SEQUENTIAL FACTORIAL ESTIMATION

by

GUANG-CHUEN LIN

B. Ed., Taiwan Normal University, 1960

M. Ed., Taiwan Normal University, 1964

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1967

Approved by:

____

Major Professor

TABLE OF CONTENTS

INTRODUCTION

Frequently in scientific investigation, particularly where an empirical approach has to be adopted, problems arise in which the effects of a number of different factors on some property or process are required to be evaluated. Such problems can usually be most economically investigated by arranging the experiments according to an ordered plan in which all the factors are varied in a regular way. Provided the plan has been correctly chosen, it is then possible to determine not only the effect of each individual factor but also the way in which each effect depends on the other factors (i.e. the interactions). This makes it possible to obtain a more complete picture of what is happening than would be obtained by varying each of the factors one at a time while keeping the others constant. Achieving this object is to decide on a set of values or levels, for each of the factors to be studied, and to carry out one or more trials of the process with each of the possible combinations of the levels of factors. Such an experiment is termed a factorial experiment.

A complete factorial experiment, in which all possible combinations of all the levels of the different factors are investigated, will involve a large number of tests when the number of factors is large. It is possible to investigate the main effects of the factors and their more important interactions in a fraction of the number of tests required for the complete factorial designs, thus enabling the size of an experiment to be reduced to a fraction of a full factorial experiment while still providing all the important information, (Box and Hunter 1961a, Davies 1963).

The 2^n factorial designs are used to study the effects of n variables upon a response $\phi = f(x_1, x_2, \ldots, x_n)$. The mathematical model initially assumed is the polynomial

$$E(y) = \phi = b_o + \sum_{i=1}^{n} b_i x_i + \sum\sum_{i<j}^{n} b_{ij} x_i x_j + \sum\sum_{i<j<k}^{n} b_{ijk} x_i x_j x_k + \ldots \quad (1)$$

where y is the observed response or the yield, E(y) the expected value of y, the b's unknown coefficients, and the x's independent variables. Least squares estimates of all the coefficients may be obtained using Yates' Algorithm (1937). If the assumptions are correct, these coefficients will measure the individual effects of the variables (Box 1957). Since the 2^n designs constrain each variable to two levels the quadratic, cubic and other coefficients associated with the powers of the variables x_i are not considered in the model. In practice the complexity of the model is reduced by postponing consideration of the third and higher order terms until the first and second order terms have been fully explored (Hunter 1964). Thus the model may become either

$$E(y) = \phi = b_o + \sum_{i}^{n} b_i x_i \quad (2)$$

or

$$E(y) = \phi = b_o + \sum_{i}^{n} b_i x_i + \sum\sum_{i<j}^{n} b_{ij} x_i x_j \quad (3)$$

the first order model and second order model, respectively.

This simplication permits the use of 2^{n-k} fractional factorial designs where k is the magnitude of fractionation.

After a 2^{n-k} fractional factorial has been completed, an experimenter may wish to analyze the results and hope that large effects may be quickly discovered. It may happen that the results of this initial block of runs fail to provide all the information expected, so that additional blocks are then added, the experimenter proceeding sequentially and pausing to review his data at the conclusion of each block. Often the individual runs comprising the block are also run sequentially. However it is the usual practice for the experimenter to wait until all runs in a block have been completed before analyzing the data.

In order to perform the analysis of sequential data, the exact least squares estimates of all the coefficients in the model can be rapidly obtained at the conclusion of each run, or any group runs, through the use of a "predictor-corrector equation", once an initial block of runs has been completed. This equation was developed by Plackett (1950). The original results are due to Gauss (1821).

This report will illustrate the factorial designs, the mathematical derivation of predictor-corrector equation, numerical examples and the applications of predictor-corrector equation.

FACTORIAL DESIGNS

1). Notation for the 2^n series

A complete 2^n factorial design requires all combinations of two levels of each of n variables. The runs comprising the experimental design are conveniently set out in either of two notations as illustrated for the eight runs comprising a 2^3 factorial in Table 1.

Table 1

Symbols for 2^3 Factorial Design

Run Number	Notation 1: Variables A B C	Notation 2: Variables 1	2	3
1	(1)	-	-	-
2	a	+	-	-
3	b	-	+	-
4	ab	+	+	-
5	c	-	-	+
6	ac	+	-	+
7	bc	-	+	+
8	abc	+	+	+

In the second notation the variables are denoted by number 1, 2, 3, and their two versions take two different values, the high level is a plus sign, the low level a minus sign. The notation using plus and minus signs will be used in this report. The list of experimental runs is called the design matrix. For a 2^n factorial, the design matrix contains n columns and $N = 2^n$ rows.

2). Estimation of effects

On the assumption that the observations are uncorrelated and have equal variance, then the 2^n factorial designs provide independent minimum variance estimates of the grand average and of the $2^n - 1$ effects.

In Table 2, where for convenience a 2^3 design is used, a matrix of independent variables X is generated from the design matrix. For example, 12 interaction column in X is obtained by multiplying the corresponding elements of the separate 1 and 2 columns. The first column of X consists entirely of plus signs and is used to provide an estimate of the mean. For a 2^n design the full matrix of independent variables X contains 2^n columns as well as 2^n rows. The estimate of effect ij...n is obtained by taking the sum of products between the elements of Y and the corresponding elements of the column $X_{ij...n}$ and dividing this product by N/2 where $N = 2^n$; e.g.,

$$ij...n = 2/N\ (X'_{ij...n}\ Y). \tag{4}$$

Table 2

2^3 Factorial Design

Design Matrix			Matrix of Independent Variables X								Observations
1	2	3	I	1	2	3	12	13	23	123	Y
-	-	-	+	-	-	-	+	+	+	-	4
+	-	-	+	+	-	-	-	-	+	+	8
-	+	-	+	-	+	-	-	+	-	+	6
+	+	-	+	+	+	-	+	-	-	-	10
-	-	+	+	-	-	+	+	-	-	+	12
+	-	+	+	+	-	+	-	+	-	-	6
-	+	+	+	-	+	+	-	-	+	-	4
+	+	+	+	+	+	+	+	+	+	+	8

Thus, from Table 2 the 12 interaction effects is

$$12 = \frac{2}{N} X'_{12}Y = \frac{2}{8} (+ - - + + - - +) \begin{bmatrix} 4 \\ 8 \\ 6 \\ 10 \\ 12 \\ 6 \\ 4 \\ 8 \end{bmatrix}$$

$$= 1/4(4 - 8 - 6 + 10 + 12 - 6 - 4 + 8)$$

$$= 2.5$$

Each estimate has variance

$$\text{Var(effect)} = \frac{4\sigma^2}{N}, \qquad (5)$$

where σ^2 is the variance of the individual observations.

The average is obtained by taking the sum of products of column X_I with the observation column Y and dividing the result by N, thus

$$\text{average} = \bar{y} = (X_I'Y)/N. \qquad (6)$$

Thus $\bar{y} = 58/8 = 7.25$ with variance σ^2/N. By this process 2^n estimates can be obtained from 2^n runs. When n is large, the wealth of such estimates becomes an embarrassment. However, in many practical situations, the three-factor and multi-factor interaction effects can often be hopefully supposed to be negligible in size (Cochran and Cox 1957, Box and Hunter 1961, John 1966). In this situation, fractional designs using a smaller number of runs may be employed.

3). 1/2 fraction of the 2^4 factorial

For illustration, the one half fraction of the 2^4 design will be first discussed. Since the design is to contain 2^{4-1} =

8 runs, a 2^3 factorial design is first written down. The + and - elements associated with the 123 interaction then are used to identify the + and - versions of variable 4. The combination of observations used to estimate the main effect 4 is identical to that used to estimate the three-factor interaction effect 123. The estimates of 4 and 123 are said to be confounded. The "4" effect really estimates the sum of the effects of 4 and 123.

The resulting eight combinations shown in Table 3 give a particular half fraction of the complete 2^4 design. A $(1/2)^k$ fraction of a 2^n factorial design is called a 2^{n-k} fractional factorial.

Table 3

2^{4-1} Fractional Factorial Design

Design Matrix				Matrix of Independent Variables X									Observations
1	2	3	123=4	I=1234	1	2	3	4	12	13	23	123	Y
-	-	-	-	+	-	-	-	-	+	+	+	-	12.1
+	-	-	+	+	+	-	-	+	-	-	+	+	21.7
-	+	-	+	+	-	+	-	+	-	+	-	+	29.0
+	+	-	-	+	+	+	-	-	+	-	-	+	25.7
-	-	+	+	+	-	-	+	+	+	-	-	+	17.3
+	-	+	-	+	+	-	+	-	-	+	-	-	17.3
-	+	+	-	+	-	+	+	-	-	-	+	-	12.9
+	+	+	+	+	+	+	+	+	+	+	+	+	36.2

It is desirable to have a general method which enables one to determine which effects are confounded. This is accomplished for this design by introducing the equality 4 = 123 where the multiplication product 123 refers to the multiplication of the individual elements in the corresponding column 1, 2, 3. It is obvious that by multiplying the elements in any column by a column of identical elements, a column of pluses corresponding to I will result. Thus it follows $1^2 = I$, $2^2 = I$, and so on. On multiplying both sides of the equation $4 = 123$ by 4:

$$4^2 = 1234 \quad \text{that is } I = 1234. \tag{7}$$

This identity is readily confirmed for if the elements in column 1, 2, 3 and 4 are multiplied together a column of plus signs is obtained, that is I. The interaction associated with I is said to be a generator pf the design. In this particular instance there is only one generator so this provides the defining relationships which exist between the effects. Thus the estimates such as 12 and 34 are confounded. Similarly the main effect 2 is confounded with three-factor interaction 134 and so on.

4). Linear combination of effects

To proceed to estimate the main effect 2 and the three-factor interaction 134, the estimate of 2 is really an estimate of the combination of the effect 2 + 134. Eight linear combinations of effects L_I, L_1, ... are available. Thus $L_1 = 1/4(X_1' Y)$ or equally $L_1 = 1/4(X_{234}' Y)$ and so on.

On studying Table 4, the two-factor interaction are mutually confounded in pairs, but assuming that the three and four factor interactions are either non-existent or negligible the estimates L_I, L_1, L_2, L_3 and L_4 can be taken to be estimate of the average and the main effect 1, 2, 3 and 4. If, further, prior knowledge is available that, for example, the 34 interaction effect is negligible, then the estimate L_{12} could be taken to estimate the 12 interaction effect alone.

Table 4

Eight Linear Combinations of Effects from a 2^{4-1} Design with Defining Relation I = 1234

L_I = average + 1234 = 21.53	L_4 = 4 + 123 = 9.05
L_1 = 1 + 234 = 7.40	L_{12} = 12 + 34 = 2.60
L_2 = 2 + 134 = 8.85	L_{13} = 13 + 24 = 4.25
L_3 = 3 + 124 = -1.20	L_{14} = 14 + 23 =-1.60

5). The alternative fraction

In the above example, in forming the 2^{4-1} design, the factor 4 was associated with the three-factor interaction 123. In standard ordering, the elements of the three-factor interaction column, and hence of factor 4, are

− + + − + − − +

The factor 4 can either use these elements as they stand, or it can be associated with negative of the 123 effect, that is with the elements

+ − − + − + + −

In the first case 4 = 123 that is I = 1234, and in the second case -4 = 123 that is I = -1234. In Table 5, the two parts together constitute a complete 2^4 factorial design.

In Table 6 eight linear combinations of effect L_I', L_1', ... associated with the fraction having defining relation I = -1234 are given. If both fraction are present, then simple addition and substraction of the L and L' linear combination will provide unconfounded estimate of all the effects.

Table 5

Design Matrix for the Two 2^{4-1} Fractional Factorials

Defining Relation $I = 1234$ 1	2	3	4	Observations Y	Defining Relation $I = -1234$ 1	2	3	4	Observations Y
-	-	-	-	12.1	-	-	-	+	16.8
+	-	-	+	21.7	+	-	-	-	18.1
-	+	-	+	29.0	-	+	-	-	10.4
+	+	-	-	25.7	+	+	-	+	32.1
-	-	+	+	17.3	-	-	+	-	12.3
+	-	+	-	17.3	+	-	+	+	25.0
-	+	+	-	12.9	-	+	+	+	35.1
+	+	+	+	36.2	+	+	+	-	27.4

Table 6

Eight Linear Combinations of Effects from a 2^{4-1} Design with Defining Relation $I = -1234$

L'_1 = average - 1234 = 22.15	$L'_4 = 4 - 123 = 10.20$
$L'_1 = 1 - 234 = 7.00$	$L'_{12} = 12 - 34 = 0$
$L'_2 = 2 - 134$ 8.20	$L'_{13} = 13 - 24 = -4.50$
$L'_3 = 3 - 124 = 5.60$	$L'_{14} = 14 - 23 = -4.40$

Solving for all the effects gives

Main Effects

1 = 7.20
2 = 8.53
3 = 2.20
4 = 9.62

Two-factor Interactions

12 = 1.30
13 = -0.12
14 = -3.00
23 = 1.40
24 = 4.38
34 = 1.30

Three-factor Interactions

123 = -0.58
124 = -3.40
134 = 0.32
234 = 0.20

Four-factor Interaction

1234 = -0.31

Average Response = 21.84

The estimates are the same as would be obtained from an analysis of a full 2^4 design.

6). The general 1/2 fraction of the 2^n designs

It is usual to use the interaction of highest order to split a full 2^n factorial into two half fractions. The generator is 123...n and the defining relation I = 123...n.

The one-half fraction of all the 2^n factorial designs are best obtained by first writing down the design matrix for a full 2^{n-1} factorial and then adding the nth variables by identifying its + and - versions with the + and - signs of the highest order interaction 123...(n - 1).

For $n > 5$ the half-replicate design permits the estimation of a plethora of linear combinations of effects, many of which are combinations of higher order interactions solely. Therefore smaller fractions of the 2^n designs will be employed, that is the 2^{n-k} fractional factorials for $k > 1$. For such designs there is not one, but k generators which combine to provide the defining relation.

7). Three type of 2^n factorials

For convenience, Box and Hunter (1961a, 1961b) divide 2^{n-k} fractional factorial designs into three types.

(i) Designs of Resolution III in which no main effect is confounded with any other main effect, but main effects are confounded with two-factor interactions and two-factor interaction with one another. The 2^{3-1} design is of Resolution III written as 2^{3-1}_{III}.

(ii) Designs of resolution IV in which no main effect is confounded with any other main effect or two-factor interaction, but where two-factor interactions are confounded with one another. For example, the 2^{4-1} design is of Resolution IV written as 2_{IV}^{4-1}.

(iii) Designs of Resolution V in which no main effect or two-factor interaction is confounded with any other main effect or two-factor interaction, but two-factor interactions are confounded with three-factor interactions. For example, the 2^{5-1} design is of Resolution V written as 2_{V}^{5-1}.

This report does not intend to further discuss the 2^{n-k} designs. It will only illustrate the design matrix for a 2_{III}^{7-4} design.

For the 2_{III}^{7-4} fractional factorial design, it requires $2^{7-4} = 2^3 = 8$ runs for testing $n = 7$ variables. This starts with the construction of design matrix with the 2^3 factorial and then associate four additional variables with the plus and minus signs of the four interaction columns. For example, set

$$4 = 12, \quad 5 = 13, \quad 6 = 23, \quad 7 = 123 \tag{8}$$

to obtain the following 2_{III}^{7-4} design(Table 7). The identifications in Eq. (8) provide the generating relations

$$I = 124, \quad I\ 135, \quad I = 236, \quad I = 1237. \tag{9}$$

The complete relation for this 2_{III}^{7-4} design is

$$\begin{aligned} I &= 124 = 135 = 236 = 1237 = 2345 = 1346 = 347 = 1256 \\ &= 257 = 167 = 456 = 1457 = 2467 = 3567 = 1234567 \end{aligned}$$

Assuming that all interactions between three of more variables are negligible, then by repeated use of the defining relations the following linear combinations of effects will be obtained:

$$
\begin{array}{ll}
L_1 = \text{average} & L_4 = 4 + 12 + 56 + 37 \\
L_1 = 1 + 24 + 35 + 67 & L_5 = 5 + 13 + 46 + 27 \\
L_2 = 2 + 14 + 36 + 57 & L_6 = 6 + 23 + 45 + 17 \\
L_3 = 3 + 15 + 26 + 47 & L_7 = 7 + 34 + 25 + 16
\end{array}
\qquad (10)
$$

Table 7

Design Matrix for a 2^{7-4}_{III} Design

1	2	3	4 = 12	5 = 13	6 = 23	7 = 123
-	-	-	+	+	+	-
+	-	-	-	-	+	+
-	+	-	-	+	-	+
+	+	-	+	-	-	-
-	-	+	+	-	-	+
+	-	+	-	+	-	-
-	+	+	-	-	+	-
+	+	+	+	+	+	+

From the above illustration, the procedure of adding fractions in sequence with suitably switched signs provides a useful method for the systematic isolation and confirmation of important effects in multi-variable systems. In the next section, this repoet will develop a predictor-corrector equation. Through the use of a predictor-corrector equation an experimenter may quickly determine the least squares estimates of all the coefficients when the 2^n factorial designs are denoted in a polynomial model.

DERIVATION OF PREDICTOR-CORRECTOR EQUATION

1). Derivation

In this section the predictor-corrector equation for estimation of the coefficients in a linear model where additional data become available will be derived.

Let Y represent a column vector of N stochastic observations $Y_1, Y_2, \ldots, Y_N$, let B represent a column vector of q unknown coefficients $b_1, b_2, \ldots, b_q$, and let the matrix of independent variables X be composed of N rows, and q columns. Then, the observational equations may be represented

$$Y = XB + e \tag{11}$$

where e is an N x 1 column vector of error components, with $E(e) = 0$, $E(ee') = \sigma^2 I_N$, and where $E(Y) = XB$. If $N \geq q$, the least squares estimates B are provided by solving the normal equations $X'X\hat{B} = X'Y$ giving $\hat{B} = (X'X)^{-1}X'Y$ under the usual assumption that X'X has rank q and hence that its inverse exists. For the situation in which the model and experimental designs have been chosen (see example Table 2) so that $X'X = rNI_q$ where r is the number of times the design is replicated and I_q is a q x q identity matrix. The variance-covariance matrix of the estimates B is $\sigma^2(X'X)^{-1}$ (Graybill 1961).

Suppose that Z be an n x q matrix of n additional row vectors z_i, i = 1, 2, ..., n added onto X and let y be the corresponding n x 1 vector of new observations. Then the model now becomes

$$\begin{bmatrix} Y \\ y \end{bmatrix} = \begin{bmatrix} X \\ Z \end{bmatrix} B + \begin{bmatrix} e_N \\ e_n \end{bmatrix}, \tag{12}$$

and the associated normal equations

$$\begin{bmatrix} X \\ Z \end{bmatrix}' \begin{bmatrix} X \\ Z \end{bmatrix} \hat{B}* = \begin{bmatrix} X \\ Z \end{bmatrix}' \begin{bmatrix} Y \\ y \end{bmatrix}$$

or

$$(X'X + Z'Z)\hat{B}* = (X'Y + Z'y).$$

Substituting for $X'X\hat{B} = X'Y$:

$$(X'X + Z'Z)\hat{B}* = (X'X\hat{B} + Z'y) \tag{13}$$

where $\hat{B}*$ is the new vector of estimates based on all $(N + n)$ observations.

Now let $\hat{y} = Z\hat{B}$ be the predicted values for the additional row vectors based on the initial estimates B and let $d = y - \hat{y}$. Then $y = d + \hat{y}$ or $y = d + Z\hat{B}$.

From Eq. (13):

$$(X'X + Z'Z)\hat{B}* = X'X\hat{B} + Z'(d + Z\hat{B})$$

$$(X'X + Z'Z)\hat{B}* = (X'X + Z'Z)\hat{B} + Z'd$$

thus

$$\hat{B}* = \hat{B} + (X'X + Z'Z)^{-1}Z'(y - \hat{y}). \tag{14}$$

From the fact $X'X = rNI_q$ it follows that

$$(X'X + Z'Z)^{-1} = (rNI_q + Z'Z)^{-1}$$

$$= 1/rN\left[I_q + 1/rN(Z'Z)\right]^{-1}.$$

Since $(I + UV)^{-1} = I - U(I + VU)^{-1}V$,

so

$$(X'X + Z'Z)^{-1} = \frac{1}{rN}\left[I_q - \frac{1}{rN}Z'(I_n + \frac{1}{rN}ZZ')^{-1}Z\right].$$

Here further requiring that the added row vectors z_i comprising Z must be row-wise orthogonal, that is $z_i z_j' = 0$ for $i \neq j$, then $ZZ' = qI_n$, so

$$(X'X + Z'Z)^{-1} = \frac{1}{rN}\left[I_q - \frac{1}{rN}Z'(I_n + \frac{1}{rN}qI_n)^{-1}Z\right]$$

$$= \frac{1}{rN}\left[I_q - \frac{1}{rN}Z'\left(\frac{rN + q}{rN}I_n\right)^{-1}Z\right]$$

$$= \frac{1}{rN}\left[I_q - \frac{1}{rN + q}Z'Z\right]. \qquad (15)$$

From Eq. (14):

$$\hat{B}* = \hat{B} + \frac{1}{rN}\left[I_q - \frac{1}{rN + q}Z'Z\right]Z'(y - \hat{y})$$

$$= \hat{B} + \frac{1}{rN}\left[Z' - \frac{1}{rN + q}Z'ZZ'\right](y - \hat{y})$$

$$= \hat{B} + \frac{1}{rN}\left[Z' - \frac{1}{rN + q}Z'qI_n\right](y - \hat{y})$$

$$= \hat{B} + \frac{1}{rN}\left[\frac{rN + q - q}{rN + q}\right]Z'(y - \hat{y}).$$

So

$$\hat{B}* = \hat{B} + \frac{1}{rN + q}Z'(y - \hat{y}). \qquad (16)$$

This equation is termed the predictor-corrector equation and is useful whenever both X'X and ZZ' are orthogonal.

For $q = N$, $\hat{B}*$ can be written

$$\hat{B}* = \hat{B} + \frac{1}{N(r + 1)}Z'(y - \hat{y}) \qquad (17)$$

Eq. (16) can be written:

$$\hat{B}* = \hat{B} + \frac{1}{rN + q}\sum_{i}^{n}(y_i - \hat{y}_i)z_i', \qquad (18)$$

or more simply

$$\hat{B}* = \hat{B} + \sum_{i}^{n}d_i \qquad (19)$$

where the corrections for the coefficients at the conclusion of the ith additional run are given by the elements of the vector

$$d_i = \frac{1}{rN + q}(y_i - \hat{y}_i)z_i'$$

where z_i = 1 x q row vector in the matrix of independent variables associated with the ith experiment, $i = 1, 2, \ldots, n \leq N$,

y_i = new observation associated with z_i,

$\hat{y}_i = z_i\hat{B}$ = predicted response for the ith experiment.

The quantity $(y_i - \hat{y}_i)/(rN + q)$ is called the "corrector constant" for the ith run.

The variance-covariance matrix for $\hat{B}*$ is $(X'X + Z'Z)^{-1} \sigma^2$. From the above assumption $X'X = rNI_q$, $ZZ' = qI_n$, and further the elements of Z consists of +1 or -1 only (as using the two-level factorials with their associated model), then the diagonal elements Of $Z'Z = n$ and by Eq. (15), the variance of any individual estimate is

$$Var(b_i) = \frac{1}{rN}\left(1 - \frac{n}{rN + q}\right)\sigma^2. \tag{20}$$

2). Analysis of variance

The analysis of variance table corresponding to the completion of r blocks of N runs each is as shown in Table 8.

Table 8

Analysis of Variance for r Blocks of N Runs

Source	DF	SS
Crude sum of squares	rN	$SST = Y'Y$
Regression sum of squares	q	$SSR = \hat{Y}'\hat{Y}$
Deviation sum of squares	$rN - q$	$ssd = (Y - \hat{Y})'(Y - \hat{Y})$

Given n additional observations, then the new AOV table is as shown in Table 9.

Table 9

Analysis of Variance for n Additional Runs

Source	DF	SS
New Crude SS	$rN + n$	$SST^* = Y'Y + y'y$
New Regression SS	q	$SSR^* = \hat{B}^{*\prime}(X'Y + Z'y)$
New Deviation SS	$rN + n - q$	$SSD^* = SSD + \frac{rN}{rN + q}(y - \hat{y})'(y - \hat{y})$

NUMERICAL EXAMPLE

1). Computational procedure

Table 10

Data and Estimates of the Coefficients for a 2^4 Design

Run Number	Design Matrix 1	2	3	4	Response (Observation)	Estimates
1	-	-	-	-	12.1	$b_0 = 21.84$
2	+	-	-	-	18.1	$b_1 = 3.60$
3	-	+	-	-	10.4	$b_2 = 4.26$
4	+	+	-	-	25.7	$b_{12} = 0.65$
5	-	-	+	-	12.3	$b_3 = 1.10$
6	+	-	+	-	17.3	$b_{13} = -0.06$
7	-	+	+	-	12.9	$b_{23} = 0.70$
8	+	+	+	-	27.4	$b_{123} = -0.29$
9	-	-	-	+	16.8	$b_4 = 4.81$
10	+	-	-	+	21.7	$b_{14} = -1.50$
11	-	+	-	+	29.0	$b_{24} = 2.19$
12	+	+	-	+	32.1	$b_{124} = -1.70$
13	-	-	+	+	17.3	$b_{34} = 0.65$
14	+	-	+	+	25.0	$b_{134} = 0.16$
15	-	+	+	+	35.1	$b_{234} = 0.10$
16	+	+	+	+	36.2	$b_{1234} = -0.31$

To illustrate the computational procedure through the use of the predictor-corrector equation, consider the data in Table 5. In Table 10 the sixteen runs of the 2^4 factorial are listed in standard factorial notation (Davies 1963, Cochran and Cox 1957). The estimates are obtained by the method of least squares.

To fit the q = 4 coefficients in the first order model $E(y) = b_0x_0 + b_1x_1 + b_2x_2 + b_3x_3$, an initial program involving only the three variables x_1, x_2, x_3, and runs 5, 2, 3 and 8 are used. First consider the N = 4 runs of a 2^{3-1}_{III} fractional defined by I = 123. The data and estimates of the coefficients are as follows:

Matrix of Independent Variables   Vector of Observations

$$X = \begin{bmatrix} x_5 \\ x_2 \\ x_3 \\ x_8 \end{bmatrix} = \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{bmatrix} + & - & - & + \\ + & + & - & - \\ + & - & + & - \\ + & + & + & + \end{bmatrix} \end{array}; \qquad Y = \begin{bmatrix} 12.3 \\ 18.1 \\ 10.4 \\ 27.4 \end{bmatrix};$$

Solutions (Vector of Estimated Coefficients)

$$\hat{B} = \begin{bmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \hat{b}_2 \\ \hat{b}_3 \end{bmatrix} \doteq \begin{bmatrix} 17.05 \\ 5.70 \\ 1.85 \\ 2.80 \end{bmatrix}$$

Fitted model:

$$\hat{y} = 17.05x_0 + 5.70x_1 + 1.85x_2 + 2.80x_3.$$

Suppose that n < N additional experiments drawn from the half replicate 2^{3-1} having defining relation $I = -123$ are now run. The least squares estimates of all the coefficients may be obtained by using the predictor-corrector equation after each run. Suppose a fifth experiment, say run 1, is performed following the completion of the initial block of four runs given above. Then $N = q = 4$, $r = 1$ and

$$z_1 = (+\ -\ -\ -); \qquad y_1 = 12.1;$$

$$\hat{y}_1 = z_1\hat{B} = (17.05) - (5.70) - (1.85) - (2.80) = 6.70.$$

The corrector constant is:

$$(y_1 - \hat{y}_1)/(rN + q) = (12.1 - 6.70)/(4 + 4) = 0.675.$$

The revised estimates of the coefficients are given by substituting in the predictor-corrector equation, Eq. (18),

$$\hat{B}^* = \begin{bmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \hat{b}_2 \\ \hat{b}_3 \end{bmatrix} = \begin{bmatrix} 17.05 \\ 5.70 \\ 1.85 \\ 2.80 \end{bmatrix} + \frac{1}{(4+4)}(12.1-6.7)\begin{bmatrix} + \\ - \\ - \\ - \end{bmatrix} = \begin{bmatrix} 17.05+0.675 \\ 5.70-0.675 \\ 1.85-0.675 \\ 2.80-0.675 \end{bmatrix} = \begin{bmatrix} 17.725 \\ 5.025 \\ 1.175 \\ 2.125 \end{bmatrix}$$

The new fitted equation is

$$\hat{y} = 17.725x_0 + 5.025x_1 + 1.175x_2 + 2.125x_3.$$

The variance of each revised coefficient is

$$\mathrm{Var}(b) = \frac{1}{rN}\left(1 - \frac{n}{rN + q}\right)\sigma^2 = 7\sigma^2/32.$$

Suppose a sixth experiment is now run, say z_6, and new estimates required. Then

$$z_6 = (+\ +\ -\ +); \qquad y_6 = 17.3;$$

$$\hat{y}_6 = z_6\hat{B} = (17.05) + (5.70) - (1.85) + (2.80) = 23.70.$$

Note here that the predicted value for every new run is computed from the coefficients obtained after the last completed block.

The corrector constant is:

$$(y_6 - \hat{y}_6)/(rN + q) = (17.3 - 23.70)/(4 + 4) = -0.800.$$

The estimates at the conclusion of runs 1 and 6 are:

$$\hat{B}^* = \begin{bmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \hat{b}_2 \\ \hat{b}_3 \end{bmatrix} = \begin{bmatrix} 17.725 + (-0.800) \\ 5.025 + (-0.800) \\ 1.175 - (-0.800) \\ 2.125 + (-0.800) \end{bmatrix} = \begin{bmatrix} 16.925 \\ 4.225 \\ 1.975 \\ 1.325 \end{bmatrix}$$

The fitted equation is:

$$\hat{y} = 16.925x_0 + 4.225x_1 + 1.975x_2 + 1.325x_3.$$

Each coefficient now has variance $6\sigma^2/32$.

At the conclusion of the seventh experiment z_7:

$$z_7 = (+ - + +); \qquad y_7 = 12.9;$$

$$\hat{y}_7 = z_7\hat{B} = (17.05) - (5.70) + (1.85) + (2.80) = 16.0.$$

The corrector constant is:

$$(y_7 - \hat{y}_7)/(rN + q) = (12.9 - 16.0)/(4 + 4) = -0.388.$$

The new fitted model is:

$$\hat{y} = 16.537x_0 + 4.613x_1 + 1.587x_2 + 0.937x_3.$$

Each coefficient has variance $5\sigma^2/32$.

The eighth experiment z_4 completes the second block of $n = 4$ runs giving

$$z_4 = (+ + + -); \qquad y_4 = 25.7;$$

$$\hat{y}_4 = (17.05) + 95.70) + (1.85) - (2.80) = 21.80.$$

The corrector constant is:

$$(y_4 - \hat{y}_4)/(rN + q) = 3.9/8 = 0.488;$$

$$\hat{B}^* = \begin{bmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \hat{b}_2 \\ \hat{b}_3 \end{bmatrix} = \begin{bmatrix} 16.537 + (0.488) \\ 4.613 + (0.488) \\ 1.587 + (0.488) \\ 0.937 - (0.488) \end{bmatrix} = \begin{bmatrix} 17.025 \\ 5.101 \\ 2.075 \\ 0.449 \end{bmatrix}$$

The new fitted equation is:

$$\hat{y} = 17.025x_0 + 5.101x_1 + 2.075x_2 + 0.449x_3. \qquad (21)$$

Each coefficient has variance $4\sigma^2/32$.

The second block of four runs completes a full 2^3 design. Suppose a third block of four runs, replicate of the earlier runs, is now added having defining relation I = 123, then the coefficients once again will be re-estimated at the conclusion of each run. Begin this third block with run z_{13}, thus

Table 11

Data and Estimates for a Third Block of 2^3 Design

Run	13	10	11	16
z_i	(+ - - +)	(+ + - -)	(+ - + -)	(+ + + +)
y_i	17.3	21.7	29.0	36.2
$\hat{y}_i$	10.298	19.602	13.55	24.65
Corrector	0.584	0.175	1.286	0.963
$\hat{b}_0$	17.609	17.784	19.070	20.033
$\hat{b}_1$	4.517	4.692	3.406	4.369
$\hat{b}_2$	1.491	1.316	2.602	3.565
$\hat{b}_3$	1.033	0.858	0.428	0.535
Var(b)	$11\sigma^2/96$	$10\sigma^2/96$	$9\sigma^2/96$	$8\sigma^2/96$
ΔSSD	32.741	2.940	158.764	89.027

$z_{13} = (+ - - +)$; $y_{13} = 17.3$.

Remembering now that $\hat{y}_{13} = z_{13}\hat{B}$ where $\hat{B}$ is the vector of estimates provided by the most recently completed block, obtaining on substituting the last fitted equation, Eq. (21),

$$\hat{y}_{13} = (17.025) - (5.101) - (2.075) + (0.449) = 10.298.$$

Remembering further that two blocks of N runs have been completed so that $r = 2$, $N = 4$, $q = 4$, the corrector constant for this run is:

$$(y_{13} - \hat{y}_{13})/(rN + q) = (17.3 - 10.298)/(2 \times 4 + 4) = 0.584.$$

The remaining run of third block are z_{10}, z_{11}, z_{16}. The revised estimates of the coefficients $\hat{B}^*$ computed after each run and the associated variance are given in Table 11. Also listed

Table 12

Data and Estimates for a Fourth Block of 2^3 Design

Run	9	14	15	12
z_i	(+ - - -)	(+ + - +)	(+ - + +)	(+ + + -)
y_i	16.8	25.0	35.1	32.1
$\hat{y}_i$	11.564	21.372	19.764	27.432
Corrector	0.327	0.227	0.959	0.291
$\hat{b}_o$	20.360	20.587	21.546	21.84
$\hat{b}_1$	4.042	4.269	3.310	3.60
$\hat{b}_2$	3.238	3.011	3.970	4.26
$\hat{b}_3$	0.208	0.435	1.394	1.10
Var(b)	$15\sigma^2/192$	$14\sigma^2/192$	$13\sigma^2/192$	$12\sigma^2/192$
ΔSSD	20.530	9.894	176.579	16.259

is ΔSSD, the increase in the deviation sums of squares resulting from the added run.

A fourth block with defining relation I = -123 consisting of z_9, z_{14}, z_{15}, z_{12} gives the results listed in Table 12.

The estimates of $\hat{B}^*$ after sixteen runs agree with those displayed in Table 10 as they must. After every run the estimates tabulated are the least squares estimates.

2). Analysis of variance

After the completion of each run redetermine the analysis of variance table associated with the model and the total number of runs. As shown in Table 8 and 9, the crude sum of squares increase with each added observation. The sum of squares of deviation for each added run will increase by ΔSSD_i

= Increase in Deviations sum of squares for ith run

$= rN/(rN + q)(y_i - \hat{y}_i)^2$

or more conveniently

$$\Delta SSD_i = rN(rN + q)\left[(y_i - \hat{y}_i)/(rN + q)\right]^2$$

$$= rN(rN + q)(\text{Corrector Constant for ith run})^2$$

The total SSD at the conclusion of the eight experiments is $(4+4)\left[(0.675)^2+(-0.800)^2+(-0.388)^2+(0.488)^2\right] = 47.499$.

Assuming the model is appropriate, an estimate of the variance σ^2 is provided by $s^2 = SSD^*/(rN + n - q)$. The estimate of variance at the conclusion of the sixteen run is $s^2 = 554.233/12 = 46.186$ with twelve degrees of freedom.

APPLICATION

It was illustrated above how, by the addition of available data to a 2_{III}^{3-1} design, the revised estimates $\hat{B}*$ may be obtained. Below are examples of the application of the predictor-corrector equation.

1). Augment of the model and block size

In the above example the mathematical model was not changed as additional data became available. However, at the end of the eighth run a full 2^3 factorial had been completed and orthogonal estimates of the first order, two-factor interactions and three-factor interaction could have been obtained, the three-factor interaction being confounded with the block effect. The data and associated estimates at the conclusion of the eighth experiment are displayed in Table 13.

Table 13

Data and Estimates for a 2^3 Design

Matrix of Independent Variables	Vector of Observations	Vector of Estimates

$$X=\begin{bmatrix}x_1\\x_2\\x_3\\x_4\\x_5\\x_6\\x_7\\x_8\end{bmatrix}=\begin{array}{c}\begin{array}{cccccccc}0&1&2&3&12&13&23&123\end{array}\\ \begin{bmatrix}+&-&-&-&+&+&+&-\\+&+&-&-&-&-&+&+\\+&-&+&-&-&+&-&+\\+&+&+&-&+&-&-&-\\+&-&-&+&+&-&-&+\\+&+&-&+&-&+&-&-\\+&-&+&+&-&-&+&-\\+&+&+&+&+&+&+&+\end{bmatrix}\end{array};\quad Y=\begin{bmatrix}12.1\\18.1\\10.4\\25.7\\12.3\\17.3\\12.9\\27.4\end{bmatrix};\quad \hat{B}=\begin{bmatrix}\hat{b}_0\\\hat{b}_1\\\hat{b}_2\\\hat{b}_3\\\hat{b}_{12}\\\hat{b}_{13}\\\hat{b}_{23}\\\hat{b}_{123}\end{bmatrix}=\begin{bmatrix}17.025\\5.101\\2.075\\0.449\\2.350\\-0.225\\0.600\\0.025\end{bmatrix}$$

The fitted model is:

$$\hat{y} = 17.025x_0 + 5.101x_1 + 2.075x_2 + 0.449x_3 + 2.350x_1x_2 - 0.225x_1x_3 + 0.600x_2x_3 + 0.025x_1x_2x_3.$$

Suppose a ninth experiment is now run, say z_9:

$$z_9 = (+ - - - + + + -);\ y_9 = 16.8;\ \hat{y}_9 = z_9\hat{B} = 12.100.$$

The number of runs in the block are now $N = 8$, $q = 8$, $r = 1$, and the corrector constant $= (y_9 - \hat{y}_9)/(rN + q) = 0.294$. Thus

$$\hat{B}^* = \begin{bmatrix} 17.025 + 0.294 \\ 5.101 - 0.294 \\ 2.075 - 0.294 \\ 0.449 - 0.294 \\ 2.350 + 0.294 \\ -0.225 + 0.294 \\ 0.600 + 0.294 \\ 0.025 - 0.294 \end{bmatrix} = \begin{bmatrix} 17.319 \\ 4.807 \\ 1.781 \\ 0.155 \\ 2.644 \\ 0.069 \\ 0.894 \\ -0.269 \end{bmatrix}$$

Information from additional replicate runs could continue to up-date these estimates. The estimates of $\hat{B}^*$ after the second block of eight runs is completed will agree with those displayed in Table 10 as they must.

2). Setting number of estimates q equal to block size N

To use the predictor-corrector equation it is only necessary that the q estimates provided by the block of N runs be mutually orthogonal and that the n additional runs produce vectors in the matrix of independent variables that are also row-wise orthogonal. If the n additional runs are to be drawn from a two-level design, then it is convenient to set $q = N$ even though this may require the addition of "slack" variables to the model (Hunter 1964). For example, to estimate the five coefficients in the first order model $E(y) = b_0x_0 + \sum_{i=1}^{4} b_ix_i$, the smallest two-level design that will provide orthogonal estimates is a 2_{IV}^{4-1} containing eight runs. In order that $q = N$, three slack variables might be x_1x_2, x_1x_3, x_1x_4. Of course, an experimenter would choose slack variables he felt might produce large effects and hence properly belong in the model.

As an example, data from Table 10 were used to construct an initial block comprising the eight runs of a 2_{IV}^{4-1} design with defining relation I = 1234. The matrix of independent variables X associated with the model, now containing the slack variables, is displayed in Table 14 along with the observations and the estimated coefficients.

Table 14

Data and Estimates for a 2^{4-1} Design

$$X = \begin{bmatrix} x_1 \\ x_{10} \\ x_{11} \\ x_4 \\ x_{13} \\ x_6 \\ x_7 \\ x_{16} \end{bmatrix} = \begin{array}{c} \begin{array}{cccccccc} 0 & 1 & 2 & 3 & 4 & 12 & 13 & 14 \end{array} \\ \begin{bmatrix} + & - & - & - & - & + & + & + \\ + & + & - & - & + & - & - & + \\ + & - & + & - & + & - & + & - \\ + & + & + & - & - & + & - & - \\ + & - & - & + & + & + & - & - \\ + & + & - & + & - & - & + & - \\ + & - & + & + & - & - & - & + \\ + & + & + & + & + & + & + & + \end{bmatrix} \end{array} ; \quad \begin{array}{c} Y \\ \begin{bmatrix} 12.1 \\ 21.7 \\ 29.0 \\ 25.7 \\ 17.3 \\ 17.3 \\ 12.9 \\ 36.2 \end{bmatrix} \end{array} ; \quad \begin{array}{c} \hat{B} \\ \begin{bmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \hat{b}_2 \\ \hat{b}_3 \\ \hat{b}_4 \\ \hat{b}_{12} \\ \hat{b}_{13} \\ \hat{b}_{14} \end{bmatrix} \end{array} = \begin{bmatrix} 21.53 \\ 3.70 \\ 4.43 \\ -0.60 \\ 4.53 \\ 1.60 \\ 2.13 \\ -0.80 \end{bmatrix}$$

Suppose now that run 2 is added, then

$z_2 = (+ + - - - - - -)$; $y_2 = 18.1$ and $\hat{y}_2 = z_2\hat{B} = 13.94$, using Eq. (18) the estimates will become

$$\hat{B}^* = \begin{bmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \hat{b}_2 \\ \hat{b}_3 \\ \hat{b}_4 \\ \hat{b}_{12} \\ \hat{b}_{13} \\ \hat{b}_{14} \end{bmatrix} = \begin{bmatrix} 21.53 \\ 3.70 \\ 4.43 \\ -0.60 \\ 4.53 \\ 1.60 \\ 2.13 \\ -0.80 \end{bmatrix} + \frac{1}{(8 + 8)}(18.1 - 13.94) \begin{bmatrix} + \\ + \\ - \\ - \\ - \\ - \\ - \\ - \end{bmatrix} = \begin{bmatrix} 21.79 \\ 3.96 \\ 4.17 \\ -0.86 \\ 4.27 \\ 1.34 \\ 1.86 \\ -1.06 \end{bmatrix}$$

DISCUSSION

In this report, the predictor-corrector equation is used to improve the estimates of all the coefficients in the assumed mathematical model. Before the predictor-corrector equation can be used two conditions must be satisfied: 1) the estimates B supplied by the prior block of N runs must be mutually orthogonal and 2) the added row vectors must be row-wise orthogonal, that is $z_i z_j' = 0$ for $i \neq j$. These conditions are met by the 2^n and 2^{n-k} designs and associated models illustrated in this report. The equation can, of course, be applied to other designs and models, which satisfy these conditions.

ACKNOWLEDGMENT

I wish to express my sincere gratitude to Dr. Arthur D. Dayton for his advice and guidance during the writing of this report. I also wish to express my thanks to Dr. A. M. Feyerherm and Dr. R. E. Williams for their valuable suggestions in the preparation of this report.

REFERENCES

Box, G. E. P. 1957. Evolutionary Operation. Applied Stat. No. 2, 3 - 23.

Box, G. E. P. and Hunter, J. S. 1961a. The 2^{k-p} Fractional Factorial Design I. Tech. 3, 311 - 352.

Box, G. E. P. and Hunter, J. S. 1961b. The 2^{k-p} Fractional Factorial Design II. Tech. 3, 449 - 458.

Cochran, W. G. and Cox, G. M. 1957. Experimental Designs. John Willey and Sons, Inc., New York.

Davies, O. L. 1963. Design and Analysis of Industrial Experiments. Hafner Co., New York.

Gauss, C. F. 1821 Theoria Combinationis Observationum Erroribus Minimis Obnoxiae. Werke 4, Gottingen.

Graybill, F. A. 1961. An Introduction to Linear Statistical Medels I. McGraw-Hill, New York.

Hunter, J. S. 1964. Sequential Factorial Estimation. Tech. 6, 41 - 57.

John, P. W. M. 1966. Augment 2^{n-1} Designs. Tech. 8, 469 - 480.

Plackett, R. L. 1950. Some Theorems in Linear Squares. Biometrika 37, 149 - 157.

Yates, F. 1937. Design and Analysis of Factorial Experiments. Imperial Bureau of Soil Science, London.

SEQUENTIAL FACTORIAL ESTIMATION

by

GUANG-CHUEN LIN

B. Ed., Taiwan Normal University, 1960

M. Ed., Taiwan Normal University, 1964

AN ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the

requirement for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1967

In industrial experimentation it is often possible to run the experiments in a factorial experiment consecutively and to observe or calculate the response at the completion of a block of runs or an added run before the next experiment is run. This has led experimenters to consider sequential planning schemes.

The factorial designs discussed are the 2^n and 2^{n-k} experiments. The numbers 1, 2, 3, ..., n are used to denote the n variables, plus and minus signs to represent high and low levels, respectively.

To analyze the sequential data, a predictor-corrector equation is developed

$$\hat{B}* = \hat{B} + \frac{1}{rN + q}\sum_{1}^{n}(y_i - y_i)z_i'$$

where $\hat{B}*$ = (q x 1) vector of revised estimates,

$\hat{B}$ = (q x 1) vector of estimates provided by prior block(s),

N = total number of runs in a block,

q = number of coefficients in the model,

r = number of blocks of N runs completed,

z_i = (1 x q) row vector in matrix of independent variables associated with the ith experiment, $i = 1, 2, ..., n \leq N$,

y_i = new observation associated with z_i,

$\hat{y}_i = z_i B$ = predicted response for the ith experiment,

by which an experimenter may quickly determine the least squares estimates of all the coefficients in a polynomial model

$$E(y) = b_o + \sum_{i=1}^{n} b_i x_i + \sum_{i<j} \sum^{n} b_{ij} x_i x_j + \ldots\ldots\ldots\ldots\ldots\ldots$$

after the conclusion of each run or any group run, given that an initial set of orthogonal estimates of the coefficients is available.

The equation is subject to mild restriction which are fully met in the usual application of the 2^n and 2^{n-k} factorial designs.

Two conditions must be satisfied before using predictor-corrector equation to perform factorial estimation: 1) the estimates B provided by the prior block of N runs must be mutually orthogonal and 2) the added row vectors must be row-wise orthogonal.