응용통계연구 제7권 2호
1994년 9월 pp. 201-211

Detection of Influential Interaction Effects in Parameter Design

Sang Ik Kim[1)]

Abstract

Ignoring interaction effects has been pointed out to be one of serious drawbacks in analysis of the parameter designs which are constructed by using orthogonal arrays. In this paper a detecting procedure of influential 2-factor interactions with minimum experimental runs is described, when each control factor has two levels. The presented method is based on the near orthogonal arrays which are very similiar to orthogonal arrays in the statistical structure. And those arrays are the same as trace-optimal balanced saturated two-level fractional factorial designs of resolution V.

1. Introduction

Recently the parameter design technique developed by Taguchi, under the general title of *quality engineering*, has been widely applied to improve products and processes in various industries. To construct parameter design in most cases in the literature, some special orthogonal arrays with strength 2 which allow to analyze only the main effects have been used, as suggested by Taguchi. This is because the number of runs needed in the experiment to estimate interaction effects in orthogonal arrays becomes unmanageably large as the number of factors increases. However, as investigated by Hunter(1985), if some interaction effects among the control factors do indeed exists, then the results of parameter design where only main effects are examined lead to conclusions which are not optimal. For more details about construction and analysis of parameter design, refer to Taguchi and Wu(1980), Taguchi and Phadke(1984), Kackar(1985), or Taguchi(1986).

This problem caused by the influential interactions among control factors in parameter design was addressed by Kim(1992). And he developed new designs by using a partially balanced array instead of othogonal array, which can be used to analyze the main effects and 2-factor interaction effects among control factors when each contol factor has two levels. The designs developed by Kim have no degree of freedom due to an error because of saturated property of the designs, even though they have nice properties such as minimum number of runs and balancedness property of the covariance matrix of the estimates.

For analysis of saturated parameter design, Taguchi suggested obtaining the error sum of

1) Department of Applied Statistics, Kon Kuk University, Seoul, 133-701, KOREA.

squares to be used in testing each effect simply by pooling small sums of squares of the negligible effects. Many authors in the literature, for example Pignatiello and Ramberg(1985), Quinlan(1985) employed the such pooling method to analyze the saturated parameter designs. It is a well known fact, however, that extreme bias can be induced by this kind of pooling method as noted in the simulation study of Box(1988). Moreover, this pooling method can be criticized to be too subjective in determining the effects of which sums of squares are to be pooled to get the error sum of squares.

The procedure outlined in this paper is directed toward construction and analysis of saturated 2^t parameter designs with which we can analyze up to 2-factor interactions among control factors when each of t control factors has two levels. The following section provides a brief description about the designs proposed by Kim(1992). Also can be found the construction of near orthogonal saturated 2^t parameter design in section II. And methods of analyzing the proposed saturated parameter design are described in section III. A simulated example and some remarks are presented in section IV and section V, respectively.

2. Near orthogonal saturated 2^t parameter design

According to terminology introduced by Taguchi, two designs for control factors and noise factors in parameter design are called to be inner array and outer array, respectively. And the inner array of a 2^t parameter design where each control factor has 2 levels is essentially the same as the usual 2^t(fractional) factorial design except that in parameter design an outer array for noise factors is combined to each run of inner array. Therefore, if there are m and n runs respectively in inner and outer array, the total number of runs in parameter design is $m \times n$.

For each run of inner array, n runs of outer array provide n (or more) repeat observations. These repeat observations are representative of the effects of noise factors. And we use them to compute some performance statistic to be analyzed by applying the traditional analysis of variance technique, and Taguchi called the statistic as signal-to-noise(SN) ratio. We can see therefore that designs with small values of m and n should be employed in constucting the inner and outer arrays in order for the parameter design to be cost-effective.

Note here that any design, for example an orthogonal array, can be used as an outer array provided that runs of such a design can cover possible range of all noise factors, and the number of runs is small. This is due to the fact that it is not important whether interactions among noise factors exist or not. Therefore the focus in this paper is restricted to development of inner array. And we consider here only a 2^t parameter design where we are interested in estimating the t main effects denoted by F_i and $t(t-1)/2$ 2-factor interactions denoted by F_iF_j $(i,j=1,2,\cdots,t,\ i\neq j)$ of t control factors, assuming that higher order interactions are all negligible. Such design which permits to estimate up to 2-factor interactions is called resolution V plan according to Box and Hunter(1961).

For a given number t, the minimum number of runs needed in the inner array for a 2^t parameter design with resolution V is obviously $m=1+t+t(t-1)/2$. And Kim(1992) suggested to use the following design D with $m=1+t+t(t-1)/2$ runs for this purpose. The design D is written in $(t\times m)$ matrix from in which each run is represented by a column $(x_1, x_2, \cdots, x_t)'$ with elements $x_i=0$ or 1, and where individual x_i, $i=1,2,\cdots,t$, represents the two levels of the ith control factor.

$$D=\{\ (x_1,x_2,\cdots,x_t)' \mid \sum_{i=1}^{t} x_i = d_1 \text{ or } d_2 \text{ or } d_3\ \}$$

$$\text{where}\quad \begin{array}{l} d_1 = 0 \text{ or } t \\ d_2 = 1 \text{ or } t-1 \\ d_3 = 2 \text{ or } t-2. \end{array}$$

Since each d_i $i=1, 2, 3$ can take 2 values, eight designs are possible to construct for a given number t, and each design has $m=1+t+t(t-1)/2$ runs(columns) so that it is a saturated design for resolution V plan. And Kim(1992) showed that each D is essentially the partially balanced array of strength 4 which can be considered as generalized orthogonal array in a certain sense. Moreover, it can be shown(Srivastava(1965), Kim(1992)) that for each design the covariance matrix of the estimates has some balanced property such that it has at most ten distinct elements shown in Table-1. This means that $Var(\widehat{F_i})$ $i=1,2,\cdots,t$ are all equal, and $Cov(\widehat{F_i}, \widehat{F_iF_j})$ $i\neq j,\ i,j=1,2,\cdots,t$ are the same, but not necessarily $Var(\widehat{F_i})=Cov(\widehat{F_i}, \widehat{F_iF_j})$, $Cov(\widehat{F_iF_j}, \widehat{F_jF_k})=Cov(\widehat{F_iF_j}, \widehat{F_kF_l})$, etc. , where $\widehat{A}$ denotes the estimate of the effect A. For that reason, Kim named the designs minimal balanced 2^t fractional factorial designs of resolution V.

Among the eight possible designs, Kim(1992) showed that the two designs D_1 with $d_1=t$, $d_2=1$, $d_3=t-2$ and D_2 with $d_1=0$, $d_2=t-1$, $d_3=2$ are especially optimal designs which have the minimum trace value of the covariance matrix. If we calculate the actual values of ten possible distinct variances and covariances of the estimates from $t=4$ to 10 for D_1 and D_2 which have the same covariance structure, the results are shown in Table 1. From Table 1 we can see that

(1) the diagonal elements $Var(\widehat{F_i})$ and $Var(\widehat{F_iF_j})$ $i, j=1,2,\cdots,t$, $i\neq j$ in the covariance matrix are all equal, if we ignore the variance $Var(\hat{\mu})$ of the estimate $\hat{\mu}$ of overall mean.

(2) the off-diagonal elements $Cov(\hat{\mu}, \widehat{F_iF_j})$ and $Cov(\widehat{F_i}, \widehat{F_iF_j})$, etc. are very small in magnitude compared to the diagonal elements, especially if we disregard the covariances including $\hat{\mu}$. That is, in most cases covariances are less than 1/10 of the variances in magnitude, so that we may regard the covariance matrix as the diagonal matrix.

(3) particularly when $t=5$, the covariances are all zeros, and the designs D_1 and D_2 are orthogonal arrays.

Table 1. Possible distinct elements in covariance matrics of optimal design $D_1(D_2)$

t	$Var(\hat{\mu})$	$Var(\widehat{F_i})$	$Var(\widehat{F_iF_j})$	$Cov(\hat{\mu}, \widehat{F_i})$	$Cov(\hat{\mu}, \widehat{F_iF_j})$	$Cov(\widehat{F_i}, \widehat{F_j})$
4	0.09722	0.13889	0.13889	-0.00694	0.00694	0.01389
5	0.06250	0.06250	0.06250	0	0	0
6	0.05500	0.05222	0.05222	0.00500	-0.00500	-0.00333
7	0.07639	0.05347	0.05347	0.00868	-0.00868	-0.00434
8	0.12755	0.05041	0.05041	0.01148	-0.01148	-0.00459
9	0.20898	0.05099	0.05099	0.01367	-0.01367	-0.00456
10	0.32099	0.05171	0.05171	0.01543	-0.01543	-0.00441

t	$Cov(\widehat{F_i}, \widehat{F_iF_j})$	$Cov(\widehat{F_i}, \widehat{F_jF_k})$	$Cov(\widehat{F_iF_j}, \widehat{F_iF_k})$	$Cov(\widehat{F_iF_j}, \widehat{F_kF_l})$
4	-0.01389	0.04861	0.01389	-0.04861
5	0	0	0	0
6	0.00333	-0.00361	-0.00333	0.00361
7	-0.00434	-0.00347	-0.00434	0.00347
8	0.00459	-0.00291	-0.00459	0.00291
9	0.00456	-0.00239	-0.00456	-0.00239
10	0.00441	-0.00197	-0.00441	0.00197

From these facts, we may infer that the structure of covariance matrix of the estimates in the design D_1 (or D_2) is similiar to that of orthogonal array, and we may call D_1 and D_2 as near orthogonal staturated designs of resolution V with 2 levels. Either design D_1 or D_2 is therefore recommended here to be used as inner array for costructing parameter designs with minimum runs, where each control factor has 2 levels and the main effects and 2-factor interactions are needed to be analyzed.

Example 1. When $t=4$ and $m=11$ runs, near orthogonal design D_1 with $d_1=t$, $d_2=1$, $d_3=t-2$ is constructed simply by taking one run $(x_1, x_2, x_3, x_4)'$ with each $x_i=1$ (this run corresponds to the condition $\sum_{i=1}^{t} x_i = t$), $t=4$ different runs with only one x_i being 1 such that $\sum_{i=1}^{t} x_i = 1$, and $t(t-1)/2=6$ different runs with only two x_i's being 0's such that $\sum_{i=1}^{t} x_i = t-2$. The result is as follows;

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

Example 2. For $t=5$ and $m=16$ runs, D_1 can be constructed by the similiar way and this design is essentially the orthogonal arry of strength 4.

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \end{bmatrix}$$

3. Analysis of the saturated parameter design

Since the designs D_1 and D_2 are saturated for resolution V plan if used as inner arrays in parameter design, we can not obtain the error mean squares in the analysis-of-variance procedure. However, using the facts that the designs D_1 and D_2 are near orthogonal designs and variances of the estimates of the main effects and 2-factor interactions are all the same, we can analyze the saturated designs by applying the following statistical methods.

The linear model corresponding to the design D_1 or D_2 can be written as

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon},$$

where $\underline{y}$ is $(m \times 1)$ vector of observations which are SN ratios in parameter design, and $\underline{\varepsilon}$ is $(m \times 1)$ vector of experimental errors, and X, $\underline{\beta}$ are respectively $(m \times m)$ model matrix and $(m \times 1)$ vector of parameters(effects).

The model matrix X can be formulated in such way ; all elements of the first column (which corresponds to overall mean μ) are all 1's, and the jth elements $j=1, 2, \cdots, m$ of the $(i+1)$th column corresponding to main effect F_i, $i=1, 2, \cdots, t$ is 1 or −1 according as the level of ith control factor appeared in the jth run is 1 or 0, respectively. Each column matching to interactions F_iF_j is obtained simply by elementwise multiplication of two columns corresponding to main effects F_i and F_j involved in the interaction.

The model matrix X then has a full rank and the best linear unbiased estimate(BLUE) $\hat{\underline{\beta}}$ is obtained as $\hat{\underline{\beta}} = (X'X)^{-1}X'\underline{y}$. The elements $\hat{\beta}_i$ (except for $\hat{\mu}$) $i=2, 3, \cdots, m$ of $\hat{\underline{\beta}}$ can be considered as nearly independent and identical samples drawn from the normal population under the usual normality assumption on $\underline{\varepsilon}$. Therefore, from the normal probability plot suggested by Daniel (1959) or the quantile-quantile plot (Q-Q plot) for $\hat{\beta}_i$, $i=2, 3, \cdots, m$ we can identity the influential effects which lie far from a straight line on the plots.

The normal probability plot can be obtained by plotting the empirical cumulative distribution of $\hat{\beta}_i$, $i=2, 3, \cdots, m$ on the (half) normal probability paper, and we can draw Q-Q plot with plotting the $(m-1)$ ordered pairs $(q_i, \hat{\beta}_{(i)})$, where $\hat{\beta}_{(i)}$ is an ordered value of β_i and q_i is the quantile of the standard normal distribution such that $q_i = \Phi^{-1}((i-1/2)/m-1)$ in which Φ^{-1} denotes the inverse cummulative standard normal distribution function.

As an another method, the stepwise regression selection procedure, especially forward selection technique can be employed for analysis of the proposed saturated parameter design. It is well known fact that the stepwise regression routine arrives at an unreasonable conclusions when the estimates are very highly correlated. But the designs proposed here have the property of near orthogonality, the results obtained from the stepwise method may be highly reliable.

The plotting methods are not particularly informative unless the number of estimates is moderate to large. Therefore they can be employed successfuly when m is greater than 20, but the number influential effects is small. And for the stepwise regression method, the problem of inaccurate type I error may be arised since too many tests for selecting influential effects are conducted. Therefore simultaneous application of the plotting methods and the stepwise regression procedure might be healthy in drawing proper conclusions.

4. Example

To illustrate the procedure, a 2^4 experiment of control factors is presented, and the inner array used here is D_1 shown in Example 1. For the purpose, we generate one observation y_i for each of eleven runs, which are considered here as SN ratios. In generating y_i, we use the following scheme:

(1) Among 4 main effects, only three of them F_1, F_2, F_4 are influential such that effects are 3.5, -4.0, -2.8 in order.

(2) Among 6 two-factor interactions, only two F_1F_2, F_2F_4 are influential with the effects being -2.5, 3.3, respectively. And we assign the small value 0.3 to F_3F_4 and zeros to the remaining three interactions F_1F_3, F_1F_4, F_2F_3 and one main effect F_3.

(3) The eleven random components of $\underline{\varepsilon}$ are generated from the standard normal distribution by using a normal random generation funtion of SAS/IML package, and we assign 20 to the overall mean μ.

The realization of above procedure is shown in Table-2 and corresponding model matrix X can be written easily following the way explained in section III, and the vector of parameters $\underline{\beta}$ is as follows.

$$\underline{\beta}' = (\mu, F_1, F_2, F_3, F_4, F_1F_2, F_1F_3, F_1F_4, F_2F_3, F_2F_4, F_3F_4)$$

Table 2. Simulated observations in 2^4 saturated design with D_1

Run	1	2	3	4	5	6	7	8	9	10	11
y_i	17.96	36.12	14.36	23.78	11.47	17.10	34.48	24.19	14.97	15.18	10.96

Table 3. Estimated effects for the simulated data in Table-2.

$\hat{\mu}$	$\widehat{F_1}$	$\widehat{F_2}$	$\widehat{F_3}$	$\widehat{F_4}$	$\widehat{F_1F_2}$	$\widehat{F_1F_3}$	$\widehat{F_1F_4}$	$\widehat{F_2F_3}$	$\widehat{F_2F_4}$	$\widehat{F_3F_4}$
19.87	3.56	-3.57	-0.28	-2.08	-2.29	-0.31	0.2	0.25	3.38	-0.03

The BLUE of $\hat{\beta} = (X'X)^{-1}X'y$ is listed in Table-3 and the Q-Q plot of $\hat{\beta}_i$'s is pictured in Figure-1. It appears from the plot that five points corresponding to $\widehat{F_2}$, $\widehat{F_1F_2}$, $\widehat{F_4}$, $\widehat{F_2F_4}$ and $\widehat{F_1}$ are significant effects which lie apart from a drawn straight line. However it may be difficult to draw a straight line on which negligible effects falls, particularly when the number of points in the plot is as small as 10 for this case. If we employ the forward selection stepwise method to supplement such plot, the effects F_1, F_2F_4, F_2, F_4, F_1F_2 enter into the model in order as influential effects. Thus, it is easy to draw a straight line in the Q-Q plot with aids of the stepwise procedure.

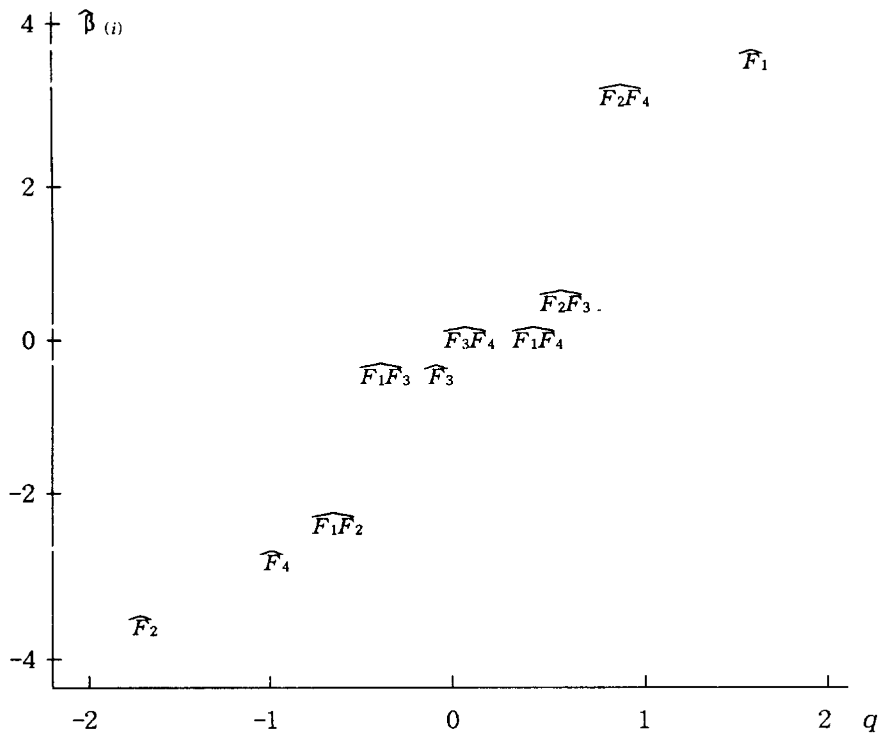

Figure 1. Q-Q plot of the estimates in Table-3

5. Concluding remarks

As noted in the section II, we can construct eight balanced saturated 2^t fractional factorial designs of resolution V for a given number t of control factors, among which two designs denoted here by D_1 and D_2 have a property of near orthogonality. Each of eight designs is partially balanced array of strength 4. (Actually the strength equals to t. That is, each design has full strength.) The structual property of a partially balanced array might be reflected in the index numbers of the array in the sense that a partially balanced array reduces to an orthogonal array if the index numbers are equal. Therefore, the examination of index numbers of other arrays except D_1 and D_2 may be helpful to conjecture the extent of orthogonality of the arrays. The index numbers of eight possible arrays(designs) are also listed in Kim(1992).

To analyze the near orthogonal arrays presented in this paper, any other plotting methods, such as Bayes-plot technique developed by Box and Meyer(1986) for example, can be employed.

References

[1] Box, G.E.P. and Hunter, J.S. (1961). The 2^{k-p} Fractional Factorial Design I, *Technometrics,* Vol. 3, 331-351.

[2] Box, G.E.P. and Meyer, R.D. (1986). An Analysis for Unreplicated Fractional Factorials, *Technometrics,* Vol. 28, 11-18.

[3] Box, G.E.P. (1988). Signal-to-Noise Ratios, Performance Criteria and Transformation (with discussion), *Technometrics,* Vol. 30, 1-40.

[4] Daniel, C. (1959). Use of Half-Normal Plots in Interpreting Factorial Two-Level Experiments, *Technometrics,* Vol. 1, 311-341.

[5] Hunter, J.S. (1985). Statistical Design Applied to Production Design, *Journal of Quality Technology,* Vol. 17, 210-221.

[6] Kackar, R.N. (1985). Off-line Quality control, Parameter Design, and the Taguchi Method, *Journal of Quality Technology,* Vol. 17, 176-209.

[7] Kim, S.I. (1992). Minimal Balanced 2^t Fractional Factorial Designs of Resolution V and Taguchi Method, *The Korean Journal of Applied Statistics,* Vol. 5, 19-28.

[8] Pignatiello, J.J. Jr. and Ramberg, J.S. (1985). Off-line Quality Control, Parameter Design, and the Taguchi Method: Discussion, *Journal of Quality Technology,* Vol. 17, 198-206.

[9] Quinlan, J. (1985). Product Improvement by Application of Taguchi Method, *American*

Supplier Institute News, 11-16.

[10] Srivastava, J.N. (1965). Optimal Balanced 2^m Fractional Factorial Designs, *S.N.Roy Memorial Volume*, Univ. of North Carolina and Indian Statistical Institute.

[11] Srivastava, J.N. and Chopra, D.V. (1971). On the Characteristic Roots of the Information Matrix of 2^m Balanced Factorial Designs of Resolution V, with Applications, *Annals of Mathematical Statistics,* Vol. 42, 722-734.

[12] Taguchi, G. and Wu, Y. (1980). *Introduction to Off-line Quality Control,* Central Japan Quality Control Association, Tokyo, Japan.

[13] Taguchi, G. and Phadke, M.S. (1984). Quality Engineering through Design Optimization, *IEEE,* 1106-1113.

[14] Taguchi, G. (1986). *Introduction to Quality Engineering,* Asian Productivity Organization, Tokyo, Japan.

파라미터 설계법에서 교호작용효과의 검출방법

김상익[2)]

요약

다구찌에 의해서 개발된 파라미터 설계법에서는 교호작용효과을 무시하고 직교배열을 이용하여 실험을 분석하는 것이 일반적인 방법으로 사용되고 있다. 본 논문에서는 직교배열과 거의 유사한 통계적 성질을 가지면서 최소의 실험횟수로 2인자 교호작용까지 분석할 수 있는 실험계획을 제시하고, 이러한 계획법에 의해 파라미터 설계법을 분석하고 유의적인 2인자 교호작용까지 검출하는 방법을 제시하였다.

2) (133-701) 서울특별시 성동구 모진동 93-1, 건국대학교 응용통계학과.