# ESTIMABLE EFFECTS AND INTERACTIONS IN AN n-WAY CROSS CLASSIFICATION WITH MISSING CELLS

By

RALPH RAYMOND LINDSTROM

Bachelor of Science
Northern Michigan University
Marquette, Michigan
1957

Master of Science
The University of Illinois
Urbana, Illinois
1960

Submitted to the faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
May 1971

ESTIMABLE EFFECTS AND INTERACTIONS IN AN

n-WAY CROSS CLASSIFICATION WITH

MISSING CELLS

Thesis Approved:

David L. Weeks

Thesis Adviser

J. Leroy Folks

Ronald W. McNew

Irvin T. Omtvedt

D. D. Durham

Dean of the Graduate College

## ACKNOWLEDGEMENTS

I wish to express my appreciation to Dr. David Weeks for serving as chairman of my advisory committee, for his leadership, teaching, patience and guidance in the work leading up to this thesis and its preparation. I am also grateful to Professor Leroy Folks, Director, Statistical Laboratory and Professors Ronald McNew and Irvin Omtvedt for their service on my advisory committee. Particular acknowledgement is made to the National Science Foundation for the support I received throughout my graduate study and during the preparation of this thesis and to Thad Smith for his work on the computer program included in the appendix.

TABLE OF CONTENTS

## LIST OF TABLES

# CHAPTER I

## INTRODUCTION

In a general n-way cross classification each observation is classified in $n$ ways. The $n$ classifications are referred to as factors and we suppose that the i-th factor has $t_i$ levels, or in all $t_1 \cdot t_2 \ldots t_n$ combinations (cells) are under consideration. The investigator is usually interested in determining the effects of changing the levels of each factor and also, in many cases, the influence that various combinations of the other factors have on these effects. If the combined effect of changing the level of several factors is not the sum of the effects of the individual factors, interaction is said to exist, and any discussion of the effect of one or several of these factors must necessarily take into consideration the influence of the other factors. This, of course, complicates the analysis considerably, but if the factors are going to occur together naturally, the information thus obtained is essential. We generally hope, however, that certain of the interactions will prove to be negligible and thereby simplify our discussion.

If an equal number of observations are obtained for each and every cell, the problems of estimation and tests of hypotheses concerning various effects and interactions are quite straight forward and well documented in many textbooks. (See for example: F. A. Graybill (6)[1]). In the event that the numbers of observations per cell are unequal the analysis gets a little more involved, but is still accomplished in much the same manner. However, if a number of the cells are not represented due to missing observations, no general method of analysis has been put forth and the investigator will probably have to write a mathematical model for the observations that are present and attempt to solve the corresponding normal equations. Because of the missing observations, the system loses much of the symmetry present when at least one observation is present in each cell and consequently the solutions are considerably more difficult to obtain. Moreover, the investigator has no assurance before he attempts to find the solutions that the effects and interactions in which he is interested are estimable. The investigator may attempt to estimate the missing data or assume that all interaction effects are zero in order to simplify the problem. But no procedure can actually recover the missing data, and the arbitrary decision that there is no interaction is obviously undesirable if some alternative exists.

The purpose of this thesis is to present a relatively simple method whereby the investigator can determine beforehand, in a n-way cross classification with missing cells, which effects and interactions are estimable free of the influence of other effects and interactions. An alternative to the assumption of no interaction is discussed, and the degrees of freedom in the analysis of variance table are separated into sets associated with confounded and unconfounded effects and interactions.

[1]Note: refers to Selected Bibliography.

To illustrate the failure of a conventional method of partitioning the degrees of freedom (and sum of squares) consider a $3^3$ factorial experiment where the factors are designated by A, B, and C. Suppose the only design points for which the experimenter was able to get at least one observation were 000, 100, 120, 220, 111, 021, 121, 211, 012, 122, 112, 202 where (i, j, k) indicates the i-th level of factor A, j-th level of factor B, and k-th level of factor C. Let us say that the ABC interaction is known to be zero, but we wish to investigate the 2 factor interactions. We will attempt to partition the degrees of freedom by combining a series of $2 \times 2$ tables.

Ignoring C a two way table for A and B yields 4 degrees of freedom associated with the interaction of A and B.

| A \ B | 0 | 1 | 2 |
|---|---|---|---|
| 0 | X | X | X |
| 1 | XX | XX | XX |
| 2 | X | X | X |

A.O.V.

| Source | df |
|---|---|
| Total | 11 |
| A | 2 |
| B | 2 |
| AB | 4 |
| Remainder | 3 |

A similar table for A and C results in 4 degrees of freedom associated with the AC interaction.

| A \ C | 0 | 1 | 2 |
|---|---|---|---|
| 0 | X | X | X |
| 1 | XX | XX | XX |
| 2 | X | X | X |

A.O.V.

| Source | df |
|---|---|
| Total | 11 |
| A | 2 |
| C | 2 |
| AC | 4 |
| Remainder | 3 |

Finally, a table for B and C yields 1 degree of freedom for the interaction of factors B and C

| B \ C | 0 | 1 | 2 |
|---|---|---|---|
| 0 | XX | | XX |
| 1 | | XX | XX |
| 2 | XX | XX | |

A.O.V.

| Source | df |
|---|---|
| Total | 11 |
| B | 2 |
| C | 2 |
| BC | 1 |
| Remainder | 6 |

If we attempt to combine these tables into a single A.O.V. table as is possible in an experiment with no missing data, we see that there are not sufficient degrees of freedom remaining, after the A, B, and C components are considered, to have 4 degrees of freedom associated with AB, 4 with AC, and 1 with BC.

A.O.V.

| Source | d.f. |
|---|---|
| Total | 11 |
| A | 2 |
| B | 2 |
| C | 2 |
| AB | 4 |
| AC | ? |
| BC | ? |

The reason for this is that these three interaction effects AB, AC and BC are confounded with each other. This occurs because due to the missing cells, it is impossible to measure the failure of the simple effects of factor A to be the same at different levels of B without changing levels of C . Thus the failure of the simple effects of A to be the same at different levels of B is confounded with the failure of the simple effects of A to be the same at different levels of C , or the AB interaction is confounded with AC . Likewise, AB is confounded with BC .

Some work has been done by Thomas (5) and Williams (4) on estimatibility of main effects for the n-way cross classification model without interaction, $Y = X\beta + e$ , where $Y$ is a $M \times 1$ vector of observations, $X$ is a $M \times \sum_{i=1}^{n} t_i$ matrix of ones and zeros, $\beta$ is a $\sum_{i=1}^{n} t_i \times 1$ vector of unknown parameters $\beta_{ij}$ , $i=1,2,\ldots,n$ , $j=1,2,\ldots,t_i$ , and $e$ is an $M \times 1$ vector of errors. This design is defined to be connected if $\beta_{ij} - \beta_{ik}$ is estimable for all $i=1,2,\ldots,n$ and for all $j,k=1,2,\ldots,t_i$ , $j \neq k$ . Williams defines a procedure for determining connectedness for main effects that is sufficient but not necessary. Thomas utilizes this procedure to show that if a $p^n$ factorial is expressed in the form $(\sum_{i=1}^{k} p_i)^n$ then the total number of connected plans obtainable by combining all combinations of the $k^n$ factorials is $(2^k - 1)^n$ .

To illustrate that this procedure is sufficient but not necessary it is presented as given by Williams below with two examples both connected in the sense of the definition. The first example, also from Williams, illustrates the procedure and demonstrates that it is sufficient while the second example shows it is not necessary. To simplify the discussion two n-tuples are defined to be nearly identical if the n-tuples are equal component-wise except for one component. It is required in the procedure that the design points corresponding to occupied cells be such that the i-th component takes on all possible values $1,2,\ldots,t_i$ over the set of all n-tuples. Otherwise parameters associated with missing values are to be eliminated from the original model.

Procedure:

1. Construct a table of all occupied cells expressing each occupied cell as an n-tuple. If a point is repeated, list it only once.
2. Select any point from the table in (1) and find all nearly identical points for the n-tuple selected. Eliminate each point from the table as it is selected.
3. Select all nearly identical points which remain in the table for each n-tuple selected in (2). Again eliminate each point from the table as it is selected.
4. Repeat step (3) for each n-tuple selected in step (3).
5. Continue this procedure until there are no points remaining in the table or until there is no n-tuple in the table which is nearly identical to any point selected in steps (2)-(4).
6. If there are points remaining in the table after step (5) the original set of design points is not connected. The n-tuples which are nearly identical form a connected subset and may be analyzed as a reduced set of design points where the parameters whose subscripts do not

appear in the subset obtained from step (5) may be eliminated from the original model. The remaining points may be divided into connected subsets by the above procedure so that each may be analyzed as a reduced set of design points.

Example 1.1: Consider Table I for a three-way cross classification where one or more observations are given for the cells containing X and no observations are contained in the other cells.

Step 1: All points corresponding to cells in Table I which contain observations are listed in Table II.

Step 2: Select any point in Table II, say (1,3,4), and take all points which are nearly identical to (1,3,4). These points are (1,3,2), (1,3,1), (1,1,4) and (2,3,4). Eliminate each of these points from Table II.

TABLE I

CELLS FOR THREE-WAY CROSS CLASSIFICATION DATA

| First Classification | Second Classification | Third Classification 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| 1 | 1 | | X | X | X |
| | 2 | X | X | | |
| | 3 | X | X | | X |
| 2 | 1 | | X | | X |
| | 2 | X | | X | X |
| | 3 | X | | | X |

TABLE II

POINTS CORRESPONDING TO OCCUPIED CELLS IN TABLE I

| | | |
|---|---|---|
| (1,1,2) | (1,3,1) | (2,2,1) |
| (1,1,3) | (1,3,2) | (2,2,3) |
| (1,1,4) | (1,3,4) | (2,2,4) |
| (1,2,1) | (2,1,2) | (2,3,1) |
| (1,2,2) | (2,1,4) | (2,3,4) |

Step 3: For each of the points selected in step 2, it is necessary to find all nearly identical points remaining in Table II. The points which are nearly identical to a particular point selected in step 2 are as follows:

| Point from step 2 | (1,3,2) | (1,3,1) | (1,1,4) | (2,3,4) |
|---|---|---|---|---|
| Nearly identical points | (1,1,2) | (1,2,1) | (1,1,3) | (2,2,4) |
| | (1,2,2) | (2,3,1) | (2,1,4) | |

Step 4: For each of the points selected in step 3, check the remaining points in Table II. The points which are nearly identical to a particular point selected in step 3 are:

| Point from step 3 | (1,1,2) | (1,2,1) | (2,2,4) |
|---|---|---|---|
| Nearly identical points | (2,1,2) | (2,2,1) | (2,2,3) |

Step 5: Since each point was eliminated as it was selected from Table II, there are no points remaining in Table II so the set of design points is connected. The vector

of unknown parameters is $\beta' = (\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}, \beta_{23}, \beta_{31}, \beta_{32}, \beta_{33}, \beta_{34})$. The set of points may now be analyzed as a three-way classification with $t_1 = 2$, $t_2 = 3$ and $t_3 = 4$.

If Table II had points remaining one could apply steps 2-5 to the remaining points and obtain other connected subsets which could be analyzed as reduced designs. It should be noted that the points obtained first would also correspond to a reduced design and could be analyzed as such.

Example 1.2: Using the same model with no interaction say, that the points corresponding to occupied cells were the following subset of Table II of example 1.1:

TABLE III

POINTS CORRESPONDING TO OCCUPIED CELLS

| | |
|---|---|
| (1,2,1) | (1,1,3) |
| (2,3,1) | (2,2,3) |
| (2,1,2) | (1,3,4) |
| (1,2,2) | (2,2,4) |

The points (2,3,1), (1,1,3), (2,1,2) and (1,3,4) in Table III are not nearly identical to any other point of the set, so according to Williams' procedure the design is not connected. Continuing with the procedure given by Williams, only two connected subsets can be found and they are {(1,2,1), (1,2,2)} and {(2,2,3), (2,2,4)}. Williams then suggests we analyze each of these sets as a reduced set of design points eliminating all other parameters from the original model. Thus according to his procedure the only estimable differences are $\beta_{31} - \beta_{32}$ and $\beta_{33} - \beta_{34}$ .

However, if we designate the observation corresponding to the design point (i,j,k) by $x_{ijk}$ , then using the observations associated with the design points of Table III we get:

$$E(x_{121} - x_{231} + x_{134} - x_{224}) = 2(\beta_{11} - \beta_{12})$$
$$E(x_{121} - x_{231} + x_{224} - x_{134}) = 2(\beta_{22} - \beta_{23})$$
$$E(x_{122} - x_{113} + x_{223} - x_{212}) = 2(\beta_{22} - \beta_{21})$$
$$E(x_{121} - x_{231} + x_{224} - x_{134} - x_{122} + x_{113} - x_{223} + x_{212}) = \beta_{21} - \beta_{23}$$
$$E(x_{121} - x_{122}) = \beta_{31} - \beta_{32}$$
$$E(x_{223} - x_{224}) = \beta_{33} - \beta_{34}$$
$$E(x_{113} + x_{223} - x_{122} - x_{212}) = 2(\beta_{33} - \beta_{32})$$
$$E(x_{121} + x_{231} - x_{224} - x_{134}) = 2(\beta_{31} - \beta_{34})$$
$$E(x_{223} + x_{212} - x_{224} - x_{224} - x_{113} + x_{122}) = 2(\beta_{32} - \beta_{34})$$
$$E(x_{113} + x_{223} - x_{121} - x_{212} - x_{121} + x_{122}) = 2(\beta_{33} - \beta_{31})$$

Thus all differences $\beta_{ij} - \beta_{ik}$ for $i=1,2,3$ and $j,k=1,2,\ldots,t_i$ $j \neq k$ are estimable; the design is connected and may be analyzed as a three-way classification with $t_1 = 2$ , $t_2 = 3$ and $t_3 = 4$ . In other examples, such as a $\frac{1}{2}$ replication of a $2^3$ factorial experiment, no two design points are nearly identical and yet, of course, all differences are estimable and the design is connected in the absence of interaction.

This thesis presents a procedure for determining not only which simple effects are estimable, but also which interactions are estimable in a general n-way cross classification design with interaction and missing cells. The criteria of the procedure are necessary and sufficient for estimability and can be modified for use in situations where certain of the interactions are known to be zero. The procedure gives easily obtainable estimates of the estimable effects and interactions, although in general they do not make use of all the observations and hence may not be the best estimates available. The estimated variances of these estimates are readily obtainable, as well as a partitioning of the degrees of freedom associated with each effect and interaction in an analysis of variance table into confounded and unconfounded sets.

# CHAPTER II

## ESTIMABILITY OF EFFECTS AND INTERACTIONS

### Introduction

Throughout this chapter we will be considering an n-way cross-classification design with interaction where, either by design or circumstance, a number of the cells have no observations. It will be understood that if there are no observations for some level of a factor, that level will be deleted from the original model. Similarly if there are no observations for any level of some factor the factor will be eliminated from the model. While it is usually not practical to consider designs for very large values of $n$, the treatment in this chapter will be entirely general.

### Definitions, Notation, and Preliminary Results

To facilitate the ensuing discussion, a brief description of the notation and definitions to be used is presented first. The $n$ factors of the cross-classification design will be designated by integers $i=1,2,\ldots,n$ and $a_i$ will denote a level of factor $i$, $a_i = 0,1,\ldots,t_i-1$. Each combination of levels of factors is then associated with an n-tuple $(a_1, a_2, \ldots, a_n)$, called a design point representing level $a_1$ of factor one, $a_2$ of factor two, etc. Whenever convenient the $n$ component vector $(a_1,\ldots,a_n)$ will be designated by $\underline{a}(n)$. We will let $X$ be the set of all observations obtained and $D_x$ be the corresponding set of distinct points for all observations in X . We can define a relation on $D_x$ as follows:

<u>Definition 2.1</u>: For all points $\underline{a}(n) = (a_1, a_2, \ldots, a_n)$ , $\underline{b}(n) = (b_1, b_2, \ldots, b_n)$ in $D_x$ , $\underline{a}(n)\ R_i\ \underline{b}(n)$ if and only if $a_j=b_j$ for $j \neq i$ . We will say that the two points are $R_i$ , (read "related in class i"), whenever this definition is satisfied.

The above definition is similar to Williams' definition of "nearly identical" discussed in the first chapter, but in addition allows a point to be related to itself, and also specifies in which component two points differ when they are equal component-wise except for one component.

It is easily seen that this relation is an equivalence relation and thus partitions the set of design points $D_x$ into disjoint subsets each containing either a single point or points that differ only in the ith coordinate. To distinguish among the equivalence classes we will let $R_i(a_1, \ldots, a_{i-1}, *, a_{i+1}, \ldots, a_n)$ be the class of points that are $R_i$ and for which the jth coordinate is $a_j$ , $j=1,2,\ldots,i-1, i+1,\ldots,n$ .

For example, in the subset D=((000),(001),(002),(010),(020),(021),(110),(112)) of the design points of a $3^3$ factorial arrangement of treatment combinations, we have:

$R_3(00*)=(\ (000),(001),(002)\ )$

$R_3(01*)=(\ (010)\ )$

$R_3(02*)=(\ (020),(021)\ )$

$R_3(11*)=(\ (110),(112)\ )$

Similarly we could partition $D$ into disjoint subsets using either $R_1$ or $R_2$ .

The interaction, in the model assumed, of the factors $i_1, i_2, \ldots i_k$ which is associated with the levels $a_{i_1}, \ldots, a_{i_k}$ will be designated by $\mu(\underline{i}(k); \underline{a}(i(k))) = \mu(i_1, i_2, \ldots, i_k ; a_{i_1}, a_{i_2}, \ldots, a_{i_k})$ and the effect of level $a_i$ of the ith factor will be $\mu(i;a_i)$ . An observation associated with the design point $\underline{a}(n)$ will be denoted by $x_{\underline{a}(n)}$ and $\overline{x}(\underline{i}(k); \underline{a}(k)) = \overline{x}(i_1, i_2, \ldots, i_k; a_1, a_2, \ldots, a_k)$ will represent the mean of all observations at level $a_1$ of factor $i_1$ , level $a_2$ of factor $i_2, \ldots,$ level $a_k$ of factor $i_k$ .

Using the notation above, an n-way cross classification model with interaction and one observation per treatment combination may be written:

$$x_{\underline{a}(n)} = \sum_{k=0}^{n} \sum_{\underline{i}(k) \varepsilon S} \mu(\underline{i}(k);\underline{a}(i(k))) + e_{\underline{a}(n)}$$

$$a_i = 0,1,\ldots,t_i-1$$

where the second summation is understood to be $\mu$ when $k=0$ , $S$ is the set of all vectors $\underline{i}(k)$ where $i_j$ is taken from $1,2,\ldots,n$ with $i_1 < i_2 < i_3 \ldots < i_k$ , and the errors $e_{\underline{a}(n)}$ are uncorrelated and all have the same mean $0$ and variance $\sigma^2$ .

For example with $N=3$ the model would be:

$$\begin{aligned} x_{a_1,a_2,a_3} = \mu &+ \mu(1;a_1) + \mu(2;a_2) + \mu(3;a_3) \\ &+ \mu(1,2;a_1,a_2) + \mu(1,3;a_1,a_3) + \mu(2,3;a_2,a_3) \\ &+ \mu(1,2,3;a_1,a_2,a_3) + e_{a_1,a_2,a_3} \end{aligned}$$

The above notation is somewhat non-standard, but is adopted to permit a completely general discussion of an n-way classification. However, since the examples of this thesis involve only small values of $n$, standard notation will be used in all examples. Thus the model corresponding to the case above for $n=3$ will be:

$$y_{ijk} = \mu+\alpha_i+\beta_j+\Gamma_k+(\alpha\beta)_{ij}+(\alpha\Gamma)_{ik}+(\beta\Gamma)_{jk}+(\alpha\beta\Gamma)_{ijk}+e_{ijk}$$
$$i=1,2,\ldots,t_1 \quad j=1,2,\ldots,t_2 \quad k=1,2,\ldots,t_3$$

The unknown constants $\mu$ $\alpha_i$, $\beta_j$, and $\Gamma_k$ are called the mean and additive treatment constants respectively, $(\alpha\beta)_{ij}$, $(\alpha\Gamma)_{ik}$ and $(\beta\Gamma)_{jk}$ are called 2 factor interaction terms and $(\alpha\beta\Gamma)_{ijk}$ is the 3 factor interaction term. The sets of parameters $\alpha_i$, $\beta_j$ and $\Gamma_k$ will be associated with factors denoted by A, B and C respectively.

The concept of interaction will be generalized by defining a component of interaction as in Mann(3).

<u>Definition 2.2</u>: Any linear form

$$\sum_{\underline{a}(k)} \ell_{\underline{a}(k)} \, x(\underline{i}(k);\underline{a}(k))$$

which is not identically zero will be termed a component of the interaction between the factors $i_1,i_2,\ldots,i_k$ if $\sum_{a_i} \ell_{\underline{a}(k)} = 0$ for $i=1,2,\ldots,k$ and for all choices $a_1,\ldots,a_{i-1},a_{i+1},\ldots,a_k$.

It is easily seen, and will be demonstrated in Theorem 2.1, that if all $\alpha$ factor interaction effects are zero for $\alpha > k$, then the expected value of a linear form satisfying Definition 2.2 is a linear combination

$$\sum_{\underline{a}(k)} \ell_{\underline{a}(k)} \, \mu(\underline{i}(k);\underline{a}(k))$$

Thus the failure of a component of interaction to be zero is a measure of interaction of factors $i_1, i_2, \ldots, i_k$ over the levels $a_{i_1}, a_{i_2}, \ldots, a_{i_k}$ for which $\ell_{\underline{a}(k)}$ is not zero in the component.

Since a main effect or interaction is always confounded with higher order interactions, throughout this thesis, to avoid repeated reference to this situation, higher order interactions will be considered to be zero whenever the estimability of a particular main effect or interaction is being discussed.

Definition 2.3: The interaction effect of factors $i_1, i_2, \ldots, i_k$ will be called partially estimable if there exists a component of interaction for the factors.

If there is at least one observation per cell and we consider the set of "true" cell means of the conceptual population of yields for each cell, then there are $\prod_{j=1}^{k} (t_{i_j} - 1)$ linearly independent functions of the "true" cell means that can be used to measure a k factor interaction . The above definition requires that at least one of these functions be estimable. Throughout the remainder of this thesis, a partially estimable interaction will be referred to simply as estimable.

Definition 2.4: If the interaction effect of classes $i_1, i_2, \ldots, i_k$ is partially estimable, the associated set of design points will be referred to as a partially connected set with respect to this k factor interaction. As in definition 2.3, the word partially will be dropped throughout the remainder of this thesis.

Let us now consider the problem of determining whether or not a particular main effect or interaction is estimable. We begin by considering a few preliminary results necessary to establish a criterion for the estimability of the highest order interaction in an n-way cross classification design.

Lemma 2.1 : If there exists a nonempty subset $D_x$ of the design points such that for each $\underline{a}(n)$ in $D_x$ there is a $\underline{b}(n)$ in $D_x$ such that $a_i \neq b_i$ and $\underline{a}(n)\ R_i\ \underline{b}(n)$ for each $i=1,2,\ldots,n$, then there is a nonempty subset $D_{x'}$ of $D_x$ such that for each $\underline{a}(n)$ in $D_{x'}$, and each $i=1,2,\ldots,n$ there is exactly one $\underline{b}(n)$ in $D_{x'}$, such that $a_i \neq b_i$ and $\underline{a}(n)\ R_i\ \underline{b}(n)$ .

Proof: For a fixed $i$, say without loss of generality $i=1$, $R_1$ determines a partition of $D_x$ . The hypothesis states that all of the subsets of this partition contain at least two design points and some may contain more than two. By selecting a combination of any two design points from each subset of this partition we can construct a subset $D_{x(1)} \subseteq D_x$ such that for each $\underline{a}(n)$ in $D_{x(1)}$ there is exactly one $\underline{b}(n)$ in $D_{x(1)}$ such that $a_1 \neq b_1$ and $\underline{a}(n)\ R_1\ \underline{b}(n)$ . We can now partition $D_{x(1)}$ using $R_2$ and in the same manner select from each subset of this partition of $D_{x(1)}$ a combination of exactly two design points, thus forming a subset $D^1_{x(2)} \subseteq D_{x(1)}$ such that for each $\underline{a}(n)$ in $D^1_{x(2)}$ there is exactly one $b(n)$ in $D^1_{x(2)}$ such that $a_2 \neq b_2$ and $\underline{a}(n)\ R_2\ \underline{b}(n)$ . Of course if any subset of the partition of $D_{x(1)}$ by $R_2$ contains only one point then that point will not be found in $D^1_{x(2)}$ . Elimination of such a point, however, will produce a set of points for which some point is not $R_1$ to exactly one different point. We therefore must partition $D^1_{x(2)}$ by $R_1$ again and eliminate all points that occur alone in some subset of the partition. Denote this subset of $D^1_{x(2)}$ by $D^2_{x(2)}$ .

We now must partition $D^2_{x(2)}$ by $R_2$ again and eliminate those points that occur again thus forming a set $D^3_{x(2)}$ . Continue this process until $D^i_{x(2)} = D^{i+1}_{x(2)} = D_{x(2)}$ (say) . Then for all $\underline{a}(n)$ in $D_{x(2)}$ there exists exactly one $\underline{b}(n)$ in $D_{x(2)}$ such that $a_i \neq b_i$ and $\underline{a}(n)\ R_i\ \underline{b}(n)$ for $i=1,2$ . We now continue with the above procedure using $R_3, R_4, \ldots, R_n$ until we obtain a set $D_{x(n)}$ that remains invariant upon partitioning by $R_1, R_2, \ldots, R_n$ and eliminating all points that occur alone in some subset of some partition. Thus $D_{x(n)}$ has the property that each $R_i = 1,2,\ldots,n$ partitions it into disjoint subsets each containing exactly two points. If we can obtain a nonempty set $D_{x(n)}$ by this process then the lemma is established, since by the construction of $D_{x(n)}$ there exists exactly one $\underline{b}(n)$ in $D_{x(n)}$ such that $a_i \neq b_i$ and $\underline{a}(n)\ R_i\ \underline{b}(n)$ $i=1,2,\ldots,n$ . Suppose then that the lemma is false, that is $D_{x(n)} = \emptyset$ for all $D_{x(n)}$ constructed in the manner described above. We have for all sets $D_{x(1)}, D_{x(2)}, \ldots, D_{x(n)}$ constructed as above

$$D_{x(n)} \subseteq D_{x(n-1)} \subseteq \cdots \subseteq D_{x(1)} \subseteq D_x$$

and hence

$$\cup D_{x(n)} \subseteq \cup D_{x(n-1)} \subseteq \cdots \subseteq \cup D_{x(1)} \subseteq D_x$$

where $\cup D_{x(1)}$ refers to the union over all subsets of $D_x$ that can be constructed as above taking two points at a time from each of the subsets of the partition of $D_x$ by $R_1$ , $\cup D_{x(2)}$ is the union over all subsets that can be constructed as above from some set in $\cup D_{x(1)}$ , etc. Let $\underline{a}(n)$ be a point in $D_x$ and suppose $\underline{a}(n)$ is not in some $D_{x(1)}$ in $\cup D_{x(1)}$ . Then there is not a $\underline{b}(n)$ in $D_x$

such that $\underline{a}(n)\ R_i\ \underline{b}(n)$, but this contradicts the definition of $D_x$. Thus $\underline{a}(n)\varepsilon D_x$ implies $\underline{a}(n)\varepsilon \cup D_{x(1)}$ and hence $\cup D_{x(1)} = D_x$. Similarly by induction we can establish that $\cup D_{x(i)} = D_x$ for $i=1,2,\ldots,n$. Now if $D_{x(n)} = \emptyset$ for all $D_{x(n)}$ then $\cup D_{x(n)}=\emptyset=D_x$ which contradicts the definition of $D_x$. Thus there exists a nonempty $D_{x(n)}$ and the lemma is established with $D_{x'} = D_{x(n)}$.

We note that for each $i=1,2,\ldots,n$, $R_i$ is now a permutation of the set of design points $D_{x'}$ of Lemma 2.1 which assigns to each point of $D_{x'}$ the unique point of $D_{x'}$ that differs only in the ith coordinate.

We will use the usual notation "o" for composition of functions to denote one mapping followed by another.

Definition 2.5: We will say $\underline{b}(n)$ is accessible from $\underline{a}(n)$ if $\underline{a}(n)\ R_{i_1} \circ R_{i_2} \circ \ldots \circ R_{i_k}\ \underline{b}(n)$ for some choice of $i_1, i_2, \ldots, i_k$ out of $1,2,3,\ldots,n$. Since each $R_i$ is an equivalence relation, if $\underline{b}(n)$ is accessible from $\underline{a}(n)$ then $\underline{a}(n)$ is also accessible from $\underline{b}(n)$ and we will say that $\underline{a}(n)$ and $\underline{b}(n)$ communicate.

Communication is also an equivalence relation and therefore will partition a set of design points into disjoint classes of communicating points.

## Estimation of the n-Factor Interaction

Theorem 2.1: In an n-way cross-classification with interaction and missing cells, the n factor interaction effect is estimable if and only if there exists a non-empty subset $D_x$ of the design points such that for each $\underline{a}(n)$ in $D_x$ there exists a $\underline{b}(n)$ in $D_x$ such that $a_i \neq b_i$ and $\underline{a}(n)\ R_i^?\ \underline{b}(n)$ for each $i=1,2,\ldots,n$.

Proof: If there exists a nonempty subset $D_x$ of the design points such that for each $\underline{a}(n)$ in $D_x$ there is a $\underline{b}(n)$ in $D_x$ such that $a_i \neq b_i$ and $\underline{a}(n)\ R_i\ \underline{b}(n)$ for $i=1,2,\ldots,n$, then by Lemma 2.1 there is a subset $D_{x'}$ of $D_x$ such that for each $\underline{a}(n)$ in $D_{x'}$ and each $i=1,2,\ldots,n$ there is exactly one $\underline{b}(n)$ in $D_{x'}$ such that $a_i \neq b_i$ and $\underline{a}(n)\ R_i\ \underline{b}(n)$. We can then partition $D_{x'}$ by the equivalence relation of communication into communicating classes. Let $\underline{a^*}(n)$ be any fixed point in $D_{x'}$ and let $C$ be the set of points of $D_{x'}$ that communicate with $\underline{a^*}(n)$.

We now define a linear contrast of the observations corresponding to points in $C$ by letting $\ell_{\underline{a}(n)} = +1$ if $\underline{a}(n)$ in $C$ communicates with $\underline{a^*}(n)$ through an even number of permutations $R_{i_1}, R_{i_2}, \ldots, R_{i_{2r}}$ and let $\ell_{\underline{a}(n)} = -1$ if $\underline{a}(n)$ in $C$ communicates with $\underline{a^*}(n)$ through an odd number of permutations $R_{i_1}, R_{i_2}, \ldots, R_{i_{2r+1}}$. Thus if $\underline{a}(n)\ R_i\ \underline{b}(n)$ then $\ell_{\underline{a}(n)} = -\ell_{\underline{b}(n)}$ and hence $\sum_{a_i} \ell_{\underline{a}(n)} = 0$ for $i=1,2,\ldots,n$ and all choices $a_1,\ldots,a_{i-1},a_{i+1},\ldots,a_n$. To make this a linear function of all the observations simply let $\ell_{\underline{a}(n)}=0$ if $\underline{a}(n)$ is not in $C$. Then:

$$E \sum_{\underline{a}(n)} \ell_{\underline{a}(n)} X_{\underline{a}(n)}$$

$$= \sum_{a_1} \cdots \sum_{a_n} \ell_{\underline{a}(n)} \sum_{\alpha=0}^{n} \sum_{1,\ ,\ldots,n} \mu(\underline{i}(\alpha);\underline{a}(i(\alpha)))$$

$$= \sum_{\alpha=0}^{n-1} \sum_{1,2,\ldots,n-1} \sum_{a_{i_1}} \cdots \sum_{a_{i_\alpha}} \sum_{a_{i_{\alpha+1}}} \sum_{a_{i_n}} \ell_{\underline{a}(n)} \mu(\underline{i}(\alpha);\underline{a}(i(\alpha)))$$

$$+ \sum_{a_1} \cdots \sum_{a_n} \ell_{\underline{a}(n)} \mu(\underline{i}(n);\underline{a}(\underline{i}(n)))$$

$$= \sum_{\alpha=0}^{n-1} \sum_{1,2,\ldots,n-1} \sum_{a_{i_1}} \cdots \sum_{a_{i_\alpha}} \ell_{\underline{a}(n)} \mu(\underline{i}(\alpha);\underline{a}(i\_(\alpha))) \sum_{a_{i_{\alpha+1}}} \cdots \sum_{a_{i_n}} \ell_{\underline{a}(n)}$$

$$+ \sum_{a_1} \ldots \sum_{a_n} \ell_{\underline{a}(n)} \; \mu(\underline{i}(n);\underline{a}(i(n)))$$

$$= \sum_{a_1} \ldots \sum_{a_n} \ell_{\underline{a}(n)} \; \mu(\underline{i}(n);\underline{a}(i(n)))$$

since
$$\sum_{a_{i_{\alpha+1}}} \ldots \sum_{a_{i_n}} \ell_{\underline{a}(n)} = 0$$

Thus by Definition 2.3 the interaction of factors $1,2,\ldots,n$ is estimable.

To prove that the conditions of the theorem are sufficient, suppose there does not exist a subset $D_x$ of the design points satisfying the condition that for every $\underline{a}(n)\varepsilon D_x$, and for each $i=1,2,\ldots,n$, there exists $\underline{b}(n)\varepsilon D_x$ such that $\underline{a}(n)\ R_i\ \underline{b}(n)$. Then for every subset $D_x$ of the design points there is some $\underline{a}^*(n)$ in $D_x$ such that for some $i=1,2,\ldots,n$, say $i=k$, there is no $\underline{b}(n)$ in $D_x$ with $\underline{a}^*(n)\ R_k\ \underline{b}^*(n)$. Thus for any linear form

$$\sum_{\underline{a}(n)} \ell_{\underline{a}(n)} \, x_{\underline{a}(n)}$$

the expected value will involve at the very least

$$\mu(1,2,\ldots,k-1,k+1,\ldots,n;a_1^*,\ldots,\ldots,a_{k-1}^*,a_{k+1}^*,\ldots,a_n^*)$$

Thus, there can be no linear unbiased estimate for an $n$ factor interaction, and hence, by Definition 2.3 the $n$ factor interaction is not estimable.

It should be noted that if Theorem 2.1 were applied to a problem of estimation, only a portion of the observations and one observation per design point would be employed in the estimator. The following corallary considers the more general case.

Corollary 2.1 : If there exists a set of design points $D_x$ satisfying the conditions of Theorem 2.1, perhaps with repeated observations on each design point, then there exists an independent component of interaction, utilizing all of the observations on each design point of the component, for each independent communicating class of design points constructed from $D_x$ as in the proof of Theorem 2.1.

Proof: Lemma 2.1 assures us that the set $D_x$ has a subset $D_{x_1'}$ in which each point is $R_i$ to exactly one other point for $i=1,2,\ldots,n$ . We can then partition $D_{x_1'}$ into disjoint communicating classes $C_i$ and define a linear contrast on each $C_i$ as in the proof of Theorem 2.1 with the modification that instead of letting $\ell_{\underline{a}(n)} = \pm 1$ we will let $\ell_{\underline{a}(n)} = \frac{\pm 1}{m}$ if there are $m$ observations on the design point $a(n)$. In effect we are replacing observations on the same treatment combination by the mean $\overline{x}(\underline{i}(n);\underline{a}(n))$ of these observations. Now $\sum_1^m \ell_{\underline{a}(n)} = \pm 1$ so if we denote this sum by $\ell'_{\underline{a}(n)}$ then as in Theorem 2.1 we have:

$$E\Big(\sum_{\substack{\underline{a}(n)\\ \underline{a}(x)\varepsilon C_i}} \ell_{\underline{a}(n)} X_{\underline{a}(n)}\Big) =$$

$$E\Big(\sum_{\substack{\underline{a}(n)\\ \underline{a}(x)\varepsilon C_i}} \ell'_{\underline{a}(x)} \overline{X}(\underline{i}(n);\underline{a}(n))\Big) =$$

$$\sum_{\substack{\underline{a}(n)\\ \underline{a}(x)\varepsilon C_i}} \ell'_{\underline{a}(n)} \mu(\underline{i}(n);\underline{a}(n))$$

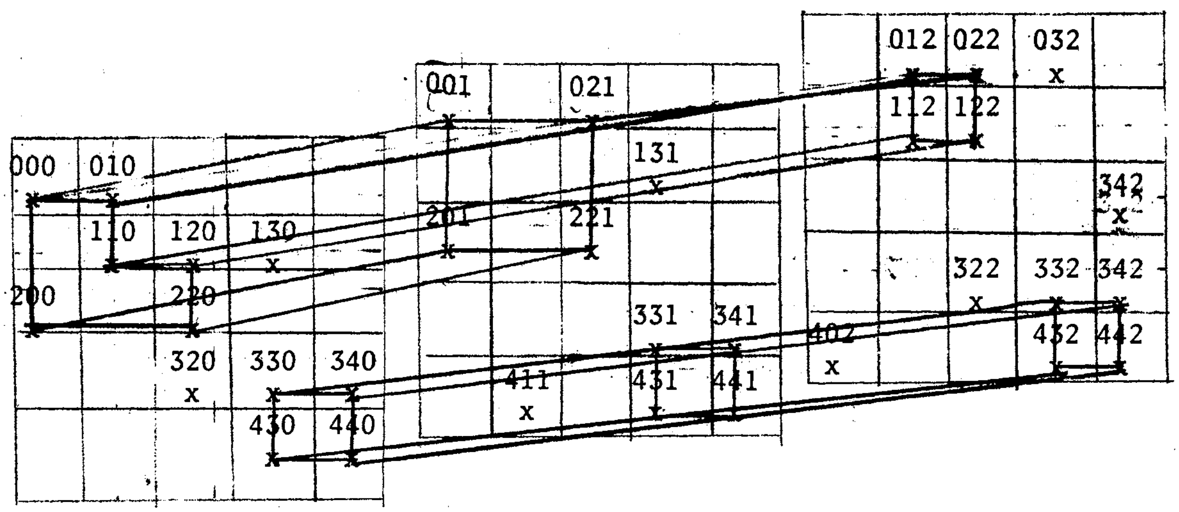
Now we can select a point $\underline{b}(n)\varepsilon D_x$ such that $\underline{b}(n)\notin C_i$ for any $i$ . We may then repeat the argument of Lemma 2.1 starting with $\underline{b}(n)$ and another point $R_i$ to it, and thus obtain another set $D_{x_2'}$ having the same properties as $D_{x_1'}$ . This is always possible since we are free to select the first pair of points, after which succeeding points must be selected so as to produce a set satisfying the requirement that each point is $R_i$ to exactly one different point of the set for $i=1,2,\ldots,n$ . Having gotten $D_{x_2'}$ we may now proceed to define a contrast of the associated observations in exactly the same manner as was done for $D_{x_1'}$ . We may continue with this procedure until all points of $D_x$ have appeared in some $D_{x_i'}$ . All contrasts obtained will be linearly independent since each involves at least one design point that is involved in no other constrast. Of course, for any constrast L (say), the quantity $\frac{L^2}{\Sigma \ell^2}$ is then a component of the sum of squares for the $a_1 \ldots, a_x$ n factor interaction, in the analysis of variance, with one degree of freedom. The number of degrees of freedom for the n factor interaction (unconfounded) will be equal to the number of linearly independent communicating sets of design points that can be obtained by the process above and the mean square for this interaction will be the total of the sums of squares of the contrasts divided by the degrees of freedom if the contrasts are orthogonal.

The following example illustrates the construction procedures discussed in the statements and proofs of Lemma 2.1 and Theorem 2.1. Suppose we had a 4×4×3 design and at least one observation was available for each of the design points:

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 000 | 010 | 001 | 021 | 012 | 022 | 032 | 110 | 120 | 131 | 112 | 122 |
| 200 | 220 | 201 | 221 | 242 | 330 | 340 | 331 | 341 | 322 | 332 | 342 |
| 430 | 431 | 411 | 440 | 441 | 402 | 432 | 442 | 320 | 130 | | |

TABLE IV

GRAPHICAL REPRESENTATION OF OCCUPIED CELLS FOR AN EXAMPLE TO ILLUSTRATE LEMMA 2.1 AND THEOREM 2.1




It is obvious from the figure above that there exists a set of points $D_{x'}$, such that each point in $D_{x'}$ has exactly one other point in $D_{x'}$ that is $R_i$ to it in each of the three mutually orthogonal directions (see traced path). In addition, all of the points in $D_{x'}$ communicate with each other. To find this set by a sorting procedure we first partition the set of design points by $R_1$, $R_2$, and $R_3$ .

| | $R_1$ | | | $R_2$ | | | $R_3$ | |
|---|---|---|---|---|---|---|---|---|
| 000 | 200 | | 000 | 010 | | 000 | 001 | |
| 010 | 110 | | 110 | 120 | 130 | 010 | 012 | |
| 120 | 220 | 320 | 200 | 220 | | 021 | 022 | |
| 130 | 330 | 430 | 320 | 330 | 340 | 032 | | |
| 340 | 440 | | 430 | 440 | | 110 | 112 | |
| 001 | 201 | | 001 | 021 | | 130 | 131 | |
| 411 | | | 131 | | | 200 | 201 | |
| 021 | 221 | | 201 | 221 | | 220 | 221 | |
| 131 | 331 | 431 | 331 | 341 | | 242 | | |
| 341 | 441 | | 411 | 431 | 441 | 320 | 322 | |
| 402 | | | 012 | 022 | 032 | 330 | 331 | 332 |
| 012 | 112 | | 112 | 122 | | 340 | 341 | 342 |
| 022 | 122 | 322 | 242 | | | 402 | | |
| 032 | 332 | 432 | 322 | 332 | 342 | 411 | | |
| 242 | 342 | 442 | 402 | 432 | 442 | 430 | 431 | 432 |
| | | | | | | 440 | 441 | 442 |

Next, all points that are alone in some set of some partition are eliminated in all partitions. These points (singly underlined above) are 411, 131, 402, 032 and 242. As a result some new points are now left alone in some set of a partition. These points (In this example only one point 130 doubly underlined above) are now similiarly eliminated from all partitions. This process is continued until no set in any partition contains a single point. The resulting set of points is the set $D_x$ referred to in the hypothesis of Lemma 2.1 and Theorem 2.1. If this set is empty, then by Theorem 2.1 the three factor interaction is not estimable. The set of points

$D_x$ for this example is listed below partitioned by $R_1$, $R_2$, and $R_3$ .

| $R_1$ | | | $R_2$ | | | $R_3$ | | |
|---|---|---|---|---|---|---|---|---|
| 000 | 200 | | 000 | 010 | | 000 | 001 | |
| 010 | 110 | | 110 | 120 | | 010 | 012 | |
| 120 | 220 | 320 | 200 | 220 | | 021 | 022 | |
| 330 | 430 | | 320 | 330 | 340 | 110 | 112 | |
| 340 | 440 | | 430 | 440 | | 120 | 122 | |
| 001 | 201 | | 001 | 021 | | 200 | 201 | |
| 021 | 221 | | 201 | 221 | | 220 | 221 | |
| 331 | 431 | | 331 | 341 | | 330 | 331 | 332 |
| 341 | 441 | | 431 | 441 | | 340 | 341 | 342 |
| 012 | 112 | | 012 | 022 | | 430 | 431 | 432 |
| 022 | 122 | 322 | 112 | 122 | | 440 | 441 | 442 |
| 332 | 432 | | 322 | 332 | 342 | 320 | 322 | |
| 342 | 442 | | 432 | 442 | | | | |

We now begin to eliminate all points except two from every set containing more than two points. As a result of the elimination of a point other points which would then be left alone in some set of some partition would also be eliminated. For example the elimination of 320 (singly underlined above) necessitates the elimination of 322 (doubly underlined). Elimination of 332 (singly underlined) results in the elimination of 342 and 432 (doubly underlined) which in turn result in the elimination of 442 (triply underlined). The remaining points (partitioned below) are now such that each set in the partitions by $R_1$, $R_2$, and $R_3$ contains exactly two points (each point is $R_i$

to exactly one other point i=1,2,3). This is the set of points $D_{x'}$, referred to in Lemma 2.1 and Theorem 2.1.

| $R_1$ | | $R_2$ | | $R_3$ | |
|---|---|---|---|---|---|
| 000 | 200 | 000 | 010 | 000 | 001 |
| 010 | 110 | 110 | 120 | 010 | 012 |
| 120 | 220 | 200 | 220 | 021 | 022 |
| 001 | 201 | 001 | 021 | 110 | 112 |
| 021 | 221 | 201 | 221 | 200 | 201 |
| 012 | 112 | 012 | 022 | 220 | 221 |
| 022 | 122 | 112 | 122 | 120 | 122 |
| 330 | 430 | 330 | 340 | 330 | 331 |
| 340 | 440 | 430 | 440 | 340 | 341 |
| 331 | 431 | 331 | 341 | 430 | 431 |
| 341 | 441 | 431 | 441 | 440 | 441 |

We now note that certain points of the remaining set communicate with each other (e.g. 000 $R_1{}^o R_2{}^o R_3$ 221) while others do not (e.g. 000 does not communicate with 330). We then partition this set into disjoint communicating subclasses;

L = (000 010 110 120 200 220 001 021 201 220
012 022 112 122) and

M = (330 340 430 440 331 341 431 441)

We can also get different communicating sets;

N = (330 340 430 440 332 342 432 442)

0 = (331 341 431 441 332 342 432 442)

by eliminating a different choice in the process of getting a set of points $D_{x'}$ such that each point in $D_{x'}$ is $R_i$, $i=1,2,3$ to exactly one other point of $D_{x'}$. But $O = (M \cup N) - (M \cap N)$ so $O$ is not independent of sets $M$ and $N$. Sets $L$, $M$ and $N$ however are independent of each other and by the contrast defined in Theorem 2.1 yield three independent estimates of three factor interaction. In addition the squares of these contrasts with the appropriate divisors will account for 3 degrees of freedom for the ABC interaction in an analysis of variance table.

## Sufficient Conditions for the Estimability of an $\alpha$ Factor Interaction

Theorem 2.2: In an n-way cross classification of cells, if a model is assumed in which all higher order interaction effects involving factors $i_1,\ldots,i_\alpha$ are zero, the $\alpha$ factor interaction of factors $i_1,\ldots,i_\alpha$ is estimable if for some combination $a^*_{i_{\alpha+1}},\ldots,a^*_{i_n}$ (fixed) of the levels of factors $i_{\alpha+1},\ldots,i_n$ there exists a subset $D_{x*}$ of the design points such that for each $a_{\underline{i}(n)}$ in $D_{x*}$, $a_{i_{\alpha+1}}=a^*_{i_{\alpha+1}},\ldots,a_{i_n}=a^*_{i_n}$ and there is a $b_{\underline{i}(n)}$ in $D_{x*}$ such that $a_i \neq b_i$ and $a_{\underline{i}(n)}\ R_i\ b_{\underline{i}(n)}$ for $i=1,2,\ldots,\alpha$.

Proof: Since we can rearrange the order of factors, let us assume, without loss of generality, that $i_1,\ldots,i_\alpha$ are the first factors $1,2,\ldots,\alpha$. If we ignore for the moment the factors $\alpha+1,\ldots,n$, then we know by Theorem 2.1 that the interaction of factors $1,2,\ldots,\alpha$ can be estimated free of the lower order effects and interactions of these factors if and only if there exists a subset $D_x$ of the design points such that for each $\underline{a}(n)$ in $D_x$ and each $i=1,2,\ldots,\alpha$ there is a $\underline{b}(n)$ in $D_x$ such that $a_i \neq b_i$ and $\underline{a}(n)\ R_i\ \underline{b}(n)$. We established in Lemma 2.1 that if such a set $D_x$ existed we could find a subset $D_{x'}$ such that for each $\underline{a}(n)$ in $D_{x'}$ there is exactly one $\underline{b}(n)$ in $D_{x'}$ such that for $i=1,2,\ldots,\alpha$, $a_i \neq b_i$ and $\underline{a}(n)\ R_i\ \underline{b}(n)$, and that $D_{x'}$ could be partitioned into disjoint communicating subsets $C_i$. On each $C_i$ we were able to define a contrast that gave us an estimate of the interaction of factors $1,2,\ldots,\alpha$ free of all main effects and lower order interactions of these factors.

In such a contrast if the coordinates $a_{\alpha+1},\ldots,a_n$ are the same for all design points involved, then the expected value of this contrast will not involve any main effects and interactions of factors $\alpha+1,\ldots,n$, since by definition $\sum_{a_i} \ell_{\underline{a}(n)} = 0$, for $i=1,2,\ldots,n$ For any interaction effect of a combination of $k < \alpha$ factors out of $1,2,\ldots,\alpha$, say $i_1,\ldots,i_k$, with a combination of $p \leq \alpha-k$ factors out of $\alpha+1,\ldots,n$, say $i_{k+1},\ldots,i_{k+p}$, the expected value of our linear combination yields the following for this interaction effect:

$$\sum_{a_1}\cdots\sum_{a_n} \ell_{\underline{a}(n)}\, \mu(i_1,\ldots,i_k,i_{k+1},\ldots,i_{k+p};a_{i_1},\ldots,a_{i_k},a^*_{i_{k+1}},a^*_{i_{k+p}})$$

$$= \left[\sum_{a_{i_1}}\cdots\sum_{a_{i_{k+p}}} \ell_{\underline{a}(n)}\, \mu(i_1,\ldots,i_{k+p};a_{i_1},\ldots,a^*_{i_{k+p}})\right]\left[\sum_{a_{i_{k+p+1}}}\cdots\sum_{a_{i_n}} \ell_{\underline{a}(n)}\right]$$

$$= 0$$

since at least one of the factors $1,2,\ldots,\alpha$ is among $i_{k+p+1},\ldots,i_n$ and the sum of the coefficients over this factor is zero by the definition of our contrast.

The following example illustrates Theorem 2.2.

Suppose we have observations for the points 1001, 1000, 1002, 1101, 1100, 1011, 1111, 1112, 0010, 0012, 0110, 0002, 0102 in a a $2^3 \times 3$ layout and designate the four factors by A, B, C and D . Let the sets of parameters be $\alpha_i$ , $\beta_j$ , $\Gamma_k$ and $\delta_\ell$ with i=0,1 , j=0,1 , k=0 , and $\ell$=0,1,2 . Applying Theorem 2.1 for the ABCD interaction we find that the sorting procedure eliminates all points, so ABCD is not estimable.

Applying Theorem 2.2 for the ABC interaction and sorting at the 0 level of D we get 1000, 1100, 0010, 0110 . This set cannot be connected since at least 8 points are needed. For ABC at the 1 level of D we get 1001, 1101, 1011, 1111 . Again this set is not connected since we need 8 points.

For ABC at the 2 level of D we get 1002, 1112, 0012, 0002, 0102 and again this set is not connected.

Similarly for ABD for fixed levels of C , no sets have 8 points and for ACD for fixed B no sets have 8 points. Thus none of ABC, ABD, or ACD are estimable.

For the 0 level of A sorting for BCD yields a set with less than 8 points and for the 1 level of D we get 1001, 1000, 1002, 1101, 1100, 1011, 1111, 1112 but there is no connected set sorting on B, C and D. Thus no 3 factor interaction is estimable.

Using the notation $XY_{ij}$ to indicate the set of points at level i of factor X and level j of factor Y , we apply Theorem 2.2 and get the following sets:

| For $AC_{10}$ | For $AC_{11}$ | For $AC_{01}$ | For $AC_{00}$ |
|---|---|---|---|
| 1001 | 1011 | 0010 | 0002 |
| 1000 | 1111 | 0012 | 0102 |
| 1002 | 1112 | 0110 | |
| 1101 | | | |
| 1100 | | | |

Ignoring AC for $AC_{10}$ we find that 01,00,11,10 are connected so BD is estimable. In fact $X_{1001}-X_{1000}-X_{1101}+X_{1100}$ estimates $\beta\delta_{01}-\beta\delta_{00}-\beta\delta_{11}+\beta\delta_{10}$ free of all other effects and interactions in the absence of 3 and 4 factor interactions.

| For $AD_{11}$ | For $AD_{10}$ | For $AD_{12}$ | For $AD_{00}$ | For $AD_{02}$ | For $AD_{01}$ |
|---|---|---|---|---|---|
| 1001 | 1000 | 1002 | 0010 | 0002 | no points |
| 1101 | 1100 | 1112 | 0110 | 0012 | |
| 1011 | | | | 0102 | |
| 1111 | | | | | |

Ignoring AD for $AD_{11}$ the points listed above are connected so BC is estimable.

| For $BC_{00}$ | $BC_{10}$ | $BC_{01}$ | $BC_{11}$ |
|---|---|---|---|
| 1001 | 0102 | 1011 | 1111 |
| 1000 | 1101 | 0010 | 1112 |
| 1002 | 1100 | 0012 | 0110 |
| 0002 | | | |

None of the above sets are connected.

For BD and AB and CD similarly we get no connected sets.

The breakdown of the degrees of freedom in the analysis of variance could be as follows:

| AOV | | |
|---|---|---|
| | A | 1 |
| | B | 1 |
| | C | 1 |
| | D | 2 |
| | BC | 1 |
| | BD | 1 |
| Confounded Interactions | | 5 |
| Tot. | | n-1=12 |

## Necessary and Sufficient Conditions for the Estimability of an n-1 Factor Interaction

The problem of finding estimable components of interaction would be considerably simplified if the conditions for estimability of an $\alpha$ factor interaction stated in the previous theorem were necessary as well as sufficient. Unfortunately this is not the case as the following example illustrates.

Suppose that observations were available on the eight design points 0001, 0101, 1000, 1100, 0010, 0110, 1011 and 1111 of a $2^4$ factorial experiment with factors A,B,C and D each at two levels. Obviously the 4 factor interaction is not estimable and it is easily verified that the 3 factor interactions are confounded with one another. Consider then, the estimability of 2 factor interactions in the absence of higher order interactions.

The expected value of

$$\frac{1}{2}(X_{0001}+X_{1100}+X_{0010}+X_{1111}-X_{0101}-X_{1000}-X_{0110}-X_{1011})$$

is $\alpha\beta_{00}+\alpha\beta_{11}-\alpha\beta_{01}-\alpha\beta_{10}$ so the AB interaction is estimable and yet the conditions of Theorem 2.2 are not satisfied for this two factor interaction. Thus the conditions of the theorem are sufficient but not necessary.

A careful look at this example will reveal that it was constructed by forming a connected set on AB at the 0 level of C and the same connected set on AB at the 1 level of C and then switching the levels of D so that AB will be connected at the 0 level of D and also at the 1 level but not for a constant level of CD.

at 0 level of C

| A | B | C | D |
|---|---|---|---|
| 0 | 0 | 0 | 1 |
| 0 | 1 | 0 | 1 |
| 1 | 0 | 0 | 0 |
| 1 | 1 | 0 | 0 |

at 1 level of C

| A | B | C | D |
|---|---|---|---|
| 0 | 0 | 1 | 0 |
| 0 | 1 | 1 | 0 |
| 1 | 0 | 1 | 1 |
| 1 | 1 | 1 | 1 |

| at 0 level of D | | | |
|---|---|---|---|
| A | B | C | D |
| 0 | 0 | 1 | 0 |
| 0 | 1 | 1 | 0 |
| 1 | 0 | 0 | 0 |
| 1 | 1 | 0 | 0 |

| at 1 level of D | | | |
|---|---|---|---|
| A | B | C | D |
| 0 | 0 | 0 | 1 |
| 0 | 1 | 0 | 1 |
| 1 | 0 | 1 | 1 |
| 1 | 1 | 1 | 1 |

This construction could not have been carried out without two factors to switch around as we did. This result is the content of the next theorem which merely states that the conditions are necessary as well as sufficient for the estimability of a $n-1$ factor interaction. The more general and complex case for any $\alpha$ factor interaction is considered in Theorem 2.4.

<u>Theorem 2.3</u>: In an n-way cross classification of cells, if a model is assumed in which the $n$ factor interaction is zero, then the interaction of any $n-1$ factors without loss of generality say classes $1,2,\ldots,n-1$, is estimable if and only if there exists a subset $D_x$ of the design points satisfying:

1. For each $\underline{a}(n)$ in $D_x$ we have $a_n = a_n^*$ (fixed).
2. For each $\underline{a}(n)$ in $D_x$, there exists $\underline{b}(n)$ in $D_x$ such that $\underline{a}(n)\ R_i\ \underline{b}(n)$ with $a_i \neq b_i$ for $i=1,2,\ldots,n-1$.

<u>Proof</u>: If conditions (1) and (2) are satisfied for some set $D_x$ then by Theorem 2.2 the interaction of factors $1,2,\ldots,n-1$ is estimable.

If the interaction of factors $1,2,\ldots,n-1$ is estimable then ignoring $a_n$ for the moment we know by Theorem 2.1 that condition (2) must be satisfied for $i=1,2,\ldots,n-1$ or else the interaction of factors $1,2,\ldots,n-1$ would be confounded with some main effect or lower order interaction of these factors. Thus there must exist a subset $D_{x'}$ of the design points satisfying condition (2). Let

$\underline{a}^*(n)$ be an element of $D_{x'}$ and suppose for some $i=1,\ldots,n-1$ and each $\underline{b}(n)$ in $D_{x'}$ such that $\underline{b}(n)\ R_i\ \underline{a}^*(n)$ with $a_i \neq b_i$, we have $b_n \neq a_n^*$. Then there can be no component of interaction of factors $1,2,\ldots,n-1$ whose expected value does not involve $\mu(1,2,\ldots,i-1, i+1,\ldots,n;\ a_1^*,\ldots,a_{i-1}^*,a_{i+1}^*,\ldots,a_n^*)$. But this is a contradiction since the interaction of factors $1,2,\ldots,n-1$ was assumed to be estimable.

Thus for each $i=1,2,\ldots,n-1$ there exists $\underline{b}(n)$ in $D_{x'}$, such that $b_n=a_n^*$ and $\underline{b}(n)\ R_i\ \underline{a}^*(n)$ with $a_i^* \neq b_i$. Since the above argument holds for any $\underline{a}(n)$ and $a_n=a_n^*$ fixed, we have established the existence of a set $D_x \subseteq D_{x'}$ satisfying conditions (1) and (2) of the theorem. Hence these conditions are necessary as well as sufficient and the theorem is proved.

The following example illustrates Theorem 2.3. Suppose we had observations for the design points pictured and listed below:

TABLE V

GRAPHICAL REPRESENTATION OF OCCUPIED CELLS FOR AN EXAMPLE TO ILLUSTRATE LEMMA 2.1 AND THEOREM 2.3

B

| A \ B | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| 0 | 000 x | | | |
| 1 | | 110 x | 120 x | |
| 2 | 200 x | | | |
| 3 | | 310 x | 320 x | 330 x |

$C_0$

| | | | |
|---|---|---|---|
| 001 x | | | 031 x |
| 101 x | 111 x | 121 x | 131 x |
| | | 221 x | 231 x |
| | | 321 x | 331 x |

$C_1$

| | | | |
|---|---|---|---|
| | | 022 x | 032 x |
| 102 x | 112 x | | |
| 202 x | 212 x | 222 x | 232 x |
| | 312 x | | |

$C_2$

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Points | 000 | 001 | 031 | 022 | 032 | 110 | 120 | 101 | 111 |
| | 121 | 131 | 102 | 112 | 200 | 221 | 231 | 202 | 222 |
| | 232 | 310 | 320 | 330 | 321 | 331 | 312 | 212 | |

all points drop out sorting for ABC so the 3 factor interaction is not estimable. For AB at $C_0$ we get the set

(1) 110, 120, 310, 320 satisfying Theorem 2.3 since 11, 12, 31, and 32 are connected. (Note: The contrast $X_{110}-X_{120}+X_{320}-X_{310}$ estimates $\alpha\beta_{11}-\alpha\beta_{12}+\alpha\beta_{32}-\alpha\beta_{31}$).

For AB at $C_1$ we get the sets

(1) 001 031 101 131 which is connected according to Theorem 2.3

(2) 121 131 221 231 which is connected according to Theorem 2.3

(3) 221 231 321 331 which is connected according to Theorem 2.3

For AB at $C_2$ we get

(1) 022 032 222 232 which is connected according to Theorem 2.3

(2) 102 202 212 112 which is connected according to Theorem 2.3

All of the 6 sets above are independent when we ignore C, so we have 6 linearly independent estimates of AB interaction.

For AC at $\beta_0$

(1) 000 200 001 101 102 202

For AC at $\beta_1$

(1) 110 310 112 312

For AC at $\beta_2$

(1) 120 320 121 321

For AC at $\beta_3$

(1) 031 231 032 232

All of the 4 above sets are connected and independent ignoring B so we have 4 linearly independent estimates of AC interaction.

For BC at $A_0$ no set exists.

For BC at $A_1$

(1) 110 120 111 121

(2) 101 111 102 112

For BC at $A_2$

(1) 221 231 222 232

For BC at $A_3$

(1) 321 331 320 330

All of the 4 sets are connected and independent ignoring A so we have 4 linearly independent estimates of BC interaction.

The breakdown of the degrees of freedom in the analysis of variance could be as follows:

| AOV | Source | d.f. |
|---|---|---|
| | A | 3 |
| | B | 3 |
| | C | 2 |
| | AB | 6 |
| | AC | 4 |
| | BC | 4 |
| | Confounded Interactions | 3 |
| | Total-mean | 25 |

Corollary 2.3: If the interaction of factors $1,2,\ldots,\alpha$ is estimable according to either of Theorems 2.1, 2.2 or 2.3 then any lower order interaction of any subset of these factors is also estimable.

Proof: Without loss of generality when considering the interaction of any $\alpha$ factors we will consider the first $\alpha$. If we let $\alpha=n$, then Theorem 2.2 is satisfied vacuously by Theorem 2.1 and if we let $\alpha=n-1$ then the conditions of Theorem 2.3 imply Theorem 2.2. Thus for all values of $\alpha$ the conditions of Theorem 2.2 are satisfied if an interaction is estimable according to either of Theorems 2.1, 2.2, or 2.3. Hence there exists a set of points $D_{x*}$ connected in the sense of Theorem 2.1 over factors $1,2,\ldots,\alpha$ for some fixed combination $a^*_{i_{\alpha+1}},\ldots,a^*_{i_n}$ of the other factors. Consider then the interaction of factors $1,2,\ldots,\alpha-1$. For any point in $D_{x*}$ let $a^*_\alpha$ be the level of factor $\alpha$ and let $D_{y*}$ be the set of points in $D_{x*}$ for which $a_\alpha = a^*_\alpha$. Now since for each $a_{\underline{i}(n)}$ in $D_{x*}$ there is a $b_{\underline{i}(n)}$ in $D_{x*}$ such that $a_i \neq b_i$ and $a_{\underline{i}(n)}\ R_i\ b_{\underline{i}(n)}$ $i=1,2,\ldots,\alpha$. Then certainly for each $a_{\underline{i}(n)}$ in $D_{y*}$ there is a $b_{\underline{i}(n)}$ in $D_{y*}$ such that $a_{\underline{i}(n)}\ R_i\ b_{\underline{i}(n)}$, $a_i \neq b_i$ $i=1,2,\ldots,\alpha-1$. Thus the points of $D_{y*}$ are connected over factors $1,2,\ldots,\alpha-1$ for some fixed combinations of the other factor and by Theorem 2.2 the $\alpha-1$ factor interaction is estimable. By induction it thus follows that all interactions of any subset of the factors $1,2,\ldots,\alpha$ are estimable.

Corollary 2.4: If the $k$ factor interaction of factors $i_1,\ldots,i_k$ is estimable then, for any level of any other factor $i_j$ $j \neq i,\ldots,k$ involved in the estimate, there must exist a connected set of points over factors $i_1,\ldots,i_k$ for which the level of factor $i_j$ remains constant.

Proof: The conditions stated in this corollary were proven to be necessary for estimability in the proof of Theorem 2.3 for $k=n-1$ and $i_j=n$.

However, the proof there did not depend on how many factors were involved in the interaction or which factor remained constant, and in no way necessitated the involvement of any other factors. Hence the same argument can be used to establish this corollary.

## Necessary and Sufficient Conditions for the Estimability of an $\alpha$ Factor Interaction

Theorem 2.4: In an n-way classification, if a model is assumed in which all higher order interaction effects are zero, then the $\alpha$ factor interaction of factors $1,2,\ldots,\alpha < n$ is estimable if and only if we can find a subset of design points $C \subseteq D_x$ satisfying the following conditions:

(1) If we ignore the coordinates $\alpha+1,\ldots,n$ then the points of $C$ are connected in the sense of Theorem 2.1 over factors $1,2,\ldots,\alpha$ .

(2) The linear combination over the points of $C$ described in Lemma 2.2 is such that if $\underline{a}^*(\alpha) = (a^*_{i_1} \ldots a^*_{i_\alpha})$ is any combination of $\alpha$ coordinates of $\underline{a}(n)$ except $a_1,a_2,\ldots,a_\alpha$ and $C(a^*_{i_1},\ldots,a^*_{i_\alpha})$ is the set of all points in $C$ having $a_{i_1} = a^*_{i_1},\ldots,a_{i_\alpha} = a^*_{i_\alpha}$ then

$$\sum_{\underline{a}(n)\ \varepsilon\ C(\underline{a}^*(\alpha))} \ell_{\underline{a}(n)} = 0$$

Proof: Condition (1) is necessary by the argument of Theorem 2.1 and if condition (2) is not fulfilled then

$$\sum_{\underline{a}(n)} \ell_{\underline{a}(n)} \mu(\underline{i}(\alpha);\underline{a}(i(\alpha))) \neq 0$$

for all combinations $a_{i_1} \ldots a_{i_\alpha}$ (except $a_1,\ldots,a_\alpha$ ) and the interaction of factors $1,2,\ldots,\alpha$ is confounded with another $\alpha$ factor interaction. Thus if either one of the conditions (1) or (2) is not satisfied then the $\alpha$ factor interaction effect of factors $1,2,\ldots,\alpha$ is not estimable. Hence the conditions are necessary.

That the conditions (1) and (2) are sufficient is also easily seen since we have already shown in Lemma 2.1 that if condition (1) is satisfied there exists a linear combination of the observations that estimates the $\alpha$ factor interaction effect of factors $1,2,\ldots,\alpha$ free of lower order interaction effects and main effects of these factors. Ignoring factors $\alpha+1,\ldots,n$, conditions (1) and (2) assure us that the expected value of this linear combination will not involve any other $\alpha$ factor interaction effects. Also since the sum of the coefficients is 0 over all coefficients associated with design points having $\alpha$ coordinates identical, for all possible combinations of $\alpha$ out of $1,2,\ldots,n$ except $1,2,\ldots,\alpha$ then on each of these sets all combinations of $k < \alpha$ coordinates will also be constant and consequently an equivalent statement to (2) holds for $k < \alpha$. Thus:

$$E\left[\sum_{\underline{a}(n)} \ell_{\underline{a}(n)} x_{\underline{a}(n)}\right] = \sum_{\underline{a}(n)} \ell_{\underline{a}(n)} \mu(\underline{\alpha};\underline{a}(\alpha))$$

and by our definition the $\alpha$ factor interaction of factors $1,2,\ldots,\alpha$ is estimable.

The following example illustrates Theorem 2.4: Suppose we had observations for the design points 0000, 0101, 1000, 1101, 0011, 0110, 1011, and 1110 . The contrasts for each interaction as defined in the proof of Theorem 2.1 are listed below:

| 0000 | 0101 | 1000 | 1101 | 0011 | 0110 | 1011 | 1110 | |
|---|---|---|---|---|---|---|---|---|
| +1 | -1 | -1 | +1 | -1 | +1 | +1 | -1 | ABC |
| +1 | +1 | -1 | -1 | -1 | -1 | +1 | +1 | ABD |
| +1 | -1 | -1 | +1 | +1 | -1 | -1 | +1 | ACD |
| no connected set | | | | | | | | BCD |
| +1 | -1 | -1 | +1 | +1 | -1 | -1 | +1 | AB |
| +1 | -1 | +1 | -1 | +1 | -1 | +1 | -1 | CD |
| +1 | +1 | -1 | -1 | -1 | -1 | +1 | +1 | AC |
| +1 | -1 | -1 | +1 | -1 | +1 | +1 | -1 | AD |
| +1 | -1 | +1 | -1 | -1 | +1 | -1 | +1 | AD |
| +1 | +1 | +1 | +1 | -1 | -1 | -1 | -1 | BD |

Obviously the set of points is not connected for ABCD since at least $2^4$ are needed. Applying Theorem 2.3 we see that the set is connected for ABC but not for a fixed level of D. Thus

$$\begin{aligned} &E(X_{0000}-X_{0101}-X_{1000}+X_{1101}-X_{0011}+X_{0110}+X_{1011}-X_{1110}) \\ &= \alpha\beta\Gamma_{000}-\alpha\beta\Gamma_{010}-\alpha\beta\Gamma_{100}+\alpha\beta\Gamma_{110}-\alpha\beta\Gamma_{001}+\alpha\beta\Gamma_{011}+\alpha\beta\Gamma_{101}-\alpha\beta\Gamma_{110} \\ &+ 2\alpha\delta_{00}+2\alpha\delta_{11}-2\alpha\delta_{01}-2\alpha\delta_{10} \end{aligned}$$

so ABC is confounded with AD. Similarly ABD is connected but not for fixed C and is confounded with AC. Likewise ACD is connected but not for fixed B and is confounded with AB, and there is no connected set for BCD.

If all higher order interactions are zero;

Checking for AB we find a connected set on AB satisfying Theorem 2.4. This is evident since the set (00, 01, 10, 11) is connected over constant levels of C (0 & 1) and D(0 & 1). Thus

$$E(X_{0000}-X_{0101}-X_{1000}+X_{1101}+X_{0011}-X_{0110}-X_{1011}+X_{1110})=2(\alpha\beta_{00}+\alpha\beta_{11}-\alpha\beta_{01}-\alpha\beta_{10})$$

Similarly checking for AC and AD we find them estimable. However checking BC, BD and CD we find connected sets that do not satisfy condition (2) of Theorem 2.4. Then for example a contrast on BC yields

$$E(X_{0000}+X_{1000}+X_{0110}+X_{1110}-X_{0101}-X_{1101}-X_{0011}-X_{1011})$$
$$=2(\beta\Gamma_{00}+\beta\Gamma_{11}-\beta\Gamma_{10}-\beta\Gamma_{01})+2\alpha\delta_{00}+2\alpha\delta_{10}-2\alpha\delta_{01}-2\alpha\delta_{11}$$
$$+2\beta\delta_{00}+2\beta\delta_{10}-2\beta\delta_{11}-2\beta\delta_{01}+2\Gamma\delta_{00}+2\Gamma\delta_{10}-2\Gamma\delta_{01}-2\Gamma\delta_{11}$$

So BC is confounded with other 2 factor interactions. Likewise BD and CD are confounded. The degrees of freedom for the analysis of variance could be partitioned as follows:

AOV

| Source | d.f. |
|---|---|
| A | 1 |
| B | 1 |
| C | 1 |
| AB | 1 |
| AC | 1 |
| AD | 1 |
| Confounded | 1 |
| Total-mean | 7 |

Suppose that our n-way cross classification design is found (or assumed) to be free of all interaction. We then know that all main effects are estimable if and only if all differences $\mu(i,a_i)-\mu(i;b_i)$ are estimable $a_i$ , $b_i=a_0,\ldots,a_{t_i-1}$, $a_i\neq b_i$ , $i=1,2,\ldots,n$ (there again we drop all parameters from the model if all observations for the corresponding factors are missing.)

The following theorem gives us necessary and sufficient conditions for the estimability of a difference $\mu(i;a_i)-\mu(i;b_i)$ , $a_i\neq b_i$ . Thus the theorem provides us with a means of determining exactly which main effects will be estimable and which will not.

## Necessary and Sufficient Conditions for the Estimability of a Simple Effect

Theorem 2.5: In an n-way cross classification if a model is assumed with no interaction let

$$D_a = \{\underline{a}(n) \mid \underline{a}(n) \varepsilon D_x \text{ and } a_k = a\}$$
$$D_b = \{\underline{a}(n) \mid \underline{a}(n) \varepsilon D_x \text{ and } a_k = b\}$$

where $a$ and $b$ are chosen from $0,1,2,\ldots,t_{k-1}$ , $a \neq b$ . Then the difference $\mu(k;a)-\mu(k;b)$ is estimable if and only if there exists constant coefficients $\ell_{\underline{a}(n)}$ for all $\underline{a}(n) \varepsilon D_a \cup D_b$ such that

$$\sum_{a_k=a} \ell_{\underline{a}(n)} = -\sum_{a_k=b} \ell_{\underline{a}(n)} = c \text{ (say) and } \sum_{a_i} \ell_{\underline{a}(n)} = 0 \text{ for } i \neq k$$

Proof: The proof is quite obvious since under the conditions of the theorem

$$E\left(\sum_{a_1} \ldots \sum_{a_n} \ell_{\underline{a}(n)} x_{\underline{a}(n)}\right) = \pm c(\mu(k;a)-\mu(k,b)) + \sum_{i \neq k} \sum_{a_i} \ell_{\underline{a}(n)} \mu(i;a_i)$$
$$= \pm c(\mu(k;a)-\mu(k;b)) + \sum_{i \neq k} \mu(i;a) \sum_{a_i} \ell_{\underline{a}(n)}$$
$$= \pm c(\mu(k;a)-\mu(k;b))$$

Hence $\pm \frac{1}{c} \sum_{a_1} \ldots \sum_{a_n} \ell_{\underline{a}(n)} x_{\underline{a}(n)}$ is an unbiased estimate of $\mu(k;a)-\mu(k;b)$.

Conversely if $\sum_{a_k=a} \ell_{\underline{a}(n)} = c_1$ and $\sum_{a_k=b} \ell_{\underline{a}(n)} = c_2$ then

$\sum_{a_k=a} - \frac{c_2}{c_1} \ell_{\underline{a}(n)} = -\sum_{a_k=b} \ell_{\underline{a}(n)} = -c_2$ so we need only consider the case where $\sum_{a_k=a} \ell_{\underline{a}(n)} = -\sum_{a_k=b} \ell_{\underline{a}(n)}$ .

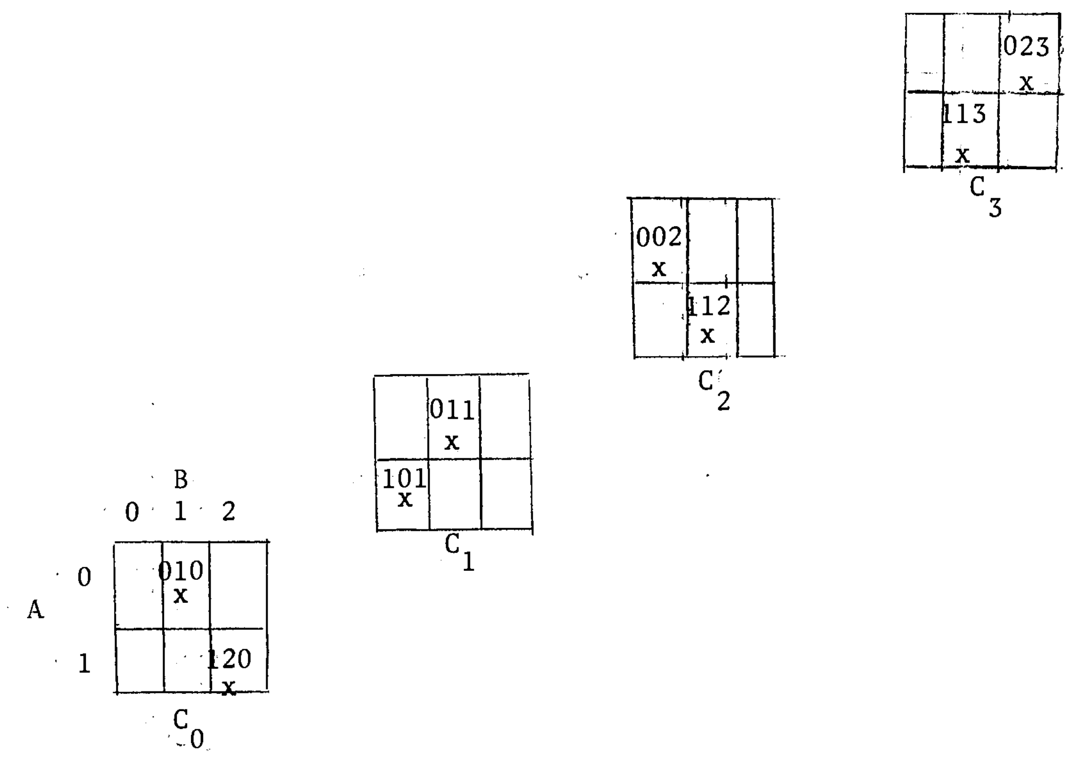
If for all sets of constant coefficients $\ni \sum_{a_k=a} \ell_{\underline{a}(n)} = -\sum_{a_k=b} \ell_{\underline{a}(n)}$ we have $\sum_{a_i=j} \ell_{\underline{a}(n)} \neq 0$ for some $j$ which is the ith coordinate of a point in $D_a \cup D_b$ then $E\left[\sum_{a_1} \ldots \sum_{a_n} \ell_{\underline{a}(n)} x_{\underline{a}(n)}\right]$ will involve $\mu(i;j)$

and hence the difference $\mu(k;a)-\mu(k;b)$ is confounded with $\mu(i;j)$.

As an example of Theorem 2.5, suppose a model with no interaction is assumed with 2 levels of a factor A, 3 levels of a factor B and 4 levels of a factor C and we had observations for the design points 010, 120, 011, 101, 002, 112, 023, and 113.

TABLE VI

GRAPHICAL REPRESENTATION OF OCCUPIED CELLS FOR AN EXAMPLE TO ILLUSTRATE THEOREM 2.5




Examining the simple effect $\alpha_0-\alpha_1$ we find that there are 4 points at the 0 level of A and 4 points at the 1 level, a 1:1 ratio. Checking levels of B ignoring C (see Table VII on next page), we find $B_0$ represented once at each of the levels of A, $B_1$, represented twice at each level and $B_2$ represented once at each level, all in the same 1:1 ratio as the number of points of $A_0$ to points $A_1$.

TABLE VII

OCCUPIED CELLS FOR THE POINTS OF TABLE VI
FIRST IGNORING C AND THEN IGNORING B

| A \ B | 0 | 1 | 2 |
|---|---|---|---|
| 0 | x | xx | x |
| 1 | x | xx | x |

Figure 1

| A \ C | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| 0 | x | x | x | x |
| 1 | x | x | x | x |

Figure 2

A check of levels of C ignoring B (figure 2 above) reveals each level of C represented once at each level of A, the same 1:1 ratio. Hence an unbiased estimate of $\alpha_0-\alpha_1$ is found by taking 1/4 of the contrast that assigns 1 as a coefficient to those points at the 0 level of A and -1 to those at the 1 level.

For the simple effect $\beta_0-\beta_1$ we find 2 points 101 and 002 at the 0 level of B and 4 points 010, 011, 112, and 113 at the 1 level. Ignoring levels of C we find $A_0$ represented 1:2 and $A_1$ represented 1:2 in the same ratio as $B_0:B_1$ (see figure 1 below). Ignoring A we find $C_0$ and $C_3$ represented at the 1 level of B but not at the 0 level. (See Table VIII).

## TABLE VIII

OCCUPIED CELLS FOR THE POINTS OF TABLE VI AT THE 0 & 1 LEVEL OF B FIRST IGNORING C AND THEN A

| A \ B | 0 | 1 |
|---|---|---|
| 0 | x | xx |
| 1 | x | xx |

Figure 1

| C \ B | 0 | 1 |
|---|---|---|
| 0 | | x |
| 1 | x | x |
| 2 | x | x |
| 3 | | x |

Figure 2

The points 010 and 113 must therefore be eliminated. The remaining set of points has 2 points at the 0 level of B and 2 at the 1 level. The 0 and 1 levels of A are now represented once each at the 0 and 1 levels of B as are the 1 and 2 levels of C. Thus $\beta_0 - \beta_1$ is estimated by 1/2 of the contrast assigning 1's to the points at the 0 level of B and -1's to points at the 1 level.

A similar analysis for the remaining simple effects yields the contrasts given in the table below.

| 010 | 120 | 011 | 101 | 002 | 112 | 023 | 113 | Effects Estimated |
|---|---|---|---|---|---|---|---|---|
| +1 | -1 | +1 | -1 | +1 | -1 | +1 | -1 | $4(A_0 - A_1)$ |
| 0 | 0 | -1 | +1 | +1 | -1 | 0 | 0 | $2(B_0^0 - B_1^1)$ |
| +1 | -1 | 0 | 0 | 0 | 0 | -1 | +1 | $2(B_1^0 - B_2^1)$ |
| +1 | +1 | 0 | 0 | 0 | 0 | -1 | -1 | $2(C_0^1 - C_3^2)$ |
| 0 | 0 | +1 | +1 | -1 | -1 | 0 | 0 | $2(C_1^0 - C_2^3)$ |
| +1 | 0 | -1 | 0 | 0 | 0 | 0 | 0 | $(C_0^1 - C_1^2)$ |
| -1 | +1 | +1 | -1 | +1 | -1 | -1 | +1 | Error |
| +1 | +1 | +1 | +1 | +1 | +1 | +1 | +1 | Mean |

The above contrasts account for all of the degrees of freedom as summarized in the table below:

AOV

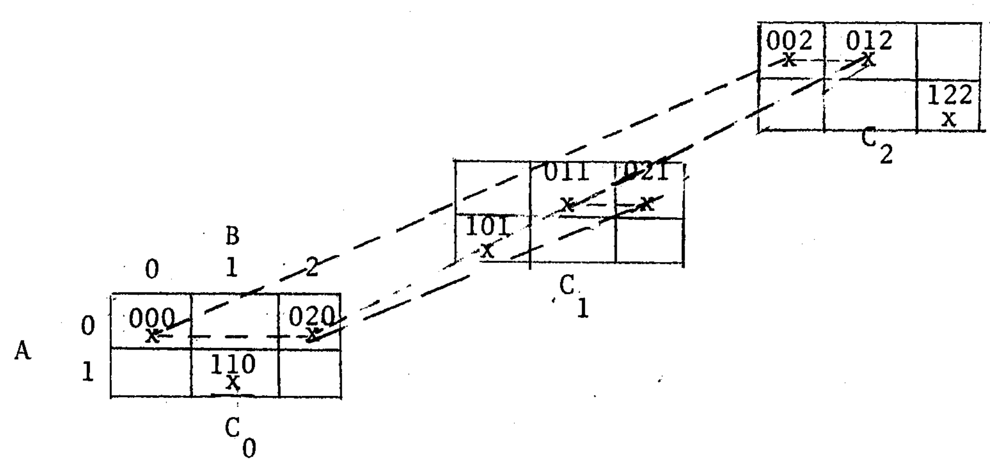
| Source | d.f. |
|---|---|
| A | 1 |
| B | 2 |
| C | 3 |
| Error | 1 |
| Tot. Mean | 7 |

If a model with interaction is assumed then, of course, the differences $\mu(k;a)-\mu(k;b)$ are not estimable free of the interaction effects. If, due to numerous missing observations, only minimal information can be obtained about the interactions, then, as throughout this chapter, we attempt to obtain as many unconfounded estimates as the data permits.

For an example, utilizing Theorem 2.5 where a model with interaction is assumed, suppose we had observations for the points 000, 020, 011, 021, 002, 012, 110, 101, 122.

TABLE IX

GRAPHICAL REPRESENTATION OF OCCUPIED CELLS FOR A SECOND EXAMPLE ILLUSTRATING THEOREM 2.5



Examining the simple effect $\alpha_0-\alpha_1$ we find the following situation:

The contrast for $\alpha_0-\alpha_1$ involves 6 points at level 0 (for $A_0$) and 3 points at level 1 (for $A_1$), a ratio of 2:1. Checking levels of B (0,1, and 2) represented at the 0 level of A we find these same levels represented at the 1 level of A and in exactly the same ratios 2:1 as the number of points at $A_0$ to points at $A_1$.

Checking levels of C represented in each set we find a similar 2:1 ratio at each level of C. The simple effect $\alpha_0-\alpha_1$ is thus estimated by taking 1/6 of the contrast that assigns +1 to points at the 0 level of A and -2 to points at the 1 level. The contrast is given in the table below.

Using the criterion of Theorem 2.2 we find BC is estimable, the estimate being $X_{000}-X_{002}+X_{021}-X_{011}+X_{012}-X_{020}$. Now selecting two points in the BC contrast which are related, for example 000 and 002 we find the difference $X_{020}-X_{021}$ estimates $\Gamma_0-\Gamma_1$. Similarly $X_{000}-X_{002}$ estimates $\Gamma_0-\Gamma_2$, $X_{002}-X_{012}$ estimates $\beta_0-\beta_1$ and $X_{000}-X_{020}$ estimates $\beta_0-\beta_2$. Thus, since BC is estimable by Theorem 2.2, we are assured by Corrollary 2.3 that all simple effects of B and C over the levels involved in the BC estimate will also be estimable. These estimates can of course be improved upon by involving all of the available points as was done in the case of the estimate of $\alpha_0-\alpha_1$, above. The contrasts for all estimates are given in the following table:

| | $6(\alpha_0-\alpha_1)$ | $3(\beta_0-\beta_1)$ | $3(\beta_0-\beta_2)$ | $3(\Gamma_0-\Gamma_1)$ | $3(\Gamma_0-\Gamma_2)$ | BC interaction |
|---|---|---|---|---|---|---|
| 000 | +1 | +1 | +1 | +1 | +1 | +1 |
| 002 | +1 | +1 | +1 | 0 | -1 | -1 |
| 020 | +1 | 0 | -1 | +1 | +1 | -1 |
| 021 | +1 | 0 | -1 | -1 | 0 | +1 |
| 011 | +1 | -1 | 0 | -1 | 0 | -1 |
| 012 | +1 | -1 | 0 | 0 | -1 | +1 |
| 101 | -2 | +1 | +1 | -1 | 0 | 0 |
| 110 | -2 | -1 | 0 | +1 | +1 | 0 |
| 122 | -2 | 0 | -1 | 0 | -1 | 0 |

The breakdown of degrees of freedom is given below:

| | |
|---|---|
| A | 1 |
| B | 2 |
| C | 2 |
| BC interaction | 1 |
| Confounded | 2 |
| Tot. Mean | 8 |

## CHAPTER III

## SUMMARY AND EXTENSIONS

In this thesis, procedures were presented to determine which main effects and interactions are estimable in a general n-way cross classification in which all observations are missing for any number of cells. Initially a definition of estimability under these circumstances was essential and in Chapter II such a definition was developed. Briefly, an interaction of a set of factors, or the main effect of a factor, was defined to be estimable if there existed a linear combination of the cell means in which at least one observation was taken that estimated a linear function of the interaction effects (or main effects) of these factors free of all other main and interaction effects. In all cases the interaction is confounded with higher order interactions. So, to avoid repeated reference to the presence of these effects, higher order interactions were considered negligible when the discussion centered around a particular interaction.

A condition of connectedness among the factors involved in the interaction, which is both necessary and sufficient for estimability of the interaction was developed. It was found that this condition is all that is needed to determine whether or not the highest order interaction is estimable and if it is, then all interactions and simple effects over the factors and levels involved were also found to be estimable. A simple algorithm was presented for determining whether or not this interaction is estimable, how many linearly independent estimates exist, what they are and exactly what they estimate.

A computer program, written for the IBM 1130 computer, which checks for estimability of the highest order interaction is presented in the Appendix. As a problem for further study, this computer program could be improved upon and expanded to determine the number of independent estimates that exist, and to check lower order interactions for estimability in the absence of higher order interactions.

With regard to the problem of estimability of lower order interactions, it was found that the connectedness criterion must still be satisfied over the factors involved but that an additional problem of confounding with interactions involving other factors was now present. It was established that if the interaction is to be estimable then when these other factors are considered one at a time, the connectedness property must be satisfied on the factors of the interaction over a set of design points in which the single factor maintains a constant level. Also if the connectedness property holds over a set of points in which all other factors simultaneously maintain a constant level, then the interaction is always estimable.

In the special case of an n-1 factor interaction there is only one other factor not involved in the interaction, so these two conditions together proved to be necessary and sufficient, and we had a simple check for estimability of the interaction of n-1 factors by applying the algorithm developed for n factors on these n-1 factors over a constant level of the remaining factor.

Necessary and sufficient conditions for the estimability of any interaction were presented and although guidelines for an algorithm to determine estimability were suggested in the examples, no such algorithm was determined. A simple algorithm for this case which could be easily programmed for computer use would be a useful extension of the work of this thesis.

Another problem suggested by the work presented here is a study of the number of possible configurations with a given number of cells missing and what proportions of these configurations permit estimation of various interactions.

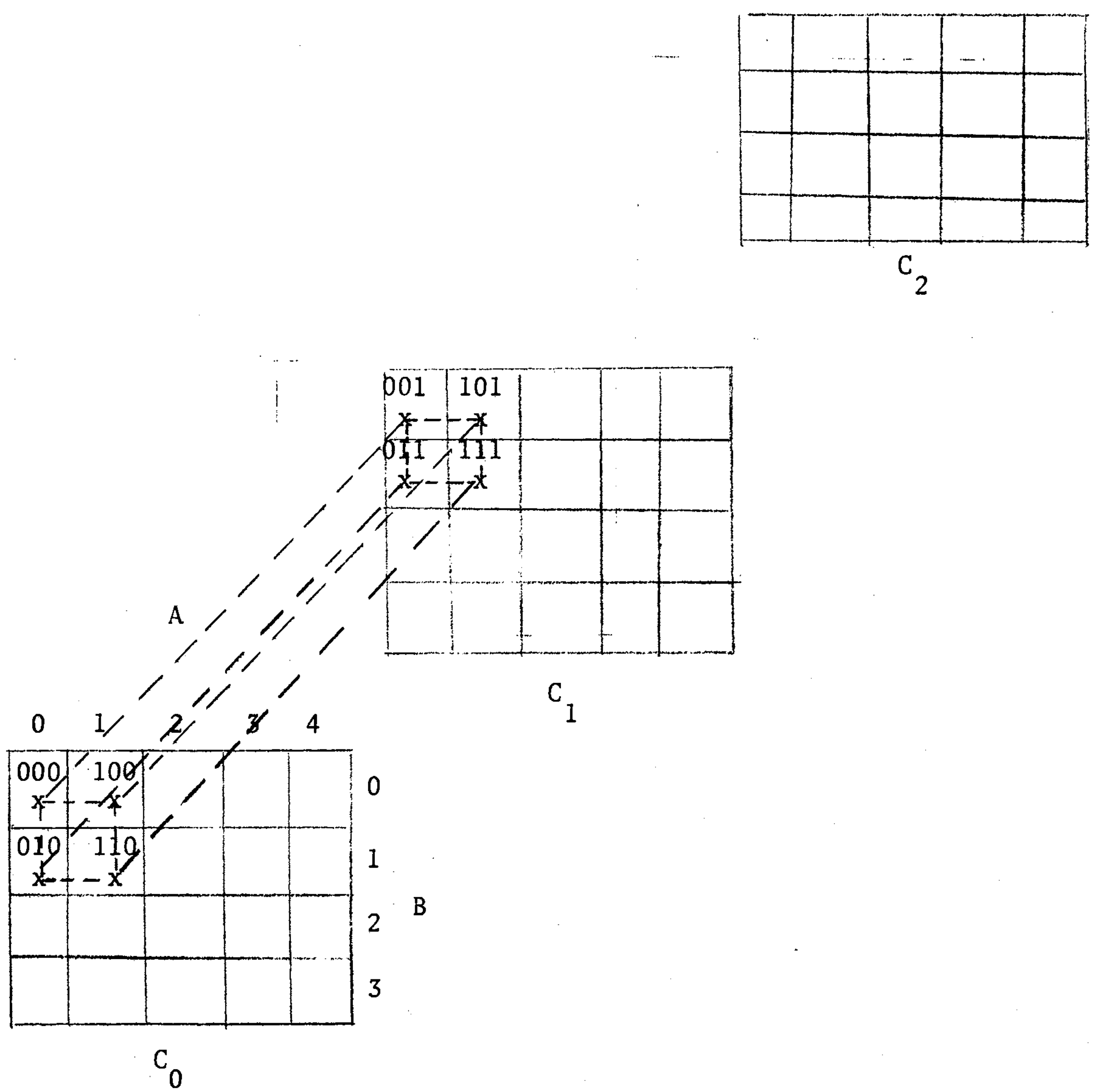
While the primary concern of the previous chapter was with problems of estimability of main effects and interactions, applications to the design of experiments are evident. Similar to situations in response surface investigations, or as with fractional replication, we may intentionally only study a portion of the entire set of treatment combinations. Utilizing the techniques of the previous chapter, however, the experimenter now has almost complete freedom to choose which main effects and interactions he wishes to investigate and how much information he wants on each. The single restriction being that, he must select points that form connected sets, as defined in Chapter II, in order to get unconfounded estimates.

These techniques can best be explained by considering several examples. In order to be able to illustrate the problems graphically, let us consider a three dimensional situation. Suppose we had a factor A at 5 levels, a factor B at 4 levels, and a factor C at 3 levels.

If we wished to estimate the ABC interaction, and consequently all other main effects and interactions, with a minimal number of points we could run an experiment using the connected set of points 000, 100, 010, 110, 001, 101, 011 and 111.

TABLE X

GRAPHICAL REPRESENTATION OF OCCUPIED CELLS FOR AN EXAMPLE DESIGNED WITH MISSING CELLS HAVING ALL EFFECTS AND INTERACTIONS ESTIMABLE



All interactions and main effects are then estimable, but of course, only over the 0 and 1 levels of each factor. Thus the simple effects $\alpha_0-\alpha_1$ , $\beta_0-\beta_1$ and $\Gamma_0-\Gamma_1$ are the only ones estimable. A breakdown of the degrees of freedom for linearly independent estimates would be as follows:

| AOV | Source | d.f. |
|---|---|---|
| | A | 1 |
| | B | 1 |
| | C | 1 |
| | AB | 1 |
| | AC | 1 |
| | BC | 1 |
| | ABC | 1 |

The estimates and quantities being estimated are:

$X_{000}-X_{010}-X_{100}+X_{110}-X_{001}+X_{011}+X_{101}-X_{111}$ estimating

$\alpha\beta\Gamma_{000}-\alpha\beta\Gamma_{010}-\alpha\beta\Gamma_{100}+\alpha\beta\Gamma_{110}-\alpha\beta\Gamma_{001}+\alpha\beta\Gamma_{011}+\alpha\beta\Gamma_{101}-\alpha\beta\Gamma_{110}$

$1/2(X_{000}-X_{010}-X_{100}+X_{110}+X_{001}-X_{011}-X_{101}+X_{111})$ estimating

$\alpha\beta_{00}-\alpha\beta_{10}-\alpha\beta_{01}+\alpha\beta_{11}$

$1/2(X_{000}+X_{010}-X_{100}-X_{110}-X_{001}-X_{011}+X_{101}+X_{111})$ estimating

$\alpha\Gamma_{00}-\alpha\Gamma_{10}-\alpha\Gamma_{01}-\alpha\Gamma_{11}$

$1/2(X_{000}-X_{010}+X_{100}-X_{110}-X_{001}-X_{011}-X_{101}+X_{111})$ estimating

$\beta\Gamma_{00}-\beta\Gamma_{10}-\beta\Gamma_{01}+\beta\Gamma_{11}$

$1/4(X_{000}+X_{010}-X_{100}-X_{110}+X_{001}+X_{011}-X_{101}-X_{111})$ estimating

$\alpha_0-\alpha_1$

$1/4(X_{000}-X_{010}+X_{100}-X_{110}+X_{001}-X_{011}+X_{101}-X_{111})$ estimating

$\beta_0-\beta_1$

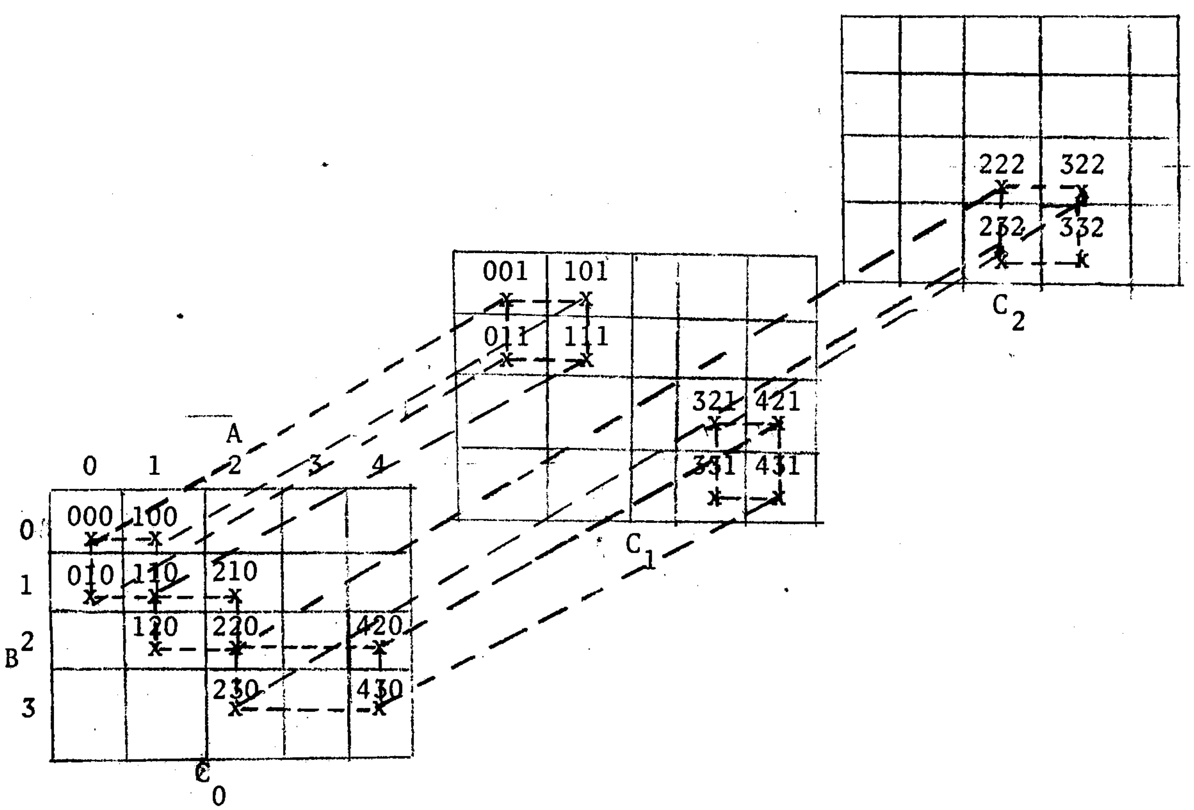
$1/4(X_{000}+X_{010}+X_{100}+X_{110}-X_{001}-X_{011}-X_{101}-X_{111})$ estimating

$\Gamma_0-\Gamma_1$

If we desired more information on all interactions and estimates of all other simple effects, we could add the connected set of points 220, 230, 420, 430, 321, 421, 331, 431, 222, 322, 232 and 332 to provide a second estimate of ABC interaction, two additional estimates of AB and BC interaction, one additional estimate of AC interaction and one each of the simple effects $\alpha_2-\alpha_3$ , $\alpha_2-\alpha_4$ , $\beta_2-\beta_3$ and $\Gamma_0-\Gamma_3$ . Connecting these two sets by adding the points 210 and 120 would provide estimates of $\alpha_1-\alpha_2$ and $\beta_1-\beta_2$ as well as an additional estimate of AB interaction. The remaining degrees of freedom unaccounted for would consist of confounded interactions and main effects.

TABLE XI

GRAPHICAL REPRESENTATION OF TABLE X WITH POINTS ADDED TO GIVE INFORMATION ON ALL SIMPLE EFFECTS AND ADDITIONAL INFORMATION ON INTERACTIONS



The estimates and quantities being estimated in addition to those of the previous set are:

$X_{220}-X_{230}-X_{420}+X_{430}-X_{321}+X_{421}+X_{331}-X_{431}-X_{222}-X_{322}+X_{232}-X_{332}$ estimating

$\alpha\beta\Gamma_{220}-\alpha\beta\Gamma_{230}-\alpha\beta\Gamma_{420}+\alpha\beta\Gamma_{430}-\alpha\beta\Gamma_{321}+\alpha\beta\Gamma_{421}+\alpha\beta\Gamma_{331}-\alpha\beta\Gamma_{431}-\alpha\beta\Gamma_{222}+\alpha\beta\Gamma_{322}$

$+ \alpha\beta\Gamma_{232}-\alpha\beta\Gamma_{332}$

$X_{110}+X_{220}-X_{210}-X_{120}$ estimating $\alpha\beta_{11}+\alpha\beta_{22}-\alpha\beta_{21}-\alpha\beta_{12}$

$X_{321}-X_{421}-X_{331}+X_{431}$ estimating $\alpha\beta_{32}-\alpha\beta_{42}+\alpha\beta_{43}-\alpha\beta_{33}$

$X_{222}-X_{322}-X_{232}+X_{332}$ estimating $\alpha\beta_{22}-\alpha\beta_{32}-\alpha\beta_{23}+\alpha\beta_{33}$

$\frac{1}{2}$ $(X_{220}+X_{230}-X_{420}-X_{430}-X_{321}+X_{421}-X_{331}+X_{431}-X_{222}+X_{322}+X_{332}-X_{232}$ estimating

$\alpha\Gamma_{20}-\alpha\Gamma_{40}+\alpha\Gamma_{41}-\alpha\Gamma_{31}+\alpha\Gamma_{32}-\alpha\Gamma_{22}$

$X_{322}-X_{321}+X_{331}-X_{322}$ estimating $\beta\Gamma_{22}-\beta\Gamma_{21}+\beta\Gamma_{31}-\beta\Gamma_{32}$

$X_{220}-X_{230}-X_{222}+X_{232}$ estimating $\beta\Gamma_{20}-\beta\Gamma_{30}-\beta\Gamma_{22}+\beta\Gamma_{32}$

$\frac{1}{2}$ $(X_{220}+X_{230}-X_{420}-X_{430})$ estimating $\alpha_2-\alpha_4$

$\frac{1}{2}$ $(X_{110}-X_{220}-X_{210}+X_{120})$ estimating $\alpha_1-\alpha_2$

$\frac{1}{2}$ $(X_{222}-X_{322}+X_{232}-X_{332})$ estimating $\alpha_2-\alpha_3$

$\frac{1}{2}$ $(X_{110}-X_{220}+X_{210}-X_{120})$ estimating $\beta_1-\beta_2$

$\frac{1}{6}$ $(X_{220}-X_{230}+X_{420}-X_{430}+X_{321}+X_{421}-X_{331}-X_{431}+X_{222}+X_{322}-X_{232}-X_{332})$

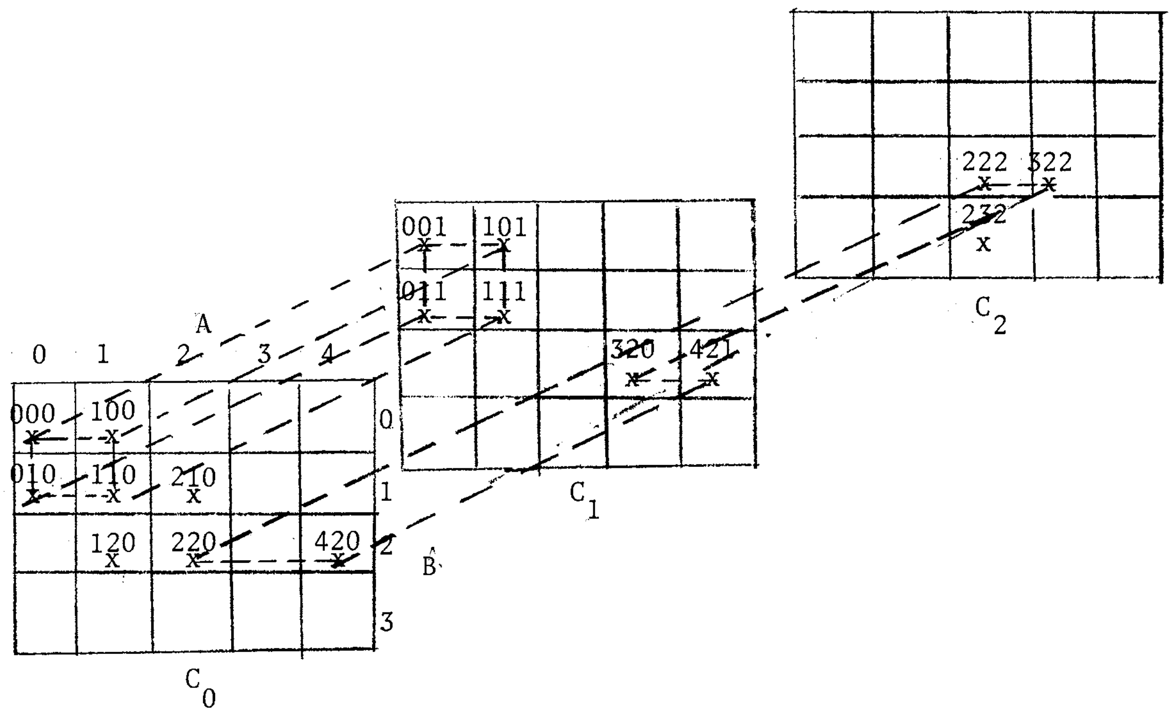
estimating $\beta_2-\beta_3$

$\frac{1}{2}$ $(X_{220}+X_{230}-X_{222}-X_{232})$ estimating $\Gamma_0-\Gamma_2$

If we felt that the ABC and BC interactions were negligible and did not desire additional information on them, but wanted estimates of all simple effects and more information on AB and AC we could use the original set of points, and add all the points of the second set except 230, 430, 331, 431, and 332.

TABLE XII

GRAPHICAL REPRESENTATION OF TABLE X WITH POINTS DELETED DUE TO NEGLIGIBLE INTERACTIONS




The breakdown of degrees of freedom for linearly independent estimates would now be:

| | d.f. |
|---|---|
| A | 4 |
| B | 3 |
| C | 2 |
| AB | 2 |
| AC | 2 |
| BC | 1 |
| ABC | 1 |
| Confounded | 1 |

The estimates and quantities being estimated, in addition to those of the original set are now as follows:

$X_{110}+X_{220}-X_{210}-X_{120}$ estimating $\alpha\beta_{11}+\alpha\beta_{22}-\alpha\beta_{21}-\alpha\beta_{12}$

$X_{220}-X_{420}-X_{321}+X_{421}-X_{222}+X_{322}$ estimating $\alpha\Gamma_{20}-\alpha\Gamma_{40}+\alpha\Gamma_{41}-\alpha\Gamma_{31}+\alpha\Gamma_{32}-\alpha\Gamma_{22}$

$X_{220}-X_{420}$ estimating $\alpha_2-\alpha_4$

$1/2(X_{110}-X_{220}-X_{210}+X_{120})$ estimating $\alpha_1-\alpha_2$

$X_{222}-X_{322}$ estimating $\alpha_2-\alpha_3$

$X_{110}-X_{220}+X_{210}-X_{120}$ estimating $\beta_1-\beta_2$

$X_{222}-X_{232}$ estimating $\beta_2-\beta_3$

$X_{220}-X_{222}$ estimating $\Gamma_0-\Gamma_2$

Obviously many other designs are possible depending on which interactions or main effects are of interest and how much information is desired on each. The previous examples are only illustrative of the possible applications of the techniques investigated in this thesis to the design of experiments.

## SELECTED BIBLIOGRAPHY


(1) Cochran, W. G., and Cox, G. *Experimental Designs.* New York; John Wiley and Sons 1957.

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

(3) Mann, H. B. *Analysis and Design of Experiments.* New York: Dover Publications, Inc., 1949.

(4) Williams, D. R. "A New Approach to Factorial Experimentation" (Unpub. Ph.D. Dissertation, Oklahoma State University, 1963).

(5) Thomas, H. L. "Partitioned Factorials" (Unpub. Ph.D. Dissertation, Oklahoma State University, 1964).

(6) Graybill, F. A. *An Introduction to Linear Statistical Models,* Vol. 1, New York: McGraw-Hill Book Company, Inc., 1961.

# APPENDIX

DEFINITION OF SYMBOLS USED IN THE FOLLOWING COMPUTER PROGRAM

| | |
|---|---|
| DIM | number of dimensions |
| GI | group index, temporary |
| POINT | contains coordinates for 1 point, temporary |
| PI | point index, temporary |
| IN | used by free-format input. contains card in A2 format, which refers to the location of data to be inputed into the program. |
| BASE | used for calculating subscript, temporary |
| GNUM | contains group of numbers for points in POINT. Index = "don't care" dimension |
| CØØRD | contains coordinate restrictions imposed on set. Index = dimension - 10000 ≡ no restriction. |
| CØRG | contains the coordinate values for each group. |
| CØRG(L) = | CORG(group number x number of dimensions + DIM - 10000 = don't care |
| FAULT | list of groups containing only 1 point; which are used to start deletion process. |
| FL | index of last entry in FAULT |
| GLIM | number of groups defined |
| GPS | This contains groups that each point belongs to |
| GPS(L) = | GPS(point number x number of dimensions + DIM) |
| MEM | contains list of points for each group. |
| GRPL | contains index of MEM containing list of points for each group. Index = group number |
| TMASK | coordinate restrictions of set to be traced |
| TV | trace value |
| CØRAN | CØRAN(1,J) - coordinate base dimension of J<br>CØRAN(2,J) - coordinate limit dimension of J<br>CØRAN(3,J) - coordinate increase dimension of J |
| PTDEL | The stak of points currently deleted in attempt to find minimal subset |

| | |
|---|---|
| STACK | in LIND1 contains points to delete, and in LIND2 contains groups which were closed. |
| CANCL | list of coordinate restrictions of successful sets. The coordinate restrictions for another set may not have any previous one for a subset. |
| CLIM | index of last entry in CANCL. |
| WØRK | in LIND1 number of groups still open. |
| TRACE | set to $\|TV\|$ when CØØRD matches TMASK. |
| Z | printer device number |
| CHAIN | contains index of next point in equivalence chain. Negative index indicated head of chain. |
| NP | number of points defined |
| IS | generally index of STACK |
| IRET | used to indicate where to branch after completing a common routine |
| NØSU | index of subsets |
| NFAIL | number of failures for a level |
| NTRY | number of trys for a level |
| NCAN | number cancelled for a level |

```
*ONE WORD INTEGERS
      SUBROUTINE LINDO
      INTEGER GI,POINT(8),PI,      IN(41),BASE,GNUM(8)
     1 INTEGER DIM,COORD(8),CORG( 600),FAULT(100),FL,GLIM,GPS( 600)
     2 INTEGER GRPL(301),KD,MEM( 600),TMASK(8),TV
     3 INTEGER CORAN(3,8),PD,PTDEL(10 ),STACK(100),CANCL( 600),DIML(8)
     4 INTEGER CLIM,WORK,HI,GI,      TRACE,Z,STO,CHAIN(100)
     5 COMMON DIM,COORD,CORG,FAULT,FL,GLIM,GPS,GRPL,KD,MEM,NP,TMASK,TV
     6 COMMON CORAN,PD,PTDEL,STACK,IS,IRET,CANCL,DIML,CLIM,WORK,HI,LGM
     7 COMMON LEVEL,TRACE,NOSU,      GI,Z,NFAIL,NTRY,NCAN,STO,LSTO,CHAIN
C        LCORG=LENTH OF CORG,     LFAUL=LENGTH OF FAULT.
C        LGPS=LENGTH OF GPS,      LMEM=LENGTH OF MEM
      LCORG=800
      LFAUL=100
      LGPS=800
      LMEM=700
C        Z IS THE PRINTER DEVICE NUMBER.
C        WIDTH OF 120 CHARS. HAS BEEN ASSUMED.
      Z=3
      IF(IRET)1,1,11
   11 WRITE(Z,510)STO
C        READ DIMENSION OF POINTS
    1 IN(41)=81
      CALL FIN(IN,A)
      DIM=A
C        EXIT IF DIM .LT. 1
      IF(DIM)51,51,102
   51 WRITE(Z,530)
      CALL EXIT
  102 KD=DIM&1
C        READ TRACE MASK
      CALL FIN(IN,A)
      TV=A
      WRITE(Z,500)DIM,TV
      IF(TV)142,142,141
  141 DO 140 I=1,DIM
      CALL FIN(IN,A)
  140 TMASK(I)=A
      WRITE(Z,519)(TMASK(I),I=1,DIM)
C        READ COORDINATE RANGE
  142 DO 10 I=1,DIM
      DO 10 J=1,3
      CALL FIN(IN,A)
   10 CORAN(J,I)=A
      WRITE(Z,550)(I,(CORAN(J,I),J=1,3),I=1,DIM)
      WRITE(Z,520)
      GLIM=0
      CORG(1)=10000
      NP=0
C        READ A POINT
   50 DO 2, I=1,DIM
      CALL FIN(IN,A)
      K=A
      IF(K&10000)2,4,2
```

```
    2  POINT(I)=K
C        PUT POINT IN GROUPS
       PI=NP*DIM
       NP=NP&1
       DO 59 I=1,DIM
   59  GNUM(I)=0
       DO 20 GI=1,GLIM
       BASE=(GI-1)*DIM
       DO 25 J=1,DIM
       L=BASE&J
       IF(CORG(L)&10000)30,26,30
   30  IF(CORG(L)-POINT(J))20,25,20
   26  KJ=J
   25  CONTINUE
C        POINT BELONGS IN GROUP GI
       L=PI&KJ
       GPS(L)=GRPL(GI)
       GNUM(KJ)=GI
       GRPL(GI)=L
   20  CONTINUE
C        *CREATE NEW GROUPS IF NEEDED
       DO 35 I=1,DIM
       IF(GNUM(I))35,40,35
C        MAKE NEW GROUP FOR DIMENSION I
   40  DO 45 J=1,DIM
       L=GLIM*DIM&J
   45  CORG(L)=POINT(J)
       L=GLIM*DIM&I
       CORG(L)=-10000
       L=PI&I
       GPS(L)=0
       GLIM=GLIM&1
       GNUM(I)=GLIM
       GRPL(GLIM)=L
       IF(GLIM-300)35,35,60
   35  CONTINUE
       WRITE(Z,532)NP,(POINT(I),GNUM(I),I=1,DIM)
       IF(PI-DIM-LGPS)50,5,5
C        DETECT END OF LIST CODE
    4  IF(I-1)7,8,7
C        *AT END OF POINT
    8  WRITE(Z,502)NP,GLIM
       IF(TV)80,81,80
   80  WRITE(Z,503)
C        SET UP ARRAYS GRPL,GPS,MEM,AND FAULT
   81  FL=0
       K=1
       DO 52 GI=1,GLIM
       J=GRPL(GI)
       KS=K
   54  K=K&1
       IF(J)53,53,55
   55  MEM(K)=(J-1)/DIM&1
       M=GPS(J)
```

```
      GPS(J)=GI
      J=M
      IF(K-LMEM)54,70,70
C        SUBSCRIPT FOR MEM IS TOO LARGE.
   70 WRITE(Z,506)
      CALL EXIT
C        STORE LENGTH OF GROUP
   53 M=K-1
      L=M-KS
      MEM(KS)=L
C        STORE POINTER TO POINT LIST FOR GROUP GI
      GRPL(GI)=KS
      IF(TV)82,83,82
   82 WRITE(Z,533)GI,(MEM(J),j=KS,M)
C        TEST FOR FAULT
   83 IF(L-1)52,63,52
C        RECORD FAULT
   63 FL=FL&1
      FAULT(FL)=GI
      IF(FL-LFAUL)52,61,61
   52 CONTINUE
C        MAKE DUMMY GROUP ENTRY AFTER LAST GROUP
      GRPL(GLIM&1)=K
C        LIST FAULTS
      IF(FL)57,57,58
   57 WRITE(Z,534)
      GO TO 65
   58 WRITE(Z,535)(FAULT(I),I=1,FL)
C        INITIALIZE COORD,DIML,LEVEL,CLIM,CANCL
   65 DO 62 I=1,DIM
      DIML(I)=0
      COORD(I)=-10000
   62 CONTINUE
      LEVEL=0
      CLIM=0
      CANCL(1)=-10001
      STO=0
      IRET=1
      RETURN
    5 WRITE(Z,506)
      CALL EXIT
    7 WRITE(Z,507)
      CALL EXIT
   60 WRITE(Z,536)
      CALL EXIT
   61 WRITE(Z,537)
      CALL EXIT
  510 FORMAT('OEND OF RUN FOR THIS DATA SET.'I20,' STACK OVERFLOWS')
  530 FORMAT('OEND OF JOB.')
  500 FORMAT('IDIMENSION='I3,5X,'TRACE VALUE='I2)
  519 FORMAT('.TRACE MASK'/(10I10))
  550 FORMAT('OCOORDINATE VALUES'/'.  DIM  BASE  LIMIT  INCR'/(4I7))
  520 FORMAT('OPOINT'5X,'COORDINATES AND (GROUPS)'/1X)
  532 FORMAT(I5,'.'8(I7.'  ('I4,')'))
```

```
502 FORMAT('0***END OF LIST'16,' POINTS AND'16,' GROUPS ARE DEFINED.')
503 FORMAT('OLIST OF GROUPS')
533 FORMAT('OGROUP ('14,')'15,' POINTS'/(1X,20I6))
534 FORMAT('ONO FAULTS.')
535 FORMAT('OLIST OF FAULTS BY GROUP NUMBERS'/(1X,20I6) )
506 FORMAT('OERROR -- DATA EXCEEDS STORAGE CAPACITY')
507 FORMAT('OERROR -- END OF LIST CODE WAS NOT AT A POINT BOUNDARY')
536 FORMAT('OERROR -- LIST OF GROUPS EXCEEDS CAPACITY')
537 FORMAT('OERROR -- LIST OF FAULTS EXCEEDS CAPACITY')
    END
// FOR                                                                  LIN10000
// DUP
*STORE    WS  UA  LINDO                                                 LIN10010
*ONE WORD INTEGERS                                                      LIN10020
      SUBROUTINE LIND1                                                  LIN10030
      INTEGER PI,BI,OBUF(10),GI,SGI
    1 INTEGER DIM,COORD(8),CORG( 600),FAULT(100),FL,GLIM,GPS( 600)
    2 INTEGER GRPL(301),KD,MEM( 600),TMASK(8),TV
    3 INTEGER CORAN(3,8),PD,PTDEL(10 ),STACK(100),CANCL( 600),DIML(8)
    4 INTEGER CLIM,WORK,HI,GI,  TRACE,Z,STO,CHAIN(100)
    5 COMMON DIM,COORD,CORG,FAULT,FL,GLIM,GPS,GRPL,KD,MEM,NP,TMASK,TV
                                                                        LIN10080
    6 COMMON CORAN,PD,PTDEL,STACK,IS,IRET,CANCL,DIML,CLIM,WORK,HI,LGM
                                                                        LIN10090
    7 COMMON LEVEL,TRACE,NOSU,  GI,Z,NFAIL,NTRY,NCAN,STO,LSTO,CHAIN
      EQUIVALENCE (WORK,NWORK)                                          LIN10105
C       LIS=LENGTH OF STACK-1, AT LEAST LFAUL&1/3*LMPT                  LIN10110
      LIS=100                                                           LIN10120
C       IRET=1 -- ENTRY FROM LINSO.  BEGIN                              LIN10130
C       IRET=2 -- CANCEL PROCESS                                        LIN10140
C       IRET=3 -- TRACE PRINT RETURN                                    LIN10150
C       IRET=4 -- NEXT RESTRICTION                                      LIN10160
      GO TO (1,167,110,902),IRET                                        LIN10170
C       THIS SECTION SEARCHES OVER THE POINTS TO FIND A CONSISTENT
        SET.                                                            LIN10180
C       THIS SUBSECTION WILL ELIMINATE ALL POINTS NOT IN THE            LIN10190
C       SUBSET DEFINED.                                                 LIN10200
C       SET TRACE IF REQUESTED ON THIS SUBSET                           LIN10210
  134   TRACE=0                                                         LIN10220
        NTRY=NTRY&1                                                     LIN10230
        IF(IV)104,101,160                                               LIN10240
  160   DO 103 I=1,DIM                                                  LIN10250
        IF(TMASK(I)&10001)158,103,158                                   LIN10260
  158   IF(TMASK(I)-COORD(I))101,103,101                                LIN10270
  103   CONTINUE                                                        LIN10280
C         TRACE IS REQUESTED FOR THIS SUBSET.                           LIN10290
  104   TRACE=IABS(TV)                                                  LIN10300
        WRITE(Z,523)                                                    LIN10310
        WRITE(Z,511)(I,COORD(I),I=1,DIM)                                LIN10320
C         CHECK FOR CANCELLATION OF SUBSET                              LIN10330
  101   DO 150 I=1,CLIM,DIM                                             LIN10340
        DO 151 J=1,DIM                                                  LIN10350
        K=I&J-1                                                         LIN10360
        IF(CANCL(K)&10000)152,151,152                                   LIN10370
```

```
  152 IF(CANCL(K)-COORD(J))150,151,150                      LIN10380
  151 CONTINUE                                              LIN10390
      NCAN=NCAN&1                                           LIN10410
      IF(TRACE)124,124,154                                  LIN10420
  154 WRITE(Z,518)                                          LIN10430
      WRITE(Z,511)(J,CANCL(J),J=I,K)                        LIN10440
      GO TO 124                                             LIN10450
  150 CONTINUE                                              LIN10460
C       SUBSET NOT CANCELLED                                LIN10470
      PD=0                                                  LIN10480
      IS=0                                                  LIN10490
  167 WORK=0                                                LIN10500
      LSTO=0                                                LIN10510
      LGI=0                                                 LIN10520
C       MASK OFF GROUPS NOT IN SET                          LIN10530
      DO 9 GI=1,GLIM                                        LIN10540
      DO 106 IDIM=1,DIM                                     LIN10550
      IF(COORD(IDIM)&10000)107,106,107                      LIN10560
  107 J=LGI&IDIM                                            LIN10570
      IF(COORD(DIM)-CORG(J))109,106,109                     LIN10580
  106 CONTINUE                                              LIN10590
C       GROUP IS INCLUDED IN SUBSET                         LIN10600
      J=IABS(GRPL(GI))                                      LIN10610
      GRPL(GI)=J                                            LIN10620
      MEM(J)=IABS(GRPL(GI&1))-J-1                           LIN10630
      WORK=WORK&1                                           LIN10640
      GO TO 9                                               LIN10650
C       GROUP IN NOT INCLUDED IN SUBSET                     LIN10660
  109 GRPL(GI)=-IABS(GRPL(GI))                              LIN10670
    9 LGI=LGI&DIM                                           LIN10680
C       CANCEL IF NO GROUPS ARE INCLUDED                    LIN10690
      IF(WORK)113,113,114                                   LIN10700
  113 IF(TRACE)121,121,115                                  LIN10710
  115 WRITE(Z,508)                                          LIN10720
C       PLACE FAULT POINTS IN POINTS-TO-DELETE STACK        LIN10730
  114 DO 165 I=1,FL                                         LIN10740
      M=FAULT(I)                                            LIN10750
      IF(GRPL(M))165,165,166                                LIN10760
  166 J=GRPL(M)                                             LIN10770
      IS=IS&1                                               LIN10780
      STACK(IS)=MEM(J&1)                                    LIN10790
  165 CONTINUE                                              LIN10800
      NDEL=0                                                LIN10810
C       PRINT REMAINING GROUPS IF TRACE .GT. 1 AND PD=0     LIN10820
      IF(PD)110,112,110                                     LIN10830
  112 IF(TRACE-1)110,110,111                                LIN10840
  111 WRITE(Z,510)WORK                                      LIN10850
      IRET=3                                                LIN10860
      RETURN                                                LIN10870
C       THIS SECTION ELIMINATES REMAINING SINGULAR POINTS.  LIN10880
C       BRANCH TO 26 IF SUCCESS (STACK EMPTY).              LIN10890
  110 BI=0                                                  LIN10080
      NDEL=0                                                LIN10930
  210 IF(IS)26,26,119                                       LIN10940
```

```
C          UNSTACK POINT TO DELETE                                  LIN10950
  119  PI=STACK(IS)                                                 LIN10960
   19  IF(TRACE-2)21,21,20                                          LIN10970
C          PRINT DELETION IF TRACE .GT. 2                           LIN10980
   20  BI=BI&1                                                      LIN10990
       OBUF(BI)=PI                                                  LIN11000
       IF(BI-10)21,22,22                                            LIN11010
   22  WRITE(Z,539)OBUF                                             LIN11020
       BI=0                                                         LIN11030
   21  IS=IS-1                                                      LIN11040
       J=(PI-1)*DIM                                                 LIN11050
       SGI=0                                                        LIN11060
C          DELETE POINT FROM EACH GROUP IT IS IN                    LIN11070
       DO 14 IDIM=1,DIM                                             LIN11080
       J=J&1                                                        LIN11090
       GI=GPS(J)                                                    LIN11100
C          SKIP TO 14 IF GROUP IS EMPTY.                            LIN11110
       IF(GRPL(GI))14,14,10                                         LIN11120
C          DELETE POINT PI FROM GROUP GI.                           LIN11130
   10  K=GRPL(GI)                                                   LIN11140
C          SEARCH FOR POINT IN GROUP                                LIN11150
       M=MEM(K)-1                                                   LIN11160
       L=M&K                                                        LIN11170
       DO 11 I=K,L                                                  LIN11180
       IF(MEM(I&1)_PI)11,12,11                                      LIN11190
   11  CONTINUE                                                     LIN11200
C          POINT WAS NOT FOUND IN GROUP                             LIN11210
       M=M&1                                                        LIN11220
C          CONSIDER DELTED IF STACK FULL                            LIN11230
       IF(IS-LIS)13,13,14                                           LIN11240
C          REMOVE POINT FROM GROUP.                                 LIN11250
   12  MEM(I&1)=MEM(L&1)                                            LIN11260
       MEM(L&1)=PI                                                  LIN11270
       MEM(K)=M                                                     LIN11280
C          TEST FOR 1 REMAINING POINT.                              LIN11290
   13  IF(M-1)31,15,14                                              LIN11300
   15  IF(IS-LIS)16,17,18                                           LIN11310
C          STACK PI AND SET STACK OVERFLOW SWITCH                   LIN11320
   17  IS=IS&1                                                      LIN11330
       STACK(IS)=PI                                                 LIN11340
       LSTO=1                                                       LIN11350
   18  PI=MEM(K&1)                                                  LIN11360
       SGI=GI                                                       LIN11370
       GO TO 19                                                     LIN11380
C          STACK REMAINING POINT AND CLOSE GROUP.                   LIN11390
   16  IS=IS&1                                                      LIN11400
       STACK(IS)=MEM(K&1)                                           LIN11410
   31  GRPL(GI)=-GRPL(GI)                                           LIN11420
       NDEL=NDEL&1                                                  LIN11430
   14  CONTINUE                                                     LIN11440
       IF(SGI)200,200,201                                           LIN11450
  201  NDEL=NDEL&1                                                  LIN11460
       GRPL(SGI)=-GRPL(SGI)                                         LIN11470
  200  IF(NDEL-WORK)210,172,172                                     LIN11480
```

```
C           *FAILURE
C           ALL POINTS WERE DELETED.
  172   STO=STO&LSTO
        IF(BI)173,173,171
  171   WRITE(Z,539)(OBUF(I),I=1,BI)
  173   IF(TRACE)120,120,174
  174   WRITE(Z,551)
C           RETURN IF THIS WAS AN ATTEMPT AT A SUBSET OF A
            CONSISTENT SET
  120   IF(PD)121,121,122
  122   IRET=4
        RETURN
C           CHALK IT UP.
  121   NFAIL=NFAIL&1
  902   IF(LEVEL)132,132,124
C           THIS SECTION GENERATES RESTRICTIONS PLACED ON THE SET
  124   I=1
  133   IDIM=DIML(I)
        K=COORD(IDIM)
C           INCREMENT COORDINATE FOR DIMENSION ON LEVEL I
C           TEST FOR RESET MARKER
        IF(K&10000)170,169,170
  170   K=K&CORAN(3,IDIM)
        IF(K-CORAN(2,IDIM))127,127,125
C           RESET MARKER FOUND. SET TO FIRST COORDINATE VALUE.
  169   K=CORAN(L,IDIM)
C           NEXT COORDINATE VALUE WAS FOUND
  127   COORD(IDIM)=K
        I=I-1
C           GO TO 134 WHEN ALL COORDINATES ARE FIXED
        IF(I)133,133,134
C           HIGHEST COORDINATE HAS BEEN USED.
C           RESET DIMENSION IDIM.
  125   COORD(IDIM)=-10000
        I=I&1
        IF(I-LEVEL)133,133,140
C           COORDINATES FOR ALL DIMENSIONS HAVE REACHED MAXIMUM
            VALUE
C           INCREMENT A DIMENSION
  140   DO 128 IG=1,LEVEL
        IF(DIML(IG)-DIM&IG-1)129,128,128
  128   CONTINUE
C           ALL DIMENSIONS ARE MAXIMUM FOR THIS LEVEL
  132   NGO=NTRY-NFAIL
        WRITE(Z,514)LEVEL,NTRY,NCAN,NFAIL,NGO
C           TEST FOR END OF RUN.
        IF(LEVEL-DIM)161,131,131
C           INCREMENT LEVEL
  161   LEVEL=LEVEL&1
    1   WRITE(Z,515)LEVEL
        NTRY=0
        NFAIL=0
        NCAN=0
        IF(LEVEL)134,134,181
```

```
  181  IG=LEVEL                                                   LIN12010
C           RESET DIMENSIONS FOR LEVELS = 1 TO IG                 LIN12020
  129  K=DIML(IG)                                                 LIN12030
       DO 130 I=1,IG                                              LIN12040
       L=IG-I                                                     LIN12050
  130  DIML(L&1)=K&I                                              LIN12060
C           NOW SET VALUES FOR THESE DIMENSIONS                   LIN12070
       GO TO 124                                                  LIN12080
C           A CONSISTANT SUBSET HAS BEEN FOUND                    LIN12090
   26  IF(BI)28,28,27                                             LIN12100
   27  WRITE(Z,539)(OBUF(I),I=1,BI)                               LIN12110
   28  IRET=2                                                     LIN12120
       RETURN                                                     LIN12130
C           END OF RUN                                            LIN12140
  131  IRET=1                                                     LIN12150
       RETURN                                                     LIN12160
  508  FORMAT('.NO GROUPS MEET COORDINATE RESTRICTIONS')          LIN12170
  510  FORMAT('.LIST OF'I6 ' GROUPS MEETING COORDINATE            LIN12180
       RESTRICTIONS')
  511  FORMAT(1X,8(I5,')='I5))                                    LIN12190
  514  FORMAT('OEND OF TESTS ON LEVEL'I3,/I8,' SETS               LIN12200
       GENERATED'I10,.' SETS CANCELLED'I10,' SETS FAILED'         LIN12210
       I10,' SETS CONSISTENT'/ .121('*'))                         LIN12220
  515  FORMAT('1***LEVEL'I3)                                      LIN12230
  518  FORMAT('.THIS SUBSET CANCELLED BY THE SUBSET WITH THE      LIN12240
       FOLLOWING RESTRICTIONS')                                   LIN12250
  523  FORMAT('OCOORDINATE RESTRICTIONS')                         LIN12260
  539  FORMAT('*POINTS DELETED '10I6)                             LIN12270
  551  FORMAT('.FAILURE. ALL POINTS DELETED'/1HO)                 LIN12280
       END                                                        LIN12290

// FOR
*ONEWORD INTEGERS
       SUBROUTINE LIND2
       INTEGER A, B
       INTEGER OBUF(20),PI
     1 INTEGER DIM,COORD(8),CORG( 600),FAULT(100),FL,GLIM,GPS( 600)
     2 INTEGER GRPL(301),KD,MEM( 600),TMASK(8),TV
     3 INTEGER CORAN(3,8),PD,PTDEL(10 ),STACK(100),CANCL( 600),DIML(8)
     4 INTEGER CLIM,WORK,HI,GI,  TRACE,Z,STO,CHAIN(100)
     5 COMMON DIM,COORD,CORG,FAULT,FL,GLIM,GPS,GRPL,KD,MEM,NP,TMASK,TV
     6 COMMON CORAN,PD,PTDEL,STACK,IS,IRET,CANCL,DIML,CLIM,WORK,HI,LGM
     7 COMMON LEVEL,TRACE,NOSU,  GI,Z,NFAIL,NTRY,NCAN,STO,LSTO,CHAIN
C           LCANC=LENGTH OF CANCL.  LMPT=LENGTH OF MPT AND PTDEL.
       LCANC=800
       LMPT=10
C           IRET=2 -- CONSISTENT SET
C           IRET=3 -- TRACE PRINT
       GO TO (2,2,100),IRET
C           *A CONSISTENT SET HAS BEEN FOUND
C           PRINT CONSISTENT SET
     2 WRITE(Z,501)(I,COORD(I),I=1,DIM)
       IF(CLIM-DIM-LCANC)156,156,157
  156  DO 155 I=1,DIM
```

```
      CLIM=CLIM&1
  155 CANCL(CLIM)=COORD(I)
      DO 90 I=1,NP
C         ADD THIS SET TO CANCEL BUFFER.
   90 CHAIN(I)=0
      IRET=1
      GO TO 3
   55 NOSU=1
      NG=GLIM
      PD=0
      IS=1
C         SEARCH FOR GROUPS CONTAINING ONLY TWO POINTS
C         FOR EACH ONE FOUND, DEFINE THE TWO POINTS TO BE EQUIVALANT.
      DO 5 GI=1,GLIM
      K=GRPL(GI)
      IF(K)8,8,6
    6 IF(MEM(K)-2)21,21,5
    8 NG=NG-1
    5 CONTINUE
      IF(TRACE-1)94,94,83
   83 DO 84 K=1,NP
      PI=-CHAIN(K)
      IF(PI)84,84,85
   85 WRITE(Z,585)PI
  585 FORMAT('OCHAIN'I3)
      CALL LIND5(PI)
   84 CONTINUE
   94 IF(NG)80,80,81
C         INITIALIZE TO NO DELETIONS
   81 PTDEL(1)=0
      HI=1
C         HUNT FOR CHAIN 1 GREATER THAN K
   31 IS=IS&1
      STACK(IS)=-PD
   73 K=PTDEL(HI)
      L=10000
      PI=0
   10 PI=PI&1
      IF(PI-NP)11,11,12
   11 PD=-CHAIN(PI)
      IF(PD)10,9,9
C         CHECK FOR PD .GT. K AND PD .LT. L
    9 IF((PD-K)*(PD-L))13,10,10
   13 L=PD
      GO TO 10
C         SEARCH FOR NEXT CHAIN IS FINISHED. GO TO 15 IF K IS HIGHEST CHAIN
   12 IF(L-10000)14,115,115
C         DELETE CHAIN PD
   14 PD=L
      WRITE(Z,506)PD
      PTDEL(HI)=PD
      PTDEL(HI&1)=PD
      PI=PD
      IRET=2
```

```
C           PI IS A POINT EQUIVALANT TO PD.  DELETE IT.
   29  J=(PI-1)*DIM
C           *DELETE POINT PI
       DO 16 IDIM=1,DIM
       J=J&1
       GI=GPS(J)
       K=GRPL(GI)
       IF(K)16,16,17
C           SEARCH FOR POINT PI IN GROUP GI
   17  M=MEM(K)-1
       L=M&K
       DO 18 I=K,L
       IF(MEM(I&1)-PI)18,19,18
   18  CONTINUE
       GO TO 20
C           PUT POINT AT END OF LIST AND DECR LENGTH
   19  MEM(I&1)=MEM(L&1)
       MEM(L&1)=PI
       MEM(K)=M
       IS=IS&2
       STACK(IS-1)=0
       STACK(IS)=GI
       WRITE(Z,510)GI
  510  FORMAT(' GROUP'I4,' DELETED')
   20  IF(M-2)21,21,16
C           TWO POINTS REMAINING
   21  A=MEM(K&1)
       B=MEM(K&2)
       CALL LIND4(A,JA,IA)
       CALL LIND4(B,JB,IB)
       IF(IA-IB)40,54,40
   40  IF(IA-PD)23,22,23
   22  IF(IB-PD)74,53,54
   23  IF(IB-PD)54,24,54
   24  IF(IA-PD)74,53,54
   27  IF(NG)28,28,16
   16  CONTINUE
       PI=CHAIN(PI)
       IF(PI)30,30,29
C           ALL POINTS IN CHAIN PD HAVE BEEN DELETED
   30  HI=HI&1
       GO TO 31
C           NO GROUPS LEFT.  TEST FOR NON-EMPTY SET
   28  K=IABS(GRPL(GI))
       A=MEM(K&1)
       CALL LIND4(A,I,J)
       IF(J-PD)32,74,32
C           MINIMAL CONSISTENT SET
   32  WRITE(Z,502)NOSU
       NOSU=NOSU&1
       K=0
   37  IF(J)33,33,34
   34  K=K&1
       OBUF(K)=J
```

```
      IF(K-20)35,36,36
   36 WRITE(Z,505)OBUF
      K=0
   35 J=CHAIN(J)
      GO TO 37
  115 HI=HI-1
      IF(HI)59,59,15
   33 IF(K)74,74,39
   39 WRITE(Z,505)(OBUF(I),I=1,K)
      GO TO 74
C       UNDO EQUIVALANCES CAUSED BY DELETION OF CHAIN PD.
   15  PD=-STACK(IS)
   74  WRITE(Z,572)PD
  572 FORMAT('.UNDO DELETION'I3)
      WRITE(3,505)(STACK(I),I=1,IS)
   72 IS=IS-1
      GI=STACK(IS)
      IF(GI)73,73,42
   42 K=GRPL(GI)
      IS=IS-1
      IF(K)117,117,114
  117 K=-K
      NG=NG&1
  114 GRPL(GI)=K
      IF(STACK(IS))72,72,116
  116 LC=IABS(GRPL(GI&1))-K
      L=MEM(K)
      IF(1-2)45,45,72
   45 M=K&L&1
      CALL LIND4(MEM(M),IA,IB)
      IF(IB-PD)43,44,43
   44 L=L&1
      IF(L-LC)45,43,43
   43 MEM(K)=L
      A=MEM(K&1)
      B=MEM(K&2)
      CALL LIND4(A,JA,IA)
      CALL LIND4(B,JB,IB)
      IF(IA-IB)72,47,72
   47 I=CHAIN(A)
      CHAIN(A)=CHAIN(B)
      CHAIN(B)=I
  113 J=I
      LO=I
   49 KJ=J
      J=CHAIN(J)
      IF(J)111,111,112
  111 I=CHAIN(A)
      GO TO 113
  112 IF(J-I)50,51,50
   50 IF(J-LO)48,49,49
   48 LO=J
      ILO=KJ
      GO TO 49
```

```
   51 CHAIN(ILO)=-LO
      WRITE(3,581)A,B,IA,LO
  581 FORMAT(' UNEQ'2I4,5X,'CHAINS'2I4)
      GO TO 72
C       ALL EQUIVANENCES UNDONE
   80 WRITE(Z,503)
   59 WRITE(Z,500)
C       GOTO RESTRICTION GENERATOR ROUTINE
      IRET=4
C       ROUTINE TO SET A=B
   54 GRPL(GI)=-GRPL(GI)
      NG=NG-1
      IF(IA-IB)52,53,76
   76 CHAIN(JA)=IA
      GO TO 77
   52 CHAIN(JB)=IB
   77 I=CHAIN(A)
      CHAIN(A)=CHAIN(B)
      CHAIN(B)=1
      STACK(IS-1)=1
      IF(TRACE-1)53,53,82
   82 WRITE(Z,507)B,A,IB,IA
   53 GO TO (5,27),IRET
C       *PRINT REMAINING POINTS.
C       FIND A FREE DIMENSION
C       (NONSTANDARD USE OF IS)
    3 DO 70 IS=1,DIM
      IF(COORD(IS)&10000)70,71,70
   70 CONTINUE
C       CHECK EACH PT FOR INCLUSION IN GROUP FOR DIMENSION IS
   71 J=0
      DO 60 PI=1,NP
      IF(LIND3(PI))60,60,61
C       POINT NOT DELETED.  ADD TO OUTPUT BUFFER.
   61 J=J&1
      OBUF(J)=PI
      CHAIN(PI)=-PI
      IF(J-20)60,63,63
   63 WRITE(Z,505)OBUF
      J=0
   60 CONTINUE
  104 IF(J)55,55,66
   66 WRITE(Z,505)(OBUF(I),I=1,J)
      GO TO 55
C       LIST ACTIVE GROUPS
  100 J=0
      DO 101 GI=1,GLIM
      IF(GRPL(GI))101,101,102
  102 J=J&1
      OBUF(J)=GI
      IF(J-20)101,103,103
  103 WRITE(Z,505)OBUF
      J=0
  101 CONTINUE
```

```
      RETURN
  157 WRITE(Z,300)
      CALL EXIT
   78 WRITE(Z,301)
      CALL EXIT
  300 FORMAT('ERROR -- LENGTH OF CANCEL BUFFER EXCEEDED.')
  301 FORMAT('OERROR -- LENGTH OF PTDEL EXCEEDED.')
  500 FORMAT('OEND OF CONSISTENT SET AND SUBSETS'/60('-'))
  501 FORMAT('0*****CONSISTENT SET FOUND WITH FOLLOWING RESTRICTIONS'/
     .1X,8(I5,')=' I5))
  502 FORMAT('OSUBSET'I3,5X,'LIST OF POINTS')
  503 FORMAT('OONLY SUBSET IS SET ITSELF.')
  505 FORMAT(1X,20I6)
  506 FORMAT('.CHAIN'I4,' DELETED')
  507 FORMAT('POINT'I4,' IS EQUIV TO POINT'I4,5X,'CHAIN'I4,'' JOINED TO
      *CHAIN'I4)
      END
// DUP
* STORE    WS UA LIND2
// FOR
* ONE WORD INTEGERS
      FUNCTION LIND3 (PI)
      INTEGER PI
    1 INTEGER DIM,COORD(8),CORG( 600),FAULT(100),FL,GLIM,GPS( 600)
    2 INTEGER GRPL(301),KD,MEM( 600),TMASK(8),TV
    3 INTEGER CORAN(3,8),PD,PTDEL(10 ),STACK(100),CANCL( 600),DIML(8)
    4 INTEGER CLIM,WORK,HI,GI,  TRACE,Z,STO,CHAIN(100)
    5 COMMON DIM,COORD,CORG,FAULT,FL,GLIM,GPS,GRPL,KD,MEM,NP,TMASK,TV
    6  COMMON CORAN,PD,PTDEL,STACK,IS,IRET,CANCL,DIML,CLIM,WORK,HI,LGM
    7 COMMON LEVEL,TRACE,NOSU, GI,Z,NFAIL,NTRY,NCAN,STO,LSTO,CHAIN
      K=(PI-1)*DIM&IS
      K=GPS(K)
C        FIND INDEX OF POINT IN GROUP K
      K=GRPL(K)
      IF(K)1,I,2
C        GROUP EMPTY
    1 LIND3=0
      RETURN
    2 L=MEM(K)&K-1
      DO 3 K=K,L
      IF(MEM(K&1)-PI)3,4,3
    3 CONTINUE
C        NOT IN GROUP
      GO TO 1
    4 LIND3=K
      RETURN
      END
// DUP
*STORE  WS  UA  LIND3
      END
//FOR
*ONE WORD INTEGERS
      SUBROUTINE LIND4(A,B,C)
      INTEGER A,B,C
```

```
      1 INTEGER DIM,COORD(8),CORG( 600),FAULT(100),FL,GLIM,GPS( 600)
      2 INTEGER GRPL(301),KD,MEM( 600),TMASK(8),TV
      3 INTEGER CORAN(3,8),PD,PTDEL(10 ),STACK(100),CANCL( 600),DIML(8)
      4 INTEGER CLIM,WORK,HI,GI,   TRACE,Z,STO,CHAIN(100)
      5 COMMON DIM,COORD,CORG,FAULT,FL,GLIM,GPS,GRPL,KD,MEM,NP,TMASK,TV
      6 COMMON CORAN,PD,PTDEL,STACK,IS,IRET,CANCL,DIML,CLIM,WORK,HI,LGM
      7 COMMON LEVEL,TRACE,NOSU,   GI,Z,NFAIL,NTRY,NCAN,STO,LSTO,CHAIN
        IA=A
      1 B=IA
        IA=CHAIN(IA)
        IF(IA)2,2,1
      2 C=-IA
        RETURN
        END
//DUP
*STORE   WS  UA  LIND4
//FOR
*ONE WORD INTEGERS
        SUBROUTINE LIND5(PI)
        INTEGER PI,OBUF(20)
      1 INTEGER DIM,COORD(8),CORG( 600),FAULT(100),FL,GLIM,GPS( 600)
      2 INTEGER GRPL(301),KD,MEM( 600),TMASK(8),TV
      3 INTEGER CORAN(3,8),PD,PTDEL(10 ),STACK(100),CANCL( 600),DIML(8)
      4 INTEGER CLIM,WORK,HI,GI,   TRACE,Z,STO,CHAIN(100)
      5 COMMON DIM,COORD,CORG,FAULT,FL,GLIM,GPS,GRPL,KD,MEM,NP,TMASK,TV
      6 COMMON CORAN,PD,PTDEL,STACK,IS,IRET,CANCL,DIML,CLIM,WORK,HI,LGM
      7 COMMON LEVEL,TRACE,NOSU,   GI,Z,NFAIL,NTRY,NCAN,STO,LSTO,CHAIN
        J=0
      4 J=J&1
        OBUF(J)=PI
        IF(J-20)1,1,2
      2 WRITE(Z,500)OBUF
        J=0
      1 PI=CHAIN(PI)
        IF(PI)3,3,4
      3 WRITE(Z,500)(OBUF(I),I=1,J)
        RETURN
  500 FORMAT(20I6)
        END
//DUP
*STORE   WS  UA  LIND5
//FOR
*NAME LINDS
*ONE WORD INTEGERS
*IOCS(CARD, 1132 PRINTER)
*LIST SYMBOL TABLE
      1 INTEGER DIM,COORD(8),CORG( 600),FAULT(100),FL,GLIM,GPS( 600)
      2 INTEGER GRPL(301),KD,MEM( 600),TMASK(8),TV
      3 INTEGER CORAN(3,8),PD,PTDEL(10 ),STACK(100),CANCL( 600),DIML(8)
      4 INTEGER CLIM,WORK,HI,GI,   TRACE,Z,STO,CHAIN(100)
      5 COMMON DIM,COORD,CORG,FAULT,FL,GLIM,GPS,GRPL,KD,MEM,NP,TMASK,TV
      6 COMMON CORAN,PD,PTDEL,STACK,IS,IRET,CANCL,DIML,CLIM,WORK,HI,LGM
      7 COMMON LEVEL,TRACE,NOSU,   GI,Z,NFAIL,NTRY,NCAN,STO,LSTO,CHAIN
        IRET=0
```

```
      1  CALL LINDO
      2  CALL LIND1
         IF(IRET-1)1,1,3
      3  CALL LIND2
         IF(IRET-1)1,1,2
// DUP
*STORE   WS  UA  LINDS
```

VITA

Ralph Raymond Lindstrom

Candidate for the Degree of

Doctor of Philosophy

Thesis: ESTIMABLE EFFECTS AND INTERACTIONS IN AN n-WAY CROSS CLASSIFICATION WITH MISSING CELLS

Major Field: Statistics

Biographical:

Personal Data: Born in Ishpeming, Michigan, March 10, 1935, the son of Turri R. and Pearl P. Lindstrom.

Education: Attended grade school in Negaunee, Michigan; graduated from Negaunee High School, Negaunee, Michigan, in 1953; received the Bachelor of Science degree from Northern Michigan University with majors in Mathematics and Industrial Arts in June 1957; received the Master of Science degree from the University of Illinois, with a major in Mathematics, in August 1960; completed the requirements for the Doctor of Philosophy degree with a major in Mathematics and Statistics in May, 1971.

Professional Experience: Employed as an Instructor, Department Mathematics, Michigan Technological University, Houghton, Michigan, 1957-1964, employed as an Assistant Professor, Department of Mathematics, Northern Michigan University, Marquette, Michigan 1966-1970.