ON ORTHOGONAL ARRAYS OF STRENGTH FOUR AND THEIR APPLICATIONS

Thesis for the Degree of Ph. D.

MICHIGAN STATE UNIVERSITY

Rita Zemach

1965

This is to certify that the

thesis entitled

ON ORTHOGONAL ARRAYS OF STRENGTH FOUR
AND THEIR APPLICATIONS

presented by

Rita Zemach

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Statistics

Esther Seiden
Major professor

Date May 21, 1964

ABSTRACT

ON ORTHOGONAL ARRAYS OF STRENGTH FOUR

AND THEIR APPLICATIONS

by Rita Zemach

An orthogonal array (N,k,s,t) of s levels, k constraints, strength t, and index Q, is a $k \times N$ matrix with entries from a set of s elements, $s \geq 2$, such that each $t \times N$ sub-matrix contains all possible $t \times 1$ column vectors with the same frequency Q. The array serves as a design for a factorial experiment, with k factors, each occurring at s levels, and N treatments. If the array is of strength $t = 2v$, all interactions involving v or fewer factors can be estimated, assuming there is no interaction of more than v factors. If $t = 2v + 1$, all interactions involving v or fewer factors can be estimated even if interactions of $v + 1$ factors are present, but estimates of interactions of $v + 1$ factors may be confounded with one another. Thus arrays of strength four have the smallest strength which allows estimation of main effects and all first-order interactions, assuming that higher order interactions are negligible.

The main problem of interest for orthogonal arrays is to determine the maximum possible number of constraints for given N, s, and t.

Chapter I is devoted to a general discussion of orthogonal arrays, and the method of analysis. An iterative bound on the maximum number of constraints for an array (Qs^t,k,s,t) is given, which depends on the maximum number of constraints for an array $(Qs^{t-1},k,s,t-1)$.

Rita Zemach

The main theorem of Chapter II gives a method of constructing an array $A = (Q2^{t+1},k+1,2,t+1)$ from an array $A' = (Q2^{t},k,2,t)$, when $t = 2v$. If k is the maximum number of constraints possible for A', then $k + 1$ is the maximum possible number of constraints for A.

In addition, the structure of arrays $(Q2^{t},t+1,2,t)$ is analyzed, and for $Q = q2^{n}$, a method of extending any array $(Q2^{t},t+1,2,t)$ to $t + n + 1$ constraints is described.

Chapter III describes a method of constructing arrays by matrix multiplication, due to Bose, for the case $s = p^{u}$, $Q = p^{v}$, p a prime. The algebraic properties of arrays of this class are discussed, and bounds on the maximum possible number of constraints are reviewed. A bound is given for arrays of strength four which can be constructed by matrix multiplication, when $s = 3$, $Q = 3^{u}$.

In Chapter IV, the maximum possible number of constraints is determined, when $s = 2$, $t = 4$, for arrays with $N = 32, 48, 64$ and 80. In each case, arrays with the maximum number of constraints are constructed. A detailed examination of the structure of most of the arrays is included.

The last Chapter is devoted to a brief discussion of the relation of orthogonal arrays to information theory.

ON ORTHOGONAL ARRAYS OF STRENGTH FOUR AND THEIR APPLICATIONS

By

Rita Zemach

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Statistics

1965

ACKNOWLEDGMENTS

I wish to express my gratitude to Dr. Esther Seiden for suggesting this area of research, and for her helpful guidance and valuable advice in the preparation of this thesis.

I appreciated very much the support of my research by the National Science Foundation through fellowships from September 1962 to September 1963, and during the summer of 1964.

Finally, I wish to thank my husband and children for their patience and encouragement during my entire period of graduate study.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

1.1 Factorial Design

Frequently in statistical investigations it is known that the characteristic being studied is affected by several different factors, $F_1, \ldots, F_k$. Each factor may assume several different levels, e.g. F_i assumes s_i levels, $i = 1, \ldots, k$. Estimation of each of the main effects is desired, as well as information about interactions among the various factors.

I shall call a complete factorial design one which includes each of the possible $s_1 \cdot s_2 \cdot \ldots \cdot s_k$ treatments exactly once. A treatment in which F_i occurs at level a_i, $i = 1, \ldots, k$ will be designated by $(a_1, a_2, \ldots, a_k)$.

In a symmetrical factorial design all factors assume the same number of levels, i.e., $s_1 = s_2 = \ldots = s_k = s$. In this case a complete factorial design consists of the set of all s^k possible k-tuples of s elements. It will be referred to as a complete s^k design.

The true yield of a treatment in a factorial design is in general regarded as the sum of an over-all mean μ, the effects of the k factors, and the effects of interactions of all orders among these factors.

The special case $s = 2$ occurs when each of the k factors assumes 2 levels, which may be the presence or absence of the factor, or its existence at high or low intensity, or in two different forms.

If the number of factors is very large, the number of treatments necessary for a complete s^k design reaches a prohibitive size. This problem arises, for example, in exploratory experiments, where the researcher cannot, for the time being, ignore the effect of any factor which may influence the characteristic under investigation. At the same time, although the complete s^k design provides estimates of interactions of all orders, in most practical situations the higher order interactions may be assumed to be negligible.

These considerations have given rise to the use of fractional replication of complete factorial designs. In a $1/s^n$ replication of a complete s^k factorial design, the s^k treatments are partitioned into blocks of s^{k-n} treatments, satisfying certain properties. This type of design is discussed in section 3.1. One block may be represented by a $k \times s^{k-n}$ matrix representing k factors and s^{k-n} treatments. The term block of a design will hereafter imply a block of a partitioning as described.

The partitioning of a complete s^k design is said to be of strength t if no interaction of t or fewer factors is confounded with the block effect. The properties of a single block as a design will be described under the more general topic of orthogonal arrays.

1.2 Orthogonal Arrays

An orthogonal array (N,k,s,t) of s levels, k constraints, strength t, and index Q, is a $k \times N$ matrix A with entries from a set S of s elements, $s \geq 2$, such that each $t \times N$ sub-matrix of A contains all possible $t \times 1$ column vectors of S with the same frequency Q. Clearly $N = Qs^t$. The set S may be taken as the set of integers $(0,1,\dots\dots,s-1)$.

If A is of strength t , then any $k' \times N$ sub-array, $t \leq k' \leq k$, is also of strength t . Hence, if no array (Qs^t, k', s, t) exists, then neither does any array (Qs^t, k, s, t) for $k > k'$.

If A is of strength t , it is also of strength t' for every $t' \leq t$.

An orthogonal array serves as a design for a factorial experiment, with k factors at s levels, and N treatments.

Arrays of the form $(s^2, k, s, 2)$ may be derived from k-2 mutually orthogonal Latin squares of side s , the rows of one square providing one row of the array. The remaining two rows are derived from the following orthogonal (but not Latin) square and its transpose:

$$\begin{bmatrix} 0 & 0 & \dots\dots & 0 \\ 1 & 1 & \dots\dots & 1 \\ 2 & 2 & \dots\dots & 2 \\ \vdots & \vdots & & \\ (s-1) & (s-1) & \dots\dots & (s-1) \end{bmatrix} .$$

Orthogonal arrays were introduced by C. R. Rao, [17], who first discussed these factorial designs in terms of "hypercubes" of strength t .

If the design of a factorial experiment is an orthogonal array of strength t , then no estimate of a main effect is confounded with any interaction of (t-1) or fewer factors, and in general, no interaction involving m factors, $m < t$, is confounded with an interaction involving (t-m) or fewer factors.

By using an array of strength $2t$, all interactions involving t or fewer factors can be estimated, assuming there are no interaction of more than t factors. With an array of strength $2t+1$, interactions of t factors can be estimated, even if interactions of $t+1$ factors are present. However, interactions of $t+1$ factors may be confounded with one another.

Thus, arrays of strength 4, which will be the subject of particular study in this thesis, allow estimation of all main effects and all first-order interactions, assuming that interactions of second or higher order are negligible. (An interaction is "of n^{th} order" if it involves $n+1$ factors.)

If no interaction of more than $(t-1)$ factors is present, a sufficient condition to measure all main effects only is that an array be of strength t.

It should be noted that while a block of a factorial design of strength t (see Sec. 1.1) is an orthogonal array of strength t, the converse is not true. The array need not satisfy all the conditions imposed on the block, and may still be of strength t.

In particular, a complete s^k factorial design is an array of strength k.

The main problem of concern with respect to an array $A=(N,k,s,t)$ is the maximum value of k for a given t, N, and s, or the maximum number of factors which can be accomodated in a design of strength t and index Q.

The maximum number of constraints for orthogonal arrays of strength 2 and 3, when $s=2$, has been determined for many cases.

For the case $t=2$, it was shown by Plackett and Burman [16] that for an array $(Qs^2,k,s,2)$

$$k \leq \frac{Qs^2 - 1}{s - 1}$$

which reduces to $k \leq 4Q - 1$ for the case $s=2$. Construction of arrays which achieve this bound for all Q up to $Q=50$, except for the cases $Q=29, 47$, has been determined by Paley [15], Williamson [23], and Baumert and Hall [3]. A method of construction is given whenever $Q = 2^n$. (See [5, 11, 15]). For the case $t=3$, Rao [17] shows that for an array $(Qs^3,k,s,3)$

$$k \leq \frac{Qs^2 - 1}{s - 1} + 1$$

which reduces to $k \leq 4Q$ when $s=2$. Again it was shown by Bose [5] that arrays can be constructed with $k=4Q$ when $Q=2^n$. Seiden [19] gives a method of construction of arrays $(Q2^3,4Q,2,3)$ whenever an array $(Q2^2,4Q-1,2,2)$ exists.

For the case $t=4$, the bound of Rao becomes

$$k \leq \frac{(s-3)+\sqrt{(s-3)^2 + 8(Qs^4-1)}}{2(s-1)}$$

which gives the following bounds for $s=2$:

	N	Q	k
(1)	16	1	5
(2)	32	2	7
(3)	48	3	9
(4)	64	4	10
(5)	80	5	12

It will be shown that for the cases listed, this bound can be achieved only for case (1). Moreover, in the remaining cases, the exact maximum for k will be established.

Bush [9] has shown that, in general, if $Q=1$,

$$k \leq s + t - 1 \qquad s \text{ even}$$

$$k \leq s + t - 2 \qquad s \text{ odd.}$$

If $s \leq t$, then the bound is attained, for $s=p^n$, p a prime.

The following inequality can be established between the maximum number of constraints k' and k of orthogonal arrays $(Qs^{t-1},k',s,t-1)$ and (Qs^t,k,s,t):

> Let k' be the maximum number of constraints possible for an array of index Q and strength $t-1$.
>
> Then for an array of index Q and strength t, $k \leq k' + 1$.

Proof: Let $A = (Qs^t,k,s,t)$ be an orthogonal array of strength t. The i^{th} row of A, $i=1,\ldots,k$, contains each element $0,1,2,\ldots,s-1$ repeated Qs^{t-1} times. Without loss of generality, consider the first row. Let A' be any sub-matrix of $t-1$ rows from among the last $k-1$ rows. Each of the s^{t-1} $(t-1)$-tuples must appear exactly Qs times in A', and these must be arranged so that each one appears exactly Q times following each of the elements $0,1,\ldots,s-1$ in the first row. Therefore the last $k-1$ rows may be partitioned into s arrays $(Qs^{t-1},k-1,s,t-1)$. Hence if k' is the maximum number of constraints possible for an array of strength $t-1$, the maximum for an array of strength t is less than or equal to $k'+1$.

1.3 Some Conditions Necessary for the Existence of Arrays

For any orthogonal array $A = (N,k,s,t)$ of index Q, let n_{ij} denote the number of columns (other than the i^{th}) having j coincidences with the i^{th} column, $i=1,\ldots,N$; $j=0,\ldots,k$. It is shown by Bose and Bush [7] that for each i, the k variables n_{ij} must satisfy a set of $t+1$ linear equations:

$$\sum_{j=0}^{k} n_{ij} = Qs^{t} - 1$$

$$\sum_{j=1}^{k} \binom{j}{h} n_{ij} = \binom{k}{h} (Qs^{t-h}-1) \quad 1 \leq h \leq t$$

where $\binom{j}{h} = 0$ for $j < h$. These equations will be used in several proofs and will be referred to as the "necessary equalities."

Some additional properties will prove useful for the construction of orthogonal arrays.

Property 1: For an array $A = (Qs^{t},k,s,t)$, let $\hat{n}_{ik}$ be the maximum value of n_{ik} among all solutions to the necessary equations. Then for any array it is necessary that

$$n_{ij} \leq \binom{k}{j} (\hat{n}_{ik} + 1) \quad j=0,1,\ldots,k-1.$$

Proof: Suppose a particular column having j coincidences with the i^{th} column is repeated more than $\hat{n}_{ik}+1$ times. Then an array would exist whose solution yields $n_{ik} > \hat{n}_{ik}$. There are at most $\binom{k}{j}$ columns having j coincidences with the i^{th} column.

Property 2: If no solution exists with $n_{ik} > 0$, then every array (Qs^t,k,s,t) has Qs^t distinct columns. This follows from property 1.

Property 3: If for some k, $n_{ik} = 0$ for all solutions, then every array (Qs^t,k',s,t) with $k' > k$ must have $n_{ik} = 0$.

For the case $t=4$, the equations take the form:

$$\sum_{j=0}^{k} n_{ij} = Qs^4 - 1$$

$$\sum_{j=1}^{k} j\, n_{ij} = k(Qs^3 - 1)$$

$$\sum_{j=2}^{k} \binom{j}{2} n_{ij} = \binom{k}{2} (Qs^2 - 1)$$

$$\sum_{j=3}^{k} \binom{j}{3} n_{ij} = \binom{k}{3} (Qs - 1)$$

$$\sum_{j=4}^{k} \binom{j}{4} n_{ij} = \binom{k}{4} (Q - 1)$$

1.4 Analysis of Orthogonal Arrays

Consider k factors $F_1,\ldots\ldots,F_k$. Let $A = (N,k,2,4)$ be the array. Let α_{ijm---} be the true yield of the treatment with F_1 at level i, F_2 at level j, F_3 at level m, etc., where each i, j, m, -, -, is either 0 or 1, and $ijm---$ takes on at most only those N distinct values indicated by the N columns of the array A.

The main effects of factors F_1, F_2, F_3, --- will be denoted by the letters a, b, c, etc., and the interactions by pairs (ab), (ac), (bc), etc. It is assumed that no interactions of three or more factors are present. The parameters to be estimated are the over-all mean μ, k main effects, and $k(k-1)/2$ first-order interactions. Assume that:

$$\mu = \alpha_{.....} = 1/N \sum_{A} \alpha_{ijm---}$$

$$a_i = \alpha_{i....} - \alpha_{.....} = 2/N \sum_{(i\ \text{fixed})} \alpha_{ijk---} - \alpha_{.....}$$

$$b_j = \alpha_{.j...} - \alpha_{.....}$$

$$\vdots$$

$$(ab)_{ij} = \alpha_{ij...} - \alpha_{i....} - \alpha_{.j...} + \alpha_{.....}$$

$$\vdots$$

(In practice it is customary, with only two levels, to define each effect at twice the value given here, so that, for example, with one factor the effect is

$$a = \alpha_1 - \alpha_0$$

rather than

$$a_1 = \alpha_1 - 1/2(\alpha_1 + \alpha_0)$$

$$= 1/2(\alpha_1 - \alpha_0) = 1/2(a)$$

$$a_0 = \alpha_0 - 1/2(\alpha_0 + \alpha_1)$$

$$= 1/2(\alpha_0 - \alpha_1) = -1/2(a)$$

as indicated above.)

Let $Y = \{y_{ijm---}\}$ denote the $1 \times N$ vector of observations for (ijm---) contained in A.

The assumptions of the model are:

(i) $y_{ijm---} = \mu + \left\{ a_i + b_j + c_m + --- \right\}$ (k main effects)

$+ \left\{ (ab)_{ij} + (ac)_{im} + -- \right\}$ ($\binom{k}{2}$ interactions)

$+ \; e_{ijm---}$

i, j, m, -, -, -, contained in (0,1)

(ijm---) contained in A

(ii) $\left\{ e_{ijm---} \right\}$ are normally distributed with mean zero and co-variance matrix $I\,\sigma^2$.

Because of the assumptions about the effects we have

$a_0 = -a_1$ for each main effect a;

$(ab)_{11} = -(ab)_{10}$

$(ab)_{11} = -(ab)_{01}$

$(ab)_{10} = -(ab)_{00}$

$(ab)_{01} = -(ab)_{00}$ for each interaction.

Hence, $(ab)_{11} = (ab)_{00} = -(ab)_{10} = -(ab)_{01}$.

Therefore the estimates may be derived as follows: Let A^* be the $k \times N$ matrix derived from A by substituting -1 for 0.

Construct the following $m \times N$ matrix X, $m = \binom{k}{2} + k + 1$, where each row of X corresponds to one of the parameters of the model:

The first row of X consists of all 1's, and corresponds to μ,

The next k rows are the rows of A^*, and correspond to the main effects of the k factors, The row corresponding to an interaction (ab) is obtained as follows: let i(a) be the i^{th} element in the row corresponding to a , and let i(b) be the i^{th} element in the row corresponding to b. Then $i(a) \cdot i(b)$ is the i^{th} element in the row corresponding to (ab), $i = 1, \ldots, N$.

Because the array A is of strength 4, A^* has the property that in each four-rowed sub-matrix, each possible column 4-tuple of 1's and -1's appears the same number of times. Thus the matrix X has the property that any two rows are mutually orthogonal, and each row except the first contains Qs^{t-1} 1's and Qs^{t-1} -1's. These rows of X therefore are the coefficients of a set of mutually orthogonal contrasts.

Let Θ be the $1 \times m$ vector

$$(\mu, a_1, b_1, c_1, ---, (ab)_{11}, (ac)_{11}, (bc)_{11}, ---).$$

Then assumption (i) may be stated

$$E(Y) = \Theta X.$$

In minimizing $(Y - \Theta X)'(Y - \Theta X)$ the normal equations, in matrix form, are

$$XX'\Theta' = XY'.$$

The matrix XX' is a diagonal matrix with N's along the diagonal and zeros elsewhere.

Solving the normal equations yields

$$\hat{\Theta}' = (XX')^{-1} XY'.$$

These least-square estimates are

$$\hat{\mu} = \frac{1}{N} \sum_{A^*} y_{ijm---} = y_{\cdot\cdot\cdot\cdot\cdot}$$

$$\left\{ \hat{a}_i = \frac{1}{N} \sum_{i \text{ fixed}} y_{ijm---} - \frac{1}{N} \sum_{-i \text{ fixed}} y_{ijm---} \right.$$

$$\left. = y_{i....} - \hat{\mu} \right\} \qquad \text{For } k \text{ main effects;}$$

$$\left\{ (\widehat{ab})_{ij} = \frac{1}{N} \sum_{i \cdot j \text{ fixed}} y_{ijm---} - \frac{1}{N} \sum_{-(i \cdot j) \text{ fixed}} y_{ijm---} \right.$$

$$\left. = y_{ij...} - y_{i....} - y_{.j...} + \hat{\mu} \right\}$$

for $\binom{k}{2}$ interactions, where i,j,m,-,-, are contained in (-1, 1).

The estimate of an interaction of three or four factors, should it be present, may be found by adding to the X matrix those rows of A^* which correspond to the factors involved. However, in arrays of strength 4, if three-factor interactions are present, two-factor and three-factor interactions may be confounded. Similarly, if a four-factor interaction is present, it may be confounded with a main effect.

Each estimate has variance σ^2/N. If the experiment is replicated m times, each effect or interaction is estimated by the mean of m individual estimates, and thus has variance $\sigma^2/(mN)$.

In a factorial experiment with k factors, each at s levels, and N treatments, each main effect carries $(s - 1)$ degrees of freedom, and each first-order interaction $(s - 1)^2$ degrees of freedom. The number of degrees of freedom for error is thus

$$N - 1 - k(s - 1) - \binom{k}{2} (s - 1)^2.$$

CHAPTER II

THEOREMS ON CONSTRUCTION OF ARRAYS

2.1 Construction of Arrays of Strength t+1 from Arrays of Strength t When $t = 2v$, $s = 2$

Lemma: Let A be an array $(Qs^t, t+1, s, t)$ with $s = 2$, of strength t and $t+1$ constraints. Then any two columns differing in an even number of coordinates will appear the same number of times, and any two columns differing in an odd number of coordinates will appear together a total of Q times.

Proof: Let $(a_1, a_2, \ldots, a_{t+1})'$ be any column $(t+1)$-tuple, where each a_i is 0 or 1. Let a_i^* be 0 if a_i is 1, and 1 if a_i is 0. Let $x(a_1, \ldots, a_{t+1})$ denote the number of times the column $(a_1, \ldots, a_{t+1})'$ appears in A. Since A is of strength t and index Q, for any two coordinates a_i and a_j,

$$x(a_1, \ldots, a_i, \ldots, a_j, \ldots, a_{t+1}) + x(a_1, \ldots, a_i^*, \ldots, a_j, \ldots, a_{t+1}) = Q;$$

$$x(a_1, \ldots, a_i^*, \ldots, a_j, \ldots, a_{t+1}) + x(a_1, \ldots, a_i^*, \ldots, a_j^*, \ldots, a_{t+1}) = Q.$$

Therefore

$$x(a_1, \ldots, a_i, \ldots, a_j, \ldots, a_{t+1}) = x(a_1, \ldots, a_i^*, \ldots, a_j^*, \ldots, a_{t+1}).$$

Hence two columns differing in two coordinates must appear the same number of times, and two columns differing in one coordinate must appear together Q times. By successive applications of this rule, two columns differing in an even number of coordinates muat appear the same number of times, and two columns differing in an odd number of coordinates must appear together Q times.

The following theorem is proved by Seiden in [19]:

> Let S be an ordered set of s elements $e_0, e_1, \ldots, e_{s-1}$. For any integer t consider the s^t different t-tuples of the elements of S. They can be divided into s^{t-1} sets, each consisting of s t-tuples and closed under cyclic permutation of the elements of S. Denote these sets by S_i, $i = 1, 2, \ldots, s^{t-1}$. Suppose that it is possible to find a scheme of r rows with elements belonging to S
>
> $$\begin{matrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{r1} & a_{r2} & \cdots & a_{rn} \end{matrix} \qquad (n = Qs^{t-1})$$
>
> such that in every t-rowed sub-matrix the number of elements belonging to each S_i is the same, say, equal to Q; then one can use this scheme in order to construct an orthogonal array (Qs^t, r, s, t). If in addition this scheme consists of an array of strength $t-1$, then one can construct an orthogonal array $(Qs^t, r+1, s, t)$.

Theorem: Let $A = (Qs^t, k, s, t)$ be an array of strength t, where $t = 2v$ and $s = 2$. Then A may be used to construct an array of strength $t+1$ with $k+1$ constraints, and index Q. If k is the maximum number of constraints possible for the array A, then $k+1$ is the maximum number of constraints possible for the new array.

Proof: Let A' be any sub-matrix of A consisting of $t+1$ rows. Let $(a_1, \ldots, a_{t+1})$ be any $(t+1)$-tuple of the elements $(0,1)$, and let $x(a_1, \ldots, a_{t+1})$ be the number of times $(a_1, \ldots, a_{t+1})$ appears in A'. By the previous lemma, since $t+1$ is odd,

$$x(a_1, \ldots, a_{t+1}) + x(a_1^*, \ldots, a_{t+1}^*) = Q.$$

The array thus forms a scheme satisfying the conditions of the theorem quoted. The second statement of the theorem follows from the inequality proved in Section 1.2.

This theorem shows that in considering the maximum number of constraints possible for orthogonal arrays, for fixed values of Q, it is sufficient to consider strengths $t = 2v$. The maximum number of constraints for strength $t = 2v+1$ will be exactly one more.

The previous theorem has the following application. It frequently happens that after an experiment has been performed, it may seem desirable to include one or more additional factors. Then either (1) an entierly new experiment may be performed and the original information disregarded, or (2) additional treatment combinations may be designed so that information on the additional factors is obtained, while information from the original treatments is preserved. For (2), it must be assumed, of course, that there is no "block effect" of the time lag between parts of the experiment.

Suppose one additional factor is to be included, where each factor is at two levels, and suppose the original design was an orthogonal array (Qs^t,k,s,t), $s = 2$, $t = 2v$. The obvious way to achieve (2) is to consider the original experiment as half of an array $(Qs^{t+1},k+1,s,t)$ and to add the remaining half-array in which the new factor will appear constant at the 1 level [1].

The problem may arise that while in the original experiment there were no $v+1$-factor interactions, the introduction of a new factor makes this assumption questionable.

The previous theorem shows that the additional treatment combinations may be designed so that the augmented experiment is of strength $2v + 1$.

2.2 Structure of Arrays of Strength t with $t+1$ Constraints

Let $A = (2^t, t+1, 2, t)$ be an array of index 1, $t \geq 2$. By a theorem of Bush [9], $t+1$ is the maximum possible number of constraints.

By the lemma of section 2.1, each column differing from the first column in an even number of places must appear once, and there are no columns differing from the first column in an odd number of places. The total number of columns differing in an even number of places is $2^t - 1$, and there are no other columns. This also proves the uniqueness of the array A up to the permutation of the elements 0 and 1.

This proves: there is exactly one array of strength t, with $t+1$ constraints, and index 1, $s=2$, up to a permutation of the elements $(0,1)$.

Hereafter this array will be denoted by D, and it will be assumed that the first column consists entirely of 0's. Then if t is even (odd), each $(t+1)$-tuple with an odd (even) number of 0's is repeated exactly once in the array, where zero is considered even.

Let D' be the array of index one derived from D by permuting the elements $(0,1)$ in any row. Since each column has the number of 0's increased or decreased by one, D' has the following structure: there is no column of all 0's; if t is even (odd), each column of D' has an even (odd) number of 0's. Since each column is distinct, D' must contain each possible column with an even (odd) number of 0's, because the total number of columns is 2^t.

When t is even, interchanging 0 and 1 in one row is equivalent to interchanging 0 and 1 in the entire array.

<u>Example</u>: For the case $t=4$, the two forms of the array $A = (2^4,5,2,4)$ are:

D:

$$
\begin{array}{cccccccccccccccc}
0&0&0&0&0&0&0&1&1&1&1&0&1&1&1&1\\
0&0&0&0&1&1&1&0&0&0&1&1&0&1&1&1\\
0&0&1&1&0&0&1&0&0&1&0&1&1&0&1&1\\
0&1&0&1&0&1&0&0&1&0&0&1&1&1&0&1\\
0&1&1&0&1&0&0&1&0&0&0&1&1&1&1&0
\end{array}
$$

D':

$$
\begin{array}{cccccccccccccccc}
0&0&0&0&0&0&0&1&1&1&1&0&1&1&1&1\\
0&0&0&0&1&1&1&0&0&0&1&1&0&1&1&1\\
0&0&1&1&0&0&1&0&0&1&0&1&1&0&1&1\\
0&1&0&1&0&1&0&0&1&0&0&1&1&1&0&1\\
1&0&0&1&0&1&1&0&1&1&1&0&0&0&0&1
\end{array}
$$

<u>Theorem</u>: Let $A = (Q2^t,t+1,2,t)$ be an array of strength t and $t+1$ constraints. A can be decomposed into Q repetitions of the array $(2^t,t+1,2,t)$, each repetition of the form D or D'.

<u>Proof</u>: It follows from the lemma in section 2.1 that each $(t+1)$-tuple differing from the first column in an even number of coordinates must appear the same number of times, say x, $1 \leq x \leq Q$, forming x arrays of type D. Each of the $(t+1)$-tuples differing by an odd number of coordinates will have to appear $Q - x$ times, forming $Q - x$ arrays of type D'.

Assume that $Q = q2^n$, q odd $(q \geq 1)$. Then any array $A = (Q2^t,t+1,2,t)$ may be extended to one with $t+n+1$ constraints in the following way. We have shown that the array consists of $q2^n$ component arrays, each of which is type D or D'. We take an n-tuple of the elements (0,1) and add it to all the columns of a particular component array. This is done for all the component arrays in such a way that each of the possible 2^n n-tuples has been added to all the columns of q such component arrays.

We show below that this gives $(Q2^t, t+n+1, 2, t)$.

For t rows all belonging to the original array A, each t-tuple appears Q times.

If $n \geq t$, it is clear that for a set of t rows, all of which are among the added rows, each t-tuple must appear with the same frequency Q.

Consider a set of t rows, of which $t-m$ belong to A and m belong to the added rows. For a t-tuple from these rows, let y denote the first $t-m$ elements, which are in A, and let z denote the m elements in the added rows. We note that the $(t-m)$-tuple y occurs exactly 2^m times in each component array of A, and each m-tuple z occurs below $q2^{n-m}$ component arrays. Thus the t-tuple (y,z) occurs $2^m \times q2^{n-m}$ times in the extended array.

It follows from a theorem of Bush [9] that if $Q = q2^n$, an array of type $(Q2^t, t+n+1, 2, t+n)$ can be constructed. However, our method describes a way of extending any $(Q2^t, t+1, 2, t)$ to $(Q2^t, t+n+1, 2, t)$.

Let $A = (Q2^4, 5, 2, 4)$ where Q is odd. If A is the array consisting of Q repetitions of the array D, then A cannot be extended.

Proof: Suppose A is extended to an array A' of six constraints. The columns of A satisfy the equations $n_{i_5} = (Q-1)$, $n_{i_4} = 0$, $n_{i_0}=0$. Then A' must have $n_{i_5} + n_{i_6} = (Q-1)$, $n_{i_0} = 0$. But in A', solving the necessary equations in terms of the dependent variables, we must have:

$$n_{io} = (Q2^4 - 1) - 6(Q2^3 - 1) + 15(Q2^2 - 1) - 20(Q2 - 1)$$

$$+ 15(Q - 1) - n_{i5} - 5n_{i6}$$

$$= 3Q - 5 - n_{i5} - 5n_{i6} \text{ .}$$

Thus $n_{io} = 0$ implies $n_{i5} + 5n_{i6} = 3Q - 5$.

The system of equations

$$n_{i5} + n_{i6} = Q - 1$$

$$n_{i5} + 5n_{i6} = 3Q - 5$$

has the solution $n_{i6} = Q/2 - 1$. There is no integer solution if Q is odd.

CHAPTER III

GROUP ARRAYS

3.1 Relation of Orthogonal Arrays to Finite Projective Geometry

Consider a factorial design for k factors, each occurring at s levels, where $s = p^n$, p a prime. The s levels may be represented by the elements of the Galois Field $GF(p^n)$. Then each treatment vector $(a_1,\ldots,a_k)$ may be taken as a point of a finite projective geometry of dimension $(k-1)$. This geometry will be denoted by $PG(k-1,p^n)$.

In this setting, R. C. Bose, in his definitive paper of 1947 [5], made a study of the problem of confounding in the blocks of a factorial design (see Sec. 1.1).

A complete symmetrical factorial design s^k is said to be of the class (s^k, s^m) if it is partitioned into s^m blocks $(m \leq k)$ of s^{k-m} treatments each. The block size is s^r, where $r = k-m$.

Bose shows that if $m_t(r,s)$ denotes the maximum number of factors that it is possible to accommodate in a symmetrical factorial experiment in which each factor is at $s = p^n$ levels, and each block of size s^r, so that no estimate of an interaction involving t or fewer factors is confounded with blocks, then $m_t(r,s)$ equals the maximum number of rows it is possible to take in a matrix of r columns of elements of $GF(p^n)$ so that no t rows are dependent.

Equivalently, $m_t(r,s)$ is given by the maximum number of points it is possible to choose in $PG(r-1,s)$ so that no t are conjoint (t points are said to be conjoint if they lie in the same $t-2$ dimensional subspace).

It was pointed out in Section 1.2 that a block of a complete factorial design of strength t is an orthogonal array of strength t. Thus, for those arrays which are blocks of a complete factorial design, Bose has made the problem of determining the maximum number of constraints a geometric problem.

The following theorem of Bose and Bush [7] gives explicitly a method of constructing orthogonal arrays of strength t.

Theorem: Let C be a matrix of k rows and r columns, such that every $t \times r$ sub-matrix is of rank t, with entries from a Galois Field $GF(p^n)$. Then an orthogonal array (s^r, k, s, t) may be constructed, where $s = p^n$.

Proof: Let B_r be the $r \times s^r$ matrix whose columns are the s^r different r-tuples of elements of $GF(p^n)$. (B_r is the complete s^r array). The $k \times s^r$ matrix $A = C \times B_r$ is the required array.

If A' is a $t \times s^r$ sub-matrix of A, and C' is the corresponding sub-matrix of C of rank t, each column $(a_1, \ldots, a_t)'$ of A' is obtained from s^{r-t} different columns of B. The array A obtained is of index s^{r-t}.

3.2 Bounds on the Number of Constraints of Arrays of Strength Four of Type $C \times B_r$

Following the notation of the previous section, let B_r be the $r \times s^r$ matrix whose columns are the s^r different r-tuples of elements of $GF(s)$, where $s = p^n$. Let C be a $k \times r$ matrix with entries from $GF(s)$, such that every $t \times r$ sub-matrix is of rank t. It follows from the theorems of Bose (Sec. 3.1) that the maximum number of constraints possible in an array (s^r, k, s, t) of the form $C \times B_r$ is equal to the maximum number of points in $PG(r-1,s)$ such that no t are conjoint. This number is denoted by $m_t(r,s)$.

When $t = 2$, all distinct points of $PG(r-1,s)$ can be chosen, so that $m_2(r,s) = (s^r - 1)/(s - 1)$. As we have indicated (Sec. 1.2), this is the maximum number of constraints possible in any array of strength 2 and s^r columns.

When $t = 3$, $s = 2$, $m_3(r,2) = 2^{r-1}$. This has been shown by Bose [5] to be the maximum number of points in $PG(r-1,2)$ such that no three are colinear. Bose and Bush [7] have also shown this to be the maximum number of constraints possible in any array of strength 3, and 2^r columns, when $s = 2$.

Bose also proves the following [5]:

$$m_3(3,s) = s + 2 \quad \text{when } s = 2^n;$$

$$m_3(3,s) = s + 1 \quad \text{when } s = p^n, p \text{ an odd prime;}$$

$$m_3(4,s) = s^2 + 1 \quad \text{when } s = p^n, p \text{ an odd prime.}$$

Seiden [24] proves $m_3(4,3) = s^2 + 1$ for the case $s = 2^2$. Qvist [25] proves $m_3(4,s) = s^2 + 1$, $s = 2^n$, $n > 1$. In particular, $s^2 + 1$ is the maximum number of constraints possible in any array when $s = 3$, $t = 3$, $N = 3^4$, i.e., (81, 10, 3, 3) [20].

A bound for $m_4(r,2)$, $r \geq 8$, and the values of $m_4(r,2)$ for $r = 4,5,6,7$, and 8 are presented in [24]. We may assume in each case that the coordinates of the first r points chosen form the rows of the $r \times r$ identity matrix I_r.

(1) In $PG(3,2)$, $m_4(4,2) = 5$.

The only point which can be added to I_4 is (1,1,1,1).

(2) In $PG(4,2)$, $m_4(5,2) = 6$.

Again, only one point may be added to I_5. This point must have either five coordinates equal to 1, or four 1's and one 0.

(3) In PG(5,2), $m_4(6,2) = 8$.

If the point (1,1,1,1,1,1) is added to I_6, no more points may be added. There are essentially two different sets of two points which may be added:

1 1 1 1 1 0	1 1 1 1 0 0
1 1 1 0 0 1	1 1 0 0 1 1

(4) In PG(6,2), $m_4(7,2) = 11$.

Nine possible sets of four points are exhibited in [21] which may be added to I_7, such that no four are dependent. In each case, the set of 11 points contains a subset of 8 points lying in the same 5 dimensional subspace.

(5) In PG(r-1,2), $r \geq 8$, $m_4(r,2) \leq 3(2^{r-6} -1) + 8$.

This bound follows from the fact that a 5-dimensional subspace in PG(r-1,2) passes through $(2^{r-6} -1)$ 6-dimensional subspaces. To the 8 points in the 5-dimensional subspace, 3 points may be added in each 6-dimensional subspace.

(6) In PG(7,2), $m_4(8,2) = 17$.

The bound in (5) is 17 when $r = 8$. Seventeen points have been found in PG(7,2) such that no four are dependent.

Using a similar method, we can show that the maximum number of points, no four in a plane, in PG(r-1,3) is at most $3(3^{r-4} -1) + 5$, for $r \geq 6$.

<u>Lemma 1</u>: The maximum number of points, no 4 in a plane, in PG(3,3) is 5.

Proof: Any point added to I_4 must be a linear combination, with non-zero coefficients, of the four basis points. Two such points can differ in at most two coordinates, or be alike in one coordinate; hence one point may be written as a linear combination of the other point and two basis points. Then exactly one point may be added. Therefore, $m_4(4,3) = 5$.

Lemma 2: The maximum number of points, no 4 in a plane, in PG(4,3) is 11.

Proof: There are 5 different sets of four basis points among the points of I_5. By the previous lemma, at most one point may be added for each such set of four basis points. At most one point can be added with five non-zero coordinates. Two such points can differ in at most two places, and can be alike in at most two places, which is impossible. Thus at most 6 points can be added to the basis points.

The following set of 11 points of PG(4,3), with no four in a plane, is the maximum number which may be found. Therefore, $m_4(5,3) = 11$.

$$\begin{matrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 2 & 0 \\ 1 & 2 & 2 & 0 & 1 \\ 1 & 1 & 0 & 2 & 2 \\ 1 & 0 & 2 & 1 & 2 \\ 0 & 1 & 2 & 2 & 1 \end{matrix} \qquad \begin{matrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{matrix}$$

This set of points is unique, up to an interchange of rows or columns, or multiplication of a column by a non-zero element. Let us assume that the point with five non-zero coordinates has a 1 in each place. Assume also that the other points added to I_5, with four non-zero coordinates, each have a 1 as the leading non-zero coordinate. There must then be two 2's and one 1 among the other non-zero coordinates. In each PG(3,3), therefore, there is a choice of three points: (1 2 1 2), (1 2 2 1), or (1 1 2 2). In addition, any pair of these points chosen must differ in one or two of the places where both are non-zero. There are then five additional ways of selecting one point from each PG(3,3). The six sets shown may be obtained from one another by interchanging columns.

1 2 1 2 0	1 2 2 1 0	1 2 2 1 0	1 1 2 2 0	1 1 2 2 0
1 1 2 0 2	1 2 1 0 2	1 1 2 0 2	1 2 2 0 1	1 2 1 0 2
1 2 0 1 2	1 1 0 2 2	1 2 0 2 1	1 2 0 1 2	1 2 0 2 1
1 0 2 2 1	1 0 2 2 1	1 0 1 2 2	1 0 1 2 2	1 0 2 1 2
0 1 1 2 2	0 1 2 1 2	0 1 2 2 1	0 1 2 1 2	0 1 1 2 2

<u>Theorem:</u> The maximum number of points, no four in a plane, in PG(r-1,3), is at most $3(3^{r-4}-1)+5$, $r \geq 6$.

<u>Proof:</u> Clearly any set of 11 points in a PG(4,3) contains five points lying in the same PG(3,3). A three-dimensional space in PG(r-1,3) is contained in exactly $(1/2)(3^{r-4}-1)$ four-dimensional spaces. Five points may be selected in the three-dimensional space, and at most six additional points may be selected in each four-dimensional space. Therefore

$$m_4(r,3) \leq 6[(1/2)(3^{r-4}-1)]+5.$$

3.3 <u>Structure of Arrays of Type C x B_r</u>

The theorem of Bose and Bush in Sec. 3.1 describes a method of constructing arrays by matrix multiplication. Let us consider the $r \times s^r$ matrix B_r used in the construction of the array A = (s^r,k,s,t). The s levels are regarded as members of the Galois Field GF(s), and the columns of B_r are the s^r r-tuples of elements of GF(s).

We may consider the columns of B_r as the elements of an r-dimensional vector space. In particular, the columns form an Abelian group under coordinate-wise addition(modulo s). Let $A = C \times B_r$ be the array constructed by the method of Bose and Bush. It follows that, if repeated elements are identified, the columns of A form a subgroup of B_k, the group of k-tuples of elements of GF(s). In addition, if A is an array (s^r,k,s,t), $k > r$, then so is each coset of A in B_k.

To show that A is a group, consider any two columns h_i,h_j of A. By the construction of A (see Sec. 3.1), $h_i = Cb_i$ and $h_j = Cb_j$ for two columns b_i, b_j of B_r. Hence,

$$\begin{aligned} h_i + h_j &= Cb_i + Cb_j \\ &= C(b_i + b_j) \\ &= Cb_m \end{aligned}$$

for some column b_m in B_r. Similarly, $-h_i = -Cb_i = C(-b_i)$. Therefore the columns of A form a group.

We will now show that every coset of A in B_k is also an array of strength t. Let A' be any $t \times s^r$ sub-matrix of A. Since A is of index s^{r-t}, the s^r columns of A' contain s^{r-t} repetitions of each t-tuple of elements of GF(s). A' thus consists of s^{r-t} repetitions of the group B_t.

Let $(a_1,\ldots,a_k)'$ be a k-tuple of B_k which is not in A. The addition of $(a_1,\ldots,a_k)'$ to each column of A leaves the group B_t in the columns of A' invariant.

It is also clear that if n_{ij} is the number of columns in A having j coincidences with the i^{th} column, then any array which is a coset of A has the same parameters $n_{io},\ldots,n_{ik}$, $i = 1,\ldots,N$, providing that the transformed columns of the coset remain in the same order as the original columns of A. For suppose the element a_i of $GF(s)$ is added to each member of the i^{th} row of A. Then only those columns which had matching elements in the i^{th} row of A will have matching elements in the i^{th} row of the new array.

We have thus shown that each array $A = (s^r,k,s,t)$, $k \geq r$, of the form $C \times B_r$, constructed from a $k \times r$ matrix C, is a subgroup of B_k which has $s^{k-r}-1$ coset arrays. A and its cosets in turn form a group of arrays of order s^{k-r} under the usual definition of coset operations. If the s^r columns of A are distinct, then the s^{k-r} elements of this group are the blocks of a class (s^k,s^{k-r}) design of strength t.

We will now show that in any array of the form $C \times B_r$, the number of times each column is repeated depends on the rank of the matrix C.

If C is of rank v, $v \leq r$, each column which appears in A is repeated exactly s^{r-v} times.

Suppose the first v rows of C are linearly independent. Without loss of generality assume the i^{th} row of C consists of the vector with 1 in the i^{th} place and 0's elsewhere, $i = 1,2,..,v$. Then the first v rows of A are identical with the first v rows of the matrix formed by B_r, the group of s^r column vectors. Hence each column v-tuple is repeated s^{r-v} times. Each succeeding row of A is a linear combination of the first v rows; therefore columns will be identical if and only if they have matching elements in the first v rows.

It follows from this proof that the array $A = C \times B_r$ will have s^r distinct columns if and only if C is of rank r. If A is to be of strength t, then the rank of C must be at least t.

We have shown previously that if A is an array of the form $C \times B_r$, the columns of A form a group, if repeated columns are identified. The theorem which follows characterizes a larger class of arrays whose columns form groups, with identification of repeated columns. Conditions will then be given under which arrays of this class are of the form $C \times B_r$.

<u>Theorem</u>: Let $A = (N,k,s,t)$ be an array with exactly s^r distinct columns, $s^r \leq N$. Then, with identification of repeated columns, the columns of A form a group if and only if A is a matrix of rank r.

<u>Proof</u>: Let A be of rank r. A set of r linearly independent columns may be selected, and every other column of the array must be a linear combination of the r linearly independent columns. Since A contains s^r distinct columns, it must contain all linear combinations of the r independent columns. Then the columns of A form the group generated by the r linearly independent columns.

Suppose A is a group, and let the rank of A be v. A set of v linearly independent columns may be chosen, and every other column must be a linear combination of the v linearly independent columns. Furthermore, since A is a group, it must contain every linear combination of the v independent columns, therefore the number of distinct columns is s^v. It follows that $s^r = s^v$, and since $s \geq 2$, $r = v$.

If A is of rank r, then r rows, say the first r, are linearly independent, and the elements of the last $k-r$ rows are linear combinations of the elements of the first r rows. Thus if A has s^r distinct columns, there are s^r distinct columns in the first r rows. It follows that the first r rows must consist of the group B_r, with possible repetition of columns. We will now show that if each column of A is repeated s^v times, then A is of the form $C \times B_m$ for some $m \geq r$.

Suppose A is of rank r, with $N = s^m$, $m \geq r$, and each column is repeated the same number of times, say s^v, where $v = m-r$. It has been shown that r rows of A, say the first r, consist of the group B_r. Since each column of A is repeated s^v times, it follows that in the first r rows of A, each column of B_r is repeated s^v times. The first r rows of A then consist of r rows of the matrix formed by the column group B_m. We now show that the array A is of the form $C \times B_m$, where C is a $k \times m$ matrix.

Let the first r rows of the matrix C consist of the unit vectors with 1 in the i^{th} place and 0 elsewhere, $i = 1,\ldots,r$. The last $k-r$ rows of the array are linear combinations of the first r rows. Let $(a_{ji},\ldots,a_{jr})$, $j = r+1,\ldots,k$, be the coefficients of these linear combinations. Then the last $k-r$ rows of C may have the $(a_{j1},\ldots,a_{jr})$ as the first r elements, and 0's elsewhere.

The first r rows of A may be written as the product of the first r rows of C with the matrix B_m. Let $(a_{j1},\ldots,a_{jr},0,\ldots,0)$ be one of the remaining $k-r$ rows of C. The a_{ji} are the coefficients of the appropriate linear combination of the first r rows which yields the j^{th} row of A, $j = r+1,\ldots,k$.

We will demonstrate in Chapter IV that there exist arrays $(2^r,k,2,t)$, $k > r$, containing 2^r distinct columns, which are neither subgroups of B_k nor cosets of subgroups. Thus they cannot be derived from construction of the form $C \times B_r$. A method of construction is described which yields such arrays, and examples will be given.

3.4 Confounding in Arrays of the Form $C \times B_r$

In any array $A = (2^r,k,2,t)$ of the form $C \times B_r$, the pattern of confounding may immediately be determined by the structure of the matrix C used to generate the array.

We will assume the rank of C is r. We may take as the first r rows of C the identity matrix I_r. The array constructed will then be a complete 2^r factorial design for the first r factors $F_1,\ldots,F_r$. Let the last $k-r$ rows of C correspond to the factors $F_{r+1},\ldots,F_k$. Among these $k-r$ rows, the structure of the i^{th} row indicates the interaction with which the estimate of the i^{th} effect is confounded, for $i = r+1,\ldots,k$.

Let $(c_1,\ldots,c_r)$, $c_i = 0$ or 1, be one of the $k-r$ rows added to I_r to form C. Then the estimate of the effect corresponding to $(c_1,\ldots,c_r)$ is confounded with the estimate of the interaction of those factors F_i for which $c_i = 1$, in $(c_1,\ldots,c_r)$. Then, using "identities" derived from the confounding with respect to factors $F_{r+1},\ldots,F_k$, all other confounding in the array may be determined.

For example, if the $(r+1)^{th}$ row of C is of the form (1,1,1,1, 0,0,...,0), the estimate of the effect of F_{r+1} is confounded with the estimate of the interaction of F_1, F_2, F_3, F_4, denoted as 1234, if such an interaction should be present. The "identity" in this case is expressed as $r+1 = \pm\ 1234$, or $I = \pm\ 1234(r+1)$. The method follows that of Box and Hunter [8].

The parity of the confounding may be determined as follows: if the row of C corresponding to F_i has an even number of 1's, the confounding is negative; if the row has an odd number of 1's, the confounding is positive. (See Sec. 1.4.)

An example may be given for the case of the array A = (128,11,2,4) constructed with a matrix C taken from [21]. Let C be the following matrix:

$$\begin{bmatrix} & & & I_7 & & & \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

The identities derived from the last four rows of C are

$$I = -\ 1234568$$

$$I = \ 123479$$

$$I = \ 12567\ 10$$

$$I = -\ 2467\ 11.$$

These identities may now be combined, using the order two property $xx = I$, so that, for example, (123479)(12567 10) = 34569 10. This yields the additional identities

I = - 56789

I = - 3478 10

I = 1378 11

I = 34569 10

I = - 1369 11

I = - 145 10 11

I = - 1289 10

I = 24589 11

I = 2368 10 11

I = - 23579 10 11

I = 146789 10 11

From these identities, we can determine which interactions are confounded. For example, from I = 12567 10 , it follows that

1 = 2567 10

12 = 567 10

125 = 67 10,

etc.

It is easily seen that since each identity contains at least five factors, no main effect is confounded with any interaction of three of fewer factors, and no interactions of two factors are confounded with one another. This is the property of arrays of strength 4, as described in Sec. 1.2.

CHAPTER IV

ARRAYS OF 32, 48, 64 and 80 COLUMNS, $s = 2$, $t = 4$

4.1 Arrays of 32 Columns, $s = 2$, $t = 4$

Theorem: The maximum number of constraints for an array $(32, k, 2, 4)$ is six.

Proof: Consider the equations whose solution is necessary for the existence of an array with 7 constraints (see Sec. 1.3). Solving them in terms of the dependent variables yields

$$n_{i0} = 3 - n_{i5} - 5n_{i6} - 15\, n_{i7}$$

$$n_{i3} = -35 + 10\, n_{i5} + 40\, n_{i6} + 105\, n_{i7}$$

According to the first equation, $n_{i6} = n_{i7} = 0$, and $n_{i5} \leq 3$. There is then no non-negative solution for n_{i3} , and thus an array of 7 constraints cannot exist.

Using the following method, two different arrays $(32, 6, 2, 4)$ may be constructed. We may assume that the first column must consist entirely of 0's. Let the first five rows consist of two arrays $(16, 5, 2, 4)$ of the form D or D', as described in Section 2.2. The columns of D are made up of all five-tuples with an odd number of 0's, and the columns of D' are made up of all five-tuples with an even number of 0's. A row of 0's may be added to one array D , and a row of 1's to the other array. The two arrays constructed are:

(1)

$$\begin{bmatrix} D & D \\ 0 & 1 \end{bmatrix}$$

(2)

$$\begin{bmatrix} D & D' \\ 0 & 1 \end{bmatrix}$$

The number i below a D or D' indicates that a row of sixteen i's is to be added to the five-rowed array.

Array (1) has solution

$n_{10} = 0$, $n_{11} = 5$, $n_{12} = 5$, $n_{13} = 10$, $n_{14} = 10$, $n_{15} = 1$, $n_{16} = 0$.

Array (2) has solution

$n_{10} = 1$, $n_{11} = 0$, $n_{12} = 15$, $n_{13} = 0$, $n_{14} = 15$, $n_{15} = 0$, $n_{16} = 0$.

These are the only solutions possible to the necessary equations for an array (32, 6, 2, 4).

4.2 Arrays of 48 Columns, s = 2, t = 4

Theorem: The maximum number of constraints for an array (48, k, 2, 4) is five.

Proof: The necessary equations of Section 1.3 may be solved in terms of the dependent variables for the case $k = 6$. We then have

$$n_{i0} = 4 - 5n_{i6} - n_{i5}.$$

Any solution must have $n_{i6} = 0$ for all i. It follows that if an array of six constraints exists, all columns are distinct.

Consider the complete array (2^6, 6, 2, 4) consisting of all 2^6 distinct 6-tuples. In any four rows, each 4-tuple appears exactly four times. If 48 distinct columns can be chosen to form an array of index 3, then in any four rows of the remaining 16 columns, each 4-tuple must appear exactly once. This would yield an array of 16 columns with $k = 6$, contradicting the fact that an array (16, k, 2, 4) can have at most five constraints [9].

It follows from the theorem of section 2.2 that every array (48, 5, 2, 4) is composed of three arrays of the form D or D'.

4.3 Arrays of 64 Columns, s = 2, t = 4

We will prove in this section that the maximum number of constraints for an array (64, k, 2, 4) is eight.

When k = 7 , no solution to the necessary equations exists with $n_{i7} > 0$. It follows that $n_{i7} = 0$ for all k greater than 7, by Property 3 of section 1.3. The following table gives all solutions which must be considered for k = 6, 7, 8 and 9

Solutions for k = 6, 7, 8, 9; N = 64.

	n_{i0}	n_{i1}	n_{i2}	n_{i3}	n_{i4}	n_{i5}	n_{i6}	n_{i7}	n_{i8}	n_{i9}
6-1:	0	10	10	20	20	2	1			
6-2:	1	5	20	10	25	1	1			
6-3:	2	0	30	0	30	0	1			
6-4:	0	11	5	30	10	7	0			
6-5:	1	6	15	20	15	6	0			
6-6:	2	1	25	10	20	5	0			
7-1:	0	4	15	5	30	6	3	0		
7-2:	1	0	20	5	25	10	2	0		
7-3:	0	5	10	15	20	11	2	0		
7-4:	0	6	5	25	10	16	1	0		
7-5:	1	1	15	15	15	15	1	0		
7-6:	0	7	0	35	0	21	0	0		
7-7:	1	2	10	25	5	20	0	0		
8-1:	0	2	9	12	15	18	7	0	0	
8-2:	0	1	14	2	25	13	8	0	0	
8-3:	0	3	4	22	5	23	6	0	0	
9-1:	0	0	9	6	18	9	21	0	0	0
9-2:	0	1	4	16	8	14	20	0	0	0

Theorem: Every array (64, 8, 2, 4) has solution 8-1.

Proof: We assume the first column of an array consists entirely of 0's.

Solution 8-2 has $n_{i1} = 1$. Let $i = 1$. Suppose the single 0 in this column is coincident with a 0 in any of the columns with two 0's. This would result in two columns with seven coincidences, contradicting the fact that if $k = 8$, $n_{i7} = 0$. Thus the single 0 in the column with one 0 cannot be coincident with a 0 in any column with two 0's. The array then has a sub-array of seven rows with $n_{10} = 1$, $n_{11} = 0$. This sub-array would have solution 7-2 with respect to the first column, with $n_{12} = 20$. Of the columns with two 0's in the sub-array, six would have to have 0's added in the extension. But in solution 8-2, $n_{i3} = 2$. Because of this contradiction, no array can exist with solution 8-2.

Suppose an array exists with solution 8-3. This solution has $n_{i1} = 3$, and the columns having one coincidence with the i^{th} column must be distinct. Thus the array would contain a sub-array of seven rows with $n_{i0} = 1$, $n_{i1} \geq 2$. The only such solution is 7-7 with $n_{i0} = 1$, $n_{i1} = 2$. But in 8-3, $n_{i5} + n_{i6} + n_{i7} + n_{i8} = 29$. In any seven-rowed sub-array, these columns have four or more coincidences with the i^{th} column. However, in 7-7, $n_{i4} + n_{i5} + n_{i6} + n_{i7} = 25$. Hence an array with solution 7-7 cannot be extended to an array with solution 8-3. Because of this contradiction, no array can exist with solution 8-3.

This leaves 8-1 as the only solution for an array of eight constraints.

Theorem: The maximum number of constraints for an array (64, k, 2, 4) is eight.

Proof: We first prove that solution 9-2 cannot correspond to an array. Such an array would have one column having one coincidence with the i^{th} column, since $n_{i_1} = 1$. It would then contain a sub-array of eight constraints with $n_{i_0} \geq 1$. No such array exists.

An array with solution 8-1 cannot be extended to an array with solution 9-1. Consider the solutions with respect to the first column of all 0's. Solution 8-1 has $n_{i_1} = 2$. These two columns with one 0 have six coincidences. Since solution 9-1 has $n_{i_1} = 0$, each of the two columns must have a 0 in the ninth row. The two columns would then have seven coincidences, contradicting the fact that $n_{i_7} = 0$ for all i.

Theorem: Arrays (64, k, 2, 4) of seven constraints with solutions 7-1, 7-2, 7-4, 7-6, and 7-7 cannot be extended to eight constraints.

Proof: We consider the solutions with respect to the first column of 0's.

(i) Suppose an array has solution 7-1. Since $n_{i_6} = 3$, we may assume there are four columns as in (a).

(a)

$$\begin{matrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{matrix}$$

(b)

$$\begin{matrix} - & - & - & - & - & - \\ - & - & - & - & - & - \\ - & - & - & - & - & - \\ - & - & - & - & - & - \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{matrix}$$

Suppose one of the last three rows of (a) is deleted. In each case the remaining six rows have $n_{16} = 1$ and $n_{15} = 2$, and hence must form an array of six constraints with solution 6-1.

The array of seven constraints with solution 7-1 must now have six columns consisting of five 0's and two 1's, since $n_{i5} = 6$. However, deleting one of the last three rows must leave each of these columns with only four 0's. Therefore, in the last three rows, these six columns have only 0's, as in (b).

Suppose the array is extended to eight rows. Four of the columns in (b) must coincide with the first column in the eighth row, since solution 8-1 has $n_{i6} = 7$. We would then have the 4-tuple (0 0 0 0) appearing five times in the last four rows. Thus the extension is impossible.

(ii) Suppose an array with solution 7-2 is extended to eight rows. According to 7-2, the extension must have $n_{i0} + n_{i1} = 1$. Since the unique solution for $k = 8$ has $n_{i0} + n_{i1} = 2$, this is impossible.

(iii) Consider an array with solution 7-4. The array must have six columns with exactly one 0. If an eighth row is added, these columns will be followed by two 1's and four 0's, as in (c).

(c)

$$\begin{matrix} 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \end{matrix}$$

(d)

$$\begin{matrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ - & - & - & - & - \\ - & - & - & - & - \\ - & - & - & - & - \\ - & - & - & - & - \\ - & - & - & - & - \\ - & - & - & - & - \end{matrix}$$

The eight-rowed array must now have five more columns with exactly two 0's. All of these columns must have 1's in the first two rows as in (d); otherwise, there would be a column having seven coincidences with at least one of the first two columns.

If any one of rows 3 through 8 is deleted, the remaining seven-rowed array would have $n_{10} = 0$, and hence $n_{11} \geq 5$. This implies that if any one of rows 3 through 6 is deleted, then at least two of the columns of (d) must have only one 0 in the remaining rows, while if row 7 is deleted, at least three of the columns of (d) must have only one 0 in the remaining rows. Therefore the five columns of (d), with two 0's in each column, must have at least $(4)(2) + 3 = 11$ 0's. This is clearly a contradiction.

(iv) An array with solution 7-6 can be extended only to an array of eight constraints with $n_{i5} + n_{i6} = 21$. Solution 8-1 has $n_{i5} + n_{i6} = 25$.

(v) Suppose an array with solution 7-7 is extended to eight rows. Solution 7-7 has $n_{i0} = 1$, $n_{i1} = 2$. Let $i = 1$. The column with no 0, and one of the columns with one 0, must both have a 0 in the eighth row. This would result in two columns of the eight-rowed array having seven coincidences, while solution 8-1 has $n_{i7} = 0$.

Construction of Arrays with Seven and Eight Constraints

An array (64, 8, 2, 4) of the form $C \times B_6$ may be constructed by the method of matrix multiplication, as described in section 3.1.

To form the rows of the matrix C, there are essentially two different sets of eight points which may be chosen in PG(5,2) such that no four are conjoint. We may assume in each case that the first six rows of C form the 6×6 identity matrix I_6. The matrices C which result are:

$$C_1: \quad \begin{bmatrix} & & I_6 & & & \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 \end{bmatrix}$$

$$C_2: \quad \begin{bmatrix} & & I_6 & & & \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

In addition, the following set of seven points may be chosen, which cannot be extended to eight points:

$$C_3: \quad \begin{bmatrix} & & I_6 & & & \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

The array $C_3 \times B_6$, of seven constraints, has solution 7-6 , and thus cannot be extended to eight constraints by any method.

If a seven-rowed matrix C is used which consists of I_6 and a row with four non-zero coordinates, for example the first seven rows of C_1 , the array formed will have solution 7-3. If the seventh row has five non-zero coordinates, as in the first seven rows of C_2, the array will have solution 7-5.

We may also construct arrays of seven constraints by the method described in section 2.2. Since $Q = 2^2$, any array of five constraints can be extended by this method to seven constraints.

Any array (64, 5, 2, 4) is composed of four arrays of the form D or D' . It may be extended by adding one of the pairs (0 0), (0 1), (1 0), (1 1) to the columns of one D or D' , the same pair being added to all the columns of a single D or D'.

We may assume that in each array constructed by this method, there is a five-rowed array of the form D followed by (0 0), so that the seven-rowed array has a column of all 0's. Using all possible arrangements of the remaining five-rowed component arrays, we may construct six different arrays by this method. These six arrays are displayed below, where

$$\begin{matrix} D \\ i \\ j \end{matrix}$$

indicates that the pair (i j) is to be added to all the columns of the array D (or D').

(1) $\begin{bmatrix} D & D & D & D \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}$ Solution 7-3

(2) $\begin{bmatrix} D & D & D & D' \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}$ Solution 7-2

(3) $\begin{bmatrix} D & D & D' & D \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}$ Solution 7-4

(4) $\begin{bmatrix} D & D' & D & D' \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}$ Solution 7-5

(5) $\begin{bmatrix} D & D' & D' & D \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}$ Solution 7-6

(6) $\begin{bmatrix} D & D' & D' & D' \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}$ Solution 7-7

This method thus yields seven-rowed arrays with each possible solution except 7-1. An array with solution 7-1 may be constructed in the following form.

(7) The first 32 columns consist of

$$\begin{bmatrix} D & D \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$$

while the last 32 columns are:

(i)	(ii)
0 0 0 0 1 1 1 1 0 0 0 0 1 1 1 1	0 0 0 0 1 1 1 1 0 0 0 0 1 1 1 1
0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1	0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1
0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1	0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1
0 1 1 0 1 0 0 1 0 1 1 0 1 0 0 1	1 0 0 1 0 1 1 0 1 0 0 1 0 1 1 0
0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

One can easily verify that (7) is an orthogonal array of strength 4. We note first that rows 1 through 5 form a five-rowed array of type $D\ D\ D\ D'$. We must now show that if we take either (a) two of the first five rows, together with rows 6 and 7, or (b) three of the first five rows together with either row 6 or row 7, each 4-tuple appears four times.

Consider (a). In any two rows of D, each possible pair of the elements 0,1 appears four times. Thus in the first 32 columns of array (7), each 4-tuple ending in (0 1) or (1 0) appears four times. In the last 32 columns, the 16 columns of (i) end in (0 0), while the 16 columns of (ii) end in (1 1). It is now sufficient to check that in any two of the first five rows, each possible pair appears four times in (i) and four times in (ii). It should be noted that rows

1, 2, and 3 of (i) and (ii) are exactly alike, while the remaining four rows are the same except for an interchange of 0 and 1.

Next consider (b). In the first 32 columns, rows 1 through 5 , together with either row 6 or row 7 , clearly form an array (32, 6, 2, 4). It remains to be shown that the same property is true for the last 32 columns. It is therefore sufficient to check that in any three of the first five rows, each possible 3-tuple appears twice in (i) and twice in (ii).

Two of the seven-rowed arrays shown above, arrays (4) and (5), with solutions 7-5 and 7-6 , are actually arrays of strength five. Using the method of the theorem of section 2.1, they can be constructed from arrays of the form (32, 6, 2, 4). We may choose one of the arrays (32, 6, 2, 4) displayed in section 4.1, adjoin the same array with 0 and 1 interchanged, and add a seventh row of 0's and 1's , as indicated in section 2.1. By this method, array (1) of section 4.1 will yield the array (4) shown above, and array (2) of section 4.1 will yield (5). It follows also, from the same theorem of section 2.1, that the maximum number of constraints for an array (64, k, 2, 5) is seven.

We will now examine the algebraic properties of the seven-rowed arrays constructed, considering 0 and 1 as members of the Galois Field GF(2).

First consider the five-rowed arrays D and D'. D consists of all columns with an odd number of 0's and an even number of 1's; the columns of D' are those with an even number of 0's and an odd number of 1's. We thus have the following property: using coordinate addition (modulo 2), a column of D added to a column of D' yields a column of D' , while the sum of two columns both from either D or D' , is a column of D. Using this property, it is easily seen that the columns of arrays (1), (4), and (5) have closure, and thus form subgroups of B_7 , the group of all possible seven-tuples of the elements 0, 1.

Using the same property, it can also be shown that the columns of arrays (2), (3), (6), and (7) do not have closure. In each of these arrays, consider the two component arrays D or D' which are followed by (0 1) and (1 0). The sum of two columns, one ending in (0 1) and one ending in (1 0), will clearly be a column ending in (1 1). To have closure, arrays (2) and (6) would have to have an array D followed by (1 1), while array (3) would have to have D' followed by (1 1). These requirements are not met. In array (7), we note that part (i) contains columns of the D' type followed by (0 0). The sum of such a column, and any column chosen from the first 32 columns of the array, will not be a column of the array.

Since each of the arrays constructed contains the identity column, arrays (2), (3), (6) and (7) are neither groups, nor cosets of groups, and thus cannot be derived from the method of construction discussed in Chapter III, using matrix multiplication. These four arrays have solutions 7-2, 7-4, 7-7, and 7-1 respectively, hence none of them can be extended to eight rows.

In section 3.3 it was proved that an array with s^r distinct columns forms a group if and only if the array is a matrix of rank r. It can easily be shown that the four arrays discussed in the previous paragraph form matrices of rank seven, while for $N = 64$, $r = 6$. Arrays (2), (3) and (6) contain the five columns:

$$\begin{matrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{matrix}$$

In addition, (2) contains (0 0 0 0 0 1 0) and (0 0 0 0 0 0 1); (3) contains (0 0 0 0 0 1 1) and (0 0 0 0 0 0 1); and (6) contains (0 0 0 0 1 1 0) and (0 0 0 0 1 0 1). In (7) we may choose the following set of independent columns:

$$\begin{matrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 1 \end{matrix}$$

The group generated by a set of seven indepndent columns contains 2^7 distinct columns, while the arrays have only 2^6 columns, indicating once again that these arrays cannot form groups.

4.4 Arrays of 80 Columns, $s = 2$, $t = 4$

In this section it will be proved that the maximum number of constraints for an array (80, k, 2, 4) is six.

The solutions to the necessary equations for an array with $N = 80$, for $k = 5$ and $k = 6$, are given in the table on the next page.

It will first be shown that only two of the arrays with five constraints can be extended to arrays with six constraints.

Solutions for $N = 80$, $k = 5$, $k = 6$.

	n_{i0}	n_{i1}	n_{i2}	n_{i3}	n_{i4}	n_{i5}	n_{i6}
5-1:	0	25	0	50	0	4	
5-2:	1	20	10	40	5	3	
5-3:	2	15	20	30	10	2	
5-4:	3	10	30	20	15	1	
5-5:	4	5	40	10	20	0	
6-1:	0	12	15	20	30	0	2
6-2:	0	13	10	30	20	5	1
6-3:	1	8	20	20	25	4	1
6-4:	2	3	30	10	30	3	1
6-5:	0	14	5	40	10	10	0
6-6:	1	9	15	30	15	9	0
6-7:	2	4	25	20	20	8	0

Lemma 1: The arrays with solutions 5-1, 5-2, and 5-5 cannot be extended to six-rowed arrays.

Proof: There is a unique array with solution 5-i, $i = 1,\ldots,5$. Since solutions 5-2 and 5-5 represent the same array up to a permutation of the elements (0,1), it is sufficient to consider either 5-1 and 5-2, or 5-1 and 5-5.

Any array of six rows obtained by extending 5-1 would have to have $n_{i0} = 0$, $n_{i5} + n_{i6} = 4$. No such solution exists.

Consider an array admitting solution 5-5 with respect to a first column of all 0's. If an array with solution 5-5 could be extended to an array of six rows, the six-rowed array would have to have $n_{16} = 0$. Solving the necessary equations in terms of the dependent variables then yields

$$n_{11} = -36 + 5\, n_{15}, \text{ which implies } n_{15} \geq 8$$

$$n_{10} = 10 - n_{15}, \text{ which implies } n_{15} \leq 10$$

(i) $\quad n_{15} = 8$ implies $n_{10} = 2$, $n_{11} = 4$

(ii) $\quad n_{15} = 9$ implies $n_{10} = 1$, $n_{11} = 9$

(iii) $\quad n_{15} = 10$ implies $n_{10} = 0$, $n_{11} = 14$.

But any extension of an array with solution 5-5 must have $n_{10} + n_{11} \leq 9$. Then (i) is the only possibility, with complete solution $n_{10} = 2$, $n_{11} = 4$, $n_{12} = 25$, $n_{13} = 20$, $n_{14} = 20$, $n_{15} = 8$, $n_{16} = 0$.

An array of type 5-5 has four columns of all 1's, one column of each type with one 0, and four columns of each type with four 0's. The first 5 rows of the following matrix are part of such an array, and the sixth row is the extension, if an extension is possible.

$$\begin{array}{ccccccccccccc}
1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\
\hdashline
1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0
\end{array}$$

The elements added to the first nine columns are determined by the solution $n_{10} = 2$, $n_{11} = 4$, for the extension. Then if the fifth row is deleted, a five-rowed array remains with $n_{10} = 3$, which can only be of type 5-4. In 5-4, each column with four 0's is repeated three times, and $n_{15} = 1$. Therefore the last four columns above must have three 1's and one 0 added.

Now suppose the fourth row is deleted. Again the same array with solution 5-4 is obtained, and should have each column with three 0's repeated twice. Since (0 0 0 1 1) appears three times in this sub-array, this leads to a contradiction. Hence the extension is not an array. Since array 5-5 cannot be extended, neither can array 5-2.

Note that solutions 5-3 and 5-4 represent the same array with 0 and 1 interchanged. We have shown that arrays with these solutions are the only ones which can be sub-arrays of a six-rowed array.

Lemma 2: Solutions 6-4 and 6-5 do not correspond to arrays.

Proof: Consider an array having solution 6-4 with respect to a first column of 0's. Since $n_{i6} = 1$ and $n_{i5} = 3$, there must be another column of all 0's , and three columns with five 0's and one 1. It follows from lemma 1 that the columns with five 0's must be distinct, since no five-rowed sub-array may have more than three columns which are all 0's. Thus the array would have the following columns:

$$\begin{matrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{matrix} .$$

To complete the requirement that the four-tuple (0 0 0 0) appear five times in each four-rowed sub-array, the columns (0 0 0 0 1 1) and (0 0 0 1 0 1) must each appear three times. However, the four-tuple (0 0 0 1) would then appear six times in rows 1, 2, 3, 6.

Suppose an array has solution 6-5. With $n_{i_5} = 10$, it must have a sub-array with solution 5-4. Consequently, it would be necessary that $n_{i_0} + n_{i_1}$ be less than or equal to 13, but 6-5 has $n_{i_1} = 14$.

Construction of Arrays (80, 6, 2, 4)

The construction of an array with solution 6-1 will now be described, and it will be shown that the construction of an array having solution 6-1 with respect to a first column of 0's is unique. Further, it will be shown that this array admits each of the remaining solutions with respect to some column.

An array having solution 6-1 with respect to the first column of 0's may be constructed as follows.

a) There are three columns with six 0's ($n_{16} = 2$).

b) The four-tuples of 0's must be completed by the columns with four 0's; hence each possible column with four 0's appears twice ($n_{14} = 30$).

c) Each four-tuple consisting of three 0's and one 1 appears in two distinct columns with four 0's, and each of these distinct columns appears twice. For example, (0 0 0 1) in rows 1, 2, 3, 4 appears in (0 0 0 1 1 0) and (0 0 0 1 0 1). To complete the four-tuples consisting of three 0's and one 1, each type of column with three 0's must appear once ($n_{13} = 20$).

d) Each four-tuple consisting of two 0's and two 1's appears in one type of column with four 0's (repeated twice) and in two distinct columns with three 0's, so it must appear once more. For example, (0 0 1 1) in rows 1, 2, 3, 4 appears in (0 0 1 1 0 0), (0 0 1 1 1 0), and (0 0 1 1 0 1). Hence, each type of column with two 0's and four 1's must appear once to complete the four-tuples with two 0's and two 1's ($n_{12} = 15$).

e) Each of the four-tuples with three 1's and one 0 appears in one of the columns with three 0's, and two of the columns with two 0's. For example, (0 1 1 1) in rows 1, 2, 3, 4 appears in (0 1 1 1 0 0), (0 1 1 1 0 1), and (0 1 1 1 1 0). Then each type of column with one 0 must appear twice to complete the four-tuples with three 1's and one 0 ($n_{11} = 12$).

f) The four-tuple consisting of four 1's appears in two distinct columns with one 0, each repeated twice, and in one column with two 0's. Thus the array is complete ($n_{10} = 0$).

It can now be shown that the array just constructed also has solutions 6-2, 6-3, 6-6, and 6-7 with respect to different columns.

Let a column with one 0, say (0 1 1 1 1 1), be the i^{th} column. This column is repeated twice; hence $n_{i6} = 1$. Since no column with five 0's and one 1 appears, $n_{i0} = 0$, and the solution must be 6-2.

Suppose a column with four 0's and two 1's, say (0 0 0 0 1 1), is the i^{th} column, so that $n_{i6} = 1$. Since (1 1 1 1 0 0) appears once, $n_{i0} = 1$ and the solution is 6-3.

Let a column with three 0's, say (0 0 0 1 1 1), be the i^{th} column. Since each column with three 0's appears once, the solution has $n_{i6} = 0$, $n_{i0} = 1$. This solution must be 6-6.

If (1 1 1 1 0 0) is the i^{th} column, the array has $n_{i6} = 0$, $n_{i0} = 2$, since columns with two 0's appear once, and columns with four 0's appear twice. The solution is therefore 6-7.

We will now prove that an array (80, k, 2, 4) with $k = 7$ does not exist. It will first be shown that an array of six constraints having solution 6-1 with respect to some column cannot be extended to seven constraints. We then show that every six-rowed array has solution 6-1 with respect to some column.

<u>Theorem</u>: The maximum number of constraints for an array (80, k, 2, 4) is six.

<u>Lemma 3</u>: An array admitting solution 6-1 with respect to some column cannot be extended to seven rows.

<u>Proof</u>: A seven-rowed extension of an array with solution 6-1 must have $n_{i6} + n_{i7} = 2$. There are only two solutions to the necessary equations which might correspond to such an extension.

	n_{i0}	n_{i1}	n_{i2}	n_{i3}	n_{i4}	n_{i5}	n_{i6}	n_{i7}
7-1:	0	4	23	0	40	10	1	1
7-2:	0	8	4	35	10	20	2	0

First consider 7-1. An array with solution 7-1 could be constructed only by extending an array with solution 6-1. But an array with solution 7-1 must have a six-rowed sub-array with $n_{i0} > 0$. This is a contradiction; hence no array can exist with solution 7-1.

An array with solution 7-2 has $n_{i0} = 0$, $n_{i1} = 8$. Therefore it could not be an extension of an array with solution 6-7, since 6-7 has $n_{i0} + n_{i1} = 6$. Moreover, since $n_{i1} = 8$, the array with solution 7-2 would have to have at least two of the columns having one coincidence with the i^{th} column alike. Crossing out the row including these coincidences would leave six rows with $n_{i0} > 1$. Hence the array could not be an extension of any of the remaining six-rowed arrays.

Since neither of the possible extensions of an array with solution 6-1 exists, the array cannot be extended.

<u>Lemma 4</u>: An array (80, 6, 2, 4) must have solution 6-1 for some value of i.

<u>Proof</u>: Solution 6-1 is the only solution with $n_{i6} = 2$. We will show that any array (80, 6, 2, 4) must have some column repeated three times. The array must then have solution 6-1 with respect to this column.

An array with solution 6-2 must have one of the columns having one coincidence with the i^{th} column repeated three times.

Solutions 6-3, 6-6, and 6-7 will be considered with $i = 1$, assuming the first column to be all 0's.

Solution 6-3 has $n_{i5} = 4$. It follows from lemma 1 that the columns with five 0's must be distinct. Therefore an array with this solution must have six columns as follows:

$$\begin{matrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{matrix}$$

Then the column (0 0 0 0 1 1) must appear three times.

An array with solution 6-7 is the same as an array with solution 6-3, with 0 and 1 interchanged.

An array with solution 6-6 must have columns as follows:

$$\begin{array}{cccccccccc c ccc}
0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \quad & 1 & 1 & 1 \\
0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & & 0 & 0 & 0
\end{array} .$$

Consider the first 10 columns shown. Deleting the first row leaves a five-rowed array with $n_{15} = 2$. Since there are three columns with five 0's, and since columns differing in an even number of coordinates must appear the same number of times, each column with three 0's must appear three times. In particular, the column (1 1 0 0 0) must appear three times in the last five rows. Further, it must be preceded each time by a 1 in the first row, since the four-tuple (0 0 0 0) already appears in rows 1, 4, 5 and 6 five times.

<u>Proof of Theorem</u>: Lemmas 3 and 4 together prove the theorem.

CHAPTER V

RELATION OF FACTORIAL DESIGN TO INFORMATION THEORY

This chapter will briefly describe the relation between the construction of orthogonal arrays by the method of matrix multiplication (described in section 3.1) and the construction of t-error correcting codes.

A signalling alphabet $A(k,v,s)$ is a set of v distinct k-place sequences of members of a set of s elements. $A(k,v,s)$ is a subset of B_k, the set of all s^k possible sequences of k elements. An encoder $E(k,v)$ is a 1-1 correspondence between a set of v distinct messages and the v sequences of $A(k,v,s)$. A message is transmitted by presenting the k elements of the corresponding sequence in succession. The output is some member of B_k.

A decoder $D(k,v)$ is a 1-1 correspondence between the members $a_0, a_1, \ldots, a_{v-1}$ of $A(k,v,s)$ and v disjoint subsets of B_k which form a partition $S_0, S_1, \ldots, S_{v-1}$ of B_k. If an element b_i in S_i is received as output, it is read as a_i. $E(k,v)$ and $D(k,v)$ together form a k-place s-ary code.

Let $s = p^n$, p a prime. Then the elements of B_k may be regarded as the group of vectors with coordinates from $GF(s)$.

For each sequence b in B_k let $w(b)$ be the number of non-zero coordinates in b. The Hamming distance [13] between two sequences b_1, b_2 is defined to be

$$d(b_1, b_2) = w(b_1 - b_2).$$

Hamming distance clearly satisfies all the properties of a metric.

Suppose a_i is an element of $A(k,v,s)$ transmitted, and b_i is the element of B_k received. Then

$$e_i = b_i - a_i$$

is called the noise vector. The number of errors in transmission is then

$$w(e_i) = d(a_i, b_i).$$

The code is said to be t-error-correcting if b_i is contained in S_i, the set corresponding to a_i, whenever $w(e_i) \leq t$, for $i = 0,1,\ldots,v-1$. This implies that the received sequence is correctly interpreted when there have been t or fewer errors of transmission.

Group Codes

Let $v = s^r$, and let $A(k,v,s)$ be a subgroup of order s^r of B_k. Assume $a_0 = (0\ 0\ 0\ldots0)$. A decoder may be constructed as follows: let the group B_k be partitioned into the s^m cosets of $A(k,v,s)$, where $m = k - r$. In the j^{th} coset, choose an element b_j with the property that $w(b_j) \leq w(b)$ for all b in the j^{th} coset, and call b_j the coset leader. In particular, $b_0 = a_0$. Now let S_j be the set

$$\left\{a_j + b_0,\ a_j + b_1,\ \ldots,\ a_j + b_{u-1}\right\}$$

$$j = 1,2,\ldots,v-1;\ v=s^r;\ u=s^m.$$

The $v-1$ sets S_j form the required partition of B_k.

The code obtained is called a (k,r) s-ary group code. This class of codes was first considered by Slepian [22] for the binary case, and by Bose [6] for the s-ary case. The scheme described may be displayed as follows:

$$\begin{array}{lccccc}
\text{Alphabet:} & a_0 & a_1 & ------ & a_{v-1} \\
 & b_1 & (b_1 + a_1) & ------ & (b_1 + a_{v-1}) \\
\text{Coset leaders} & b_2 & (b_2 + a_1) & ------ & (b_2 + a_{v-1}) \\
 & - & - & ------ & - \\
 & - & - & ------ & - \\
 & b_{u-1} & (b_{u-1} + a_1) & ------ & (b_{u-1} + a_{v-1}) \\
 & \uparrow & \uparrow & & \uparrow \\
 & (S_0) & (S_1) & ------ & (S_{v-1})
\end{array}$$

The transmitted message is correctly received if and only if the error vector is a coset leader. Thus the code is t-error-correcting if and only if for each b in B_k , $w(b) \leq t$ implies b is a coset leader.

Bose proves that a (k,r) s-ary group code is t-error-correcting if and only if $w(a) \geq 2t + 1$ for every a in $A(k,v,s)$, except a_0. This was first proved by Hamming for the binary case. Finding a t-error-correcting (k,r) s-ary group code is thus equivalent to finding a subgroup B_r of B_k , of order s^r , such that each element of B_r has weight at least $2t + 1$. Naturally it is of interest to maximize r.

Relation of Group Codes and Orthogonal Arrays

Let C be a $k \times r$ matrix of rank r, with entries from $GF(s)$, $r \leq k$. Suppose C has property P_d that any $d \times r$ submatrix of C is of rank d , $d \leq r$. It was shown in section 3.1 that C generates an array $A = (s^r,k,s,d)$.

Without loss of generality, let C be the matrix

$$\begin{bmatrix} I_r \\ M \end{bmatrix}$$

where I_r is the $r \times r$ identity matrix, and M is the $(k-r) \times r$ matrix of rows added to I_r which satisfy the property P_d stated above.

Consider a sequence $b = (x_1,\ldots,x_k)$ in B_k which is orthogonal to the columns of C. The sequence b defines a linear dependence among the rows of C,

$$bC = \sum_{i=1}^{k} x_i \ (c_{i1},\ldots,c_{ir}).$$

Since every set of d or fewer rows of C are linearly independent, $bC = 0$ implies that b is a sequence of weight $d + 1$ or greater. Therefore the group of sequences orthogonal to C is an alphabet of minimum weight $d + 1$. The matrix C thus serves to define two mutually orthogonal subspaces of B_k, one constituting an array of strength d, and the other a group alphabet of minimum weight $d + 1$.

The alphabet may be derived from C as follows:

Let

$$C^* = \begin{bmatrix} M' \\ -I_{k-r} \end{bmatrix}$$

C^* is a $k \times (k-r)$ matrix of rank $(k-r)$, and

$$[(C^*)' C] = [M \ \ -I_{k-r}] \begin{bmatrix} I_r \\ M \end{bmatrix} = 0.$$

The columns of C^* generate a subgroup A of sequences of minimum weight $d + 1$.

The existence of a matrix C with property P_d is shown by Bose [5,6] to be sufficient for the existence of an array (s^r,k,s,d) and necessary and sufficient for the existence of a t-error-correcting group code with alphabet A, where $d = 2t$.

BIBLIOGRAPHY

1. Addelman, Sidney, "Augmenting Factorial Plans to Accomodate Additional Two-level Factors", Biometrics, Sept. 1962, pp. 308-322.

2. Baumert, L. D., S. W. Golomb, and M. Hall, "Discovery of an Hadamard Matrix of Order 92", Bulletin of the American Mathematical Society (68), 1962, pp. 237-238.

3. Baumert, L. D., and M. Hall, "A New Construction for Hadamard Matrices", Bulletin of the American Mathematical Society (71), 1965, pp. 169-170.

4. Birkhoff, G., and S. MacLane, "A Survey of Modern Algebra", Macmillan Co., 1959.

5. Bose, R. C., "Mathematical Theory of the Symmetrical Factorial Design", Sankhya, Vol. 8, Part 2, 1947.

6. Bose, R. C., "On Some Connections Between the Design of Experiments and Information Theory", Bulletin of the International Statistical Institute, Vol. 38, Part 4, 1961.

7. Bose, R. C., and K. A. Bush, "Orthogonal Arrays of Strength Two and Three", Annals of Mathematical Statistics, Vol. 23, No. 4, Dec. 1952.

8. Box, G. E. P., and J. S. Hunter, "The 2^{k-p} Fractional Factorial Designs", Technometrics, Vol. 3, No. 3, August 1961.

9. Bush, K. A., "Orthogonal Arrays of Index Unity", Annals of Mathematical Statistics, Vol. 23, No. 4, Dec. 1952.

10. Carmichael, Robert, "Groups of Finite Order", Dover Publications, Inc., 1956. Copywright 1937.

11. Fisher, R. A., "The Theory of Confounding in Factorial Experiments in Relation to the Theory of Groups." Ann. Eugen. Lond., (11) 1942. pp. 341-353.

12. Hall, Marshall, Jr., "The Theory of Groups", Macmillan Co., 1959.

13. Hamming, R. W., "Error Detecting and Error Correcting Codes", Bell System Technical Journal (29), 1950, pp. 147-160.

14. Kempthorne, Oscar, "The Design and Analysis of Experiments", John Wiley and Sons, N.Y., 1952.

15. Paley, R. E. A. C., "On Orthogonal Matrices", Journal of Mathematics and Physics (12), 1933, pp. 311-320.

16. Plackett, R. L., and J. P. Burman, "The Design of Optimum Multifactorial Experiments", Biometrika, Vol. 33, 1946.

17. Rao, C. R., "Factorial Experiments Derivable from Combinatorial Arrangement of Arrays", Journal of the Royal Statistical Society, Supplement, Vol. 9, No. 1, 1947.

18. Scheffe, Henry, "The Analysis of Variance", John Wiley and Sons, N.Y., 1959.

19. Seiden, Esther, "On the Problem of Construction of Orthogonal Arrays", Annals of Mathematical Statistics, Vol. 25, No. 1, March 1954.

20. Seiden, Esther, "Further Remarks on the Maximum Number of Constraints of an Orthogonal Array", Annals of Mathematical Statistics, Vol. 26, No. 4, December 1955.

21. Seiden, Esther, "On the Maximum Number of Points No Four on One Plane in a Projective Space PG(r-1,2)", Technical Report RM-117, ES-7, Department of Statistics, Michigan State University, 1964.

22. Slepian, D., "A Class of Binary Signalling Alphabets", Bell System Technical Journal (35), 1956, pp. 203-234.

23. Williamson, J., "Hadamard's Determinant Theorem and the Sum of Four Squares", Duke Mathematical Journal (11), 1944, pp. 65-81.

24. Seiden, Esther, "A Theorem in Finite Projective Geometry and an Application to Statistics", Proc. Amer. Math. Soc., Vol. 1 (1950), pp. 282-286.

25. Qvist, B., "Some Remarks Concerning Curves of the Second Degree in a Finite Plane", Ann. Acad. Sci. Fennicae. Series A, I: Math.-Phys. (134). 1952.