<Reseach Paper>

Journal of the Korean
Statistcal Society
Vol. 13, No. 1, 1984

Nonestimability of Missing Values for 2^k and 3^k Factorial Designs

Jung W. Sim* and Sung H. Park**

ABSTRACT

A method of missing value estimation for a general design is described. In particular, the cases of missing value estimation for 2^k and 3^k factorial designs are studied. The cases where the missing values are not estimable for 2^k and 3^k design are explored and discussed. Some examples are illustrated to show the missing value estimation and the nonestimable cases.

1. Introduction

Suppose a linear regression model is

$$\underset{\sim}{y}=X\underset{\sim}{\beta}+\underset{\sim}{e} \tag{1}$$

where $\underset{\sim}{y}'=(y_1, \cdots, y_n)$, $\underset{\sim}{\beta}'=(\beta_0, \cdots, \beta_k)$, $\underset{\sim}{e}\sim(\underset{\sim}{0}, I_n\sigma^2)$, and X is an $n\times(k+1)$ matrix with rank $k+1$. Suppose that the matrix X' is divided into (X_1', X_2') in such a way that X_1 is associated with $n-m$ response values $\underset{\sim}{y}_1$ that are observed, and X_2 is associated with m response values $\underset{\sim}{y}_2$ that are missing.

The residual sum of squares when we fit the model (1) is expressed as

$$\begin{aligned} S^2 &= \underset{\sim}{y}'\underset{\sim}{y}-\hat{\underset{\sim}{\beta}}'X'\underset{\sim}{y} \\ &= \underset{\sim}{y}'(I_n-X(X'X)^{-1}X')\underset{\sim}{y} \\ &= \underset{\sim}{y}'H\underset{\sim}{y} \end{aligned}$$

where

$$H=I_n-X(X'X)^{-1}X', \tag{2}$$

and H is an $n\times n$ symmetric matrix. Letting

* Department of Computer Science and Statistics, Chonnam National University, Kwangju 500, Korea
** Department of Computer Science and Statistics, Seoul National University, Seoul 151, Korea

$$H=(\underline{h}_1, \cdots, \underline{h}_n),$$

where $\underline{h}_i$ is the ith column vector of H, and partially differentiating S^2 with respect to y_i and setting the result equal to zero to satisfy the condition for a minimum, we find that

$$\underline{h}_i' \underline{y}=0, \text{ or } h_{i1} y_1+\cdots+h_{in} y_n=0. \tag{3}$$

Since

$$(\partial/\partial y_i)(\underline{h}_i' \underline{y})=h_{ii}>0,$$

because H is positive definite when $S^2>0$ this equation gives a minimum.

Solution of the equation (3) for y_i provides us with the estimate y_i of a single missing value y_i. If m observations $y_{n-m+1}, \cdots, y_n$ are missing, we must solve the simultaneous equations

$$\underline{h}'_{n-m+1} \underline{y}=0, \cdots, \underline{h}_n' \underline{y}=0 \tag{4}$$

for the missing values.

Note that

$$\begin{aligned} H&=I_n-X(X'X)^{-1} X' \\ &=I_n-\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}(X' X)^{-1}(X_1', X_2') \\ &=\begin{bmatrix} I_{n-m}-X_1(X' X)^{-1} X_1' & -X_1(X' X)^{-1} X_2' \\ -X_2(X' X)^{-1}X_1' & I_m-X_2(X' X)^{-1} X_2' \end{bmatrix} \end{aligned}$$

Thus if

$$\underline{y}_2'=(y_{n-m+1}, \cdots, y_n)$$

is the vector of missing values, we obtain the estimates $\hat{\underline{y}}_2$ from

$$-X_2(X' X)^{-1}X_1' \underline{y}_1+(I_m-X_2(X' X)^{-1} X_2')\hat{\underline{y}}_2=0$$

If

$$G=(I_m-X_2(X' X)^{-1} X_2')^{-1} \tag{5}$$

exists,

$$\hat{\underline{y}}_2=GX_2(X' X)^{-1}X_1' \underline{y}_1. \tag{6}$$

This formulae for missing value estimations is derived by Tocher (1952) and Draper (1961).

Let $\underline{b}^*$ be the estimates of $\underline{\beta}$ using $\underline{z}'=(\underline{y}_1', \hat{\underline{y}}_2')$, the observed and estimated responses together. It is not difficult to show the estimates $\underline{b}=(X_1' X_1)^{-1} X_1' \underline{y}_1$ from the observed responses are identical to the estimates $\underline{b}^*=(X' X)^{-1} X' \underline{z}$. Note that

$$\underset{\sim}{b}=(X'X-X_2'X_2)^{-1}X_1'\underset{\sim}{y}_1$$
$$=(X'X)^{-1}(I_{k+1}-X_2'X_2(X'X)^{-1})^{-1}X_1'\underset{\sim}{y}_1$$
$$=(X'X)^{-1}(I_{k+1}+X_2'GX_2(X'X)^{-1})X_1'\underset{\sim}{y}_1$$

where the identity $(I+AB)^{-1}=I-A(I+BA)^{-1}B$ has been employed.

Therefore,

$$\underset{\sim}{b}^*=(X'X)^{-1}X'\underset{\sim}{z}$$
$$=(X'X)^{-1}(X_1'\underset{\sim}{y}_1+X_2'\hat{\underset{\sim}{y}}_2)$$
$$=(X'X)^{-1}(X_1'\underset{\sim}{y}_1+X_2'GX_2(X'X)^{-1}X_1'\underset{\sim}{y}_1)$$
$$=(X'X)^{-1}(I_{k+1}+X_2'GX_2(X'X)^{-1})X_1'\underset{\sim}{y}_1$$
$$=\underset{\sim}{b}. \tag{7}$$

The expected value and the variance of $\underset{\sim}{b}^*$ are

$$E(\underset{\sim}{b}^*)=E(\underset{\sim}{b})=\underset{\sim}{\beta}$$
$$\mathrm{Var}(\underset{\sim}{b}^*)=\mathrm{Var}(\underset{\sim}{b})=(X_1'X_1)^{-1}\sigma^2.$$

Let s^2 be the mean squared error of the regression model whose response values are composed of all the observations, and s^{*2} be the mean squared error of the regression model whose response values are composed of all the observations and the estimated missing values.

Then

$$s^{*2}=s^2.$$

Since

$$(n-m-k)s^{*2}=\underset{\sim}{z}'\underset{\sim}{z}-\underset{\sim}{b}^{*\prime}X'\underset{\sim}{z}$$
$$=\underset{\sim}{y}_1'\underset{\sim}{y}_1+\hat{\underset{\sim}{y}}_2'\hat{\underset{\sim}{y}}_2-\underset{\sim}{b}'(X_1'\underset{\sim}{y}_1+X_2'\hat{\underset{\sim}{y}}_2)$$
$$=\underset{\sim}{y}_1'\underset{\sim}{y}_1-\underset{\sim}{b}'X_1'\underset{\sim}{y}_1$$
$$=(n-m-k)s^2.$$

In the following sections, we discuss how to estimate missing values for 2^k and 3^k factorial designs, and nonestimable cases are studied for these designs. Some authors such as Box et. al. (1970), Healy and Westmacott (1956), Preece (1971), Shearer (1973) and Wilkinson (1958) have studied missing value estimation. However, nonestimable cases for 2^k and 3^k factorial designs have not been treated.

2. 2^k Factorial Design

Consider the linear regression model with k independent variables in the 2^k factorial design

$$y=\beta_0+\sum_{i=1}^{k}\beta_i x_i+e \tag{8}$$

or

$$y=\beta_0+\sum_{i=1}^{k}\beta_i x_i+\sum_{i<j}^{k}\beta_{ij} x_i x_j+e. \tag{9}$$

The model without the intercept term β_0 is not considered here.

Note that the model (9) has additional $\binom{k}{2}$ interaction terms compared with the model (8).

In this design, if missing observation occur, the solutions of the equations (4) for missing values can be obtained by using the matrix H. We now examine the matrix H and the cases of non-estimability for the 2^k factorial design without replication.

2.1 2^2 Factorial Design

The regression model (8) in the 2^2 factorial design may be written as

$$\underline{y}=x\underline{\beta}+\underline{e},$$

where $\underline{y}'=(y_1, y_2, y_3, y_4)$, $\underline{\beta}=(\beta_0, \beta_1, \beta_2)$, $\underline{e}=N(\underline{0}, I_4\sigma^2)$

and

$$X=\begin{bmatrix}1 & -1 & -1\\ 1 & -1 & 1\\ 1 & 1 & -1\\ 1 & 1 & 1\end{bmatrix}$$

For this design the matrix H is given by

$$H=\begin{bmatrix}a & b & b & a\\ & a & a & b\\ \text{(sym)} & & a & b\\ & & & a\end{bmatrix},$$

where $a=1/4$ and $b=-1/4$.

Let m be the number of missing values. If $m=1$, the missing value is estimable. We can easily obtain that the single missing value y_i is estimated by $\hat{y}_1=y_2+y_3-y_4$, $\hat{y}_2=y_1+y_4-y_3$, $\hat{y}_3=y_1+y_4-y_2$ and $\hat{y}_4=y_2+y_3-y_1$. If $m=2$, missing values are not estimable. since the number of regression parameters to be estimated is greater than the number of nonmissing obervations. We can observe such nonestimability in a different way.

Suppose the last two observations y_3 and y_4 are missing. Then the simultaneous equaltions (4) become

$$\begin{bmatrix} a & b \\ b & a \end{bmatrix} \begin{bmatrix} \hat{y}_3 \\ \hat{y}_4 \end{bmatrix} = \begin{bmatrix} -by_1 - ay_2 \\ -ay_1 - by_2 \end{bmatrix}$$

Let

$$\begin{bmatrix} a & b \\ b & a \end{bmatrix} = Q.$$

Then since the matrix Q is singular, y_3 and y_4 are not estimable.

The regression model (9) in the 2^2 factorial design

$$y = \hat{\beta}_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + e$$

has 4 regression parameters. Therefore, 2^2 design is a saturated design for this model, and if there is a single missing value, it is nonestimable. Note that, since $X'X = XX' = 4I_4$, the matrix H becomes 0_4, 4×4 null matrix. Accordingly, the corresponding matrix Q for any missing value estimation becomes a null matrix.

2.2 2^3 Factorial Design

In the 2^2 factorial design for the regression model (8) the matrices X and H are given by

$$X = \begin{bmatrix} 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \qquad H = \begin{bmatrix} c & d & d & f & d & f & f & e \\ & c & f & d & f & d & e & f \\ & & c & d & f & e & d & f \\ & & & c & e & f & f & d \\ & & & & c & d & d & f \\ & \text{(sym)} & & & & c & f & d \\ & & & & & & c & d \\ & & & & & & & c \end{bmatrix}, \qquad (10)$$

where $c = 4/8$, $d = -2/8$. $e = 2/8$, and $f = 0$.

If $m \leq 3$, all the missing values are estimable. If $m = 4$, four missing values on a plane or on a diagonal plane in a cube as shown in Fig. 1 are not estimable, because the corresponding matrix Q is singular. But four missing values in other cases are estimable. For example in Fig.1 four missing values

(1), c, b, bc; (1), c, a, ac; and (1), c, ab, abc; etc.,

are not estimable, but

(1), c, a, ab; (1), c, b, ab; (1), c, ac, abc;

and

(1), c, bc, abc; etc.,

are estimable. If $m \geq 5$, all the missing values are not estimable because at least four of them are on a plane or on a diagonal plane

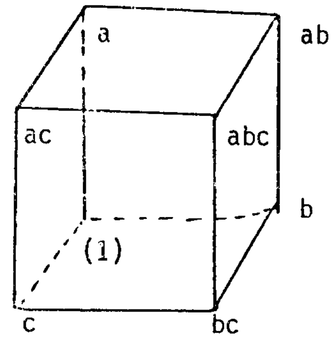

Fig. 1. 2^3 Factorial Design

The model (9) for the 2^3 factorial design is

$$y=\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_3+\beta_{12} x_1 x_2+\beta_{13} x_1 x_3+\beta_{23} x_2 x_3+e$$

which has 7 regression parameters. Hence, if $m=1$, the missing value is estimable. However, if $m \geq 2$, the missing values are nonestimable. For this design, the matrix H is given by

$$H=\begin{bmatrix} a & b & b & a & b & a & a & b \\ & a & a & b & a & b & b & a \\ & & a & b & a & b & b & a \\ & & & a & b & a & a & b \\ & & & & a & b & b & a \\ & \text{(sym)} & & & & a & a & b \\ & & & & & & a & b \\ & & & & & & & a \end{bmatrix},$$

where $a=1/8$ and $b=-1/8$.

For example, if $abc(=y_8)$ is missing, we can obtain from the equation (4) that

$$\hat{abc}=\hat{y}_8=(1)+bc+ac+ab-a-b-c.$$

2.3 2^4 Factorial Design

The matrix H for the model (8) is given by (11) below, where $p=11/16$, $q=-3/16$, $s=-1/16$, and $t=1/16$.

If $m \leq 7$, all the missing values are estimable. If $m=8$, all the missing values on a 3-dimensional cube or on the other level of that 3-dimensional cube points are not estimable,

$$H=\begin{bmatrix}
p & q & q & s & q & s & s & t & q & s & s & t & s & t & t & r \\
 & p & s & q & s & q & t & s & s & q & t & s & t & s & r & t \\
 & & p & q & s & t & q & s & s & t & q & s & t & r & s & t \\
 & & & p & t & s & s & q & t & s & s & q & r & t & t & s \\
 & & & & p & q & q & s & s & t & t & r & q & s & s & t \\
 & & & & & p & s & q & t & s & r & t & s & q & t & s \\
 & & & & & & p & q & t & r & s & t & s & t & q & s \\
 & & & & & & & p & r & t & t & s & t & s & s & q \\
 & & & & & & & & p & q & q & s & q & s & s & t \\
 & & & & & & & & & p & s & q & s & q & t & s \\
 & & & & & & & & & & p & q & s & t & q & s \\
 & & & & & & & & & & & p & t & s & s & q \\
 & & & & & & & & & & & & p & q & q & s \\
 & & \text{(sym)} & & & & & & & & & & & p & s & q \\
 & & & & & & & & & & & & & & p & q \\
 & & & & & & & & & & & & & & & p
\end{bmatrix}, \qquad (11)$$

but the missing values of all other cases are estimable. For example, 8 cube points

(1), c, b, bc, a, ac, ab, abc; (1), c, d, cd, b, bd, bc, bcd,

or high level of the 8 cube points about factor d

d, cd, bd, bcd, ad, acd, abd, $abcd$;

a, ac, ad, acd, ab, abd, abc, $abcd$; etc.,

are not estimable. If $m \geq 9$, when a subset of the missing points are composed of the missing points which are not estimable for $m=8$, the missing values are not estimable.

Example 1. In a 2^4 factorial design, suppose the observations are given by

$$\underset{\sim}{y}'=((1), d, c, cd, b, bd, bc, bcd, a, ad, ac, acd, ab, abd, abc, abcd)$$
$$=(15, 26, 18, cd, 28, 22, 11, 19, a, 17, 20, 24, 29, 22, 16, 23).$$

The letters 'cd' and 'a' represent missing values which are to be estimated. Missing values are the 4th and 9th observations. Using the matrix H in (11), we can see that

$$\underset{\sim}{h}_4'=(s, q, q, p, t, s, s, q, t, s, s, q, r, t, t, s)$$
$$\underset{\sim}{h}_9'=(q, s, s, t, s, t, t, r, p, q, q, s, q, s, s, t).$$

The estimation equations $\underset{\sim}{h}_4'\underset{\sim}{y}=0$ and $\underset{\sim}{h}_9'\underset{\sim}{y}=0$ are given by

$$p\widehat{cd}+t\hat{a}=-84q-29r-108s-66t$$
$$t\widehat{cd}+p\hat{a}=-81q-19r-134s-56t.$$

Therefore, we obtain

$$\begin{bmatrix}\widehat{cd}\\ \hat{a}\end{bmatrix}=\begin{bmatrix}17.6\\ 22.4\end{bmatrix}$$

For the model (9) of the 2^4 factorial design, there are 11 regression parameters. If $m \leq 5$, the missing values are estimable, and otherwise, they are not. Since the treatment is similar to the 2^3 design, the details are omitted here.

3. 3^k Factorial Design

3.1 3^2 Factorial Design

For a 3^2 factorial design the regression model is assumed

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{11} x^2_{1i} + \beta_{22} x^2_{2i} + \beta_{12} x_{1i} x_{2i} + e_i, \quad i = 1, \cdots, 9, \qquad (12)$$

where

$$X = \begin{bmatrix} 1 & -1 & -1 & 1 & 1 & 1 \\ 1 & -1 & 0 & 1 & 0 & 0 \\ 1 & -1 & 1 & 1 & 1 & -1 \\ 1 & 0 & -1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & -1 & 1 & 1 & -1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

For the model (12), the matrix H is given by

$$H = \begin{bmatrix} r & t & v & t & u & u & v & u & w \\ & s & t & u & t & u & u & t & u \\ & & r & u & u & t & w & u & v \\ & & & s & t & t & t & u & u \\ & & & & s & t & u & t & u \\ & & & & & s & u & u & t \\ & & & & & & r & t & v \\ & & & & & & & s & t \\ & & & & & & & & r \end{bmatrix}$$

where $r = 7/36$, $s = 16/36$, $t = -8/36$, $u = 4/36$, $v = 1/36$, and $w = -5/36$.

Let us consider the estimability for the missing values in the 3^2 factorial design. If $m \leq 2$, all the missing values arees timable. If $m = 3$, three missing values on a real line or on a diagonal dotted line as in Fig.2 are not estimable, since in the simultaneous estimation equations the corresponding matrix Q is singular. That is, in Fig. 2,

$$y_{00}, y_{10}, y_{20};\ y_{01}, y_{11}, y_{21};\ y_{02}, y_{12}, y_{22};$$

$$y_{00}, y_{01}, y_{02};\ y_{10}, y_{11}, y_{12};\ y_{20}, y_{21}, y_{22};$$

$$y_{00},\ y_{11},\ y_{22};\ y_{02},\ y_{11},\ y_{20}$$

are not estimable. If $m \geq 4$, all the missing values are not estimable.

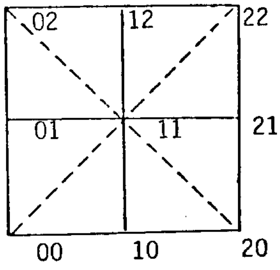

Fig. 2. Factorial Design

Example 2. In a 3^2 factorial design, suppose the observations are

$$\underline{y}' = (y_{00},\ y_{01},\ y_{02},\ y_{10},\ y_{11},\ y_{12},\ y_{20},\ y_{21},\ y_{22})$$
$$= (5,\ 7,\ 8,\ *,\ 6,\ 6,\ 9,\ 8,\ *),$$

where *'$_2$s are the missing values which are to be estimatcd. Using the matrix H, we can see that

$$\underline{h}_4' = (t,\ u,\ u,\ s,\ t,\ t,\ t,\ u,\ u),$$
$$\underline{h}_9' = (w,\ u,\ v,\ u,\ u,\ t,\ v,\ t,\ r).$$

The estimation equations $\underline{h}_4'\underline{y} = 0$ and $\underline{h}_9'\underline{y} = 0$ are thus

$$\hat{y}_{10}\, s + \hat{y}_{22}\, u = -26t - 23u$$
$$\hat{y}_{10}\, u + \hat{y}_{22}\, r = -14t - 13u - 17v - 5w.$$

Hence, we can obtain that

$$\hat{\underline{y}}_2 = \begin{bmatrix} \hat{y}_{10} \\ \hat{y}_{22} \end{bmatrix} = \begin{bmatrix} 5.625 \\ 6.5 \end{bmatrix}$$

3.2 3^3 Factorial Design

In the 3^3 factorial design, the regression model is given by

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_{11} x^2{}_{1i} + \beta_{22} x^2{}_{2i} + \beta_{33} x^2{}_i$$
$$+ \beta_{12} x_{1i} x_{2i} + \beta_{13} x_{1i} x_{3i} + \beta_{23} x_{2i} x_{3i} + e_i,\ i = 1, \cdots, 27,$$

where the matrix H is

$$
H=\left[\begin{array}{ccccccccccccccccccccccccccc}
e & g & h & g & q & p & h & p & p & g & q & p & q & u & u & p & u & t & h & p & p & p & u & t & p & t & w \\
 & f & g & q & q & q & p & h & p & q & s & q & u & q & u & u & q & u & p & h & p & u & q & u & t & v & t \\
 & & e & p & q & g & p & p & h & p & q & g & u & u & q & t & u & p & p & p & h & t & u & p & w & t & p \\
 & & & f & s & h & g & q & p & q & u & u & s & q & q & q & u & u & p & u & t & h & q & v & p & u & t \\
 & & & & r & s & q & s & q & u & q & u & q & s & q & u & q & u & u & q & u & q & s & q & u & q & u \\
 & & & & & f & p & q & g & u & u & q & q & q & s & u & u & q & t & u & p & v & q & h & t & u & p \\
 & & & & & & e & g & h & p & u & t & q & u & u & g & q & p & p & t & w & p & u & t & h & p & p \\
 & & & & & & & f & g & u & q & u & u & q & u & q & s & q & t & v & t & u & q & u & p & h & p \\
 & & & & & & & & e & t & u & p & u & u & q & p & q & g & w & t & p & t & u & p & p & p & h \\
 & & & & & & & & & f & s & h & s & q & q & h & q & v & g & q & p & q & u & u & p & u & t \\
 & & & & & & & & & & r & s & q & s & q & q & s & q & q & s & q & u & q & u & u & q & u \\
 & & & & & & & & & & & f & q & q & s & v & q & h & p & q & g & u & u & q & t & u & p \\
 & & & & & & & & & & & & r & t & s & s & q & q & q & u & u & s & q & q & q & u & u \\
 & & & & & & & & & & & & & r & t & q & s & q & u & q & u & q & s & q & u & q & u \\
 & & & & & & & & & & & & & & r & q & q & s & u & u & q & q & q & s & u & u & q \\
 & & & & & & & & & & & & & & & f & s & h & p & u & t & q & u & u & g & q & p \\
 & & & & & & & & & & & & & & & & r & s & u & q & u & u & q & u & q & s & q \\
 & & & & & & & & & & & & & & & & & f & t & u & p & u & u & q & p & q & g \\
 & \text{(sym)} & & & & & & & & & & & & & & & & & e & g & h & g & q & p & h & p & p \\
 & & & & & & & & & & & & & & & & & & & f & g & q & s & q & p & h & p \\
 & & & & & & & & & & & & & & & & & & & & e & p & q & g & p & p & h \\
 & & & & & & & & & & & & & & & & & & & & & f & s & h & g & q & p \\
 & & & & & & & & & & & & & & & & & & & & & & r & s & q & q & q \\
 & & & & & & & & & & & & & & & & & & & & & & & f & p & q & g \\
 & & & & & & & & & & & & & & & & & & & & & & & & e & g & h \\
 & & & & & & & & & & & & & & & & & & & & & & & & & f & g \\
 & & & & & & & & & & & & & & & & & & & & & & & & & & e
\end{array}\right],
$$

where $e=53/108$, $f=71/108$, $g=-25/108$, $h=-7/108$, $p=5/108$, $q=-4/108$, $r=80/108$, $s=-16/108$, $t=-1/108$, $u=8/108$, $v=-13/108$, and $w=-19/108$.

Next, we consider the estimability for the missing values in the 3^3 factorial design. If $m\leq 8$, all the missing values are estimable. If $m=9$, nine missing values on a plane or on a diagonal plane are not estimable. That is, in Fig.3, nine missing values on a plane

$$y_{000},\ y_{001},\ y_{002},\ y_{010},\ y_{011},\ y_{012},\ y_{020},\ y_{021},\ y_{022};$$

$$y_{100},\ y_{101},\ y_{102},\ y_{110},\ y_{111},\ y_{112},\ \beta_{120},\ y_{121},\ y_{122};$$

$$y_{200},\ y_{201},\ y_{202},\ y_{210},\ y_{211},\ y_{212},\ y_{220},\ y_{221},\ y_{222};\ \text{etc.},$$

or nine missing values on a diagonal plane

$$y_{000},\ y_{001},\ y_{002},\ y_{110},\ y_{111},\ y_{112},\ y_{220},\ y_{221},\ y_{222};$$

$$y_{000},\ y_{010},\ y_{020},\ y_{101},\ y_{111},\ y_{121},\ y_{202},\ y_{212},\ y_{222};\ \text{etc.},$$

are not estimable. If $m \geq 10$, when a subset of the missing points is composed of the missing points which are not estimable for $m=9$, all of the missing values are not estimable.

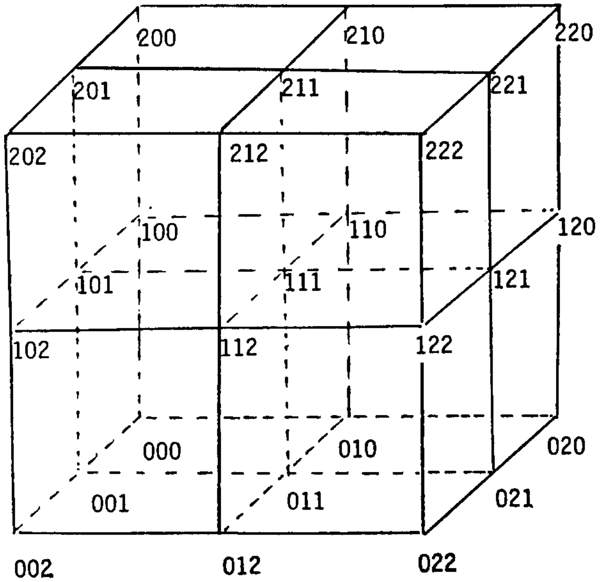

Fig. 3. 3^3 Factorial Design

4. Remarks

In this paper missing value estimation was considered for 2^k and 3^k factorial designs. If missing observations occur in such factorial designs, some good properties such as orthogonality and easy use of Yates' algorithms are not applicable. However, by inserting the estimated values into the missing observations, such good properties can be recovered and the usual statistical analysis can be performed. Moreover, the residual sum of squares is unaltered whether the estimated missing values are inserted or not.

There are some cases where missing values are not estimable. The cases were studied in detail for 2^k and 3^k factorial designs. In general, the nonestimable cases could be found in such cases that subsets of 2^k or 3^k factorial points have special structures as shown before.

Further studies are desired in missing value estimation. In particular, missing value estimation in fractional factorial designs, central composite designs, etc., should be of interest for further study.

5. Acknowledgment

We would like to express our gratitude to a referee for his constructive criticism and valuable suggestions which improved the original manuscript.

REFERENCES

(1) Box, M.J., Draper, N.R. and Hunter, W.G. (1970). Missing Values in Multiresponse Nonlinear Model Fitting, *Technometrics*, Vol. 12, No. 3, 613-620.

(2) Draper, N.R. (1961). Missing Values in Response Surface Designs, *Technometrics*, Vol. 3, No. 3, 389-398.

(3) Healy, M.J. and Westmacott, M.H. (1956). Missing Values in Experiments Analysed on automatic computers, *Applied Statistics*, Vol. 5, 203-206.

(4) Preece, D.A. (1971). Iterative Procedures for Missing Values in Experiments, *Technometrics*, Vol. 13, No. 4, 743-753.

(5) Shearer, P.R. (1973). Missing Data in Quantitatipe Designs, *Applied Statistics*, Vol. 22, 135-140.

(6) Tocher, K.D. (1952). The Design and Analysis of Block Experiments, *J. Roy. Statist. Soc. Ser.* B, Vol. 14, 45-100.

(7) Wilkinson, G.N. (1958). Estimation of Missing Values for the Analysis of Incomplete Data, *Biometrics*, Vol. 14, 257-286.