0278 \sum x_2 y \quad b_3 = .0278 \sum x_3 y,$

$b_{12} = .0247 \sum x_2 x_1 y,\ b_{13} = .0247 \sum x_1 x_3 y,\ b_{23} = .0495 \sum x_2 x_3 y$

$b_0 = .25\ 0 - .1111\ (\sum x_2^2 y + \sum x_3^2 y)$

$b_{11} = .0247\ x_1^2 y - .0124\ (\sum x_2^2 y + \sum x_3^2 y)$

$b_{22} = -.1111\ 0 + .0741 \sum x_2^2 y + .0494 \sum x_3^2 y - .0124 \sum x_1^2 y$

$b_{33} = -.1111\ 0 + .0494 \sum x_2^2 y + .0741 \sum x_3^2 y - .0124 \sum x_1^2 y$

For obtaining error variance as also the sum of squares due to lack of fit here we use the analysis of variance:

	d.f
Due to B locks	3
Due to fitted constants	9
Due to lack of fit	15
Error	8
	35

The design can be used with advantage as we can determine the lack of fit sum of squares with large number of degrees of freedom

Design No.4

Block 1 x_1	x_2	x_3	x_4	Block 2 x_1	x_2	x_3	x_4
a	q	0	0	-a	q	0	0
a	-q	0	0	-a	-q	0	0
-a	0	q	0	a	0	q	0
-a	0	-q	0	a	0	-q	0
a	0	0	q	-a	0	0	q
a	0	0	-q	-a	0	0	-q
-a	0	0	0	a	0	0	0
-a	0	0	0	a	0	0	0
0	p	-p	-p	0	-p	-p	-p
0	-p	p	-p	0	p	p	-p
0	-p	-p	p	0	-p	p	p
0	p	p	p	0	p	-p	p

Here we have as a result of the conditions of the design:

$$a^2 = 1.5, \; p^2 = 1.5, \; q^2 = 3.6$$

The solutions of various regression coefficients are

$$b_i = 1/24 \sum x_i y, \; b_{ji} = 1/18 \sum x_j x_i y \quad i, j = 2...4$$

$$b_0 = 7/8 \; G - 1/6 \sum x_1^2 y - 1/6 \left(\sum x_2^2 y - \sum x_3^2 y - \sum x_4^2 y \right)$$

$$b_{11} = -1/6 \; G + 1/2 \sum x_1^2 y + 1/18 \left(\sum x_2^2 y + \sum x_3^2 y + \sum x_4^2 y \right)$$

$$b_{22} = -1/6 \; G + 1/18 \sum x_1^2 y + 1/18 \sum x_2^2 y + 1/9 \left(\sum x_3^2 y + \sum x_4^2 y \right)$$

$$b_{33} = -1/6 \; G + 1/18 \sum x_1^2 y + 1/9 \sum x_2^2 y + 1/18 \sum x_3^2 y + 1/9 \sum x_4^2 y$$

$$b_{44} = -1/6 \; G + 1/18 \sum x_1^2 y + 1/9 \left(\sum x_2^2 y + \sum x_3^2 y \right) + 1/18 \sum x_4^2 y$$

Variance of the estimated response Y_0 at the point $(x_{10}, x_{20}, x_{30}, x_{40})$ is

$$\frac{V(Y_0)}{\sigma^2} = 7/8 - 15/24\, x_{10}^2 - 7/24\, d^2$$

$$+ 1/2\, x_{10}^4 + 1/18\, d^4 + \frac{x_{10}^2 d^2}{6}$$

where $d^2 = x_{20}^2 + x_{30}^2 + x_{40}^2$

For obtaining the error variance as also the sum of squares due to lack of fit we may apply analysis of variance technique as given below, in case of one replication

Due to blocks	1
Due to fitted constants	14
Due to lack of fit	8
	23

Here we do not get any error sum of squares at all. But in case of two replications we have the following partition of d.f

Due to replications	1
Block within replications	2
Due to fitted constants	14
Due to lack of fit	8
Error	22
	47

Design No. 5

Block 1 x_1	x_2	x_3	x_4	x_5	Block 2 x_1	x_2	x_3	x_4	x_5
0	p	-p	p	-p	0	p	-p	p	p
0	p	p	-p	-p	0	p	p	-p	p
0	-p	p	p	-p	0	-p	p	p	p
0	p	-p	-p	p	0	p	-p	-p	-p
0	-p	-p	p	p	0	-p	-p	-p	-p
0	-p	p	-p	p	0	p	p	p	-p
0	p	p	p	p	0	-p	-p	-p	p
0	-p	-p	-p	-p	0	-p	-p	p	-p
a	q	0	0	0	-a	q	0	0	0
a	-q	0	0	0	-a	-q	0	0	0
-a	0	q	0	0	a	0	q	0	0
-a	0	-q	0	0	a	0	-q	0	0
a	0	0	q	0	-a	0	0	q	0
a	0	0	-q	0	-a	0	0	-q	0
-a	0	0	0	q	a	0	0	0	q
-a	0	0	0	-q	a	0	0	0	-q
0	0	0	0	0	0	0	0	0	0

In this case $a^2 = 2.125$, $p^2 = 1.245$, $q^2 = 3.521$

The solutions of various regression coefficients are:

$b_i = .0294 \Sigma x_i y$, $i = 1, \ldots, 5$

$b_{1i} = .0333 \Sigma x_1 x_i y$, $i = 2, \ldots, 5$.

$b_{ij} = .0403 \Sigma x_i x_j y$, $i, j = 2, \ldots, 5$. $i \neq j$

$b_0 = .50 - .0689 \Sigma x_1^2 y - .1004(\Sigma x_2^2 y + \ldots + \Sigma x_5^2 y)$

$$b_{11} = -.0689\ G + .0301\Sigma x_1^2y + .0097(\Sigma x_2^2y + \ldots + \Sigma x_5^2y)$$

$$b_{ii} = -1.004\ G + .0097\Sigma x_1^2y + .0379\Sigma x_i^2y + .0170\sum_{k \neq i}\Sigma x_k^2y,$$

$$k, i = 2, \ldots, 5.$$

Variance of the estimated response Y_0 at the point $(x_{10}, x_{20}, x_{30}, x_{40})$ is

$$\frac{V(Y_0)}{\sigma^2} = .5000 - .1034x_{10}^2 - .1714d^2 + .0301x_{10}^4 + .0379d^4 + .0528x_{10}^2d^2$$

where $d^2 = x_{20}^2 + x_{30}^2 + x_{40}^2 + x_{50}^2$

For obtaining the error variance as also the sum of squares due to lack of fit we may apply analysis of variance technique as given below, in case of one replication:

Due to blocks	1
Due to fitted constants	20
Due to lack of fit	12
Total	33

Here also we do not get any error sum of squares. But in case of two replications we have:

Due to replications	1
Block within replications	2
Due to fitted constants	20
Due to lack of fit	12
Error	32
	67

Design No. 6

Block 1				Block 2			
x_1	x_2	x_3	x_4	x_1	x_2	x_3	x_4
a	-b	p	-p	a	-b	p	p
a	b	-p	-p	a	b	-p	p
-a	b	p	-p	-a	b	p	p
+a	-b	-p	p	a	-b	-p	-p
-a	-b	p	p	-a	+b	-p	-p
-a	b	-p	+p	a	b	p	-p
a	b	p	p	-a	-b	-p	p
-a	-b	-p	-p	-a	-b	p	-p
b	-a	-q	0	-b	-a	-q	0
-b	a	-q	0	b	a	-q	0
-b	a	q	0	b	-a	q	0
b	-a	q	0	-b	a	q	0
+b	-a	0	-q	b	a	0	-q
-b	a	0	-q	-b	-a	0	-q
-b	+a	0	q	b	-a	0	q
b	-a	0	q	-b	a	0	q
c	0	0	0	c	0	0	0
-c	0	0	0	-c	0	0	0
0	c	0	0	0	c	0	0
0	-c	0	0	0	-c	0	0
0	0	0	0	0	0	0	0

$\Sigma x_i^2 = N$ gives , $16(a^2+b^2)+4c^2 = 42$ (1)

and $16p^2+8q^2 = 42$ (2)

Also $\Sigma x_1^4 = \Sigma x_2^4 = 3\Sigma x_2^2 x_3^2$ gives

$$16(a^4+b^4)+4c^4 = 96a^2b^2 \quad (3)$$

$\Sigma x_3^4 = \Sigma x_4^4 = 3\Sigma x_3^2 x_4^2$ gives

$$16p^4+8q^4 = 48p^4 \quad (4)$$

By (2) and (4)

$$p^2 = 1.312; \quad q^2 = 2.624$$

From (1) and (3) we get

$$2a = 1/32\sqrt{4452-592c^2+80c^4}+1/32\sqrt{924+80c^2-80c^4}$$

$$2b = 1/32\sqrt{4452-592c^2+80c^4}-1/32\sqrt{924+80c^2-80c^4}$$

Here $0 < c^2 \leqslant 2.8985$. Thus there will be an infinite series of designs for various values of c^2 between 0 and 2.8985. The estimates of various regression coefficients are

$$b_i = 1/[16(a^2+b^2)+4c^2]\,\Sigma x_i y; \quad b_j = 1/16p^2+8q^2\,\Sigma x_j y$$

$$b_{ij} = 1/8q^2(a^2+b^2)\,\Sigma x_i x_j y; \quad i = 1,2, \; j = 3,4$$

$$b_{12} = 1/32a^2b^2\,\Sigma x_1 x_2 y; \quad b_{34} = 1/16p^4\,\Sigma x_3 x_4 y$$

$$b_0 = \tfrac{1}{4}\,G - \frac{1}{2c^2}(\Sigma x_1^2 y + \Sigma x_2^2 y) + \frac{a^2+b^2-c^2}{2c^2}(\Sigma x_3^2 y + \Sigma x_4^2 y)$$

$$b_{11} = -\frac{1}{2c^2}\,G - \frac{5c^2}{8c^4}(\Sigma x_1^2 y + \Sigma x_2^2 y) + \frac{(\Sigma x_1^2 y - \Sigma x_2^2 y)}{2[16(a^2-b^2)^2+4c^4]}$$

$$-\frac{5(a^2+b^2)-4c^2}{8c^4}(\Sigma x_3^2 y + \Sigma x_4^2 y)$$

Replacing x_1 by x_2 and x_2 by x_1 in b_{11} we get b_{22}

$$b_{33} = \frac{a^2+b^2-c^2}{2c^2}\, G - \frac{5(a^2+b^2)-4c^2}{8c^4}\left(\Sigma x_1^2y + \Sigma x_2^2y\right)$$

$$+ \left[\frac{-21(a^2+b^2-c^2)}{32c^2q^4} - \frac{\left[5(a^2+b^2)-4c^2\right]\times\left[4(a^2+b^2)-c^2\right]}{32c^4q^4}\right]\left(x_3^2y - x_4^2y\right)$$

$$+ \frac{1}{16q^4}\left(\Sigma x_3^2y - \Sigma x_4^2y\right)$$

Replacing x_3 by x_4 and x_4 by x_3 in b_{33} we get b_{44}

Variance at estimated response Y_0 at the point $(x_{10} \ldots x_{40})$ is

$$\frac{\mathrm{Var}(Y_0)}{\sigma^2} = \frac{1}{2} + d_1^2\left[\frac{1}{16(a^2+b^2)-4c^2} - \frac{1}{c^2}\right]$$

$$+ d_2^2\left[\frac{1}{16p^2+8q^2} - \frac{a^2+b^2-c^2}{c^2}\right] + d_1^4\left[\frac{5q^2}{8c^4} - \frac{1}{2\left[16(a^2-b^2)^2+4c^4\right]}\right]$$

$$+ d_2^4\left[\frac{-21(a^2+b^2-c^2)}{32c^2q^4} + \frac{1}{16q^4} + \frac{\left[5(a^2+b^2)-4c^2\right]\left[4(a^2+b^2)+c^2\right]}{32c^4q^4}\right]$$

$$+ d_1^2d_2^2\left[\frac{1}{8q^2(a^2+b^2)} - \frac{5(a^2+b^2)-4c^2}{4c^4}\right]$$

where $d_1^2 = x_{10}^2 + x_{20}^2$, $d_2^2 = x_{30}^2 + x_{40}^2$

In particular when $c^2 = 2$, $a^2 = 1.873$, $b^2 = .254$ the solutions of the regression coefficients are

$$b_i = .0238\,\Sigma x_i y; \quad b_j = .0238\,\Sigma x_j y$$

$$b_{ij} = .0224\,\Sigma x_i x_j y; \quad i = 1,2, \quad j = 3,4$$

$$b_{12} = .0697\,\Sigma x_1 x_2 y; \quad b_{34} = .0363\,\Sigma x_3 x_4 y$$

$$b_0 = .5\,G - .25(\Sigma x_1^2 y + \Sigma x_2^2 y) + .0318(\Sigma x_3^2 y + \Sigma x_4^2 y)$$

$$b_{11} = -.25\,G + .4186\,\Sigma x_1^2 y + .4014\,\Sigma x_2^2 y - .0823(\Sigma x_3^2 y + \Sigma x_4^2 y)$$

Replacing x_1 by x_2 and x_2 by x_1 in b_{11} we get b_{22}.

$$b_{33} = .0318\,G - .0823\,(\Sigma x_1^2 y + \Sigma x_2^2 y) + .0466\,\Sigma x_3^2 y + .0284\,\Sigma x_4^2 y.$$

Replacing x_3 by x_4 and x_4 by x_3 in b_{33} we get b_{44}.

$$\frac{Var(Y_0)}{\sigma^2} = .5 - .4762\,d_1^2 + .0874\,d_2^2 + .4186\,d_1^4 + .0466\,d_2^4 - .1422\,d_1^2 d_2^2$$

The analysis of variance in this case for one replication is:

	d.f
Due to blocks	1
Due to fitted constants	14
Due to lack of fit	22
Error	4
	41

In case of two replications the analysis of variance is:

Due to replication	1
Due to block within replication	2
Due to fitted constants	14
Due to lack of fit	22
Error	44
	83

As an illustration we have presented below the complete analysis of design No.1 with two replications presented earlier in the chapter. As no real data for the design are available, we have taken some fictitious data for the purpose of illustration.

The following data shows the data along with the reatment combinations of coded doses:

Treatment combinations x_1	x_2	x_3	Response Rep. 1	Rep. 2	Treatment combinations x_1	x_2	x_3	Response Rep. 1	Rep. 2
1.5	1.5	0	2	3	-1.5	1.5	0	4	3
1.5	-1.5	0	3	2	-1.5	-1.5	0	2	1
-1.5	0	1.5	1	1	1.5	0	1.5	4	4
-1.5	0	-1.5	2	3	1.5	0	-1.5	1	2
0	1.1	1.1	1	3	0	1.1	1.1	2	1
0	101	-1.1	2	4	0	1.1	-1.1	3	2
0	-1.1	1.1	4	2	0	-1.1	1.1	2	3
0	-1.1	-1.1	3	2	0	-1.1	-1.1	3	3
0	0	0	1	3	0	0	0	1	2
Total			19	23	Total			22	21

Here G = 85; $\Sigma x_1 y = 6.00$; $\Sigma x_2 y = 1.76$; $\Sigma x_3 y = -1.24$

$\Sigma x_1 x_2 y = -9.00$; $\Sigma x_1 x_3 y = 18.00$; $\Sigma x_2 x_3 y = -4.24$

$\Sigma x_1^2 y = 85.50$; $\Sigma x_2^2 y = 89.94$; $\Sigma x_3^2 y = 85.44$.

Using the expressions for solutions, the various regression coefficients are:

$b_0 = 1.7700$; $b_{11} = -.0628$; $b_{22} = .3816$; $b_{33} = .2704$

$b_1 = .1668$; $b_2 = .0489$; $b_3 = -.0344$; $b_{12} = -.2222$

$b_{13} = .4444$; $b_{23} = -.2099$.

Using the expression for fitted constants sum of squares given in page 33 we have

Fitted constants sum of squares = 13.83

To find error sum of squares:

The analysis of variance table No.I of page 34 is

	d.f	s.s
Between blocks	3	1.35
Due to treatments	4	4.80
Error sum of squares	12	10.40
Total	19	16.55

The analysis of variance table No.II of page 34 is

	d.f	s.s
Between blocks	1	.125
Due to treatments	3	3.375
Error sum of squares	3	1.375
Total	7	4.875

The analysis of variance table No.III of page 35 is

	d.f	s.s
Between blocks	1	.125
Due to treatments	3	10.375
Error sum of squares	3	1.375
Total	7	11.875

The s.s for 2.d.f as explained earlier is equal to .45.

The error sum of squares with 20d.f in the analysis of variance table for the whole data is equal to

$$10.40-1.375 - 1.375 - .45 = 13.60.$$

The analysis of variance table for testing the significance of lack of fit sum of squares is:

Due to:	d.f	s.s	M.ss:	F
Replication	1	.25		
Block within replication	2	.73		
Fitted constants	9	13.83		
Lack of fit	3	5.90	1.96	2.88
Error	20	13.60	.68	
Total	35	34.31		

In this case since the F ratio with d.f (3,20) from the table at the 5% level of significance is 3,10, a second degree surface fits well to the data, though this conclusion has no real significance as only fictitious responses have been used in the illustration.

MISSING VALUES IN RESPONSE SURFACE DESIGN

The problem of analysis with missing observations is equally important in response surface design. Accordingly we have attempted to work out a method of analysis of some of the response surface designs with one missing observation. Draper (1961) attempted the analysis of a rotatable design with one missing observation. His approach was to estimate the missing observation by minimising the sum of squares due to (lack of fit + error). In the present case we have attempted to estimate the missing observation by minimising the sum of squares due to lack of fit only. An investigation has also been made further to assess the accuracy of estimates of regression coefficients when they are estimated after the missing value is substituted by its estimate and subsequently treating the estimate as if it is one of the existing observations. The investigation in respect of three designs has been presented below. Two of these designs are asymmetrical and one rotatable in three factors.

__Design 1:__ We first consider a rotatable design with blocking where all the factors have 3 levels each:

Block 1			Response	Block 2			Response	Block 3			Response
x_1	x_2	x_3		x_1	x_2	x_3		x_1	x_2	x_3	
-a	-a	0	Y_1	-a	0	-a	Y_9	0	-a	-a	Y_{17}
-a	a	0	Y_2	-a	0	a	Y_{10}	0	-a	a	Y_{18}
a	-a	0	Y_3	a	0	-a	Y_{11}	0	a	-a	Y_{19}
a	a	0	Y_4	a	0	a	Y_{12}	0	a	a	Y_{20}
0	0	a	Y_5	0	a	0	Y_{13}	a	0	0	Y_{21}
0	0	-a	Y_6	0	-a	0	Y_{14}	-a	0	0	Y_{22}
0	0	a	Y_7	0	a	0	Y_{15}	a	0	0	Y_{23}
0	0	-a	Y_8	0	-a	0	Y_{16}	-a	0	0	Y_{24}

Let the block total be B_1, B_2 and B_3 where $B_1+B_2+B_3 = G$

Let us suppose that the observation from the design point (x'_1, x'_2, x'_3) is missing. Let the unknown value p denote the missing observation. Let it be in the jth block.

The missing point may be

(1) a point which is not repeated

or (2) a point which is repeated.

Case 1: A point which is not repeated:

Here the block sum of squares is

$$B_1^2/8 + B_2^2/8 + B_3^2/8 - G^2/24$$

The fitted constants sum of squares is given by

$$b_0 G + \dots\dots\dots + b_{33} \sum x_3^2 y - G^2/24$$

The error sum of squares is given by

$$(Y_5-Y_7)^2/2 + (Y_6-Y_8)^2/2 + \dots (Y_{21}-Y_{23})^2/2 + (Y_{22}-Y_{24})^2/2$$

The total sum of squares is $\sum Y_i^2 - G^2/24$.

The lack of fit sum of squares is obtained by subtraction. In this case, when we express the sum of squares due to the lack of fit as a function of p, an unknown substituted for the missing value, we get the lack of fit sum of squares as

$$p^2 - 9/24\,(G' + p)^2 - 1/8(p+B'_j)^2$$

$$-1/24\left[(\textstyle\sum_1 x_1 y + x'_1 p)^2 + (\sum_1 x_2 y + x'_2 p)^2 + (\sum_1 x_3 y + x'_3 p)^2\right.$$

$$-\frac{1}{16}\left[\begin{array}{l}(\Sigma_1 x_1^2y+x_1'^2p)^2+(\Sigma_1 x_2^2y+x_2'^2p)^2+(\Sigma_1 x_3^2y+x_3'^2p)^2\\ +(\Sigma_1 x_1^2y+x_1'^2p)(\Sigma_1 x_2^2y+x_2'^2p)+(\Sigma_1 x_1^2y+x_1'^2p)(\Sigma_1 x_3^2y+x_3'^2p)\\ +(\Sigma_1 x_2^2y+x_2'^2p)(\Sigma_1 x_3^2y+x_3'^2p)+(\Sigma_1 x_1x_2y+x_1'\,x_2'p)^2\\ +(\Sigma_1 x_1x_3y+x_1'\,x_3'\,p)^2+(\Sigma_1 x_2x_3y+x_2'\,x_3'\,p)^2\end{array}\right]$$

$$-(G'+p)/4\left[\Sigma_1 x_1^2y+\Sigma_1 x_2^2y+\Sigma_1 x_3^2y+p(x_1'^2+x_2'^2+x_3'^2)\right]$$

+ other terms not involving p.

Where Σ_1 is summation over (N-1) points which are not missing and the missing value is in the jth block.

Differentiating the above sum of squares with respect to p and equating it to zero we get

$$\hat{p} = \frac{\begin{array}{l}4(2B_j'-B_k-B_l)+(40-24a^2)\,G'+(3d_1^2-12)\Sigma_1(x_1^2+x_2^2+x_3^2)y\\ +6\left[x_1'\,x_2'\Sigma_1 x_1x_2y+x_1'\,x_3'\Sigma_1 x_1x_3y+x_2'\,x_3'\Sigma_1 x_2x_3y\right]\\ +3(x_1'^2\Sigma_1 x_1^2y+x_2'^2\Sigma_1 x_2^2y+x_3'^2\Sigma_1 x_3^2y)\end{array}}{48+20d_1^2-6d_1^4}$$

$= M/N$ (say)

where $d_1^2 = x_1'^2+x_2'^2+x_3'^2$ and $B_j'+B_k+B_l = G'$ where $j \neq k \neq l = 1,2,3$

<u>Case 2</u>: When the missing point is one which is repeated we have 3 methods of estimating p viz.

<u>Method 1</u>:

Minimising error sum of squares alone

Here $\hat{p}$ = Response of the corresponding point which is repeated

$= Y_j'$ (say)

<u>Method 2:</u>

Minimising (lack of fit + error) sum of squares we get the same estimate as in case 1 i.e. M/N.

<u>Method 3:</u>

Minimising lack of fit sum of squares alone we have

$$p = (M - Y_j')/N-1$$

It can be shown that p is an unbiassed estimate of the missing response and it is independent of block effects.

Variance of estimated response at the point (x_1, x_2, x_3) is given by:

$$\begin{aligned}
Var(Y) = {} & Var(b_0) + x_1 cov(b_0, b_1) + x_2 cov(b_0, b_2) \\
& + x_1^2 \left[Var(b_1) + cov(b_0, b_{11}) \right] + x_2^2 \left[Var(b_2) + cov(b_0, b_{22}) \right] \\
& + x_1 x_2 \left[cov(b_0, b_{12}) + cov(b_1, b_2) \right] + x_1^2 x_3^2 \left[Var(b_{13}) - cov + cov(b_{11}, b_{33}) \right] \\
& + x_2^2 x_3^2 \left[Var(b_{23}) + cov(b_{22}, b_{33}) \right] \\
& + x_3^2 \left[Var(b_3) - cov(b_0, b_{33}) \right] + x_1^4 Var(b_{11}) + x_2^4 Var(b_{22}) \\
& + x_3^4 Var(b_{33}) + x_1^3 cov(b_1, b_{11}) + x_2^3 cov(b_2, b_{22}) \\
& + x_1^3 x_2 cov(b_{12}, b_{11}) + x_1 x_2^3 cov(b_{12}, b_{22}) \\
& + x_1^2 x_2 \left[cov(b_1, b_{12}) + cov(b_2, b_{11}) \right] \\
& + x_1 x_2^2 \left[cov(b_2, b_{12}) + cov(b_1, b_{22}) \right]
\end{aligned}$$

The values of the various coefficients in the above expression, when any one of the points (a, a, 0),(-a, a, 0), (a,-a, 0) and (-a, -a, 0) is missing is tabulated in table 1 below. The same method of collecting coefficients as described earlier has been adopted in finding the various coefficients:

TABLE I.

Coefficients of various terms in $\frac{Var(Y)}{\sigma^2}$

Coefficient of	Missing point (a, a, 0)	(-a, a, 0)	(a, -a, 0)	(-a,-a, 0)
Constant	.6875	.6875	.6875	.6875
x_1	-.0295	.0295	-.0295	.0295
x_2	-.0295	-.0295	.0295	.0295
$x_1^2+x_2^2$	-.2930	-.2930	-.2930	-.2930
x_1x_2	.0469	.0469	.0469	.0469
$x_1^2x_3^2+x_2^2x_3^2$	.1250	.1250	.1250	.1250
x_3^2	-.2083	-.2083	-.2083	-.2083
$x_1^4+x_2^4$	.0742	.0742	.0742	.0742
x_3^4	.0625	.0625	.0625	.0625
$x_1^2x_2$	.0663	.0663	-.0663	-.0663
$x_1x_2^2$	.0663	-.0663	.0663	-.0663
x_1^3	.0221	-.0221	.0221	-.0221
x_2^3	.0221	.0221	-.0221	-.0221
$x_1^3x_2+x_1x_2^3$	.0469	-.0469	-.0469	.0469

The variances of estimated response at various points when any one of the points (a, a, 0), (a, -a, 0) (-a, a, 0) and (-a, -a, 0) is missing and when no point is missing are tabulated in table II below:

TABLE II

Var(Y)/σ^2 at various points.

Points	Missing points (a, a, 0)	(-a, a, 0)	(a,-a,0)	(-a,-a,0)	None Missing
0 0 0	.6875	.6875	.6875	.6875	.4167
1 0 0	.4613	.4761	.4613	.4761	.2709
0 1 0	.4613	.4613	.4761	.4761	.2709
0 0 1	.5417	.5417	.5417	.5417	.2709
1 1 0	.5084	.2030	.2030	.2728	.2501
1 0 1	.4405	.4553	.4405	.4553	.2501
0 1 1	.4405	.4405	.4553	.4553	.2501
1 1 1	.6126	.3072	.3072	.3770	.3543

A comparison of the values presented in the last column with those in the other columns shows that if the analysis is conducted by substituting the missing observation by its estimate and then treating it as one of the existing observation as has been done by Draper (1961) there is some risk in that the actual variance of the estimated response may be quite different from that obtainable from the non-missing case:

Design 2:

Block 1 x_1	x_2	x_3	Response	Block 2 x_1	x_2	x_3	Response
a	q	0	Y_1	-a	q	0	Y_{10}
a	-q	0	Y_2	-a	-q	0	Y_{11}
-a	0	q	Y_3	a	0	q	Y_{12}
-a	0	-q	Y_4	a	0	-q	Y_{13}
0	p	p	Y_5	0	p	p	Y_{14}
0	p	-p	Y_6	0	p	-p	Y_{15}
0	-p	p	Y_7	0	-p	p	Y_{16}
0	-p	-p	Y_8	0	-p	-p	Y_{17}
0	0	0	Y_9	0	0	0	Y_{18}
Total			B_1				B_2

Let $B_1 + B_2 = G$. Let us suppose that the observation of the design point (x_1', x_2', x_3') is missing. Let k denote the missing observation. Let it be in jth block.

As in the previous case the point (x_1', x_2', x_3') may be

(1) a point which is not repeated

or (2) a point which is repeated.

Case 1: A point which is not repeated.

Expressing the lack of fit sum of squares as a function of k (the unknown substituted for missing value) and differentiating it with respect to k and equating it to zero we have:

$$\hat{k} = \frac{\begin{array}{l} .1111(B'_j - B_k) + .9999\,G' + .1111(x'_1\Sigma_1 x_1 y + x'_2\Sigma_1 x_2 y + x'_3\Sigma_1 x_3 y) \\ + .0988(x'_1 x'_2 \Sigma_1 x_1 x_2 y + x'_1 x'_3 \Sigma_1 x_1 x_3 y) + .1980\, x'_2 x'_3 \Sigma_1 x_2 x_3 y \\ + .0988\, x'^2_1 \Sigma_1 x_1^2 y + .2964\, x'^2_2 \Sigma_1 x_2^2 y + .2964\, x'^2_3 \Sigma_1 x_3^2 y \\ - .0494(x'^2_1 \Sigma_1 x_2^2 y + x'^2_2 \Sigma_1 x_1^2 y + x'^2_1 \Sigma_1 x_3^2 y + x'^2_3 \Sigma_1 x_1^2 y) \\ + .1976(x'^2_2 \Sigma_1 x_3^2 y + x'^2_3 \Sigma_1 x_2^2 y) - .4444\, \Sigma_1 (x_2^2 + x_3^2) y \\ - .4444(x'^2_3 + x'^2_2)\, G' \end{array}}{\begin{array}{l} 1 - .1111(x'^2_1 + x'^2_2 - x'^2_3) + .8888(x'^2_2 + x'^2_3) - .5932\, x'^2_2 x'^2_3 \\ - .0988\, x'^4_1 - .2964(x'^4_2 + x'^4_3) \end{array}}$$

$= A/B$ (say).

Where $B'_j + B_k = G'$, $j \neq k$, $= 1,2$

<u>Case 2</u>: When missing point is one which is repeated

Here we have again 3 methods of estimating k viz.

<u>Method 1</u>:

Minimise error alone. Let Y_5 be the missing value

Then $\hat{K} = Y_{14} + 1/4(Y_6 + Y_7 + Y_8 + Y_9 - Y_{15} - Y_{16} - Y_{17} - Y_{18})$

<u>Method 2</u>:

Minimising (lack of fit - error) we have the same expression for k as in case 1 i.e. $\hat{k} = A/B$.

<u>Method 3</u>:

Minimising lack of fit sum of squares alone we have

$$\hat{k} = \frac{1}{B - .8}\left(A - .2(Y_6 + Y_7 + Y_8 + Y_9) - .8\, Y_{14} + .2(Y_{15} + Y_{16} + Y_{17} + Y_{18})\right)$$

Here variance of estimated response at the point (x_1, x_2, x_3) is given by

$$\begin{aligned}Var(Y) = {} & Var(b_0)+ x_1 \left[cov(b_0, b_1)\right]+x_2 cov(b_0,b_2) \\ & +x_1x_2\left[(cov(b_0,b_{12})+cov(b_1,b_2)\right] \\ & +x_1^2x_2^2 \left[Var(b_{12})+cov(b_{11},b_{22})\right] \\ & +x_1^2 \left[Var(b_1)+cov(b_0,b_{11})\right]+x_2^2 \left[Var(b_2)+cov(b_0,b_{22})\right] \\ & +x_3^2 \left[Var(b_3)+cov(b_0,b_{33})\right]+x_1^4 Var(b_{11})+x_2^4 Var(b_{22}) \\ & +x_3^4 Var(b_{33})+x_1^2x_3^2 \left[Var(b_{13})+cov(b_{11},b_{33})\right] \\ & +x_2^2x_3^2 \left[Var(b_{23})+cov(b_{22},b_{33})\right] \\ & +x_1^2x_2 \left[Cov(b_1,b_{12})+cov(b_2,b_{11})\right] \\ & +x_1x_2^2 \left[cov(b_2,b_{12})+cov(b_1,b_{22})\right] \\ & +x_1^3 cov(b_1,b_{11})+x_2^3 cov(b_2,b_{22}) \\ & +x_2x_3^2 cov(b_2,b_{33})+x_1x_3^2 cov(b_1,b_{33}) \\ & +x_1^3x_2 cov(b_{12},b_{11})+x_2^3x_1 cov(b_{12},b_{22}) \\ & +x_1x_2x_3^2 cov(b_{12}, b_{33}).\end{aligned}$$

The values of the various coefficients in the above expression when any one of the points $(a, q, 0)$, $(-a, q, 0)$ $(a, -q, 0)$ and $(-a, -q, 0)$ is missing is tabulated in table 3:

TABLE III

Coefficients of various terms in $\frac{Var(\hat{Y})}{\sigma^2}$

Coefficient of	Missing point $(a,q,0)$	$(-a,q,0)$	$(a,-q,0)$	$(-a,-q,0)$
Constant	.5000	.5000	.5000	.5000
x_1x_2	.0558	-.0558	-.0558	.0558
$x_1^2x_2^2$	.0799	.0799	.0799	.0799
x_1^2	.0834	.0834	.0834	.0834
$x_2^2 + x_3^2$	-.3610	-.3610	-.3610	-.3610
x_1^4	.0618	.0618	.0618	.0618
$x_2^4 + x_3^4$	.1606	.1606	.1606	.1606
$x_1^2x_3^2$	-.0248	-.0248	-.0248	-.0248
$x_2^2x_3^2$	.2222	.2222	.2222	.2222
$x_1^2x_2$	.1115	.1115	-.1115	-.1115
x_1^3	.0372	-.0372	.0372	-.0372
x_2^3	.0372	.0372	-.0372	-.0372
$x_1x_3^2$	-.0370	.0370	-.0370	.0370
$x_2x_3^2$	-.0370	-.0370	.0370	.0370
$x_1^3x_2$	.0496	-.0496	-.0496	.0496
$x_1x_2^3$	.0496	-.0496	-.0496	.0496
$x_1x_2x_3^2$	-.0494	.0494	.0494	-.0494

The variances of estimated response at various points when any one of the points $(a,q,0)$, $(-a,q,0)$, $(a,-q,0)$ & $(-a,-q,0$

is missing are tabulated in table 4 below:

TABLE XV

Var Y at various points
σ^2

Points	Missing points (a,q,0)	(-a,q,0)	(a,-q,0)	(-a,-q,0)	None Missing
0 0 0	.5000	.5000	.5000	.5000	.5000
1 0 0	.6824	.6824	.6824	.6824	.6649
0 1 0	.3368	.3368	.3368	.3368	.2594
1 1 0	.9771	.3697	.3697	.3803	.5495
1 -1 0	.3697	.9771	.3803	.3697	.5495
0 1 1	.3216	.3216	.3216	.3216	.3152
1 1 1	.8507	.4161	.4161	.4039	.7905

Here also we draw the same conclusion as in previous case.

Design 3:

Block 1 x_1	x_2	x_3	Response	Block 2 x_1	x_2	x_3	Response
a	-p	-p	Y_1	-a	-p	-p	Y_{10}
-a	p	-p	Y_2	a	p	-p	Y_{11}
-a	-p	p	Y_3	a	-p	p	Y_{12}
a	p	p	Y_4	-a	p	p	Y_{13}
0	q	0	Y_5	0	q	0	Y_{14}
0	-q	0	Y_6	0	-q	0	Y_{15}
0	0	q	Y_7	0	0	q	Y_{16}
0	0	-q	Y_8	0	0	-q	Y_{17}
0	0	0	Y_9	0	0	0	Y_{18}
Total			B_1				B_2

Here we get the same expressions for estimates and variance at estimated response as in design 2 since as pointed out in chapter 2, the normal equations for these two designs are the same.

SUMMARY

After the introduction of rotatable design by Box and Hunter (1957) considerable amount of study has been made in the theory of fitting multifactor response surfaces. But till now only designs of the symmetric type i.e. designs with factors all with the same number of levels have been studied. In this thesis we have suitably modified the idea of rotatability and presented two series of asymmetric designs together with a design in four factors which gives rise to an infinite series of designs. Some of these designs are found to be yielding more accurate estimates of the response at particular points than that obtainable through rotatable designs. These designs are also available with lesser number of design points. Along with the designs the expressions for estimates of various regression coefficients and their variance has also been presented.

To use such designs for agricultural experimentation it is necessary that they are split up into blocks of equal size. In this thesis we have given a number of designs belonging to the asymmetric type with blocking. Along with estimates of regression coefficients and their variances, the method of analysis in case of one or more replications has also been described.

The problem of analysis of the design through estimation of a missing value in response surface designs

has been cosidered by Draper (1961). A slightly different approach through minimising the lack of fit component of s.s has been adopted and the estimate of one missing value in case of 3 response surface designs has been given. A comparison of the variances of estimated response at various points when the data are analysed (1) after substitution of missing value and proceeding further as usual and (2) taking into account all consequences of the missing observation, has been made and it has been found that there is considerable amount of difference in the variances of any estimated response.

REFERENCES

1. Bose R.C. and Draper N.R. — Second order rotatable designs in three dimensions. Ann.Math. Stat.Vol.30(1959), 1097-1112
2. Box G.E.P. — The exploration and exploitation of response surfaces. Some general considerations and examples, Biometrics, Vol.10(1954), 16-60
3. Box G.E.P and Behnken D.W. — Some new three level designs for the study of quantitative variables. Technometrics Vol.2 (1960), 455 - 475.
4. Box G.E.P and Draper N.R. — A basis for selection of response surface design. Jour.Amer.Stat.Assoc.Vol.54 (1959)
5. Box G.E.P. and Hunter J.S. — Multifactorial experimental designs for exploring response surfaces. Ann.Math.Stat.Vol.28 (1957), 195 - 241.
6. Das M.N. — Construction of rotatable designs from factorial designs Jour.Ind.Soc.Agr.Stat.Vol. 13 (1961) 169 - 294
7. Das M.N. — On construction of second order rotatable designs through BIBD with blocks of unequal sizes Cal.Stat.Assoc.Bult.Vol.12 (1963) 31 - 46.

8. Das M.N. and Narasimham V.L. Construction of rotatable designs through BIBD. Ann.Math.Stat. Vol.33 (1962) 1421 - 1439.

9. De Baun R. Response surface designs for three factors at three levels. Technometrics Vol.1(1959) 1 - 8.

10. Draper N.R. Second order rotatable designs in four or more dimensions. Ann.Math. Stat.Vol.31 (1960) 23 - 33.

11. Gardiner D.A., Grandage A.H.E. and Hader R.J. Third order rotatable designs for exploring response surfaces. Ann. Math.Stat.Vol. 30(1959) 1082 -1096.

12. Draper N.R. Missing values in response surface design. Technometrics Vol.3 (1961) 389 - 398.

13. Narasimham V.L. On construction of second and third order rotatable designs. Thesis submitted to I.A.R.S., New Delhi (1961)

14. Rao C.R. Advanced Statistical Methods in Biometric research (1952), 75-82.

15. Thaker P.J. On investigation of response surfaces. Thesis submitted to I.A.R.S., New Delhi (1960).