COMMUN. STATIST.—THEORY METH., 29(5&6), 1143–1154 (2000)

RESOLUTION V.2 SEARCH DESIGNS

K. Chatterjee

Department of Statistics, Asutosh College, Calcutta - 700 026, INDIA

Lih-Yuan Deng

Department of Mathematical Sciences, The University of Memphis

Memphis, TN 38152, U.S.A.

Dennis K. J. Lin

Department of Management Science and Information Systems

The Pennsylvania State University, University Park, PA 16802-1913, U.S.A.

Key Words: level combination; three-factor interaction; two-factor interaction.

ABSTRACT

A fractional factorial design is called a resolution V.2 plan if it is capable of estimating all main effects and two-factor interaction effects, plus two three-factor interaction effects. In this paper, a necessary and sufficient condition for such a resolution V.2 plan is given. Furthermore, a new class of two-level resolution V.2 designs is proposed. We prove that the proposed design always satisfies such a necessary and sufficient condition. A compar-

ison of run size between designs of resolutions VII and V.2 is made. It is shown that run size for design of resolution V.2 is significantly smaller.

1. INTRODUCTION

The search design, first proposed by Srivastava (1975), is an important class of screening designs. Since then, a substantial amount of research has been done in the field of search designs. Recent research trends in this area exhibit a considerable interest in the development of search designs which, in addition to ensuring estimability of the parameters known to be present, are capable of searching and estimating K possibly present parameters. See, for example, Ghosh (1980), Shirakura (1991), Srivastava and Arora (1987), Srivastava (1992), Chatterjee and Mukerjee (1986), and Mukerjee and Chatterjee (1994).

Much work has been done in constructing search designs that are capable of estimating all main effects and identifying K two-factor interaction effects, assuming all higher-order interactions are negligible. Practical experience, however, suggests that three-factor interactions do occur in some situations. Srivastava and Ghosh (1976) proposed search designs of resolution V which allow search and estimation of *one* extra unknown effect. In this paper, we suggest a new class of search designs for two-level factorials which is capable of estimating the grand mean, all the main and two-factor interaction effects, and allows the searching and estimating of at least two three-factor interactions among all the three-factor interactions, assuming that four- and higher-order interactions are absent. In the spirit of Srivastava and Ghosh (1976), we call such a design of resolution V.2. Namely, a design is of resolution V.2 if it is a design of resolution V, and in addition, it is capable to search and estimate two three-factor interactions.

Consider the linear model below: Let $\mathbf{Y}$ be an $N \times 1$ vector of observations with

$$E(\mathbf{Y}) = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2, \quad Disp(\mathbf{Y}) = \sigma^2\mathbf{I}_N \tag{1}$$

where $\mathbf{X}_1, \mathbf{X}_2$ are the design matrices of order $N \times n_1$ and $N \times n_2$, respectively; $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2$ are the vectors of parameters of order $n_1 \times 1$ and $n_2 \times 1$, respectively; and $\mathbf{I}_N$ is an identity matrix of order N. Furthermore, only K elements of $\boldsymbol{\beta}_2$ are assumed to be non-zero, where K is relatively small compared to n_2. Here, the objective is to provide a design which will estimate the grand mean and all the main-effects ($\boldsymbol{\beta}_1$) and permit the detection and estimation of the non-zero elements of $\boldsymbol{\beta}_2$. In this paper, we will consider the "noiseless" case with $\sigma^2 = 0$, as in Srivastava (1975). The main theorem in this field is provided by Srivastava (1975), as stated in Theorem 1.

Theorem 1. A necessary and sufficient condition that the above problem will be completely solved is that for any submatrix $\mathbf{X}_2^*$ of $\mathbf{X}_2$ of order $N \times 2K$, we have

$$rank[\mathbf{X}_1, \mathbf{X}_2^*] = n_1 + 2K.$$

2. NOTATIONS AND PRELIMINARIES

Consider a factorial experiment involving m factors $F_1, F_2, ..., F_m$ each at two levels. Let the levels of the factors be coded as $0, 1$ and a typical level combination of these factors be denoted by $(f_1, f_2, ..., f_m)$, $f_i \in \{0, 1\}$, for

$1 \leq i \leq m$. For any positive integer n, let $\mathbf{I}_n$ denote the identity matrix of order n, and let $\mathbf{0}_n$ and $\mathbf{1}_n$ denote the n column vectors with all elements 0 and 1, respectively. Let $\mathbf{O}_{n\times m}$ denote the matrices of order $n \times m$ with entry value 0. For the simplicity of presentation, we may drop the subscripts hereafter.

Suppose prior information is available regarding the absence of all interactions involving four or more factors and it is known that among the three-factor interactions at most two are non-negligible, although these interactions are not known a priori. Then, under model (1) by taking $\boldsymbol{\beta}_1$ as the grand mean, main effects, and two-factor interactions, and $\boldsymbol{\beta}_2$ as the three-factor interactions, it is easy to see that if one observation is made for each of the $\nu = (2^m)$ level combinations of $F_1, F_2, ..., F_m$, the resulting design matrix will be of the form

$$\mathbf{X} = [\mathbf{1}_\nu, \mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_m, \mathbf{u}_{12}, ..., \mathbf{u}_{m-1,m}, \mathbf{u}_{123}, ..., \mathbf{u}_{m-2,m-1,m}], \tag{2}$$

where $\mathbf{u}_i$ is a vector of ± 1 for $1 \leq i \leq m$ and

$$\mathbf{u}_{ij} = \mathbf{u}_i * \mathbf{u}_j, \quad 1 \leq i < j \leq m;$$

$$\mathbf{u}_{ijk} = \mathbf{u}_i * \mathbf{u}_j * \mathbf{u}_k, \quad 1 \leq i < j < k \leq m,$$

and '$*$' denotes the Hadamard product. The Hadamard product of two vectors $\mathbf{u} = (u_1, u_2, ..., u_n)'$ and $\mathbf{v} = (v_1, v_2, ..., v_n)'$ is defined as $\mathbf{u} * \mathbf{v} = (u_1v_1, u_2v_2, ..., u_nv_n)'$. Observe that in (2), $\mathbf{1}_\nu$ corresponds to the grand mean, $\mathbf{u}_i$ corresponds to the main effect $F_i (1 \leq i \leq m)$, $\mathbf{u}_{ij}$ corresponds to the two-factor interaction $F_iF_j (1 \leq i < j \leq m)$, and finally, $\mathbf{u}_{ijk}$ corresponds to the three-factor interaction $F_iF_jF_k (1 \leq i < j < k \leq m)$. It follows that

$$\mathbf{X}_1 = [\mathbf{1}_\nu, \mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_m, \mathbf{u}_{12}, ..., \mathbf{u}_{m-1,m}] = [\mathbf{1}_\nu, \mathbf{U}_1, \mathbf{U}_2] \quad \text{and}$$

$$\mathbf{X}_2 = [\mathbf{u}_{123}, ..., \mathbf{u}_{m-2,m-1,m}].$$

Let $\mathbf{Q}$ be an $m \times N$ matrix, where N is the number of runs and the columns in $\mathbf{Q}$ represent N level combinations of $F_1, F_2, ..., F_m$. Based on the matrix $\mathbf{Q}$, we denote the corresponding submatrices of $\mathbf{X}_1$ and $\mathbf{X}_2$ as

$$\mathbf{X}_1^* = [\mathbf{1}_N, \mathbf{u}_1^*, \mathbf{u}_2^*, ..., \mathbf{u}_m^*, \mathbf{u}_{12}^*, ..., \mathbf{u}_{m-1,m}^*] = [\mathbf{1}_N, \mathbf{U}_1^*, \mathbf{U}_2^*] \quad \text{and}$$

$$\mathbf{X}_2^* = [\mathbf{u}_{123}^*, ..., \mathbf{u}_{m-2,m-1,m}^*]$$

respectively. Clearly, a minimum requirement for choosing N runs is that $rank(\mathbf{X}_1^*) = 1 + m + \binom{m}{2}$. Then, analogous to Theorem 1, we have the following theorem. In the next section, we will provide a new class of the search designs that satisfies such a requirement.

Theorem 2. A necessary and sufficient condition that $\mathbf{Q}$ represents a search design for a resolution V.2 plan is if the following holds: for any submatrix $\mathbf{X}_2^{**}$ of $\mathbf{X}_2^*$ of order $N \times 4$, we have a full-rank matrix $[\mathbf{X}_1^*, \mathbf{X}_2^{**}]$. That is,

$$rank[\mathbf{X}_1^*, \mathbf{X}_2^{**}] = n_1 + 4,$$

where $n_1 = 1 + m + \binom{m}{2}$ which is the number of parameters for grand mean, main-effects and two-factor interactions.

3. A NEW CLASS OF TWO-LEVEL SEARCH DESIGNS

Define the following $m \times N$ matrix

$$\mathbf{Q} = [\mathbf{0}_m, \mathbf{I}_m, \mathbf{A}_m, \mathbf{B}_m], \tag{3}$$

where $\mathbf{A}_m, \mathbf{B}_m$ are $m \times \binom{m}{2}$ matrices with their column vector

$$\mathbf{a}_{ij} = \mathbf{e}_i + \mathbf{e}_j, \quad \text{and} \quad \mathbf{b}_{ij} = \mathbf{1}_m - \mathbf{e}_i - \mathbf{e}_j,$$

where $\mathbf{e}_i, \mathbf{e}_j$ are the i-th and j-th unit vectors. Note that the matrix $\mathbf{A}_m(\mathbf{B}_m)$

consists of columns with 1(0) in the i,j-th position and 0(1) in the other positions for $1 \le i < j \le m$. Consequently, the number of columns in $\mathbf{Q}$ is $N = 1 + m + m(m-1) = m^2 + 1$. For example, when $m = 5$, $\mathbf{Q}$ is the 5×26 matrix

$$\begin{array}{c|ccccc|cccccccccc|cccccccccc}
0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0.
\end{array}$$

We next present some properties of such a $\mathbf{Q}$ matrix and then prove that a search design based on $\mathbf{Q}$ will always satisfy Theorem 2. The design matrix corresponding to the matrix $\mathbf{Q}$ in (3) can be written as

$$\mathbf{X}^* = [\mathbf{1}_N, \mathbf{U}_1^*, \mathbf{U}_2^*, \mathbf{U}_3^*], \tag{4}$$

where

- $\mathbf{U}_1^* = [\mathbf{u}_1^*, \mathbf{u}_2^*, \ldots, \mathbf{u}_m^*]$, where $\mathbf{u}_i^* = 2\mathbf{q}_i - \mathbf{1}$ and $\mathbf{q}_i$ is a column vector of $\mathbf{Q}'$.
- $\mathbf{U}_2^* = [\mathbf{u}_{12}^*, \ldots, \mathbf{u}_{m-1,m}^*]$, where $\mathbf{u}_{ij}^* = \mathbf{u}_i^* * \mathbf{u}_j^*$.
- $\mathbf{U}_3^* = [\mathbf{u}_{123}^*, \ldots, \mathbf{u}_{m-2,m-1,m}^*]$, where $\mathbf{u}_{ijk}^* = \mathbf{u}_i^* * \mathbf{u}_j^* * \mathbf{u}_k^*$.

Since $\mathbf{u}_i^* = 2\mathbf{q}_i - \mathbf{1}$, we have $\mathbf{q}_i = (\mathbf{u}_i^* + \mathbf{1})/2$. Therefore, there exist matrices $\mathbf{P}_1, \mathbf{P}_2$ such that

$$\mathbf{U}_1^* = \mathbf{Q}'\mathbf{P}_1 \quad \text{and} \quad \mathbf{Q}' = \mathbf{U}_1^*\mathbf{P}_2.$$

Thus, when $\mathbf{U}_1^*$ is replaced by $\mathbf{Q}'$ in the design matrix, the rank property is preserved, where

$$\mathbf{Q}' = [\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_m] = \begin{bmatrix} \mathbf{0}_m' \\ \mathbf{I}_m \\ \mathbf{A}_m' \\ \mathbf{B}_m' \end{bmatrix}.$$

Note also that $\mathbf{u}^*_{ij} = (2\mathbf{q}_i - \mathbf{1}) * (2\mathbf{q}_j - \mathbf{1}) = 4\mathbf{q}_i * \mathbf{q}_j - 2(\mathbf{q}_i + \mathbf{q}_j) + \mathbf{1}$. Therefore, $(\mathbf{u}^*_{ij} + 2(\mathbf{q}_i + \mathbf{q}_j) - \mathbf{1})/4 = \mathbf{q}_i * \mathbf{q}_j$. The entry value of $\mathbf{q}_i * \mathbf{q}_j$ is 1 only when the corresponding entries of both $\mathbf{q}_i$ and $\mathbf{q}_j$ are 1. Similarly, the rank property is again preserved when $\mathbf{U}^*_2$ is replaced by $\mathbf{Q}'_2$ which is

$$\begin{bmatrix} \mathbf{0}'_{\binom{m}{2}} \\ \mathbf{O}_{m\times\binom{m}{2}} \\ \mathbf{I}_{\binom{m}{2}} \\ \mathbf{V}_1 \end{bmatrix} = [\mathbf{q}_1 * \mathbf{q}_2, \ldots, \mathbf{q}_{m-1} * \mathbf{q}_m],$$

where $\mathbf{V}_1 = [\mathbf{v}_{12}, \ldots, \mathbf{v}_{m-1,m}]$ is a $\binom{m}{2} \times \binom{m}{2}$ matrix and the value of the entry of $\mathbf{v}_{ij} (1 \le i < j \le m)$ will be 1 when both i-th and j-th factors are set to be 1, otherwise $\mathbf{v}_{ij} = 0$. Note that there are exactly $\binom{m-2}{2}$ entries of each $\mathbf{v}_{ij}$ vector that are 1 with the remaining entries equal to 0.

Similarly, it is straightforward to see that

$$(\mathbf{u}^*_{ijk} + 4(\mathbf{q}_i * \mathbf{q}_j + \mathbf{q}_i * \mathbf{q}_k + \mathbf{q}_j * \mathbf{q}_k) - 2(\mathbf{q}_i + \mathbf{q}_j + \mathbf{q}_k) + 1)/8 = \mathbf{q}_i * \mathbf{q}_j * \mathbf{q}_k.$$

We can see that $\mathbf{q}_i * \mathbf{q}_j * \mathbf{q}_k$ value is 1 when all of $\mathbf{q}_i, \mathbf{q}_j, \mathbf{q}_k$ are 1, and is 0 otherwise. Therefore, within this vector all the entries corresponding to $\mathbf{0}'_m$, $\mathbf{I}_m$, and $\mathbf{A}_m$ in $\mathbf{Q}' = [\mathbf{0}_m, \mathbf{I}_m, \mathbf{A}_m, \mathbf{B}_m]'$ will be 0. Similarly, the rank property is preserved when $\mathbf{U}^*_3$ is replaced by

$$\begin{bmatrix} \mathbf{0}'_{\binom{m}{3}} \\ \mathbf{O}_{m\times\binom{m}{3}} \\ \mathbf{O}_{\binom{m}{2}\times\binom{m}{3}} \\ \mathbf{V}_2 \end{bmatrix},$$

where $\mathbf{V}_2 = [\mathbf{v}_{123}, \ldots, \mathbf{v}_{m-2,m-1,m}]$ is a $\binom{m}{2} \times \binom{m}{3}$ matrix and the value of the entry of $\mathbf{v}_{ijk} (1 \le i < j < k \le m)$ will be 1 when all three of i-th, j-th and k-th factors are set to be 1, otherwise $\mathbf{v}_{ijk} = 0$.

Performing the elementary column operations as described above on the matrix in (4), the resulting matrix is

$$\mathbf{M} = \begin{bmatrix} 1 & \mathbf{0}'_m & \mathbf{0}'_{\binom{m}{2}} & \mathbf{0}'_{\binom{m}{3}} \\ \mathbf{1}_m & \mathbf{I}_m & \mathbf{O} & \mathbf{O} \\ \mathbf{1}_{\binom{m}{2}} & \mathbf{A}'_m & \mathbf{I}_{\binom{m}{2}} & \mathbf{O} \\ \mathbf{1}_{\binom{m}{2}} & \mathbf{B}'_m & \mathbf{V}_1 & \mathbf{V}_2 \end{bmatrix}. \tag{5}$$

From the structure of the matrix $\mathbf{M}$ and Theorem 2, we need to show that any selection of 4 columns from $\mathbf{V}_2$ is of full column rank.

4. MAIN THEOREM

The following theorem shows that $\mathbf{Q}$ as defined in (3) satisfies Theorem 2 and thus can be used to estimate the grand mean and all main effects and two-factor interactions plus it can correctly identify at least two of the interaction effects among all three-factor interactions.

Theorem 3. The columns of the matrix $\mathbf{Q}$ as defined in (3), interpreted as the level combinations of the factors $F_1, F_2, ..., F_m$, gives the required search design in N runs.

Proof. Suppose that the columns selected are

$$\mathbf{V} = [\mathbf{v}_{i_1 j_1 k_1}, \mathbf{v}_{i_2 j_2 k_2}, \mathbf{v}_{i_3 j_3 k_3}, \mathbf{v}_{i_4 j_4 k_4}]. \tag{6}$$

In this case, we need to show that we can always find certain identified rows from the given 4 columns so that the submatrix has rank 4.

The rows of $\mathbf{V}_2$ corresponding to the rows of $\mathbf{B}'_m$ can be identified by the indexes $\{i, j\}$. The corresponding row of $\mathbf{B}'_m$ is $\mathbf{b}'_{ij} = \mathbf{1}'_m - \mathbf{e}'_i - \mathbf{e}'_j$ with '0' in the i-th and j-th position and '1' in the other positions.

We will denote $C = A - B$ as the set of all elements in set A and not in set B.

Let us first consider the column indexed by the set $\{i_1, j_1, k_1\}$. Since $\{i_1, j_1, k_1\} \neq \{i_2, j_2, k_2\}$, we can always find $i \in \{i_2, j_2, k_2\} - \{i_1, j_1, k_1\}$. Define $Z = \{i_3, j_3, k_3\} - \{i_1, j_1, k_1\}$. Next, we are going to choose j so that the row in $\mathbf{V}_2$ indexed by $\{i, j\}$ has the entry values indicated.

- If $Z - \{i\} \neq \emptyset$, the empty set, then choose $j \in Z - \{i\}$. Therefore, the entry associated with $\{i_1, j_1, k_1\}$ is 1 because $\{i, j\} \cap \{i_1, j_1, k_1\} = \emptyset$. In addition, the entries associated with $\{i_2, j_2, k_2\}$ and $\{i_3, j_3, k_3\}$ are 0, because $\{i, j\} \cap \{i_2, j_2, k_2\} \neq \emptyset$ and $\{i, j\} \cap \{i_3, j_3, k_3\} \neq \emptyset$.
- If $Z - \{i\} = \emptyset$, then we can choose any $j \notin \{i_1, j_1, k_1\}$. Therefore, the entry associated with $\{i_1, j_1, k_1\}$ is 1 because $\{i, j\} \cap \{i_1, j_1, k_1\} = \emptyset$. In addition, the entries associated with $\{i_2, j_2, k_2\}$ and $\{i_3, j_3, k_3\}$ are 0, because $i \in \{i_2, j_2, k_2\}$ *and* $i \in \{i_3, j_3, k_3\}$.

Using a similar argument to the columns indexed by $\{i_2, j_2, k_2\}$, $\{i_3, j_3, k_3\}$ and $\{i_4, j_4, k_4\}$, we can find the following rows

$$\begin{array}{cccc} 0 & 1 & b & 0 \\ c & 0 & 1 & 0 \\ 0 & 0 & d & 1 \\ (i_1 j_1 k_1) & (i_2 j_2 k_2) & (i_3 j_3 k_3) & (i_4 j_4 k_4), \end{array}$$

where the values of b, c, d can be 0 or 1.

Therefore, we find a submatrix

$$\mathbf{S} = \begin{bmatrix} 1 & 0 & 0 & a \\ 0 & 1 & b & 0 \\ c & 0 & 1 & 0 \\ 0 & 0 & d & 1 \end{bmatrix}$$

Table 1. Run size consideration for Resolution VII and V.2.

Number of factors (m)	Resolution VII+ design	Resolution V.2 ++
5	32	26
6	64	37
7	64	50
8	128	65
9	256	82
10	256	101
11	512	122
12	512	145

+ Resolution VII plan can estimate all main effects and all two- and three-factor interactions. (These run sizes are derived from Draper and Lin, 1990, Table 4.)

++ Design corresponding to matrix **Q** in (3) can estimate all main effects, two-factor interactions and correctly identify two three-factor interaction effects.

whose determinant is $1 + acd$, which is non-zero regardless of the values of a, b, c, d. $\square$

Theorem 3 suggests that the matrix **Q** presented in this section provides a search design in $N = 1 + m^2$ runs, m being the number of factors involved which allow the estimation of the grand mean and all the main effects and two-factor interactions, plus it will detect and estimate at least two of the three-factor interactions.

5. CONCLUSION

Much work has been done on the "main effect plus K" search design after the important work of Srivastava (1975). In this paper, we first give the necessary and sufficient conditions for the search design of resolution V.2 and then provide a new class of search designs that fulfills such a condition.

If a larger number of factors is under investigation and if only a small portion (say, two) of three-factor interaction effects are expected the search design given in this paper can be very useful. To illustrate the run size issue, a brief run size comparison with a two-level resolution VII plan is given in Table 1. Resolution VII is needed in order to estimate all two- and three-factor interactions. It is clear that the design corresponding to the matrix **Q** in (3) can result in considerable experimental cost.

BIBLIOGRAPHY

Chatterjee, K. and Mukerjee, R. (1986). "Some search designs for symmetric and asymmetrical factorials." *J. Statist. Plann. Inf.*, vol 13, pages 357-363.

Draper, N. R. and Lin, D. K. J. (1990). "Capacity consideration for two-level fractional factorial design." *J. Statist. Plann. Inf.*, vol 24, pages 25-35.

Ghosh, S. (1980). "On main effect plus one plans for 2^m factorials." *Annals of Statistics*, vol 8, pages 922-930.

Mukerjee, R. and Chatterjee, K. (1994). "An application of Hadamard matrices for the construction of main effect plus two plans for 2^m factorials." *Utilities Math.*, vol 45, pages 213-218.

Shirakura, T. (1991). "Main effect plus one or two plans for 2^m factorials." *J. Statist. Plann. Inf.*, vol 27, pages 65-74.

Srivastava, J. N. (1975). "Design for searching non-negligible effects." In: *A Survey of Statistical Design and Linear Models*, (Editor J. N. Srivastava), North-Holland, Amsterdam, pages 507-519.

Srivastava, J. N. (1992). "A 2^8 factorial main effect plus two plan with good revealing power." *Sankhyā*, vol 54, pages 461-474.

Srivastava, J. N. and Arora, S. (1987). "A minimal search design of resolution 3.2 for 2^4 factorial experiment." *Ind. J. Math.*, vol 3, pages 309-320.

Srivastava, J. N. and Ghosh, S. (1976). "A series of balanced 2^m factorial designs of resolution V which allow search and estimation of one extra unknown effect." *Sankhyā*, vol 38, pages 280-289.

Received June, 1999; Revised January, 2000.