International Journal of Mathematics and Statistics Studies
Vol.2, No.2, PP.13-27, June 2014
Published by European Centre for Research Training and Development UK (www.ea-journals.org)

A TECHNIQUE FOR $n2^k$ FACTORIAL DESIGNS

Ukwu Chukwunenye
Department of Mathematics, University of Jos, P.M.B 2084, Jos, Postal Code: 930001, Plateau State, Nigeria.

ABSTRACT: *This article developed an appealing technique for $n2^k$ factorial designs that would generate more compact and more efficient computational results on $n2^k$ complete experiments that would be of immense benefit to students and researchers. The article leveraged on existing body of knowledge on 2^k factorial designs to contrive and exploit a series of orthogonal and block diagonal matrices, which formed the basis for the statements and proofs of envisaged results on complete $n2^k$ experiments. The research effort culminated into statements and proofs of what the researcher referred to as Ukwu's theorem and its corollary. These would elucidate the design process, offer computational advantages on the prosecution of complete $n2^k$ experiments, as well as enhance their mathematical appreciation. The utility and applicability of the results of the investigation should be multi-disciplinary in nature and scope.*

KEYWORDS: Block, Designs, Effects, Estimated, Factorial, Orthogonal

INTRODUCTION

Factorial Designs

Factorial designs can be likened to dials whose level settings control process characteristics. A large number of processes are impinged upon and impacted by myriad factors, some of which are controllable, and others characterized by complete randomness and lack of controllability. The kernel of factorial designs is the random implementation of "standard" level combinations of the controllable factors, with a view to "optimizing" desired specified process characteristics via statistical estimation. Factorial designs are appropriate experimental designs when there are several factors to be investigated at two or more levels, under the assumption that interaction of factors may be important. Design matrices constitute the basic structure whose entries or components represent treatment combinations of several factors, their interactions and output responses, the latter of which are usually organized in contiguous output response columns. This structure facilitates the investigation of several factors at several levels by enabling the running of all combinations of factors and levels.

Factorial designs are extensions of the two-way ANOVA designs that seek to accomplish the following tasks:

(*i*) Estimation and comparison of effects of several factors usually referred to as the main effects of these factors. These effects are usually denoted by α_i's and β_j's in two-way ANOVA designs.

(*ii*) Estimation of possible interaction effects, denoted by $(\alpha\beta)_{ij}$'s with replicate observations or measurements.

(*iii*) Estimation of variance via MSE, SS and SS_{int}

(*iv*) Testing the significance of main and interaction effects and establishing confidence intervals for those effects, using obtained estimates in conjunction with the *F*- and t- tests. See Juran and Gryna (1988), Farnum (1994) and Ukwu (2014) for these and further discussions on factorial designs.

Factorial designs have widespread applicability in research, quality control circles and in Industry. The pursuance of the tasks set forth in (i) through (iv) is referred to as a complete factorial experiment.

THEORETICAL UNDERPINNING

Experimental runs (experimental trials)

A treatment combination (a single realization of joint level settings of the factors) is called an experimental run. For example, if pressure, *P*, temperature, *T* and relative density, *G* are factors whose effects on the shrinkage factors of condensate of some liquid hydrocarbon are being investigated, then the vector setting (*G*, *P*, *T*) = (0.7, 100 bars, 225 ^0F) is an experimental run or trial for the three-factor experiment-investigation of the effects of pressure, temperature and relative density on shrinkages of the liquid hydrocarbon.

Minimum number of experimental runs for a complete factorial experiment
The minimum number of such runs is the product of all the levels of the factors. Thus, if factor (F_j) has l_j levels; $j = 1, 2, \cdots, k,$ the minimum number of runs equals:

$$N = \prod_{j=1}^{k} l_j, \tag{1}$$

assuming that $F_1, F_2, \cdots, F_k$ are the only factors of interest in the experiment. The factorial experiment is then referred to as $N = \prod_{j=1}^{k} l_j$ factorial design.

(An l_1 by l_2 by $\cdots$ by l_k factorial design).

For example, a 4 by 5 by 6 (4*5*6) factorial design identifies a factorial experiment with 3 factors: F_1, F_2 and F_3 say, and 4, 5, and 6 levels for F_1, F_2 and F_3 respectively.

Prohibitive number of runs and need for pruning

Ideally, the largest feasible set of factors should be used in initiating an experimental investigation. However, the impracticality of using the full set is better appreciated by the multiplicative rule for the number of runs expressed in (1). Evidently, for large problem sizes, the computing complexity associated with prosecuting such an envisaged exercise would consume or exhaust available resources for the study, even before its completion (given several factors with high level sizes).

The need to prune to size, the number of runs becomes imperative. This can be achieved, among other possibilities, by using exactly two levels for each factor. In this case, (1) evaluates to

$$N = \prod_{j=1}^{k} 2 = 2^k.$$

Such a factorial experiment is commonly referred to as a 2^k factorial design. In other words, a factorial experiment with k factors, each of which has 2 levels is called a 2^k factorial design. As established already, this design has 2^k runs or trials.
In a situation where n replications are allowed for each cell, the design is referred to as $n2^k$ factorial design. In this case, suppose the j^{th} treatment sample is of size n_j, then the total number of treatments in all samples is given by $N = \sum_{j=1}^{k} n_j$

Identification of individual runs (coding scheme for factor levels)
There are four coding schemes popularly used to identify individual runs or vectors of treatment combinations. These are:
(*i*) Geometric coding scheme
The convention adopted in this scheme is to assign a plus (+) sign to the "higher" level setting of each factor and a minus (–) sign to the other ("lower level"). Thus, each main effect column intersecting the 2^k rows constituting the 2^k runs has exactly 2^{k-1} minus signs. The order in which these signs are distributed in each of the columns of main effects is of extreme importance; it is at the heart of such design considerations. The following standard order, Table 1, also called Yate's order is prevalent in the literature.

Run number	Factors							Response	
	F_1	F_2	F_3	F_4	F_5	$\cdots$	F_k	y	
1	−	−	−	⋯	⋯	⋯	−	y_1	↓
2	+	−	−	⋯	⋯	⋯	−	y_2	Top
3	−	+	−	⋯	⋯	⋯	−	y_3	half
4	+	+	−	⋯	⋯	⋯	⋮	y_4	(TH)
⋮	−	−	⋮	⋯	⋯	⋯	⋮	⋮	
2^{k-1}	−	−	+	⋯	⋯	⋯	−	$y_{2^{k-1}}$ ⋮	↑
$2^{k-1}+1$	+	⋮	+	⋯	⋯	⋯	+	$y_{2^{k-1}+1}$	↓
$2^{k-1}+2$	−	+	+	⋯	⋯	⋯	+	$y_{2^{k-1}+2}$	Bottom
⋮	+	+	+	⋯	⋯	⋯	+		half
⋮	⋮	−	−	⋯	⋯	⋯	+	⋮	(BH))
⋮	⋮	−	−	⋯	⋯	⋯	⋮	⋮	
⋮	⋮	+	−	⋯	⋯	⋯	⋮	⋮	
⋮	⋮	+	−	⋯	⋯	⋯	⋮	⋮	
⋮	⋮	−	⋮	⋯	⋯	⋯	⋮	⋮	
⋮	⋮	−	+	⋯	⋯	⋯	⋮	⋮	
⋮	⋮	⋮	+	⋯	⋯	⋯	⋮	⋮	
2^k-1	−	+	+	⋯	⋯	⋯	⋮	y_{2^k-1}	
2^k	+	+	+	⋯	⋯	⋯	+	y_{2^k}	↑

Table 1: Standard order (Yate's order) Design matrix for an unreplicated 2^k factorial design.

"–"and "+" denote low and high level settings respectively. Here, the F_j columns;
$j = 1, 2, \cdots, k$
are the main effects columns. The F_j column is characterized by a string of alternating 2^{j-1} (–) and (+) signs for $j = 1, 2, \cdots, k$. The (–) sign initiates the process. Note that there is only one response column (response vector) owing to the fact that the observations are unreplicated (not repeated).
The 2^k runs should be conducted in random order to minimize the effect of possible biases due to non-randomization of the runs. Put another way, the running of the design matrix should be randomized in order to minimize the effect of extraneous variables from being mistaken for, or confounded with the main and interaction effects. With n replications per cell in a 2^k factorial design, there are exactly $n2^k$ runs (treatment combinations or level settings) with n response columns y_c; $c = 1, 2, \cdots, n$
yielding the response matrix

$$(y_{rn});\ r = 1, 2, \cdots, 2^k;\ c = 1, 2, \cdots, n$$

with corresponding row sums R_r; $r = 1, 2, \cdots, 2^k$. This situation is depicted as in the table below.

Response matrix				Row sums
y_{11}	y_{12}	$\cdots$	y_{1n}	R_1
y_{21}	y_{22}	$\cdots$	y_{2n}	R_2
$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$
$y_{2^k 1}$	$y_{2^k 2}$	$\cdots$	$y_{2^k n}$	R_{2^k}

Table 2: Response matrix and corresponding row sums for a replicated factorial design in standard (Yates) order.

This is an n by 2^k matrix, with row r sum, R_r given by:

$$R_r = \sum_{c=1}^{n} y_{rc},\ r = 1, 2, \cdots, 2^k. \quad (2).$$

Given such a replicated design, the run numbers and factor columns are preserved but the response vector y is replaced by Table 2 to obtain the appropriate design matrix in an increasing order of run numbers, matrix of response columns and a vector, R of matrix row sums(R is a 2^k by 1 matrix). See Farnum (1994).

Contrasts:
These are the signed sums of elements of the row aggregated response columns after the signs of the specified effects column have been assigned the response row sums.

Following the Geometric notation, + and – signs are used to denote the levels of each factor. A "+" denotes the higher-level setting, and a "–" the lower. For example, for temperature settings at levels 100 °C and 90 °C, "+" denotes the temperature at 100 °C and "–" that at 90 °C.

Main effects:
The main effect of any factor is the average effect of that factor: the average effect of the factor at the higher-level settings minus the average effect at the lower-level settings.

Interaction effects:
The interaction between two factors F_i and F_j (average effect of F_i for the higher-level setting of F_i minus the average effect of F_j for the lower level setting of F_j.

$$\text{Main effect or interaction} = \frac{\text{Contrast for factor or interaction}}{n2^{k-1}} \quad (3).$$

= (sum of the components of the aggregated response columns after being appended with the signs of the factor or interaction column) divided by $n2^{k-1}$, (where, n is the number of replicates)

$$= \operatorname{sgn}(\text{Factor})^T.y = \operatorname{sgn}(\text{Interaction})^T.y,$$

where Factor is the vector of signs in the factor's column of the design matrix. Ditto for $\operatorname{sgn}(\text{Interaction})^T.y$,
for the interaction column of interest.

Precise form of each estimated effect:
Each estimated effect is a statistic of the form $\bar{c}_j^+ - \bar{c}_j^-$, where:

$\bar{c}_j^+$ = effect's column average (column j), at the higher-level settings.

c_j^- = effects column average at the lower level settings.

There are altogether 2^{k-1} main and interaction effects and these are assumed (following the standard order) to be located in columns j = 1 through 2^{k-1} of the design matrix.

Error estimation:
The need to estimate the error associated with the estimates $\bar{c}_j^+ - \bar{c}_j^-$ cannot be overemphasized. In this regard, the following hypotheses need to be tested:

$$H_{0j}: \mu_j^+ - \mu_j^- = 0 \text{ against } H_{1j}: \mu_j^+ - \mu_j^- \neq 0; \; j \in \{1, 2, \cdots, 2^{k-1}\}, \quad (4)$$

where μ_j^+ and μ_j^- are the population parameters corresponding to $\bar{c}_j^+$ and $\bar{c}_j^-$ respectively, for each j.
Needless to say that:

$$H_{0j}: \mu_j^+ = \mu_j^- = \mu \quad (5).$$

where μ is the grand population mean.

Above tests trigger the following program:
(i) Choosing a level of significance, α, for the tests.
(ii) Estimating the population variance, σ^2.
(iii) Getting the standard errors for the estimates $\bar{c}_j^+$ and $\bar{c}_j^-$ and hence for the differences $\bar{c}_j^+ - \bar{c}_j^-$

Call these the standard errors of estimated effects and denote each of them by

SE(estimated effect j) $= SE(\bar{c}_j^+ - \bar{c}_j^-)$.

(iv) Calculating the t-statistics $t_{\alpha/2;\,\nu}$, where ν = associated df for the estimates and subsequently comparing these with the critical t value.

(v) Making a decision on the H_{0j}'s.

(vi) Obtaining interval estimates for associating population statistics/parameters with respect to the significant ones (for which H_{0j}'s are false).

How can be above program be achieved?

The answer to this question is reflected in the following steps/procedures:

(i) Obtain the grand sample average, $\bar{y}_G$ and the main effect and interaction estimates. Choose a level of significance, α.

$$\bar{y}_G = \frac{1}{n2^k}\sum_{r=1}^{2^k}\sum_{t=1}^{n} y_{rt} \qquad (6).$$

(ii) Estimating σ^2.

Case (a) : $n = 1$.

Here, the measurements or observations are unreplicated (non-repeated). In this case, the sample does not yield an estimate of the experimental error against which the main and interaction effects can be evaluated. This problem can be resolved in one of at least three ways:

a(i) using the average of the sum of squares associated with interaction effects of at least three factors as an estimate of σ^2. Call this sum of squares $SS_{\text{int}3+}$.

a(ii) using an independent estimate of the error variance, if available.

a(iii) replicating the experiment (2^k design experiment) n times, where, $n \geq 2$. The use of $n = 2$ is quite common.

The choice of option a(i) provides a "near" ready-made estimate of variance due to experimental error, since the estimates of the main and interaction effects are already secured. The following reasoning paves the way for the desired estimate:

$$SS_j = \frac{\left(\text{Contrast}_j\right)^2}{N} = \frac{c_j^2}{N} = \frac{\left(c_j^+ - c_j^-\right)^2}{N}, \qquad (7).$$

where N is the total number of observations. $N = 2^k$; with this option contrast j is the contrast for the factor or interaction located in column j.

Number of third and higher order interactions:

$$= 2^k - \left(K_{C_0} + K_{C_1} + K_{C_2}\right)$$

$$= 2^k - \frac{\left(K^2 + K + 2\right)}{2} = N_{3^+} \qquad (8).$$

$$\left(\begin{array}{l} K^2 + K + 2 \text{ is divisible by } 2 \text{ since, } K^2 + K + 2 = K(K+1) + 2 \\ \text{and } K(K+1) \text{ is even, as well as 2; so } N_{3^+} \text{ is well defined} \end{array}\right).$$

These effects are located in columns $\frac{\left(K^2 + K + 2\right)}{2}$ through $2^k - 1$.

Therefore:

$$SS_{int3+} = \sum_{j=\frac{K^2+K+2}{2}}^{N-1} (ss_j)^2, \qquad (9).$$

$$(N-1 = 2^k - 1).$$

$$\text{Effect}_j = \bar{c}_j = \frac{c_j}{2^{k-1}} = \frac{2c_j}{2^k} = \frac{2c_j}{N}$$

(using 7), yielding

$$c_j = \frac{N}{2} \text{ Effect}_j.$$

$$(ss_j)^2 = \frac{1}{N}\left(\frac{N}{2}\right)^2 (\text{Effect}_j)^2$$

$$= \frac{N}{4}(\text{Effect}_j)^2$$

and (9) becomes:

$$SS_{int3+} = \frac{N}{4} \sum_{j=\frac{k^2+k+2}{2}}^{N-1} (\text{Effect}_j)^2 \qquad (10).$$

The variance is now estimated by:

$$s^2 = \frac{SS_{int3+}}{N_{3^+}} \qquad (11).$$

The associated degrees of freedom,

$$df = \nu = N_{3^+}.$$

In summary,

$$s^2 = \frac{N}{4}\begin{pmatrix}\text{average squared effect for three} \\ \text{or more factor interactions}\end{pmatrix},$$

where:

$$N = 2^k. \qquad (12)$$

or:

$$s^2 = \frac{1}{N}\begin{pmatrix}\text{average squared contrast for} \\ \text{three or more factor interactions}\end{pmatrix} \qquad (13).$$

Number of squared effects

$= $ number of squared contrasts

$$= df = \nu = N_{3+} = 2^k - \frac{(k^2+k+2)}{2} \qquad (14).$$

(*iii*) Getting

$$SE(\text{estimate effect}_j) = SE(\bar{c}_j^+ - \bar{c}_j^-), \qquad (15).$$

3.0 Methodology

3.1 Orthogonal vectors and orthogonal matrices

The distinct nontrivial column vectors

$$v_1,\ v_2,\ \cdots,\ v_N,$$

are said to be mutually orthogonal (pair-wise orthogonal) if :

$$v_i^t\, v_j \;=\; 0 \quad \text{for } i \neq j,\ i,\ j \in \{1,\ 2,\ \cdots,\ N\}, \qquad (16)$$

where N is any positive integer ≥ 2 and $(.)^t$ denotes the transpose of $(.)$.

Assume that each vector is r-dimensional (has r-rows). Then the following relation is immediate:

$$D = V^t\, V = \mathrm{Diag}\left(\|v_1\|^2,\ \|v_2\|^2,\ \cdots,\ \|v_m\|^2\right), \qquad (17).$$

where

$$V = (v_1,\ v_2,\ \cdots,\ v_N),\ m = \min\{r,\ N\} \text{ and } \|v_j\| \text{ is the norm of } v_j;\ \|v_j\|^2 = \sum_{i=1}^{r} v_{rj}^2,$$

the scalar (dot) product of v_j with itself or the sum of the squared components. D is an m by m diagonal matrix with the d^{th} diagonal entry given by

$$\|v_j\|^2;\ d \in \{1,\ 2,\ \cdots,\ m\}.$$

$$\left(\text{The vectors}\left\{w_j = \frac{v_j}{\|v_j\|} \text{ are orthornormal}\right\}\right).$$

In other words,

$$w_j^{\,t}.w_l = \begin{cases} 1;\ \text{if } j = l \\ 0;\ j \neq l \end{cases}$$

$$\text{or} \qquad W^t W = I_m,\ \text{where: } W = (w_1,\ \cdots,\ w_N)$$

RESULTS

Main Results

Denote the 2^k design and interaction matrix of signs by V. Then,

$\|v_j\|^2 = 2^k$, for each $j \in \{1,\ 2,\ \cdots,\ 2^{k-1}\}$ and V is a 2^k by 2^{k-1} orthogonal matrix.

Consequently:

$$D = V^t V = 2^{k-1} I_{k-1}, \qquad (18)$$

where, I_{2^k-1} is the identity matrix of order 2^{k-1}.

If V is extended to the column of row sums of the response matrix (in case of n replications of column of response; for a single replicate, $n = 1$), then, V is a square matrix of order 2^k with a vector of 1's in the last column (response column signs being + all through). Therefore,

$$D = V^t V = 2^k I_{2^k}. \qquad (19)$$

Set $U = V$ in (18), so that

$$D = U^t U = 2^{k-1} I_{k-1}.$$

Theorem

Let $c_1, c_2, \cdots, c_{2^k-1}$ be the contrasts of a 2^k replicated design with n replicates per cell in standard order. Let T_G be the grand sample total.

Let $\bar{c}_1, \bar{c}_2, \cdots, \bar{c}_{2^k-1}$ and $\bar{y}_G$ be the corresponding estimated effects and the grand mean.

Let Y be the vector of the row sums of the response matrix.

Let M be the extended 2^k by 2^k matrix of the columns of signs corresponding to the main factors, interactions, and the vector Y, (note that sgn $Y = (+, +, \ldots, +)^t$) with every occurrence of " + " replaced by " 1 " and " – " replaced by " –1 ".

Let $SE(\bar{c}_j^+)$, and $SE(\bar{c}_j^-)$ be the standard errors of the estimates of the main and interaction effects at the higher and lower level settings respectively where:

$$\bar{c}_j = \bar{c}_j^+ - \bar{c}_j^- . \quad (20)$$

Let d_{ii} be the i^{th} diagonal element of the diagonal matrix,

$$\frac{2}{n}\left(M^t M\right)^{-1},$$

where t denotes transposition.

Set:

$$N = n2^k. \quad (21).$$

Then:

(i) $$(c_1, c_2, \cdots, c_{2^k-1}, T_G)^t = M^t Y. \quad (22)$$

(ii) $$(\bar{c}_1, \bar{c}_2, \cdots, \bar{c}_{2^k-1}, 2\bar{Y}_G)^t = \frac{2}{n}\left(M^t M\right)^{-1} M^t Y. \quad (23).$$

(iii) $$d_{ii} = \frac{1}{n2^{k-1}} = \frac{2}{N} \quad (24).$$

uniformly in:

$$i \in \{1, 2, \cdots, 2^k\}. \quad (25).$$

(iv)

$$SE(\bar{c}_j^+) = SE(\bar{c}_j^-) = s\sqrt{d_{ii}}, \text{ for } j \in \{1, 2, \cdots, 2^k - 1\}. \quad (26)$$

(v) $$SE(\text{estimated } j^{th} \text{ effect}) = SE(\bar{c}_j^+ - \bar{c}_j^-) = \frac{2s}{\sqrt{N}}, \quad (27)$$

where: s^2 is the estimated population variance.

Proof

(i) M^t is the 2^k by 2^k matrix with the signs for the main and interaction effects in rows $1, 2, \cdots, 2^k - 1$

in that order, and the sign (1,1, …, 1) of the vector , Y^t in row 2^k. Therefore, the operation $M^t Y$ results in each row or factor and interaction signs and the last row vector (1, 1, … ,1) being multiplied by total response vector Y, yielding the contrasts and the grand total,

$$\begin{pmatrix} c_1 & c_2 & \cdots & c_{2^k-1} & T_G \end{pmatrix}^t, \text{ as desired.}$$

(ii) By (18),

$$M = V \text{ and } (M^tM)^{-1} = 2^{-k} I_{2^k} = \frac{n}{N} I_{\frac{N}{n}} \qquad (28).$$

so that

$$\frac{2}{n}\left(M^tM\right)^{-1} = \frac{1}{n2^{k-1}} I_{2^k} = \frac{2}{N} I_{\frac{N}{n}}. \qquad (29).$$

Therefore:

$$\frac{2}{n}\left(M^tM\right)^{-1} M^tY = \frac{1}{n2^{k-1}}\left(c_1, c_2, \cdots, c_{2^k-1}, T_G\right)^t = \left(\bar{c}_1, \bar{c}_2, \cdots, \bar{c}_{2^k-1}, \frac{2T_G}{N}\right)^t$$
$$= \left(\bar{c}_1, \bar{c}_2, \cdots, \bar{c}_{2^k-1}, 2\bar{Y}_G\right) \qquad (30).$$

as required.

(iii) $$\frac{2}{n}\left(M^tM\right)^{-1} = \frac{1}{n2^{k-1}} I_{2^k} = \text{Diag}\left(\frac{1}{n2^{k-1}}, \frac{1}{n2^{k-1}}, \cdots, \frac{1}{n2^{k-1}}\right) \equiv \text{Diag}\left(\frac{1}{n2^{k-1}}\right), \qquad (31).$$

a diagonal matrix with value $\frac{1}{n2^{k-1}}$ in each diagonal entry and 0 elsewhere. Therefore:

$$d_{ii} = \frac{1}{n2^{k-1}}, \qquad (32).$$

by the definition of d_{ii}.

The following preparation is needed to accomplish (iv) and (v).

M^tY is a matrix with the factors and interactions in rows 1 through $2^k - 1$ and the transpose of the total response vector in row 2^k. Each of the rows 1 to $2^k - 1$ has exactly 2^{k-1} positive signs and 2^{k-1} negative signs.

Let:

$$X_j^+ = \begin{cases} 1, & \text{if the level setting in column } j \text{ is high} \\ 0, & \text{otherwise.} \end{cases} \qquad (33).$$

Similarly,

$$X_j^- = \begin{cases} -1, & \text{if the level setting in column } j \text{ is low} \\ 0, & \text{otherwise.} \end{cases} \qquad (34).$$

Then the 2^k factorial design in standard order induces the following definition:

$$X_{ij}^{+} = \begin{cases} 1, & \text{if the level setting in row } i \text{ is high and is} \\ & \text{taken from column } j;\ 1+(j-1)2^k \le i \le j2^k \\ 0, & \text{elsewhere.} \end{cases} \qquad (35).$$

Similarly,

$$X_{ij}^{-} = \begin{cases} -1, & \text{if the level setting in row } i \text{ is low and is} \\ & \text{taken from column } j;\ 1+(j-1)2^k \le i \le j2^k \\ 0, & \text{elsewhere.} \end{cases} \qquad (36).$$

Let J_j^+ and J_j^- denote the columns corresponding to X_{ij}^+ and X_{ij}^- respectively, each having exactly 2^{k-1} nonzero entries and 2^{k-1} zeros.
Let U^+ and U^- be the block diagonal matrices:

$$U^+ = \text{Diag}[\,J_1^+,\ J_2^+,\ \cdots,\ J_{2^k-1}^+\,]$$

and

$$U^- = \text{Diag}[\,J_1^-,\ J_2^-,\ \cdots,\ J_{2^k-1}^-\,] \qquad (37).$$

In particular, for $j = 1,\ 2,\ \cdots,\ k$, J_j^+ is an alternating block of 2^{j-1} 0's and 2^{j-1} 1's with exactly 2^k entries.

Observe that for $j \in \{1,\ 2,\ \cdots,\ 2^k - 1\}$, J_j^+ is just column j of the design and interaction matrix with + replaced by 1 and 0 elsewhere. Also, J_j^- is just column j of the design and interaction matrix with – replaced by – 1 and 0 elsewhere.

$$\left(U^+\right)^t U^+ = \text{Diag}[\overbrace{2^{k-1},\ 2^{k-1},\ \cdots, 2^{k-1}}^{2^k-1 \text{ copies}}] = \left(U^-\right)^t U^- = 2^{k-1} I_{2^k-1} \qquad (38).$$

Hence,

$$\left[\left(U^+\right)^t U^+\right]^{-1} = \frac{1}{2^{k-1}} I_{2^k-1} = \left[\left(U^-\right)^t U^-\right]^{-1} \qquad (39).$$

Clearly, for $j = 1,\ 2,\ \cdots, 2^k - 1$,

$$[\ J_1^+ + J_1^-,\ J_2^+ + J_2^-,\ \cdots, J_{2^k-1}^+ + J_{2^k-1}^-\ ]$$

$$= \underset{1 \le j \le 2^k-1}{\text{Augmented matrix}} \left[\left(\text{block } j \text{ of } U^+ + \text{block } j \text{ of } U^-\right)\right]$$

$$= \underset{1 \le j \le 2^k-1}{\text{Augmented matrix}} \left[\left(\text{block } j \text{ of } (\ U^+ + U^-\ )\right)\right] \qquad (40).$$

$$= \text{Singly replicated } M, \text{ with row } 2^k \text{ and column } 2^k \text{ deleted.} \qquad (41).$$

For the n-fold replicated design in standard order, the following standard errors are immediate:

(iv) $SE(\bar{c}_j^{\,+}) = s\sqrt{j^{th} \text{ diagonal element of } \left[n\left(U^+\right)^t U^+\right]^{-1}} = s\sqrt{\frac{1}{n2^{k-1}}} = s\sqrt{d_{jj}} = s\sqrt{\frac{2}{N}}.$ (42).

$SE(\bar{c}_j^{\,-}) = s\sqrt{j^{th} \text{ diagonal element of } \left[n\left(U^-\right)^t U^+\right]^{-1}} = s\sqrt{\frac{1}{n2^{k-1}}} = s\sqrt{d_{jj}} = s\sqrt{\frac{2}{N}}.$ (43).

(v) Hence:

$$SE(\text{estimated } j^{th} \text{ effect})$$
$$= SE(\bar{c}_j^{\,+} - \bar{c}_j^{\,-}) = \sqrt{SE^2(\bar{c}_j^{\,+}) + SE^2(\bar{c}_j^{\,+})}$$
$$= s\sqrt{\frac{4}{n2^k}} = \frac{2s}{\sqrt{N}}. \qquad (44).$$

A curious reader will have ploughed through the following facts:

$$\text{Diag}\left[n\left(U^+\right)^t U^+\right]^{-1} = \text{Diag}\left[n\left(U^-\right)^t U^-\right]^{-1}$$
$$= \text{Diag}\left[\frac{2}{n}\left(M^t M\right)^{-1}\right] \qquad (45)$$

with row 2^k and column 2^k deleted.

Each diagonal entry in the $2^k - 1$ by $2^k - 1$ matrix has the value:

$$\frac{1}{n2^{k-1}} = \frac{2}{n2^k} = \frac{2}{N}. \qquad (46).$$

Done!

4.3 Corollary

Let $c_1, c_2, \cdots, c_{2^k-1}$ be the contrasts of a 2^k replicated design with n replicates per cell in standard order. Let T_G be the grand sample total.

Let $c_1, c_2, \cdots, c_{2^k-1}$ and $\bar{y}_G$ be the corresponding estimated effects and the grand mean.

Let Y be the vector of the row sums of the response matrix.

Let M be the extended 2^k by 2^k matrix of the columns of signs corresponding to the main factors, interactions, and the vector Y, (note that sgn $Y = (+, +, \ldots, +)^t$) with every occurrence of " + " replaced by " 1 " and " – " replaced by " –1 ".

Let $SE(\bar{c}_j^{\,+})$, and $SE(\bar{c}_j^{\,-})$ be the standard errors of the estimates of the main and interaction effects at the higher and lower level settings respectively where:

$$\bar{c}_j = \bar{c}_j^{\,+} - \bar{c}_j^{\,-}. \qquad (47)$$

Let d_{ii} be the i^{th} diagonal element of the diagonal matrix,

$$\frac{2}{n}\left(M^t M\right)^{-1},$$

where t denotes transposition.

Set:

$$N = n2^k \qquad (48).$$

Let U^+ and U^- be defined as in the proof of theorem 3.1

Set

$$V = U^{+} + U^{-} \text{ and } \tilde{V} = \text{Diag}(V,\ J_{2^k}), \text{where } J_{2^k} = (\overbrace{1,\ 1,\ \cdots,\ 1}^{2^k \text{ places}})^t$$

Let R be the response matrix:

$$R = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ y_{2^k,1} & y_{2^k,2} & \cdots & y_{2^k,n} \end{pmatrix}$$

Set $R_i = \sum_{j=1}^{n} y_{i_j}$ for $i = 1,\ 2,\ \cdots,\ 2^k$, so that $Y = (R_1,\ R_2,\ \cdots,\ R_{2^k})^t$

Set $\tilde{Y} = \text{Diag}\,\overbrace{(Y,\ Y,\ \cdots,\ Y)}^{2^k \text{ places}}$.

Then

(i) $\tilde{V}^t\,\tilde{Y} = M^t\,Y = (c_1,\ c_2,\ \cdots,\ c_{2^k-1},\ T_G)^t$.

(ii) $\tilde{V}^t\,\tilde{V} = M^t\,M = 2^k I_{2^k}$, so that $\frac{2}{n}\left(\tilde{V}^t\,\tilde{V}\right)^{-1}\tilde{V}^t\,\tilde{Y} = \frac{2}{n}\left(M^t\,M\right)^{-1}M^t\,Y$

$$= (\bar{c}_1,\ \bar{c}_2,\ \cdots,\ \bar{c}_{2^k-1},\ 2\bar{Y}_G)^t.$$

(iii) $d_{ii} = \frac{1}{n\,2^{k-1}} = \frac{2}{N}$ where: d_{ii} is the i^{th} diagonal element of $\frac{2}{n}(\tilde{V}^t\,\tilde{V})^{-1}$, unifo rmly in $i \in \{1,\ 2,\ \cdots,\ 2^k\}$.

(iv) $SE(c_j^{+}) = SE(c_j^{-}) = s\sqrt{d_{jj}}$, for $j = 1,\ 2,\ \cdots,\ 2^k - 1$.

(v) $SE(\text{estimated } j^{th} \text{ effect}) = \frac{2s}{\sqrt{N}}$.

Proof

(i) Set $J_i = J_i^{+} + J_i^{-}$, for $i = 1,\ 2,\ \cdots,\ 2^k$. Then $\tilde{V}^t\,\tilde{Y} = \text{Diag}(J_1^t,\ \cdots,\ J_{2^k}^t)\,Y$

$$= \text{Diag}(J_1^t,\ \cdots,\ J_{2^k}^t)\,\text{Diag}(Y,\ \cdots,\ Y) = \begin{pmatrix} J_1^t\,Y \\ \vdots \\ J_{2^k}^t\,Y \end{pmatrix} = (c_1,\ c_2,\ \cdots,\ c_{2^k-1},\ T_G)^t.$$

(ii) $\tilde{V}^t\,\tilde{Y} = \text{Diag}(J_1^t,\ \cdots,\ J_{2^k}^t)\ \text{Diag}(J_1,\ \cdots,\ J_{2^k}) = (J_1^t\,J_1,\ J_2^t\,J_2,\ \cdots,\ J_{2^k}^t\,J_{2^k})$

$$= 2^k\,I_{2^k} = 2^k\,I_{2^k} = M^t M, \text{ as desired.}$$

(iii) By (ii), $\frac{2}{n}(\tilde{V}^t\,\tilde{V})^{-1} = \frac{2}{n} * \frac{1}{2^k}(I_{2^k})^{-1} = \frac{2}{n\,2^k}\,I_{2^k} = \frac{1}{n\,2^{k-1}}\,I_{2^k}$.

Hence, $d_{ii} = \frac{1}{n\,2^{k-1}} = \frac{2}{n\,2^k} = \frac{2}{N}$.

(iv) For $j = 1,\ 2,\ \cdots,\ 2^k$, $SE(\bar{c}_j^{+}) = s\sqrt{d_{jj}} = s = \frac{2}{n}\left[(U^{-})^t\,U^{-}\right]^{-1}$,

noting that for a single response column

$SE(c_j^+) = s\sqrt{d_{jj}^+}$

where: d_{jj}^+ is the j^{th} diagonal element of $\left[(U^+)^t U^+\right]^{-1}$.

Clearly, $d_{jj}^+ = \frac{1}{2^{k-1}}$.

For n replicates,

$$SE(\overline{c}_j^+) = s\sqrt{\text{Diag}_j\left[n(U^+)^t U^+\right]^{-1}} = s\sqrt{\frac{1}{n} * \frac{1}{2^{k-1}}} = s\sqrt{\frac{2}{n2^k}} = s\sqrt{\frac{2}{N}}$$

$$SE(\overline{c}_j^-) = s\sqrt{\text{Diag}_j\left[n(U^-)^t U^-\right]^{-1}} = s\sqrt{\frac{1}{n} * \frac{1}{2^{k-1}}} = s\sqrt{\frac{2}{N}}$$

$$SE(\overline{c}_j) = \sqrt{SE^2(\overline{c}_j^+) + SE^2(\overline{c}_j^-)} = s\sqrt{\frac{4}{N}} = \frac{2s}{\sqrt{N}}, \text{ as desired.}$$

Alternatively,

$$SE(\overline{c}_j) = s\sqrt{\text{Diag}_j(n\tilde{V}^t\tilde{V})^{-1}} = \frac{2s}{\sqrt{N}}.$$

Observe that for n replications of the standard order 2^k design, the B matrix: $B = \tilde{V}^t\tilde{V}$ is n-fold replicated. Thus, $\tilde{B} = nB$ is used in place of B, leading to

$$SE(\overline{c}_j) = s\sqrt{\text{Diag}_j\,\tilde{B}} = \frac{2s}{\sqrt{N}}.$$

IMPLICATION TO RESEARCH AND PRACTICE

Remarks and notes:

$\text{Diag}_j(\text{Mat})$ is the j^{th} diagonal block/element of the matrix Mat.

$\tilde{Y}$ is a 2^{2k} by 2^k matrix

Y is a 2^k by 1 matrix (a column vector).

U^+, U^- and V are $(2^k - 1)\ 2^k$ by $2^k - 1$ matrices.

$\tilde{V}$ is a 2^{2k} by 2^k matrix.

$\text{Diag}(J_1^t, J_2^t, \cdots, J_{2^k}^t)$ is a 2^k by 2^{2k} matrix as are

$\text{Diag}\left[(J_1^+)^t, (J_2^+)^t, \cdots, (J_{2^k}^+)^t\right]$ and $\text{Diag}\left[(J_1^-)^t, (J_2^-)^t, \cdots, (J_{2^k}^-)^t\right]$ are 2^k by 2^{2k} matrices

$\text{Diag}\left[(J_1^t), (J_2^t), \cdots, (J_{2^k}^t)\right]$ $\text{Diag}(Y, \cdots, Y)$ is a 2^k by 1 matrix (a column vector).

The reader who could take the time to verify the multiplication conformability of the various expressions in theorem 3.1 and corollary 3.2 would appreciate the corollary better.

CONCLUSION

The author has developed a technique for $n2^k$ factorial designs by leveraging on existing body of knowledge on 2^k factorial designs and using orthogonality concepts and appropriate decision variables; these culminated into formulations and proofs of more compact and efficient computational results on $n2^k$ complete experiments that would be of immense benefit to students and researchers.

FUTURE RESEARCH

The results in this article will be extended to $n3^k$ factorial designs.

REFERENCES

Juran, J.M. and Gryna, F.M. (1988) Juran's Quality Control HAND BOOK, 4th Ed. McGraw-Hill International Editions, Industrial Engineering Series, NY.1988

Farnum, N.R. (1994) Statistical Quality Control and Improvement, Duxbury Press, Belmont, California, 1994.

Ukwu, C. (2014) "Regression Approach to ANOVA Designs" Manuscript