Screening designs for poly-factor experimentation

Kleijnen, J.P.C.

Published in:
Technometrics

Publication date:
1975

TECHNOMETRICS©, VOL. 17, NO. 4, NOVEMBER 1975

Screening Designs for Poly-Factor Experimentation

Jack P. C. Kleijnen

Katholieke Hogeschool
Tilburg, Netherlands

A survey is given of the following types of screening designs: Incomplete 2^k designs, supersaturated and group-screening designs. These designs are compared with each other. Some new results for group-screening are derived.

KEY WORDS

Experimental design
Screening
Sequential designs
Simulation

1. INTRODUCTION

At the beginning of an investigation there may be very many, say k, conceivably important factors. We may suppose that not all k factors are important but only a few, say k', factors. We do not know the value of k' and neither do we know *which* k' factors among the k factors are important and therefore we have to *screen* for these factors. Recently Ott and Wehrfritz (1972) proposed a class of designs requiring only N runs to test $2^N - 1$ factors, e.g. 4 runs for 15 factors. They remark that other strategies exist but do not give any references. Therefore our purpose is to survey some other design types; we compare the appropriateness of these types and add some new results for group-screening designs. Though the designs presented here are more than 10 years old, they are not discussed in the textbooks on experimental design. Yet we think that they may be very useful, especially in computer simulation experiments where very many factors can be studied.

2. INCOMPLETE 2^k DESIGNS OF RESOLUTION III AND IV

We assume that the reader has at least some familiarity with basic experimental design so that we do not have to explain in detail 2^{k-p} designs, etc. Many textbooks on these designs are available; for a brief résumé we refer to e.g. Box and Hunter [3]. Remember that all k factors are studied at only 2 levels; of the resulting 2^k possible level combinations only a fraction is actually run. A proper choice of this selection may yield:

Received July 1973; revised January 1975

(*i*) *Designs of resolution III:*
No main effect is confounded with any other main effect, but main effects are confounded with two-factor interactions and two-factor interactions with each other. If we assume absence of interactions, then designs exist that permit the estimation of the k main effects and the overall mean in only $N = k + 1$ runs (provided N is a multiple of 4; otherwise N is rounded upwards to the next multiple of 4, e.g. for $k = 2$ $N = 4$). These designs were tabulated by Plackett and Burman [10] for $N \leq 100$; their table is reproduced in Kleijnen [6]. Note that the familiar 2^{k-p} designs of resolution III are a subclass of the Plackett–Burman designs.

(*ii*) *Designs of resolution IV:*
No main effect is confounded with any other main effect *or* two-factor interaction, but two-factor interactions are confounded with each other. So these designs permit estimation of the main effects even if two-factor interactions exist. This class of designs can be constructed from the resolution III designs by combining the resolution III design with its "reversed" partner. In the reversed part + and − levels of the resolution III design are simply switched to − and + (moreover, a new factor can be associated with the column of the overall mean in the resolution III design). For instance, if factor 1 is studied at the levels + − + − in the resolution III design, then it is now studied at the levels (+ − + −), (− + − +). Consequently, resolution IV designs exist for k factors and $N = 2\ k$ runs (N being a multiple of 8).

For a further discussion of incomplete 2^k designs we refer to the literature. We point out that these designs are not well-suited to situations with "very" many factors since even a resolution III design requires at least $k + 1$ runs. Therefore, we proceed to the next type.

3. Supersaturated Designs

Booth and Cox [2] considered two-level factorials, where each of the k factors is at the "low" level $N/2$ times and the "high" level $N/2$ times (N being even). It is supposed that $N \leq k$. Let $\mathbf{d}_j$ denote the design column of factor j, i.e. the column with elements $x_{ij} = +1$ if factor j is at its high level, and $x_{ij} = -1$ if factor j is at its low level ($i = 1, \cdots, N$) ($j = 1, \cdots, k$). Not all k design columns can be orthogonal since $N \leq k$. Nevertheless we can try to obtain a "nearly" orthogonal design. This was quantified by Booth and Cox as

$$\min \{\max_{h \neq j} |\mathbf{d}_h' \mathbf{d}_j|\} \qquad (1)$$

If several designs have the same value for (1) then that design is chosen in which the number of pairs of columns attaining (1) is smallest. Booth and Cox tabulated the resulting designs for $N = 12$, $k \leqq 16$, 20, 24; $N = 18$, $k \leqq 24$, 30, 36; $N = 24$, $k \leqq 30$.

The authors showed that their designs are more "efficient" than "random" designs. They measured *efficiency* by the variance of z when z is the inner-product of two design columns, or

$$z = \mathbf{d}_j' \mathbf{d}_h \qquad (j \neq h) \qquad (2)$$

Random designs are designs where all or some of the elements x_{ij} of the design matrix are chosen by sampling. We do not further discuss these designs since there are many advantages in using other alternatives; see Kleijnen [6] or the elaborate discussion in *Technometrics* in 1959.

Booth and Cox [2] state "We have no experience of practical problems where such designs are likely to be useful; the conditions that interactions should be unimportant and that there should be a few dominant main effects seem very severe." However, the existence of only a few main effects is what a scientist often hopes for since it enables him to explain and control a complicated system; see also Ackoff [1], (p. 149–150). So assuming a screening situation as explained at the beginning of Section 1 seems realistic for certain situations. The scientist may also be willing to assume the absence of interactions. Yntema and Torgerson [12], (p. 22) discuss the appropriateness of approximating a response curve by main effects only. Observe that the orthogonality of the design columns j and h would vanish in general if the fundamental variables were not x_j and x_h but some transformation of these variables.

If $k > 36$ or we want to make fewer runs than listed, then no tabulated supersaturated designs are available. We could derive a design using the iterative computer procedure in Booth and Cox [2], (p. 492–494). The time needed to write and to run the computer program for the generation of such a design might very well be prohibitive. In that case we might use the following type of design.

4. Group-Screening Designs

In these designs the k factors are grouped into g groups, each group being considered as a single factor. The g group-factors may be tested in an incomplete 2^k design. The assumptions of group-screening imply that if a group-factor is found to be insignificant, then all original factors within that group-factor are insignificant and can be dropped from further investigation. If a group-factor is found to be significant, then one or more original factors in that group are significant. So in the next stage these original factors should be further investigated. In a two-stage procedure the g group-factors are tested in the first stage and the original factors of the significant groups only are tested individually in the second stage. Such a procedure was introduced by Watson in 1961. In multi-stage group-screening the groups found to be significant in the first stage (or more generally in stage i) are repartitioned into smaller groups that are tested in the next stage. These multi-stage designs were introduced independently of each other by Patel and Li in 1962. Let us consider two-stage and multi-stage group-screening in some more detail.

(a) *Two-Stage Group-Screening*

Watson (1961, p. 372) used the following assumptions (the possibility to relax certain assumptions will be discussed later):

"(*i*) all factors have, independently, the same prior possibility of being effective, $p (q = 1 - p)$,
(*ii*) effective factors have the same effect, $\Delta > 0$,
(*iii*) there are no interactions present,
(*iv*) the required designs exist,
(*v*) the directions of possible effects are known,
(*vi*) the errors of all observations are independently normal with a constant known variance, σ^2,
(*vii*) $k = gf$ where g = number of groups, and f = number of factors per group".

(In this quotation, we use our symbol k to denote the number of original factors rather than Watson's symbol. We let f denote the group size.)

Because of assumption (v) we can define the upper level of a factor as the level producing the highest response. The upper level of a *group*-factor is then defined as the level where all factors in that group are at their high level. Together with assumption (iii) this ensures that no *cancellation* of effects can occur. So a group-factor containing one or more effective factors gives a non-zero effect. We shall

discuss an example of group-screening before investigating these assumptions.

Suppose there are 9 original factors, $A, B, \cdots, I$, partitioned into 3 groups labelled X, Y, Z. Hence each group consists of 3 original factors. Testing 3 (group-)factors for their main effects can be done in a resolution III design with 4 runs: $N_1 = 2^{3-1}$. A possible fraction is given in Table 1. If X consists of A, B and C; Y consists of D, E and F, etc., then in run 1 the original factors $A, B, \cdots, I$ are at the levels $-1, -1, -1, -1, -1, -1, 1, 1, 1$. Suppose that the first stage shows that only group-factor X is effective. (Because of assumption vi ANOVA can be used, provided we have an independent estimator of σ^2.) In the second stage we then have to test only the factors within X, namely A, B and C. These 3 factors can again be tested in 4 runs: $N_2 = 2^{3-1}$.

Next we shall show that the *assumptions* (i) through (vii) are not very restrictive. Assumption (iv) was introduced in order to derive the group-size that minimizes the number of runs in the stages 1 and 2. This assumption means that in the first stage a design exists for testing g group-factors in $(g + 1)$ runs. Actually we have seen that a resolution III design in $N_1 = (g + 1)$ runs is available only if N_1 is a multiple of 4. Moreover, this assumption means that in the second stage a design exists with $N_2 = fs$ runs where s is the number of group-factors found to be significant in the first stage. This implies that one run of the previous stage is used and that $(fs + 1)$ is a multiple of 4. In practice N_1 and $(N_2 + 1)$ may not be a multiple of 4. Therefore the derived optimal values for the group-size will not be exactly optimal. However, this does not invalidate the procedure. Further if $(g + 1)$ and $(fs + 1)$ are not a multiple of 4, then the number of observations is larger than the number of parameters and degrees of freedom remain to estimate the experimental error variance. Assumption (ii) is also needed to derive the optimal group-size and therefore is not crucial. Assumption (v) can be weakened for Watson [11], (p. 385) derived: "As optimumsized groups, for $p \leq 0.15$, have only a chance of 0.06 or less of containing two or more effective factors, we will in practice not need to know the directions of all the possible effects"; also compare our comment on equation (4) below. Assumption (vi) is again needed for the derivation of the optimal design and also makes ANOVA possible; ANOVA is robust with regard to nonnormality and heteroscedasticity. Assumption (i) should be interpreted as follows. We need some *prior rough estimate* as to how many factors are thought to be effective among the total of k factors. Then p is equal to the ratio of the likely number of effective factors and the total number of factors. This p determines the optimal group-size f, derived by Watson to be

$$f_0 \approx 1/\sqrt{(1 - \alpha_1)p} \qquad (3)$$

where α_1 is the significance level in the first stage. Equation (3) shows that the group-size decreases as p increases; for high p the optimal group-size becomes 1, i.e. the factors are tested individually. This is a reasonable result since a high p means that many groups contain effective factors so in the second stage all or nearly all original factors must be tested and nothing has been gained by the grouping procedure. Of course a high value of p is contradicting the definition of a screening situation. If we have no firm estimate of p, this means that our grouping will not be optimal. However, formula (3) for the optimal group-size is only approximate and because the groupsize f must be an integer, f_0 will not fluctuate much with varying p (and α_1). For instance, if there is no experimental error ($\alpha_1 = 0$) then for $0.03 \leq p \leq 0.30$, Watson's table 2 shows $6 \geq f_0 \geq 3$. If we do not suppose that all factors have the *same* prior probability p, we can form classes of factors each class having its own prior probability. Equation (3) shows that factors with a high probability p should be placed in small groups; if $p > .30$ we take group-size 1, i.e. these factors are tested individually. So different estimates of p (which are more realistic) can be incorporated into the group-screening method. Moreover they make the grouping more flexible. The group-size is no longer a constant, but can be varied and assumption (vii) is replaced by the less restrictive assumption

$$k = g_1 \cdot f_1 + g_2 \cdot f_2 + \cdots + g_J \cdot f_J \qquad (4)$$

where g_j is the number of groups of size f_j $(j = 1, \cdots, J)$. Further the number of group-factors tested in the first stage, i.e. $\sum g_j$, can be so chosen that $N_1 = \sum g_j + 1$ is a multiple of 4; compare Watson [11], (p. 383–385). We point out that *unequal group-sizes* make it possible to test a factor *individually* when we do not know the direction of its effect (and are afraid of cancellations of various effects notwithstanding the remark we made above concerning the low probability of such cancellations). Finally, in the appendix below we weakened assumption (iii) about *interactions*. For we derived that a two-factor interaction β_{zw} biases the estimator of the

TABLE 1: *A* 2_{III}^{3-1} *design in the* 3 *group-factors X, Y and Z.*

run	X	Y	Z(= XY)
1	−1	−1	+1
2	+1	−1	−1
3	−1	+1	−1
4	+1	+1	+1

main effect of a factor p, only if the factors z, w and p belong to 3 different group factors (z, w, p = 1, $\cdots$, k). (Pure quadratic effects β_{jj} never bias the estimator of the main effect.) Therefore, if we assume that two-factor interactions exist only between particular (related) factors, then we should place those *related* factors in the *same* group. Then their two-factor interactions will not bias any estimated main effect. We further derived that if we examine the g *group*-factors in a *resolution IV* design, then main effects are not biased by any two-factor interaction (but of course are still confounded with each other within the same group-factor).

(b) Multi-Stage Group-Screening

Patel generalized Watson's procedure to more than 2 stages in order to achieve a further reduction of the number of runs. It can be shown that he used the same kind of assumptions as Watson did. Patel [9], (p. 214) derived that the expected number of runs over all $(n + 1)$ stages is minimized by choosing the *number of groups* as in (5) and (6) where g_i does not denote the total number of groups in stage i, but the number of group into which g_{i-1} is split.

$$g_1 \approx kp^{n/(n+1)} \tag{5}$$

$$g_2 = g_3 = \cdots = g_n = g_{n+1} \approx p^{-1/(n+1)} \tag{6}$$

With (5) and (6) the following *group-sizes* correspond

$$f_i \approx p^{-[n-(i-1)]/(n+1)} \tag{7}$$

Equation (7) shows that the group-sizes are geometrically decreasing with ratio $p^{1/(n+1)}$. He further showed that the total number of runs is minimized by choosing an n-stage procedure instead of an $(n - 1)$-stage procedure if

$$p < \left(1 - \frac{1}{n}\right)^{n(n-1)} \tag{8}$$

Another author, Li [7], (p. 463), also tabulated the resulting optimal number of stages as a function of p. He derived a similar multistage procedure, using the same type of assumptions as Patel. Li pointed out that it may be attractive to choose a procedure with *less stages than the optimum* if the increase in the total number of runs is only small. He derived that the per cent change in the number of runs when choosing an $(n + 1)$ stage procedure instead of an n-stage procedure is

$$\frac{N_{n+1} - N_n}{N_n} = \left(\frac{n+1}{n}\right) \cdot p^{1/n(n+1)} - 1 \tag{9}$$

We would suggest using formula (9) as follows. Given p, we determine the optimal number of stages, n_0 . Put $n_0 = n + 1$. If (9) gives only a small negative percentage then it is attractive to use $(n_0 - 1)$ stages.

Next we shall consider the problem of *estimating the prior probability* p.

(i) Li assumes that all important factors show up in separate groups and he neglects experimental error. Consequently, after the first stage we know exactly P, the number of important factors, and hence p. He further shows that the number of observations over all n stages, increases only slightly when our estimate $\hat{p}$ is wrong. For instance, if our initial estimate of p is wrong by a factor a, i.e. $\hat{p} = ap$, and $a = 1$ or $\frac{1}{2}$, then the total number of observations increases less than 6%.

(ii) Suppose, contrary to Li, that we take into account the, albeit small, probability that more than one important factor occurs within a group. Then we might use

$$\hat{P}_i = s_i \tag{10}$$

where s_i is the stochastic number of significant groups in stage i. As i increases, the group-size decreases, and (10) will be a better approximation. For the smaller the group-size, the smaller the probability that 2 or more important factors occur together in a single group. A summary of the way various quantities influence each other is given in figure 1. Note that, contrary to Patel, Li defines g as the *total* number of groups. Figure 2 shows how after each stage we can recalculate the optimal number of stages n_0 , the group-size f_0 and the resulting number of groups g_0 . When calculating these quantities we use the most recent information on P, and act as if we were to start with the *first* stage of an n-stage procedure characterized by the *current* P and the current k, which is the number of factors after elimination of insignificant groups.

A closer look at the blocks 8, 9 and 2 in figure 2 reveals that in stage i we estimate p by (11) where P_{i-1} is the number of significant factors in the previous stage.

$$\hat{p}_i = \frac{P_{i-1}}{k_{i-1}} = \frac{P_{i-1}}{P_{i-1} \cdot f_{i-1}} = \frac{1}{f_{i-1}} \tag{11}$$

The true p in stage i is equal to (12) (assuming that none of the P important factors is erroneously eliminated).

$$p_i = \frac{P}{k_i} = \frac{P}{P_{i-1} \cdot f_{i-1}} \tag{12}$$

So $\hat{p}_i$ will be a correct estimate if $P_{i-1} = P$. The probability that the latter equality holds, increases as the group-size f_{i-1} decreases. For the smaller the group-size, the smaller the probability that 2 or more

important factors occur together within 1 group. We can easily prove that the group-size does decrease as the procedure goes on, i.e. $f_i < f_{i-1}$.

Li [7], (p. 465–466) shows that the total number of observations is not affected when we divide the total group of k factors into b *subgroups*, estimate the prior probability in each subgroup by the *same* quantity P/k, and use his procedure for each subgroup separately. We would go one step further and estimate the prior probability for each subgroup by its *own* estimator if we used prior knowledge to put related factors together in a subgroup. Then in the first stage, we could save observations as our estimate $\hat{p}_i (j = 1, \cdots, b)$ is more correct.

To conclude this discussion of group-screening, the most *crucial assumptions* are: There are no interactions and the directions of the main effects are known. Otherwise a group-factor may show no effect whereas actually effects do exist. Elimination of this group-factor means that the effects of the original factors in that group are not detected. Regarding the assumption of *known directions* of the *main* effects, we repeat that factors with unknown directions may be put into separate groups of size 1, or we may rely on the low probability of more than 2 important factors occuring within 1 group and moreover having opposite effects that cancel out exactly (or having opposite effects of nearly the same magnitude such that the experimental error and the combined effects of minor factors mask the net result of the important factors with opposite signs). We further add some examples demonstrating that the experimenter may indeed know which signs the effects will have. As we mentioned in section 1 a fruitful application area of screening designs seems to be computer simulation experiments. Let us give just two examples: 1) Simulation of the effect of the quality of information: If we have "better" information (e.g. more timely or more accurate information), then we expect to obtain a better response. 2) Simulation of queuing systems: If we increase the number of service stations then less waiting time is expected. If we switch to the queuing discipline "small jobs first served" again smaller average waiting time is ex-

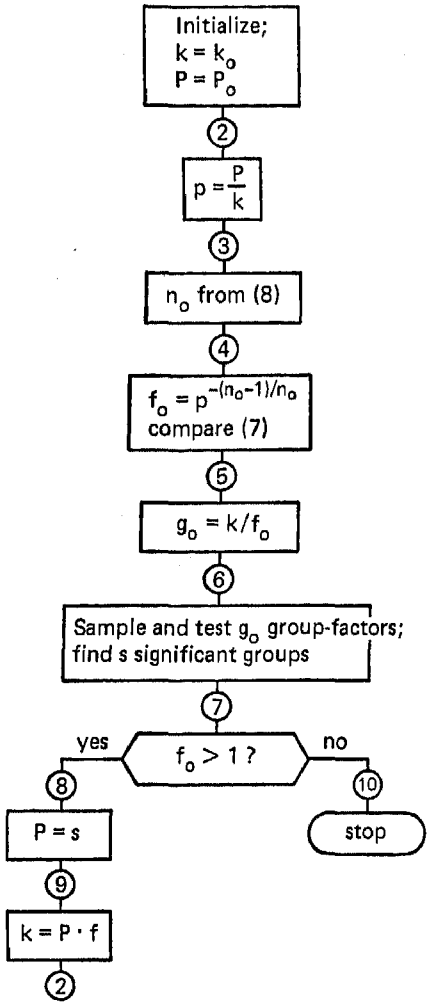
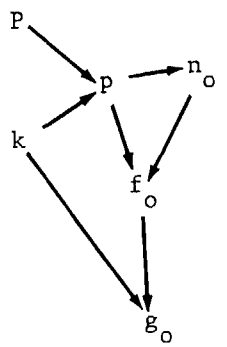


FIGURE 2—Flow diagram for the sequentialized estimation of the parameters of a multi-stage group-screening design.



P = number of important factors

k = number of factors

p = probability of being an important factor

n_0 = optimal number of stages

f_0 = optimal group-size

g_0 = optimal number of groups

FIGURE 1—Interdependence of the various parameters of a multi-stage group-screening design.

pected. In these examples prior theoretical or "common sense" reasoning makes us expect a certain direction of the effect. The experiment serves to confirm this expectation (i.e. there is an important effect in the expected direction) or to show that a possible effect is so small that it can be ignored. A case study with known direction of effects can be found in Hunter and Mezaki (1964). Regarding the assumption about the *interactions* we repeat that main effects are not masked by two-factor interactions if these interactions occur only among factors of the same group or if we examine the group-factors in a resolution IV design. We observe that independently of Li, Patel and Watson, group-screening (or "sequential bifurcation") was proposed by Jacoby and Harrison (1962, p. 131–133). At each stage they propose to split each significant group into exactly 2 subgroups (or in Patel's notation $g_i = 2$).

5. Conclusion

In this section we hope to aid the reader in evaluating the appropriateness of the various designs.

(*i*) $N \geq k + 1$

As we mentioned in Section 2 incomplete 2^k designs exist for the estimation of the k main effects in only $N = k + 1$ runs if N is a multiple of 4; otherwise, N is rounded upwards to the next multiple of 4. No interactions are assumed. If the experimenter can afford so many observations, then these resolution III designs are recommended. If he can afford $2k$ observations, then two-factor interactions do not disturb the main effect estimators (resolution IV design).

(*ii*) $N < k + 1$

In this case the experimenter may be advised to apply group-screening unless the following situation occurs.

(*iii*) $N = 12\,k \leq 24$ or $N = 18\,k \leq 36$ or $N = 24$ $k \leq 30$

For these cases, Booth and Cox tabulated supersaturated designs assuming no interactions. If the experimenter does not accept this assumption, then he may apply group-screening following the advice in the last paragraph of Section 4.

(*iv*) If the experimenter is not satisfied with the above designs, he should consult a professional statistician since other designs have been published or can be derived; for details see Kleijnen [6].

6. Appendix
Interactions in Group-Screening

Suppose the true model contains two-factor interactions β_{jh} or

$$E(y_i) = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + \sum_j \sum_h \beta_{jh} x_{ij} x_{ih} \qquad (i = 1, \cdots, N) \qquad (1.1)$$

where if all the factors are quantitative we may permit $j = h$, i.e. pure quadratic effects β_{jj}. Note that usually the main effects and interactions, say α, are defined as twice the regression parameters β; see the factor 2 in (1.13). Let the k factors x_j be grouped into g groups $X_1, \cdots, X_g$. Test these g groups in a resolution III design. Then we know that

$$\sum_{i=1}^{N} x_{ij} = 0 \qquad j = 1, \cdots, k \qquad (1.2)$$

and if the factors j and j' belong to the *same* group

$$\sum_i (x_{ij} x_{ij'}) = \sum_i (+1) = N \qquad (1.3)$$

If the factors j and j' belong to 2 *different* group-factors then

$$\sum_i (x_{ij}, x_{ij'}) = 0 \qquad (1.4)$$

since the 2 group-factors are orthogonal in a resolution III design. Next let us consider the sum of products of 3 factors, i.e. $\sum_i x_{ij} x_{ij'} x_{ij''}$. There are 3 possibilities:

(*i*) The factors j, j' and j'' belong to the same group-factor. Then

$$(x_{ij} x_{ij'}) = +1 \qquad (i = 1, \cdots, N) \qquad (1.5)$$

which, together with (1.2), yields

$$\sum_i (x_{ij} x_{ij'}) x_{ij''} = \sum_i x_{ij''} = 0 \qquad (1.6)$$

(*ii*) The factor j and j' belong to one group-factor and j'' belongs to another group. Then as in (1.5) and (1.6) we find

$$\sum_i x_{ij} x_{ij'} x_{ij''} = 0 \qquad (1.7)$$

(*iii*) The 3 factors, j, j' and j'' all belong to different group-factors, say X_1, X_2 and X_3. In a 2_{III}^{g-p} design 1 group-factor must be confounded with the interaction between 2 other group-factors, say $\vec{\mathbf{X}}_3 = \mathbf{X}_1 \vec{} \mathbf{X}_2$. Consequently

$$x_{ij''} = x_{ij} x_{ij'} \qquad (i = 1, \cdots, N) \qquad (1.8)$$

So

$$\sum_i x_{ij} x_{ij'} x_{ij''} = \sum_i (x_{ij''})^2 = \sum_i (+1) = N \qquad (1.9)$$

In a Plackett–Burman design, not a 2_{III}^{g-p} design, the two-factor interaction can be expressed as a linear combination of the g main effects and the grand mean. Hence

$$(\mathbf{x}_i\mathbf{x}_{i'}) = \sum_{l=0}^{g} a_l\mathbf{x}_l \tag{1.10}$$

or

$$x_{ij}x_{ij'} = \sum_{l=0}^{g} a_l x_{il} \tag{1.11}$$

Consequently

$$\sum_i x_{ij}x_{ij'}x_{ij''} = \sum_i \sum_{l=0}^{g} a_l x_{il}x_{ij''}$$
$$= \sum_l a_l \sum_i x_{il}x_{ij''} = Na_{j''} \tag{1.12}$$

where the last equality follows from (1.3) and (1.4).

The main effect of factor p ($p = 1, \cdots, k$) is estimated by

$$\hat{\alpha}_p = \frac{2}{N}\sum_{i=1}^{N} x_{ip}y_i \tag{1.13}$$

Hence

$$E(\hat{\alpha}_p) = \frac{2}{N}\sum_i x_{ip}E(y_i)$$
$$= \frac{2}{N}\sum_i x_{ip}\Big(\beta_0 + \sum_{j=1}^{n}\beta_j x_{ij} + \sum_i\sum_h \beta_{ih}x_{ij}x_{ih}\Big)$$
$$= \frac{2}{N}\Big[\beta_0\sum_i x_{ip} + \sum_j \beta_j \sum_i x_{ip}x_{ij} + \sum_j\sum_h \beta_{jh}\sum_i x_{ip}x_{ij}x_{ih}\Big] \tag{1.14}$$

Consider the 3 sumterms within the brackets.

(*i*) Because of (1.2) the first term reduces to 0.

(*ii*) Because of (1.3) and (1.4) the second term reduces to $N\sum_s \beta_s$ where factor s belongs to the same group as p (or is factor p itself if $p = s$).

(*iii*) Because of (1.6), (1.7), (1.9) and (1.12) the last term reduces to $Na_p\sum_z\sum_w \beta_{zw}$ where the factors z, w and p belong to 3 different group-factors (and $a_p = 1$ in a 2_{III}^{g-p} design). So

$$E(\hat{\alpha}_p) = 2\sum_s \beta_s + 2a_p\sum_z\sum_w \beta_{zw} \tag{1.15}$$

Hence if all factors have two-factor interactions then the (confounded) main effect of a factor is biased. However, if we assume that two-factor interactions exist only between particular (related) factors then we can place these factors in the same group. Their two-factor interactions do not bias the main effect of a factor in that group nor do these interactions bias main effects of factors outside that group since z and w in (1.15) must correspond with 2 different groups. The last argument also implies that pure quadratic effects β_{ii} do not bias main effects.

If we test the g group-factors in a *resolution IV* design then main effects of the group-factors are orthogonal to interactions between 2 group-factors. Hence (1.9) and (1.12) are replaced by

$$\sum_i x_{ij}x_{ij'}x_{ij''} = 0 \tag{1.16}$$

and (1.15) reduces to

$$E(\hat{\alpha}_p) = 2\sum_s \beta_s \tag{1.17}$$

7. Acknowledgement

The author is indebted to the referees for their comments that helped to improve the presentation of this paper. The original version was rewritten while the author was a Postdoctoral Fellow at IBM Research, San Jose, California, Dec. 1973–Dec. 1974.

References

[1] Ackhoff, R. L. (1967). Management misinformation systems. *Management Science, 14*, B 147–B156.

[2] Booth, K. H. V. and Cox, D. R. (1962). Some systematic supersaturated designs. *Technometrics, 4*, 489–495.

[3] Box, G. E. P. and Hunter, J. S. (1961). The 2^{k-p} fractional factorial designs. Part I. *Technometrics, 3*, 311–351.

[4] Hunter, W. E. and Mezaki, R. (1964). Catalyst selection by group screening. *Industrial and Engineering Chemistry*, 38–40.

[5] Jacoby, J. E. and Harrison, S. (1962). Multi-variable experimentation and simulation models. *Naval Research Logistics Quarterly, 9*, 121–136.

[6] Kleijnen, J. P. C. (1974). *Statistical Techniques in Simulation*, Marcel Dekker, Inc., New York.

[7] Li, C. H. (1962). A sequential method for screening experimental variables. *J. Amer. Statist. Assoc., 57*, 455–477.

[8] Ott, E. R. and Wehrfritz, F. W. (1972). A special screening program for many treatments. *Statistica Neerlandica, 26*, 165–170.

[9] Patel, M. S. (1962). Group-screening with more than two stages. *Technometrics, 4*, 209–217.

[10] Plackett, R. L. and Burman, J. P. (1946). The design of optimum multifactorial experiments. *Biometrika, 33*, 305–325.

[11] Watson, G. S. (1961). A study of the group-screening method. *Technometrics, 3*, 371–388.

[12] Yntema, D. B. and Torgerson, W. S. (1961). Man-computer cooperation in decisions requiring common sense. *IRE Transactions on Human Factors in Electronics*, 20–26.