MRC Technical Summary Report # 2746

ANALYZING TWO-LEVEL FRACTIONAL FACTORIAL EXPERIMENTS FOR POSSIBLE DISPERSION EFFECTS

George E.P. Box and R. Daniel Meyer

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53705

September 1984

(Received June 29, 1984)

Sponsored by

U.S. Army Research Office
P.O. Box 12211
Research Triangle Park
North Carolina 27709

UNIVERSITY OF WISCONSIN-MADISON
MATHEMATICS RESEARCH CENTER

ANALYZING TWO-LEVEL FRACTIONAL FACTORIAL EXPERIMENTS FOR POSSIBLE DISPERSION EFFECTS

George E.P. Box and R. Daniel Meyer

Technical Summary Report #2746
September 1984

ABSTRACT

After considering the concept of effect sparsity as a justification for the use of unreplicated fractional factorial designs, we discuss the situation where factors may influence not only the location but also the dispersion of the data. The aliasing of location and dispersion effects is explored and methods for identifying an appropriate location - dispersion model are considered.

AMS (MOS) Subject Classifications: 62K15

Key Words: Fractional Factorials, Location Effects, Dispersion Effects, Effect Sparsity, Model Identification.

Work Unit Number 4 (Statistics and Probability)

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

SIGNIFICANCE AND EXPLANATION

Unreplicated fractional factorial designs are frequently employed as screening designs when it is believed that a condition of effect sparsity will ensure that only a few of the possible effects are likely to be large. Suppose it is believed that only an (unidentified) few of a number of candidate design variables such as temperature x_1, pressure x_2, ..., speed x_k affect a quality characteristic y such as observed tensile strength. In the past the emphasis has been on determining the effect of such variables on the mean value or location of y in terms of the usual main effects and interactions which we will here call location effects. We consider in this paper the possibility of also determining the effect of variables on the variance or more generally the dispersion of y. We call such effects dispersion effects. The nature of the alias relationships between location and dispersion effects is discussed and a method developed for identification of dispersion effects when location effects are also present. An example is used to illustrate these ideas.

ANALYZING TWO-LEVEL FRACTIONAL FACTORIAL EXPERIMENTS FOR POSSIBLE DISPERSION EFFECTS

George E.P. Box and R. Daniel Meyer

1. INTRODUCTION

Table 1 shows in summary a highly fractionated two-level factorial design employed* as a screening design in an off-line welding experiment performed by the National Railway Corporation of Japan (Taguchi and Wu, 1980). In the column to the right of the table is shown the observed tensile strength of the weld, one of several quality characteristics measured.

The design was chosen on the assumption that in addition to main effects only the two-factor interactions AC, AG, AH, and GH were expected to be present. On that supposition, all nine main effects and the four selected two-factor interactions can be separately estimated by appropriate orthogonal contrasts and the two remaining contrasts corresponding to the columns labelled e_1 and e_2 measure only experimental error. Below the table are shown the grand average, the fifteen effect contrasts, and the effects plotted on a dot diagram. When the effects are plotted on normal probability paper, thirteen plot roughly as a straight line but the remaining two, corresponding to the main effects for factors B and C, fall markedly off the line, suggesting that over the ranges studied, only factors B and C affect tensile location by amounts not readily attributed to noise.

If this conjecture is true, then, at least approximately, the sixteen runs could be regarded as four replications of a 2^2 factorial design in factors B and C only. However when the results are plotted in Figure 1 so as to reflect this, inspection suggests the existence of a dramatic effect of a different kind - when factor C is at its

*To facilitate later discussion we have set out the design and labelled the levels somewhat differently from Taguchi.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

A: Kind of Welding Rods
B: Period of Drying
C: Welded Material
D: Thickness
E: Angle
F: Opening
G: Current
H: Welding Method
J: Preheating

Factor			D	E	e_1	G	F	GH	AC	A	E	AH	e_2	AG	J	B	C	Tensile strength
Column	Number	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	kg/mm^2
	1	+	-	-	+	-	+	+	-	-	+	+	-	+	-	-	+	43.7
	2	+	+	-	-	-	-	+	+	-	-	+	+	+	+	-	-	40.2
	3	+	-	+	-	-	+	-	+	-	+	-	+	+	-	+	-	42.4
	4	+	+	+	+	-	-	-	-	-	-	-	-	+	+	+	+	44.7
	5	+	-	-	+	+	-	-	+	-	+	+	-	-	+	+	-	42.4
	6	+	+	-	-	+	+	-	-	-	-	+	+	-	-	+	+	45.9
	7	+	-	+	-	+	-	+	-	-	+	-	+	-	+	-	+	42.2
Run	8	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	40.6
	9	+	-	-	+	-	+	+	-	+	-	-	+	-	+	+	-	42.4
	10	+	+	-	-	-	-	+	+	+	+	-	-	-	-	+	+	45.5
	11	+	-	+	-	-	+	-	+	+	-	+	-	-	+	-	+	43.6
	12	+	+	+	+	-	-	-	-	+	+	+	+	-	-	-	-	40.6
	13	+	-	-	+	+	-	-	+	+	-	-	+	+	-	-	+	44.0
	14	+	+	-	-	+	+	-	-	+	+	-	-	+	+	-	-	40.2
	15	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	42.5
	16	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	46.5
Effect		43.0	.13	-.15	-.30	-.15	.40	-.03	.38	.40	-.05	.43	.13	.13	-.38	2.15	3.10	

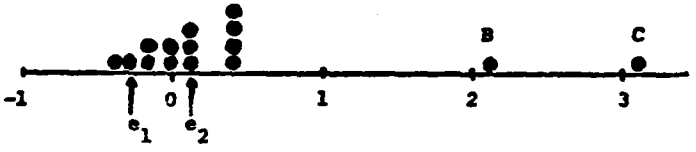


TABLE 1. A fractional two-level design used in a welding experiment showing observed tensile strength and effects. Below the estimated effects are plotted as a dot diagram.

plus level the spread of the residuals appears much larger* than when it is at its minus level. Thus in addition to detecting shifts in location due to B and C, the experiment may also have detected what we will call a dispersion effect due to C. The example raises the general possibility pursued in the remainder of this paper of analyzing unreplicated designs for dispersion effects as well as for the more usual location effects.

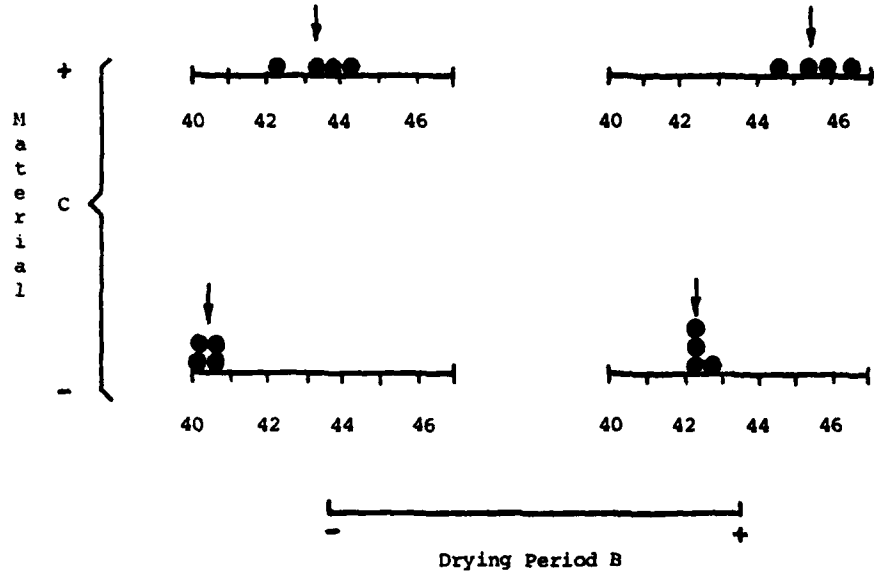


Figure 1. Tensile data as four replicates of a 2^2 factorial design in factors B and C only.

*Data of this kind might be accounted for by the effect of one or more variables other than B that affected tensile strength only at the "plus level" of C (only when the alternative material was used). Analysis of the eight runs made at the plus level of C does not support this possibility, however.

2. RATIONALE FOR USING SCREENING DESIGNS

Before proceeding we need to consider the question, "In what situations are screening designs, such as highly fractionated factorials, useful?"

2.1. Effect Sparsity

A common industrial problem is to find from a larger number of factors those few that are responsible for large effects. The idea is comparable to that which motivates the use in quality control studies of the "Pareto diagram." (See, for example, Ishikawa 1976). The situation is approximated by postulating that only a small proportion of effects will be "active" and the rest "inert". We call this the postulate of effect sparsity. For studying such situations, highly fractionated designs and other orthogonal arrays (Tippet (1934) Finney (1945), Plackett & Burman (1946), Rao (1947), Taguchi and Wu (1980)) which can screen moderately large numbers of variables in rather few runs are of great interest. Two main rationalizations have been suggested for the use of these designs; both ideas rely on the postulate of effect sparsity but in somewhat different ways.

2.2. Rationale Based on Prior Selection of Important Interactions

It is argued (see for example Davies, 1954) that in some circumstances physical knowledge of the process will make only a few interactions likely and that the remainder may be assumed negligible. For example, in the welding experiment described above there were 36 possible two-factor interactions between the nine factors, but only four were regarded as likely, leaving 32 such interactions assumed negligible. The difficulty with this idea is that in many applications the picking out of a few "likely" interactions is difficult if not impossible. Indeed the investigator might justifiably protest that, in the circumstance where an experiment is needed to determine which first order (main) effects are important, it is illogical that he be expected to guess in advance which effects of second order (interactions) are important.

2.3. Projective Rationale Factor Sparsity

A slightly different notion is that of factor sparsity. Thus suppose that, of the k factors considered, only a small subset of unknown size d, whose identity is also

unknown, will be active in providing main effects and interactions within that subset. Arguing as in Box and Hunter (1961) a two level design enabling us to study such a system is a fraction of resolution $R = d + 1$ (or in the terminology of Rao (1947) an array of strength d) which produces complete factorials (possibly replicated) in every one of the $\binom{k}{d}$ spaces of $d = R - 1$ dimensions. For example, we have seen that on the assumption that only factors B and C are important, the welding design could be regarded as four replicates of a 2^2 factorial in just those two factors. But because the design is of resolution $R = 3$ the same would have been true for any of the 36 choices of two out of the nine factors tested. Thus the design would be appropriate if it were believed that not more than two of the factors were likely to be "active."

For further illustration we consider again the sixteen-run orthogonal array of Table 1. In Table 2 adopt a roman subscript to denote the resolution R of the design, which associates factors with the dotted columns in the manner shown.

	Columns	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
(a)	2^{15-11}_{III}	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
(b)	2^{8-4}_{IV}	•	•		•			•	•			•		•	•	
(c)	2^{5-1}_{V}	•	•		•				•							•
(d)	2^4	•	•		•				•							

TABLE 2. Some alternative uses of the orthogonal array of Table 1.

It will be seen that:

(a) if we associated the fifteen contrast columns of the design with fifteen factors, we would generate a 2^{15-11}_{III} design providing four-fold replication of 2^2 factorials in every one of the 105 two-dimensional projections.

(b) if we associated only columns 1, 2, 4, 7, 8, 11, 13, and 14 with eight factors we would generate a 2^{8-4}_{IV} design providing two-fold replication of 2^3 factorials in every one of the 56 three-dimensional projections.

(c) if we associated only columns 1, 2, 4, 8, and 15 with five factors we would generate a 2_V^{5-1} design providing a 2^4 factorial in every one of the four-dimensional projections.

(d) if we associate only columns 1, 2, 4 and 8 with four factors we would obtain the complete 2^4 design from which this orthogonal array was in fact generated.

Designs (a), (b) & (c) would thus be appropriate for situations where we believed respectively that not more than 2, 3, or 4 factors would be active*. Notice that intermediate numbers of factors could be accommodated by suitably omitting certain columns. Thus the welding design is a 2_{III}^{9-5} arrangement which can be obtained by omitting 6 columns from the complete 2_{III}^{15-11}. Notice finally that for intermediate designs we can take advantage of both rationales by arranging as was done for the welding design, that particular interactions are isolated.

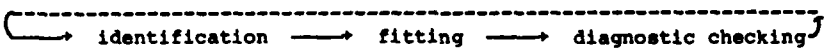
2.4. Clues and Leads: Not Final Conclusions

In the past some misunderstanding about the value of fractional designs has occurred because they were not always considered in an appropriate context. For instance, while the above analysis of the welding data screening design does not lead to unequivocal conclusions, it does suggest the possibility of location effects in B and C and a dispersion effect in C which are worth checking out. Typically the experimenter will be following, with the help of a statistician, an iterative path of investigation which cannot be forecast at the outset and he must behave very much like a detective investigating a mystery. He proceeds by following clues leading to further searchings (experiments) which provide further clues and so on. The prime objective is to achieve reasonably rapid convergence of the investigatory process. While it is true that more formal statistical procedures may be needed, to confirm finally that the investigation has

*The designs give partial coverage for a larger number of factors, for example (Box & Hunter (1961)) 56 of the 70 four-dimensional projections of the 2_{IV}^{8-4} yield a full factorial.

really reached the objective claimed, this later confirmatory stage is usually only a small part of the experimental effort.

A discussion of the iterative model building process by Box & Jenkins (1970) characterized three steps in the iterative data analysis cycle indicated below.



Most of the present paper is concerned with model identification--that is the selection of a model worthy to be entertained and fitted more formally by an efficient process such as maximum likelihood.

The situation we now address, therefore, concerns the identification of factor effects from fractional designs in the circumstance of effect sparsity, where dispersion effects as well as location effects may be present.

3. DISPERSION EFFECTS

We again use the design of Table 1 for illustration. There are 16 runs from which 16 quantities--the average and 15 effect contrasts--have been calculated. Now if we were also interested in possible dispersion effects we could also calculate 15 variance ratios. For example, in column 1 we can compute the sample variance $s^2(1-)$ for those observations associated with a minus sign and compare it with the sample variance $s^2(1+)$ for observations associated with a plus sign, to provide the ratio $F_1 = s^2(1-)/s^2(1+)$. If this is done for the welding data we obtain values for $\ln F_i$* given in Figure 2(a). It will be recalled that in the earlier analysis a large dispersion effect associated with factor C (column 15) appeared to be present but in Figure 2(a) the effect for this factor is not especially extreme, instead the dispersion effect for factor D (column 1) stands out from all the rest. This misleading indication occurs b[illegible] have not so far taken account of the aliasing of location and dispersion efe[illegible] Since sixteen linearly independent location effects have already been calculated for the original data, calculated dispersion effects must be functions of these. The general nature of the location-dispersion aliasing is explained in the section that follows. For illustration equation (1) shows the identity that exists for the dispersion effect that is the F ratio associated with factor D and hence for column 1 of the design.

In the expression $\hat{i}$ is used to indicate the usual location effect (contrast)

$$F_D = F_1 = \frac{(\hat{2}-\hat{3})^2 + (\hat{4}-\hat{5})^2 + (\hat{6}-\hat{7})^2 + (\hat{8}-\hat{9})^2 + (\hat{10}-\hat{11})^2 + (\hat{12}-\hat{13})^2 + (\hat{14}-\hat{15})^2}{(\hat{2}+\hat{3})^2 + (\hat{4}+\hat{5})^2 + (\hat{6}+\hat{7})^2 + (\hat{8}+\hat{9})^2 + (\hat{10}+\hat{11})^2 + (\hat{12}+\hat{13})^2 + (\hat{14}+\hat{15})^2} \qquad (1)$$

Now (see Table 1) $\hat{B} = \hat{14} = 2.15$ and $\hat{C} = \hat{15} = 3.10$ are the two largest location effects, standing out from all the others. The extreme value of F_D associated with an

*In this figure familiar normal theory significance levels are also shown. Obviously the necessary assumptions are not satisfied in this case, but these percentages provide a rough indication of magnitude.

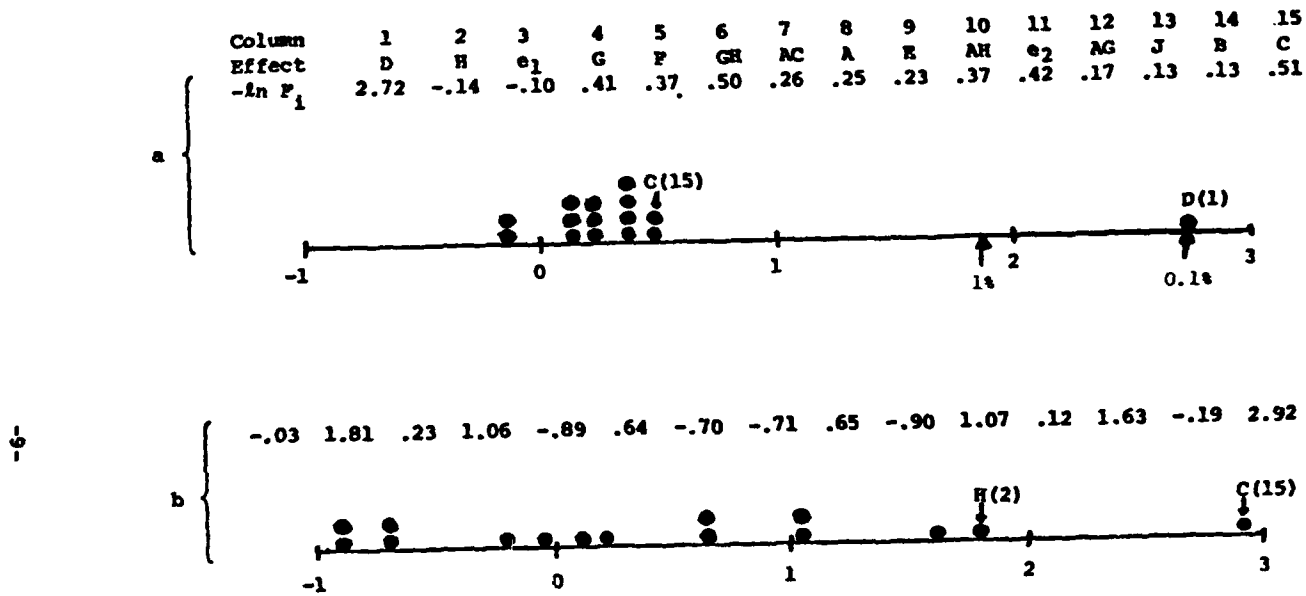


Figure 2. Welding experiment log dispersion effects (a) before, and (b) after elimination of location effects for B and C.

apparent dispersion effect for factor D is thus largely accounted for by the squared sum and squared difference of the location effects $\hat{B}$ and $\hat{C}$ which appear respectively as the last terms in the denominator and numerator of equation 1. A natural way to proceed is to compute variances from the residuals obtained after eliminating large location effects. After such elimination the alias relations of equation 1 remain the same except that location effects from eliminated variables drop out. That is, zeros are substituted for eliminated variables. Variance analysis for residuals after eliminating effects of B and C are shown in Figure 2(b). The dispersion effect associated with C (factor 15) is now correctly indicated as extreme.

4. DISPERSION AND LOCATION ALIASING

4.1. Identities Existing Between Dispersion and Location Effects

In order to study the identity relations existing between location and disperson effects consider an $n \times n$ orthogonal array with $n = 2^q$ columns of +1's and -1's labelled $\underset{\sim}{x}_0, \underset{\sim}{x}_1, \ldots, \underset{\sim}{x}_{n-1}$. Let $\underset{\sim}{x}_0 = I$ be a column of +1's and the remaining columns delineate the usual contrasts for the main effects and interactions of a 2^q factorial design. In general we suppose that the array is to be used as a 2^{k-p} fractional or full factorial to test k factors, so that $q = k - p$ with $p \geq 0$.

To illustrate ideas we will employ an eight run orthogonal array although in practice this would often be too small a design to allow variance effects to be studied. Setting $q = 3$ the columns of the resulting 2^3 design in factors A, B, and C can alternatively be labelled with numbers or letters as follows:

0	1	2	3	4	5	6	7
I	A	B	AB	C	AC	BC	ABC
+1	-1	-1	+1	-1	+1	+1	-1
+1	+1	-1	-1	-1	-1	+1	+1
+1	-1	+1	-1	-1	+1	-1	+1
+1	+1	+1	+1	-1	-1	-1	-1
+1	-1	-1	+1	+1	-1	-1	+1
+1	+1	-1	-1	+1	+1	-1	-1
+1	-1	+1	-1	+1	-1	+1	-1
+1	+1	+1	+1	+1	+1	+1	+1

As is well known the array may be used as a full factorial or as a fractional design. For example, associating three factors with columns 1, 2, 4 reproduces the 2^3 factorial, four factors associated with columns 1, 2, 4, 7 produces a 2^{4-1}_{IV} fractional, seven factors associated with columns 1 through 7 produces a 2^{7-4}_{III} fractional.

In general the columns $\underset{\sim}{x}_0, \underset{\sim}{x}_1, \ldots, \underset{\sim}{x}_{n-1}$ will form a group closed under multiplication defined such that product column $\underset{\sim}{x}_{ij}$ has for its u^{th} element $x_{iju} = x_{iu}x_{ju}$ $(u = 1, \ldots, n)$. Now suppose we are interested in contrasting variances at the lower and upper levels of factor i. Consider the elements of a column $\frac{1}{2}(\underset{\sim}{x}_0 \pm \underset{\sim}{x}_i)(i \neq 0)$; these are

$$\tfrac{1}{2}(x_{0u} - x_{iu}) = \begin{cases} +1 & \text{if } x_{iu} = -1 \\ 0 & \text{if } x_{iu} = +1 \end{cases}$$

$$\tfrac{1}{2}(x_{0u} + x_{iu}) = \begin{cases} 0 & \text{if } x_{iu} = -1 \\ +1 & \text{if } x_{iu} = +1 \end{cases} \qquad (2)$$

Thus the elements of a column $\tfrac{1}{2}(\underset{\sim}{x}_j \pm \underset{\sim}{x}_{ij})$ are

$$\tfrac{1}{2}(x_{ju} - x_{iju}) = \tfrac{1}{2}x_{ju}(x_{0u} - x_{iu}) = \begin{cases} x_{ju} & \text{if } x_{iu} = -1 \\ 0 & \text{if } x_{iu} = +1 \end{cases}$$

$$\tfrac{1}{2}(x_{ju} + x_{iju}) = \tfrac{1}{2}x_{ju}(x_{0u} + x_{iu}) = \begin{cases} 0 & \text{if } x_{iu} = -1 \\ x_{ju} & \text{if } x_{iu} = +1 \end{cases} \qquad (3)$$

Returning for illustration to the 2^3 design, if for example we wished to compare variances at the lower and upper level of factor $C = \underset{\sim}{x}_4$ (i.e. for the first and last four observations) then setting $i = 4$ it will be useful to generate the columns $\tfrac{1}{2}(\underset{\sim}{x}_j - \underset{\sim}{x}_{4\cdot j})$ $j = 0,1,\ldots,7$ thus:

j	0	4	1	5	2	6	3	7
	+1	−1	−1	+1	−1	+1	+1	−1
	+1	−1	+1	−1	−1	+1	−1	+1
	+1	−1	−1	+1	+1	−1	−1	+1
	+1	−1	+1	−1	+1	−1	+1	−1
$\tfrac{1}{2}(\underset{\sim}{x}_j - \underset{\sim}{x}_{4\cdot j})$	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0

Note that for every i the columns $(\underset{\sim}{x}_j - \underset{\sim}{x}_{i\cdot j})$ will in general appear in $\frac{n}{2} = 2^{q-1}$ identical (apart from sign) pairs. Now suppose data $\underset{\sim}{y} = (y_1,\ldots,y_u,\ldots,y_n)$ are available and let $\hat{j} = \underset{\sim}{y}'\underset{\sim}{x}_j$ from which the estimated effect of factor j may be obtained by dividing by an appropriate constant. Then for every i the quantities $\underset{\sim}{y}'(\underset{\sim}{x}_j - \underset{\sim}{x}_{i\cdot j}) = \hat{j} - \hat{i\cdot j}$ provide an exhaustive set of $n/2$ linearly independent

contrasts of those $n/2$ observations y_u for which $x_{iu} = -1$. Correspondingly, the columns $x_j + x_{i \cdot j}$ provide a similar set of contrasts for the remaining observations for which $x_{iu} = +1$. Denote by $S(i-)$ and $S(i+)$ the sums of squares of the y_u for which $x_{iu} = -1$ and $+1$ respectively. Then

$$S(i-) = \frac{1}{n} \sum_{j=0}^{n-1} [\tfrac{1}{2} x'(x_j - x_{i \cdot j})]^2 = \frac{1}{n} \sum_{j=0}^{n-1} \left(\frac{\hat{j} - \hat{i \cdot j}}{2}\right)^2$$

$$S(i+) = \frac{1}{n} \sum_{j=0}^{n-1} [\tfrac{1}{2} x'(x_j + x_{i \cdot j})]^2 = \frac{1}{n} \sum_{j=0}^{n-1} \left(\frac{\hat{j} + \hat{i \cdot j}}{2}\right)^2 . \tag{4}$$

For example, for the 8×8 array derived from the 2^3 factorial,

$$S(4-) = y_1^2 + y_2^2 + y_3^2 + y_4^2$$

$$= \frac{1}{8} \left[\left(\frac{\hat{0-4}}{2}\right)^2 + \left(\frac{\hat{1-5}}{2}\right)^2 + \left(\frac{\hat{2-6}}{2}\right)^2 + \left(\frac{\hat{3-7}}{2}\right)^2 + \left(\frac{\hat{4-0}}{2}\right)^2 + \left(\frac{\hat{5-1}}{2}\right)^2 + \left(\frac{\hat{6-2}}{2}\right)^2 + \left(\frac{\hat{7-3}}{2}\right)^2\right]$$

$$= \frac{1}{4} \left[\left(\frac{\hat{0-4}}{2}\right)^2 + \left(\frac{\hat{1-5}}{2}\right)^2 + \left(\frac{\hat{2-6}}{2}\right)^2 + \left(\frac{\hat{3-7}}{2}\right)^2\right]$$

$$= \frac{1}{4} \left[\left(\frac{\hat{I-C}}{2}\right)^2 + \left(\frac{\hat{A-AC}}{2}\right)^2 + \left(\frac{\hat{B-BC}}{2}\right)^2 + \left(\frac{\hat{AB-ABC}}{2}\right)^2\right] . \tag{5}$$

4.2. Elimination of Location Effects

The sums of squares in (4), (5) would be appropriate to compute dispersion effects only if it could be assumed that all the location effects, including the overall mean, were *known* to be zero. If this were not the case then the sums of squares $S(i-)$ and $S(i+)$ could be inflated by location effects. To remove such effects we can replace the y_u's in (5) by residuals $y_u - \hat{y}_u$ obtained after eliminating all suspected location effects including the mean by least squares.

Now the vector of residuals is orthogonal to each column vector corresponding to an eliminated variable. It follows that sums of squares calculated from such residuals will

have the same form as (4) but with all estimated effects which correspond to eliminated variables set equal to zero.

Further understanding is gained by considering the expected values of $S(i-)$ and $S(i+)$ under various circumstances. Suppose a difference in variance might exist associated with the level of the single column x_i and the sums of squares $S(i-)$ and $S(i+)$ are computed from (4) but with y_u replaced by residuals after a number of location effects have been eliminated. Then after setting to zero all the elements $\hat{j}$ and $\hat{i \cdot j}$ in (5) which correspond to eliminated variables, suppose there are ℓ cases where bracketed pairs $(\hat{j}, \hat{i \cdot j})$ have been eliminated and m cases where only one element of a bracketed pair has been eliminated so that there remains $\frac{n}{2} - \ell - m$ complete bracketed pairs.

Now for a bracketed pair

$$E\left[\frac{2}{n}\left\{\frac{1}{2}(\hat{j} - \hat{i \cdot j})\right\}^2\right] = \sigma^2(i-) \tag{6}$$

and for a single element

$$E\left[\frac{2}{n}\left\{\frac{1}{2}\hat{j}\right\}^2\right] = \frac{1}{4}\left(\sigma^2(i-) + \sigma^2(i+)\right) . \tag{7}$$

It follows that

$$E[S(i-)] = (\tfrac{1}{2}n - \ell - \tfrac{3}{4}m)\sigma^2(i-) + \tfrac{1}{4}m\sigma^2(i+) \tag{8}$$

$$E[S(i+)] = (\tfrac{1}{2}n - \ell - \tfrac{3}{4}m)\sigma^2(i+) + \tfrac{1}{4}m\sigma^2(i-) . \tag{9}$$

If we define

$$s^2(i-) = S(i-)/(\tfrac{1}{2}n - \ell - \tfrac{1}{2}m) , \tag{10}$$

then

$$E[s^2(i-)] = \sigma^2(i-) + \frac{m}{2n - 4\ell - 2m}\left(\sigma^2(i+) - \sigma^2(i-)\right) \tag{11}$$

and similarly for $s^2(i+)$ with the roles of $\sigma^2(i-)$ and $\sigma^2(i+)$ reversed.

It should be noted that, in the circumstances of effect sparsity here considered, the bias term in (11) involving $\sigma^2(i+) - \sigma^2(i-)$ would be rather small. For example, suppose, with a design having $n = 16$ runs, that $\ell = 2$ and $m = 1$, then the bias term will be

$\{\sigma^2(i+) - \sigma^2(i-)\}/22$. It seems reasonable to conclude that for purposes of model identification the elimination of location effects by simply taking residuals is unlikely to mislead.

4.3. Illustrations with the 8 × 8 Array

The general situation may be better understood by considering a few special cases again using for illustration the 8 × 8 factorial array. Setting i = 4 = C, suppose we wish to obtain the dispersion effect $s^2(4-)/s^2(4+)$ which contrasts the variances of the first four and last four observations.

Elimination of Grand Mean

Elimination of the mean which would usually be unknown results in the removal of $\hat{0}$ in equations (4). For the 8 × 8 array; n = 8, ℓ = 0, m = 1

$$s^2(4-) = \{\tfrac{1}{16}\left[(\hat{4})^2 + (\hat{1} - \hat{5})^2 + (\hat{2} - \hat{6})^2 + (\hat{3} - \hat{7})^2\right]\}/(7/2)$$

and using (11)

$$E[s^2(4-)] = \sigma^2(4-) + \tfrac{1}{14}\left[\sigma^2(4+) - \sigma^2(4-)\right]$$

The slight bias in the variance estimate arises because the isolated effect $\hat{4}$ is a function of all eight observations.

Elimination of the Mean and Effect $\hat{4}$

If now the location effect associated with factor 4 is eliminated as well as the overall mean then a complete pair is removed in (5) and in this example

$$s^2(4-) = \tfrac{1}{16}\left\{(\hat{1} - \hat{5})^2 + (\hat{2} - \hat{6})^2 + (\hat{3} - \hat{7})^2\right\}/3$$

$$E[s^2(4-)] = \sigma^2(4-)$$

No bias now occurs because elimination of 0 and 4 is equivalent to eliminating means separately from the first four and the last four observations, and $s^2(4-)$ is a function of only the first four observations. Similar effects are found with all bracketted

pairs. Thus if we eliminate factor 2 and the interaction $2 \cdot 4 = 6$ the bias term does not appear because allowance is being made for different effects of factor 2 at the two levels of factor 4.

4.3. Dispersion Interactions

Since more than one dispersion effect might be present we need to consider the possibility of interaction. If the effect of changing from the minus level to the plus level of a factor i is to multiply the variance by ϕ_i irrespective of whether the plus or minus level of factor j is employed we shall say that there is no dispersion interaction between i and j. In such a case the variances for the various factor combinations are as follows

j \ i	$-$	$+$
$+$	$\sigma^2(i-,j+) = \phi_j\sigma^2$	$\sigma^2(i+,j+) = \phi_i\phi_j\sigma^2$
$-$	$\sigma^2(i-,j-) = \sigma^2$	$\sigma^2(i+,j-) = \phi_i\sigma^2$

Equivalently for the logged variances the dispersion effects will be additive and in this metric dispersion interactions of all orders may be defined in the usual way. It shall be noted that when there is no dispersion interaction the ratio of the average variance at the plus and minus levels for factor i is

$$\frac{\sigma^2(i+,j-) + \sigma^2(i+,j+)}{\sigma^2(i-,j-) + \sigma^2(i-,j+)} = \frac{\phi_i(1 + \phi_j)\sigma^2}{(1 + \phi_j)\sigma^2} = \phi_i$$

and similarly for factor j and ϕ_j. Thus even when there is more than one dispersion effect the simple analysis described above could still be of value as a preliminary analytical device for indicating which factors needed further study. In particular if two factors i and j appeared to exhibit dispersion effects, then further analysis would be appropriate to consider the general evidence for activity of these effects taking account also of possible interaction. This could be done by considering general differences among

the sums of squares associated with the four cells S(i-,j-), S(i-,j+), S(i+,j-), S(i+,j+) of the two-way table for the two factors. As before these sums of squares would be calculated from residuals after eliminating location effects. The consequences of doing this is explored in the Appendix which gives a matrix generalization of earlier results.

A convenient function for comparing a set of variances $s_1^2,\ldots,s_k^2$ having $\nu_1,\ldots,\nu_k$ degrees of freedom respectively is Bartlett's criterion,

$$M = N\,\ell n\Big(N^{-1}\sum_{t=1}^{k}\nu_t s_t^2\Big) - \sum_{t=1}^{k}\nu_t\,\ell n\, s_t^2\,, \qquad \text{where } N = \sum_{t=1}^{k}\nu_t\,.$$

When, as would frequently be the case, the screening design is of only moderate size one could not expect to study simultaneously a large number of factors in this way. For example, for $n = 16$, the individual cells from which S(i-,j-) S(i-,j+), etc. would be calculated will each contain only four observations. However when, in circumstances of effect sparsity, only a very few such effects are likely to be of appreciable magnitude, the above analysis could be of value.

We again illustrate with the welding data. Figure 3(a) shows the 35 distinct values of M computed for the data. There are $\binom{15}{2} = 105$ ways of choosing two columns from the 15 columns of the design but these are aliased in sets of three (any column is the product of two other columns). Thus the largest value is associated with columns 15 = C, 2 = H, and 13 = J. This effect could equally well be attributed to factors C and H with interaction CH = J or to C and J with interaction CJ = H or to H and J with interaction HJ = C. It is noteworthy that the seven largest values of M which stand out from the rest* all include factor C in their triplets. Also if the dispersion effect of C is eliminated by rescaling the residuals the plot (Fig. 3(b)) no longer shows outstanding points.

*For rough comparison the normal theory 5% and 1% significance levels of M are shown although as before their precise validity is doubtful.

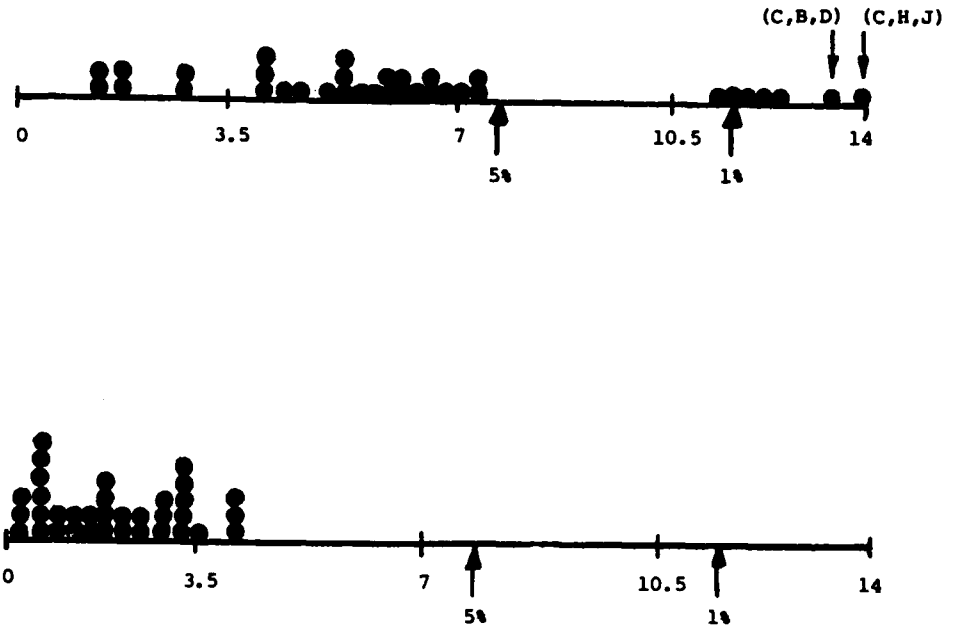


Figure 3. Values of M for distinct column triplets (a) before (b) after elimination of the possible dispersion effect due to factor C.

5. MAXIMUM LIKELIHOOD ESTIMATES OF LOCATION AND DISPERSION EFFECTS

Once a model has been identified a more precise fitting is possible using maximum likelihood. Hartley and Jayatillake (1973) have shown that the following method will give convergence to a stationary point of the likelihood. Conditional on the dispersion effects, location effects may be obtained by weighted least squares; the dispersion effects may now be recomputed from the residuals and the iteration continued until covergence is achieved. It is often convenient to assume initially that there are no dispersion effects.

For illustration the following table shows maximum likelihood estimates for the welding data assuming location effects for B and C and a dispersion effect for C. The earlier approximate estimates are indicated for comparison.

	$\hat{\mu}$	$\hat{B}$	$\hat{C}$	$\hat{\sigma}^{2}_{(C+)}$	$\hat{\sigma}^{2}_{(C-)}$	$\frac{\hat{\sigma}^{2}_{(C+)}}{\hat{\sigma}^{2}_{(C-)}}$
Maximum likelihood estimates	42.96	2.04	3.10	.469	.021	22.3
Earlier approximate estimates	43.00	2.15	3.10	.564	.031	18.2

Table 3. Estimates of location and dispersion affects: welding data.

Appendix

The results of section 4 can be generalized using matrix algebra. Again, let $\underset{\sim}{X}$ be a matrix of $\pm$ 1's with orthogonal columns $\underset{\sim}{x}_0,\ldots,\underset{\sim}{x}_{n-1}$ and $\underset{\sim}{x}_0$ a column of 1's. If $\underset{\sim}{y}$ denotes the $n \times 1$ vector of observations, we define $\hat{j} = \underset{\sim}{x}_j'\underset{\sim}{y}$. Suppose we wish to compute $S(i-)$ = sum of squares of the y_u at the minus level of column $\underset{\sim}{x}_i$. Without loss of generality, we assume $\underset{\sim}{x}_i = (-1,-1,\ldots,-1,1,1,\ldots,1)'$, and let $\underset{\sim}{I}$ be the $n \times n$ identity matrix. Define the $n \times n$ matrices $\underset{\sim}{I}_1$, $\underset{\sim}{I}_2$ by

$$\underset{\sim}{I}_1 = \begin{pmatrix} \underset{\sim}{I}_{n/2} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0} \end{pmatrix} \qquad \underset{\sim}{I}_2 = \begin{pmatrix} \underset{\sim}{0} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{I}_{n/2} \end{pmatrix}$$

where $\underset{\sim}{I}_{n/2}$ is the $\frac{n}{2} \times \frac{n}{2}$ identity matrix.

Then S(i-) can be written

$$S(i-) = (\underset{\sim}{I}_1\underset{\sim}{Y})'(\underset{\sim}{I}_1\underset{\sim}{Y}) .$$

Noting that $\frac{1}{n}\underset{\sim}{X}\underset{\sim}{X}' = \underset{\sim}{I}$ and $\underset{\sim}{I}_1\underset{\sim}{X} = \frac{1}{2}[\underset{\sim}{X} - (\underset{\sim}{I} - 2\underset{\sim}{I}_1)\underset{\sim}{X}]$

we have

$$S(i-) = \frac{1}{4n}[(\underset{\sim}{X} - (\underset{\sim}{I} - 2\underset{\sim}{I}_1)\underset{\sim}{X})'\underset{\sim}{Y}]'[(\underset{\sim}{X} - (\underset{\sim}{I} - 2\underset{\sim}{I}_1)\underset{\sim}{X})'\underset{\sim}{Y}] .$$

Now observe that $(\underset{\sim}{I} - 2\underset{\sim}{I}_1)\underset{\sim}{X}$ is just the matrix $\underset{\sim}{X}$ with every column multiplied by x_i. Therefore,

$$(\underset{\sim}{X} - (\underset{\sim}{I} - 2\underset{\sim}{I}_1)\underset{\sim}{X})'y = (\hat{0}-i\hat{\cdot}0,\ \hat{1}-i\hat{\cdot}1,\ldots,\hat{n-1} - i\hat{\cdot}n-1)'$$

and

$$S(i-) = \frac{1}{4n}\sum_{j=0}^{n-1}(\hat{j} - i\hat{\cdot}j)^2$$

as was shown in Section 4, equation (4). Similarly, we can write

$$S(i+) = \frac{1}{4n}\sum_{j=0}^{n-1}(\hat{j} + i\hat{\cdot}j)^2 .$$

As was also shown in section 4, eliminated location effects will drop out of the above expressions.

To compute the expectation of S(i-), let $\sigma^2(i-)$ and $\sigma^2(i+)$ be the variance of y_u at the minus and plus levels of x_i, and let the matrix $\underset{\sim}{Z}$ be those columns of $\underset{\sim}{X}$ corresponding to location effects included in the model (i.e. eliminated to obtain residuals). The least squares estimates of location effects are given by $\underset{\sim}{I} = \frac{1}{n}\underset{\sim}{Z}'\underset{\sim}{Y}$. Then

$$\begin{aligned} S(i-) &= (\underset{\sim}{Y} - \underset{\sim}{Z}\hat{\underset{\sim}{I}})'\underset{\sim}{I}_1(\underset{\sim}{Y} - \underset{\sim}{Z}\hat{\underset{\sim}{I}}) \\ &= \underset{\sim}{Y}'(\underset{\sim}{I} - \tfrac{1}{n}\underset{\sim}{Z}\underset{\sim}{Z}')\underset{\sim}{I}_1(\underset{\sim}{I} - \tfrac{1}{n}\underset{\sim}{Z}\underset{\sim}{Z}')\underset{\sim}{Y} \\ &= \underset{\sim}{Y}'(\underset{\sim}{I}_1 - \tfrac{1}{n}\underset{\sim}{Z}\underset{\sim}{Z}'\underset{\sim}{I}_1 - \tfrac{1}{n}\underset{\sim}{I}_1\underset{\sim}{Z}\underset{\sim}{Z}' + \tfrac{1}{n^2}\underset{\sim}{Z}\underset{\sim}{Z}'\underset{\sim}{I}_1\underset{\sim}{Z}\underset{\sim}{Z}')\underset{\sim}{Y} \end{aligned}$$

Assuming $E[y] = \underset{\sim}{Z}\underset{\sim}{\beta}$ and using the identity

$$E[\underset{\sim}{y}'\underset{\sim}{A}\underset{\sim}{y}] = \text{trace}\left(\underset{\sim}{A}\cdot E[\underset{\sim}{y}\underset{\sim}{y}']\right)$$

for $\underset{\sim}{A}$ symmetric

$$E[S(i-)] = \text{trace}\left[\left(\underset{\sim}{I}_1 - \frac{1}{n}\underset{\sim}{Z}\underset{\sim}{Z}'\underset{\sim}{I}_1 - \frac{1}{n}\underset{\sim}{I}_1\underset{\sim}{Z}\underset{\sim}{Z}' + \frac{1}{n^2}\underset{\sim}{Z}\underset{\sim}{Z}'\underset{\sim}{I}_1\underset{\sim}{Z}\underset{\sim}{Z}'\right)\cdot \left(\sigma^2(i-)\underset{\sim}{I}_1 + \sigma^2(i+)\underset{\sim}{I}_2 + \underset{\sim}{Z}\underset{\sim}{I}\underset{\sim}{I}'\underset{\sim}{Z}'\right)\right]$$

After some algebraic reduction, we have

$$E[S(i-)] = \left(\frac{n}{2} - p\right)\sigma^2(i-) + \frac{1}{n^2}\left[K\sigma^2(i-) + \left(\frac{n^2p}{2} - K\right)\sigma^2(i+)\right],$$

where p = number of columns of $\underset{\sim}{Z}$ and $K = \text{trace}(\underset{\sim}{Z}\underset{\sim}{Z}'\underset{\sim}{I}_1\underset{\sim}{Z}\underset{\sim}{Z}'\underset{\sim}{I}_1)$. K can be simplified to give the following expression:

$$K = \sum_{u=1}^{\frac{n}{2}}\sum_{v=1}^{\frac{n}{2}}\left(\sum_{j=1}^{p} z_{ju}z_{jv}\right)^2$$

where z_{ju} is the uth element of the jth column of $\underset{\sim}{Z}$. Now let ℓ = number of pairs of columns $\underset{\sim}{Z}_j$, $\underset{\sim}{Z}_k$ related by $\underset{\sim}{Z}_k = \underset{\sim}{Z}_i\cdot\underset{\sim}{Z}_j$. Expanding the above expression for K gives

$$K = \sum_{u=1}^{\frac{n}{2}}\sum_{v=1}^{\frac{n}{2}}\left(\sum_{j=1}^{p} z_{ju}^2z_{jv}^2 + 2\sum_{j<k}\sum z_{ju}z_{jv}z_{ku}z_{kv}\right)$$

$$= \frac{n^2}{4}p + 2\sum_{j<k}\sum\left(\sum_{u=1}^{\frac{n}{2}} z_{ju}z_{ku}\right)\left(\sum_{v=1}^{\frac{n}{2}} z_{jv}z_{kv}\right)$$

The sums $\sum_{u=1}^{\frac{n}{2}} z_{ju}z_{ku}$ and $\sum_{v=1}^{\frac{n}{2}} z_{jv}z_{kv}$ are equal to -1 whenever $\underset{\sim}{Z}_j = \underset{\sim}{Z}_i\cdot\underset{\sim}{Z}_k$ and they are equal to 0 whenever $\underset{\sim}{Z}_j \neq \underset{\sim}{Z}_i\cdot\underset{\sim}{Z}_k$.

Therefore

$$K = \frac{n^2}{4} p + 2\ell \ .$$

and

$$E[S(i-)] = \left(\frac{n}{2} - p + \frac{p+2\ell}{4}\right)\sigma^2(i-) + \left(\frac{n^2 p}{4} - \frac{2\ell}{4}\right)\sigma^2(i+)$$

$$= \left(\frac{n}{2} - \ell - \frac{3}{4} m\right)\sigma^2(i-) + \frac{m}{4}\sigma^2(i+)$$

where $m = p - 2\ell$.

To extend these derivations to computations involving two variables, define matrices $\underset{\sim}{I}_3$, $\underset{\sim}{I}_4$, $\underset{\sim}{I}_5$, $\underset{\sim}{I}_6$ as

$$\underset{\sim}{I}_3 = \begin{bmatrix} \underset{\sim}{I}_{n/4} & & \underset{\sim}{0} \\ & \underset{\sim}{0} & \\ \underset{\sim}{0} & & \underset{\sim}{0} \end{bmatrix} \qquad \underset{\sim}{I}_4 = \begin{bmatrix} \underset{\sim}{0} & & \underset{\sim}{0} \\ & \underset{\sim}{I}_{n/4} & \\ \underset{\sim}{0} & & \underset{\sim}{0} \end{bmatrix}$$

$$\underset{\sim}{I}_5 = \begin{bmatrix} \underset{\sim}{0} & & \underset{\sim}{0} \\ & \underset{\sim}{0} & \\ \underset{\sim}{0} & & \underset{\sim}{0} \end{bmatrix} \qquad \underset{\sim}{I}_6 = \begin{bmatrix} \underset{\sim}{0} & & \underset{\sim}{0} \\ & \underset{\sim}{0} & \\ \underset{\sim}{0} & & \underset{\sim}{I}_{n/4} \end{bmatrix}$$

where $\underset{\sim}{I}_{n/4}$ is the $\frac{n}{4} \times \frac{n}{4}$ identity matrix. Now suppose we wish to compute the sum of squares of the y_u at the minus levels of $\underset{\sim}{x}_i$ and $\underset{\sim}{x}_j$, S(i-,j-). Without loss of generality, assume $\underset{\sim}{x}_i = (-1,-1,\ldots,-1,1,1,\ldots,1)$ and $x_j = (-1,\ldots,-1;1,\ldots,1;-1,\ldots,-1;1,\ldots,1)$. Then

$$S(i-,j-) = (\underset{\sim}{I}_3 \underset{\sim}{y})'(\underset{\sim}{I}_3 \underset{\sim}{y})$$

Now use the facts $I = \frac{1}{n} \underset{\sim}{X}\underset{\sim}{X}'$,

$$\underset{\sim}{I}_3 = \frac{1}{4}\left[\underset{\sim}{I} - \left(\underset{\sim}{I} - 2(\underset{\sim}{I}_3 + \underset{\sim}{I}_4)\right) - \left(\underset{\sim}{I} - 2(\underset{\sim}{I}_3 + \underset{\sim}{I}_5)\right) + \left(\underset{\sim}{I} - 2(\underset{\sim}{I}_4 + \underset{\sim}{I}_5)\right)\right]$$

and

$$[I - (I - 2(I_3 + I_4)) - (I - 2(I_3 + I_5)) + (I - 2(I_4 + I_5))]X'X =$$

$$(\hat{0} - i\hat{\cdot}0 - j\hat{\cdot}0 + i\hat{\cdot}j\hat{\cdot}0,\ \hat{1} - i\hat{\cdot}1 - j\hat{\cdot}1 + i\hat{\cdot}j\hat{\cdot}1, \ldots,$$

$$n\hat{-}1 - i\hat{\cdot}n\hat{-}1 - j\hat{\cdot}n\hat{-}1 + i\hat{\cdot}j\hat{\cdot}n-1)'$$

to obtain

$$S(i-,j-) = \frac{1}{16n} \sum_{k=0}^{n-1} (\hat{k} - i\hat{\cdot}k - j\hat{\cdot}k + i\hat{\cdot}j\hat{\cdot}k)^2 \ .$$

Similarly,

$$S(i-,j+) = \frac{1}{16n} \sum_{k=0}^{n-1} (\hat{k} - i\hat{\cdot}k + j\hat{\cdot}k - i\hat{\cdot}j\hat{\cdot}k)^2$$

$$S(i+,j-) = \frac{1}{16n} \sum_{k=0}^{n-1} (\hat{k} + i\hat{\cdot}k - j\hat{\cdot}k - i\hat{\cdot}j\hat{\cdot}k)^2$$

$$S(i+,j+) = \frac{1}{16n} \sum_{k=0}^{n-1} (\hat{k} + i\hat{\cdot}k + j\hat{\cdot}k + i\hat{\cdot}j\hat{\cdot}k)^2 \ .$$

To compute the expectation of $S(i\pm,j\pm)$, let $\underset{\sim}{Z}$ be defined as previously, and by the same sort of calculations as in the one variable case

$$E[S(i-,j-)] = \frac{n - 2p}{4}\sigma^2(i-,j-) + \frac{1}{n^2}[K_3\sigma^2(i-,j-) + K_4\sigma^2(i-,j+)$$

$$+ K_5\sigma^2(i+,j-) + K_6\sigma^2(i+,j+)]$$

where $K_i = \text{trace}\ (\underset{\sim}{Z}\underset{\sim}{Z}'I_3\underset{\sim}{Z}\underset{\sim}{Z}'I_i)$.

To compute K_i, we must divide the columns of $\underset{\sim}{Z}$ into four groups:

Group 4: Those columns z_k such that $x_i \cdot z_k$, $x_j \cdot z_k$ and $x_i \cdot x_j \cdot z_k$ are also in $\underset{\sim}{Z}$.

Group 3: Those columns z_k such that exactly two of $x_i \cdot z_k$, $x_j \cdot z_k$, $x_i \cdot x_j \cdot z_k$ are in $\underset{\sim}{Z}$.

Group 2: Those columns z_k such that only one of $x_i \cdot z_k$, $x_j \cdot z_k$, $x_i \cdot x_j \cdot z_k$ is in $\underset{\sim}{Z}$.

Group 1: Those columns not in previous three groups.

Let m_k = number of columns in group k; note that m_k is a multiple of k. Further subdivide group 2 into three subsets

Group 2.1: those pairs with $z_g = x_i \cdot z_k$

Group 2.2: those pairs with $z_g = x_j \cdot z_k$

Group 2.12: those pairs with $z_g = x_i \cdot x_j \cdot z_k$.

Let $m_{2.k}$ be the number of columns in group 2.k; $m_2 = m_{2.1} + m_{2.2} + m_{2.12}$.
Thus it can be shown that

$$E[S(i-,j-)] = \sigma^2(i-,j-)\left[\frac{4n - 7p + 3m_4 + 2m_3 + m_2}{16}\right]$$

$$+ \sigma^2(i-,j+)\left[\frac{p - m_4 - \frac{2}{3}m_3 - m_2 + 2m_{2.1}}{16}\right]$$

$$+ \sigma^2(i+,j-)\left[\frac{p - m_4 - \frac{2}{3}m_3 - m_2 + 2m_{2.2}}{16}\right]$$

$$+ \sigma^2(i+,j+)\left[\frac{p - m_4 - \frac{2}{3}m_3 - m_2 + 2m_{2.12}}{16}\right]$$

Note that $S(i-,j-)$ will be unbiased (up to a scale factor) if all columns of $\underset{\sim}{Z}$ are in group 4 ($p = m_4$, $m_3 = m_2 = m_1 = 0$) i.e. for each variable x_k eliminated, variables $\underset{\sim}{x}_{i \cdot k}$, $\underset{\sim}{x}_{j \cdot k}$, $\underset{\sim}{x}_{i \cdot j \cdot k}$ are also eliminated. Similar expressions for expectations of $S(i-,j+)$, $S(i+,j-)$ and $S(i+,j+)$ can be worked out quite easily from the above formula by switching signs on $i\pm,j\pm$.

REFERENCES

Taguchi, G. and Wu, Y. (1980). Introduction to Off-Line Quality Control. Central Japan Quality Control Association, Nagoya, Japan.

Ishikawa, K. (1976). Guide to Quality Control. Asian Productivity Organization, Tokyo.

Tippett, L.H.C. (1934). Applications of Statistical Methods to the Control of Quality in Industrial Production. Manchester Statistical Society.

Finney, D.J. (1945). The Fractional Replication of Factorial Arrangements. Annals of Eugenics 12, 4, 291-301.

Plackett, R.L. and Burman, J.P. (1946). Design of Optimal Multifactorial Experiments. Biometrika 23, 305-325.

Rao, C.R. (1947). Factorial Experiments Derivable from Combinatorial Arrangements of Arrays. J. Roy. Statist. Soc., B9, 128-140.

Davies, O.L. ed. (1954). The Design and Analysis of Industrial Experiments. Oliver and Boyd, London.

Box, G.E.P. and Hunter, J.S. (1961). The 2^{k-p} Fractional Factorial Designs. Technometrics 3, 311-351, 449-458.

Box, G.E.P. and Jenkins, G.M. (1970). Time Series Analysis, Forecasting and Control. Holden-Day, San Francisco.

Hartley, H.O. and Jayatillake, K.S.E. (1973). Estimation for Linear Models With Unequal Variances. J. Amer. Statist. Assoc. 68, 189-192.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER #2746	2. GOVT ACCESSION NO. AD-A149455	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) ANALYZING TWO-LEVEL FRACTIONAL FACTORIAL EXPERIMENTS FOR POSSIBLE DISPERSION EFFECTS		5. TYPE OF REPORT & PERIOD COVERED Summary Report - no specific reporting period 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) George E.P. Box and R. Daniel Meyer		8. CONTRACT OR GRANT NUMBER(s) DAAG29-80-C-0041
9. PERFORMING ORGANIZATION NAME AND ADDRESS Mathematics Research Center, University of Wisconsin 610 Walnut Street Madison, Wisconsin 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Work Unit Number 4 - Statistics and Probability
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office P. O. Box 12211 Research Triangle Park, North Carolina 27709		12. REPORT DATE September 1984 13. NUMBER OF PAGES 25
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Fractional Factorials, Location Effects, Dispersion Effects, Effect Sparsity, Model Identification

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

After considering the concept of effect sparsity as a justification for the use of unreplicated fractional factorial designs, we discuss the situation where factors may influence not only the location but also the dispersion of the data. The aliasing of location and dispersion effects is explored and methods for identifying an appropriate location - dispersion model are considered.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)