Iğdır Üniversitesi Fen Bilimleri Enstitüsü Dergisi
Iğdır University Journal of the Institute of Science and Technology

Determination of The Smallest Set of Treatment Combinations for Testing Main and Interaction Effects in a 2^k Factorial Design

Hülya OLMUŞ[1] Semra ERBAŞ[1]

ABSTRACT: The aim in analyzing a 2^k factorial design is to estimate the 2^k main and interaction effects. If some of these main and interaction effects are known to be zero or negligible, it is not necessary to estimate all the main and interaction effects in 2^k factorial design. When S main and interaction effects are non-zero, all possible sets of S treatment combinations are not sufficient for estimating these main and interaction effects. For this reason, a method is introduced to obtain the smallest set of the S treatment combinations. In this study, two smallest sets are obtained for all possible scenarios of interest for 2^3 factorial design using this method given by Tsao and Wibowo. An illustration of this method is solved for 2^3 factorial design by using SPSS 13.0 package program.

Keywords: 2^k factorial design, mean response, linear programming, Simplex method

2^k Faktöriyel Düzende Etki ve Etkileşimlerin Testi İçin Deneme Kombinasyonlarının En Küçük Kümesinin Belirlenmesi

ÖZET: Bir 2^k faktöriyel düzeni analiz etmekteki amaç, 2^k tane etki ve etkileşimi tahmin etmektir. Bu etki ve etkileşimlerin bazıları sıfır veya önemsiz olarak biliniyorsa, 2^k faktöriyel düzende yer alan tüm etki ve etkileşimi tahmin etmek gerekli değildir. Genelde, S tane etki ve etkileşim sıfırdan farklı olduğu zaman, sadece S tane deneme kombinasyonu, bu etki ve etkileşimlerin tahmini için gereklidir. S tane etki ve etkileşim sıfırdan farklı olduğu zaman, S tane deneme kombinasyonun mümkün her kümesi, bu etki ve etkileşimlerin tahmini için yeterli değildir. Bu nedenle, S tane deneme kombinasyonlarının en küçük kümesini elde etmek için bir yöntem tanıtılmıştır. Bu çalışmada, tanıtılan bu yöntem ile 2^3 faktöriyel düzenin tüm mümkün durumları için iki tane en küçük küme elde edilmiştir. Ayrıca, bu kümelerin elde edilmesinde yararlanılan doğrusal programlama modelinin çözümünde WINQSB paket programı kullanılmıştır.

Anahtar kelimeler: 2^k faktöriyel düzen, ortalama yanıt, doğrusal programlama, Simpleks yöntemi

Cilt: 2, Sayı: 2, Sayfa: 51-56, 2012
Volume: 2, Issue:2, pp: 51-56, 2012

[1] Gazi University, Faculty of Science, Department of Statistic, Ankara, Turkey
Sorumlu yazar/Corresponding Author : Hülya OLMUŞ, hulya@gazi.edu.tr

Geliş tarihi / *Received:* 02.02.2012
Kabul tarihi / *Accepted:* 13.03.2012

INTRODUCTION

In a 2^k factorial design, there are k factors and each factor has two levels. When the number of factors k is large, the number of 2^k treatment combinations will be large as well. Another method for minimizing the number of treatment combinations is fractional factorial designs. For example, in a 2^7 factorial design, there are 128 treatment combinations. In this design, 7 degrees of freedom corresponds to the main effects, 21 degrees of freedom corresponds to first order interactions, 31 degrees of freedom corresponds to second order interactions and 35 degrees of freedom corresponds to third order interactions. The degrees of freedom for the remaining interactions (fourth, fifth and sixth order interactions) add up to 29. For this reason, this situation will get more complicated as the number of factors and factor levels increase. Even if the high order interaction terms are not included in the analysis or they are confounded with blocks, the degrees of freedom for the estimation of error will still be large. In this case, instead of applying whole replications of 128 observations, we could get the necessary information by using half of the observations. When only a part of an experiment is applied, it is called the fractional factorial design. These factorial designs are widely used in quality control and industry. This method saves time and money, however, it has the disadvantage of not estimating the main and interaction effects separately since these effects are confounded with other effects (Cochran and Cox, 1992).

In the literature, minimizing the number of treatment combinations is very important. The minimization procedure is based on determining the relation between all possible factors and the response variable. In a 2^k factorial design, if an interaction term is known to be zero or negligible, this term can not be estimated. So, the number of treatment combinations decreases by 1. In general, when only S main and interaction effects are non-zero, only S treatment combinations are needed to estimate these main and interaction effects. In this study, when only S main and interaction effects are non-zero, all possible sets of S treatment combinations are not sufficient to estimate these effects. For this reason, a method is introduced to obtain the smallest set of S treatment combinations. With this method, two smallest sets for all possible scenarios of interest for 2^3 factorial design is obtained (Jacob Tsao and Wibowo, 2005). For example, in a 2^3 factorial design, when AB, ABC interactions effects are assumed to be zero, S=6 main and interaction effects, general mean (ì), A, B, C, AC, BC, is tested. For estimating these main and interaction effects, it is necessary to choose a set of S=6 treatment combinations (for example, (1), a, b, c, ac, bc). However in this study, assuming AB, ABC interactions effects are zero, all possible sets of treatment combinations {(1), a, b, c, ac, bc}, {(1), b, c, ac, bc, abc}…, are not sufficient for estimating these main and interaction effects. For this reason, a method given by Tsao and Wibowo is introduced for obtaining the best set with 6 treatment combinations (Jacob Tsao and Wibowo, 2005). This method is thought to be more practical for most of the problems therefore this method is introduced with an application.

INTRODUCING THE METHOD FOR 2^3 FACTORIAL DESIGN

There are 2^3=8 treatment combinations in 2^3 factorial design. The coefficients table for this design is

Table 1. Coefficients table in 2^3 factorial design

Treatment combinations			Main effects and interactions								
			μ	A	B	AB	C	AC	BC	ABC	Mean response
(0,0,0)	(1)	1	+	-	-	+	-	+	+	-	μ_1
(1,0,0)	a	2	+	+	-	-	-	-	+	+	μ_2
(0,1,0)	b	3	+	-	+	-	-	+	-	+	μ_3
(1,1,0)	ab	4	+	+	+	+	-	-	-	-	μ_4
(0,0,1)	c	5	+	-	-	+	+	-	-	+	μ_5
(1,0,1)	ac	6	+	+	-	-	+	+	-	-	μ_6
(0,1,1)	bc	7	+	-	+	-	+	-	+	-	μ_7
(1,1,1)	abc	8	+	+	+	+	+	+	+	+	μ_8

μ_i , $(i = 1,2,...,8)$ represents the mean responses and they are defined as following.

$\mu_1 = \mu_{---}$, $\mu_2 = \mu_{+--}$, $\mu_3 = \mu_{-+-}$, $\mu_4 = \mu_{++-}$

$\mu_5 = \mu_{--+}$, $\mu_6 = \mu_{+-+}$, $\mu_7 = \mu_{-++}$, $\mu_8 = \mu_{+++}$

given in Table 1. In this table, (-) is used to represent the low levels and (+) is used to represent the high levels of the factors and ì is the overall mean.

The estimation of these eight mean responses is an important step in the estimation of the eight main and interaction effects. The regression model for a 2^3 factorial design is given as the following (Montgomery,1984; Wang, 2005).

$$Y=\mu+\frac{A}{2}x_1+\frac{B}{2}x_2+\frac{C}{2}x_3+\frac{AB}{2}x_1x_2+\frac{AC}{2}x_1x_3+\frac{BC}{2}x_2x_3+\frac{ABC}{2}x_1x_2x_3+\varepsilon \quad (1)$$

In this model, $x_i=\pm 1,\ i=1,2,3$. Since this study is based on mean responses, the following equation plays a key role in the rest of the chapter

$$y=\mu+\frac{A}{2}x_1+\frac{B}{2}x_2+\frac{C}{2}x_3+\frac{AB}{2}x_1x_2+\frac{AC}{2}x_1x_3+\frac{BC}{2}x_2x_3+\frac{ABC}{2}x_1x_2x_3 \quad (2)$$

where $x_i=\pm 1,\ i=1,2,3$. When the mean responses μ_1, μ_2, μ_3, μ_4, μ_5, μ_6, μ_7 and μ_8 are estimated, μ, A, B, C, AB, AC, BC, ABC is solved as the functions of these mean responses. For example, if the AB, AC, BC and ABC interaction effects are known to be zero, the model is

$$y=G+\frac{A}{2}x_1+\frac{B}{2}x_2+\frac{C}{2}x_3 \quad (3)$$

The unknown parameters in equation (3) are general mean (μ) and A, B, C main effects. Only S=4 mean responses are needed to estimate these unknown parameters.

Under the assumption that interaction effects AB, AC, BC and ABC are zero, the equations for each of the interaction effects given below can be written using Table 1.

$$\begin{aligned}
&\mu_1-\mu_2-\mu_3+\mu_4+\mu_5-\mu_6-\mu_7+\mu_8=0\\
&\mu_1-\mu_2+\mu_3-\mu_4-\mu_5+\mu_6-\mu_7+\mu_8=0\\
&\mu_1+\mu_2-\mu_3-\mu_4-\mu_5-\mu_6+\mu_7+\mu_8=0\\
&-\mu_1+\mu_2+\mu_3-\mu_4+\mu_5-\mu_6-\mu_7+\mu_8=0
\end{aligned} \quad (4)$$

These four equations show the linear restrictions on the eight mean responses. The aim is to revise equation (4) in a canonical form as in equation (5). A method is introduced to obtain this canonical form.

$$\begin{aligned}
&\mu_2=\mu_1+\mu_4+\mu_6-\mu_7\\
&\mu_3=\mu_1+\mu_4-\mu_6+\mu_7\\
&\mu_5=\mu_1-\mu_4+\mu_6+\mu_7\\
&\mu_8=-\mu_1+\mu_4+\mu_6-\mu_7
\end{aligned} \quad (5)$$

In obtaining equation (5) from equation (4), Phase I Simplex Method is used. In this section, some concepts are introduced based on this result. According to equation (5), μ_2, μ_3, μ_5 and μ_8 mean responses are given as a linear function of μ_1, μ_4, μ_6 and μ_7 mean responses. According to this equations set, {(1), ab, ac, bc} which corresponds to (1, 4, 6, 7) are called as treatment combinations and from these treatment combinations, μ_1, μ_4, μ_6 and μ_7 mean responses should be estimated. μ_2, μ_3, μ_5 and μ_8 are called as *redundant mean responses*. Equation (5) can be revised to obtain equation (6).

$$\begin{aligned}
&\mu_2-\mu_1-\mu_4-\mu_6+\mu_7=0\\
&\mu_3-\mu_1-\mu_4+\mu_6-\mu_7=0\\
&\mu_5-\mu_1+\mu_4-\mu_6-\mu_7=0\\
&\mu_8+\mu_1-\mu_4-\mu_6+\mu_7=0
\end{aligned} \quad (6)$$

In equation (6), the coefficients related to μ_2, μ_3, μ_5 and μ_8 form an identity matrix. At the same time, these coefficients construct the canonical form (Taha, 1982; Winston, 2004). In equation (6), μ_2, μ_3, μ_5 and μ_8 are basic variables while the remaining terms are non-basic variables.

THE SOLUTION OF THE METHOD USING LINEAR PROGRAMMING MODEL

The most widely used method for solving linear programming models is the Phase I Simplex Method. Using this method, a canonical form as in equation (5) is obtained from equation (4).

Linear programming model consists of n variables and m equations (n≥m). The objective function is expressed as the sum of artificial variables. This objective function is used to obtain m basic variables and (n-m) non-basic variables. In this study, basic variables represent the artificial variables and non-basic variables represent the mean responses. Therefore, using Y_1, Y_2, Y_3 and Y_4 artificial variables, the linear programming model will be as the following (Taha, 1982; Winston, 2004).

Min Z $Y_1+Y_2+Y_3+Y_4$

$$\begin{aligned}
&\mu_1-\mu_2-\mu_3+\mu_4+\mu_5-\mu_6-\mu_7+\mu_8+Y_1=0\\
&\mu_1-\mu_2+\mu_3-\mu_4-\mu_5+\mu_6-\mu_7+\mu_8+Y_2=0\\
&\mu_1+\mu_2-\mu_3-\mu_4-\mu_5-\mu_6+\mu_7+\mu_8+Y_3=0\\
&-\mu_1+\mu_2+\mu_3-\mu_4+\mu_5-\mu_6-\mu_7+\mu_8+Y_4=0\\
&\mu_i\in R,\ Y_j\geq 0
\end{aligned} \quad (7)$$

Note that all regular variables μ_i can be any real number instead of being restricted to non-negative values. Unfortunately, this linear programming has a trivial solution, which is Y_1=0, Y_2=0, Y_3=0 and Y_4=0, and this solution does not serve our purpose (Jacob Tsao, 2005). Due to the property of the canonical form, the values on the right side of the equation (7) do not really play any role. Therefore, the zero values on the right side of the equalities can be exchanged with any positive constants due to our purpose (Taha, 1982; Winston, 2004). In this case, we get a solution related with the objective of the study. If the values on the right side of equality (7) are chosen to be 1, 2, 3 and 4, the linear programming model will be like as the following:

Min Z $\quad Y_1+Y_2+Y_3+Y_4$

Restrictions

$$\begin{aligned}
&\mu_1-\mu_2-\mu_3+\mu_4+\mu_5-\mu_6-\mu_7+\mu_8+Y_1=1\\
&\mu_1-\mu_2+\mu_3-\mu_4-\mu_5+\mu_6-\mu_7+\mu_8+Y_2=2\\
&\mu_1+\mu_2-\mu_3-\mu_4-\mu_5-\mu_6+\mu_7+\mu_8+Y_3=3\\
&-\mu_1+\mu_2+\mu_3-\mu_4+\mu_5-\mu_6-\mu_7+\mu_8+Y_4=4
\end{aligned} \tag{8}$$

$\mu_i \in R,\ Y_j \geq 0$.

This model is solved by simplex algorithm. The Phase I linear programming model used for obtaining two smallest sets is solved by using WINQSB. The optimum solutions of this linear programming model are given in Table 2. In this table, BV represents the basic variables.

From Table 2, the restrictions can be written as

$$\begin{aligned}
&-\mu_1+2\mu_2-\mu_4-\mu_6+\mu_7=2.0\\
&-\mu_1+2\mu_3-\mu_4+\mu_6-\mu_7=1.0\\
&-\mu_1+\mu_4+2\mu_5-\mu_6-\mu_7=0\\
&\mu_1-\mu_4-\mu_6-\mu_7+2\mu_8=5.0
\end{aligned} \tag{9}$$

As a result, due to Table 2, the treatment combinations set {a, b, c, abc} corresponding to {2,3,5,8} are redundant treatment combinations. The treatment combinations of interest are {(1), ab, ac, bc} corresponding to {1,4,6,7}. Therefore, if in order to obtain a relation between mean responses $\{\mu_2,\mu_3,\mu_5,\mu_8\}$ and $\{\mu_1,\mu_4,\mu_6,\mu_7\}$ the values on the right side of the equality (9) are taken to be zero, the equations will become as in equations (10).

$$\begin{aligned}
&\mu_2=\mu_1+\mu_4+\mu_6-\mu_7\\
&\mu_3=\mu_1+\mu_4-\mu_6+\mu_7\\
&\mu_5=\mu_1-\mu_4+\mu_6+\mu_7\\
&\mu_8=-\mu_1+\mu_4+\mu_6-\mu_7
\end{aligned} \tag{10}$$

Table 2. The optimum table for linear programming problem

μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_7	μ_8	Y_1	Y_2	Y_3	Y_4	BV
0	0	0	0	0	0	0	0	0	0	0	0	Z=0
-0.5	1.0	0	-0.5	0	-0.5	0.5	0	-0.25	-0.25	0.25	1.0	μ_2=1.0
-0.5	0	1.0	-0.5	0	0.5	-0.5	0	-0.25	0.25	-0.25	0.25	μ_3=0.5
-0.5	0	0	0.5	1.0	-0.5	-0.5	0	0.25	-0.25	-0.25	1.0	μ_5=0
0.5	0	0	-0.5	0	-0.5	-0.5	1.0	0.25	0.25	0.25	0.25	μ_8=2.5

Table 3. An alternative optimum table for linear programming problem

μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_7	μ_8	Y_1	Y_2	Y_3	Y_4	BV
0	0	0	0	0	0	0	0	0	0	0	0	Z=0
1	0	0	-1	0	-1	-1	2	0.5	0.5	0.5	0.5	μ_1=5.0
0	1	0	-1	0	-1	0	1	0	0	0.5	1.25	μ_2=3.5
0	0	1	-1	0	0	-1	1	0	0.5	0	0.5	μ_3=3.0
0	0	0	0	1	-1	-1	1	0.5	0	0	1.25	μ_5=2.5

Table 4. Two smallest sets for all possible combinations in a 2^3 factorial design

Two-way interaction is zero	Three-way interaction is zero	Minimum number of treatment combinations	Redundant treatment combinations	Minimal sets of treatment combinations
None	ABC	7	a	(1), b, ab, c, ac, bc, abc
			b	(1), a, ab, c, ac, bc, abc
AB	ABC	6	a, c	(1), b, ab, ac, bc, abc
			b, c	(1), a, ab, ac, bc, abc
AB, AC	ABC	5	b, c, abc	(1), a, ab, ac, bc
			(1), b, c	a, ab, ac, bc, abc
AB, AC, BC	ABC	4	a, b, c, abc	(1), ab, ac, bc
			(1), a, b, c	ab, ac, bc, abc

If the reduced cost of any of the non-basic variables of the optimal solutions is 0, then an alternative optimal solution exists (Taha, 1982;Winston, 2004) From here, an alternative set of four redundant treatment combinations is obtained. In the simplex algorithm, substituting μ_8 with μ_1 gives an alternative solution. The redundant set of treatment combinations is {(1), a, b, c} corresponding to {1, 2, 3, 5}. An alternative optimum table is given in Table 3.

Two smallest sets for all possible combinations in a 2^3 factorial design are given in Table 4.

Table 5. The data set of time to swim 100 m for male swimmers

		Physical fitness			
		Low		High	
		Weights			
		<70	≥ 70	<70	≥ 70
Age	<35	5	9	3	5
		6	10	4	5
	≥ 35	6	12	5	7
		7	13	4	6

APPLICATION

A study is done to determine which factors affect the time to swim 100m for male swimmers. In this study, there are three factors with two levels which affect the response variable which are age of swimmers(A), weights (B) and physical fitness(C). The levels for age factor are low (<35) and high (≥35), the levels for weight factor are low (<70) and high (≥70) and the levels of physical fitness factor are low and high. This study is an example for a 2^3 factorial design. There are $2\times2\times2=2^3=8$ treatment combinations in the design: (1), a, b, ab, c, ac, bc, abc. The data table for the design with three factors with two levels and two replications is given in Table 5 (Erbaş and Olmuş, 2005).

The analysis of variance table for this data set is given in Table 6. The results given in the analysis of variance table below are found by using SPSS 13.0 package program.

Table 6. The analysis of variance table for male swimmers' time to swim 100m

Source of variation	Degrees of freedom	Sum of squares	Mean squares	F
Age (A)	1	10.56	10.56	21.12*
Weight (B)	1	45.06	45.06	90.12*
AB	1	1.56	1.56	3.12
Physical fitness (C)	1	52.56	52.56	105.12*
AC	1	0.56	0.56	1.12
BC	1	10.56	10.56	21.12*
ABC	1	0.56	0.56	1.12
Error	8	4.02	0.50	
Total	15	125.44		

Table 7. The data set of time to swim 100 m for male swimmers

		Physical fitness			
		Low		High	
		Weight			
		<70	≥ 70	<70	≥ 70
Age	<35	5	-	-	5
		6			5
	≥ 35	-	12	5	-
			13	4	

Since $F_{1,8,0.05}$=5.32, age, weight and physical fitness are found to be significant factors in explaining the time for male swimmers to swim 100m. In addition, the weight and physical fitness interaction effect is found to be significant.

In this design, when the AB, AC, BC and ABC interactions effects are assumed to be zero or negligible, we have obtained {(1), ab, ac, bc} as one of the smallest treatment combination sets for testing main effects A, B, C. According to this result, under the two replica-

Table 8. The analysis of variance table of time to swim 100m for male swimmers

Source of variation	Degrees of freedom	Sum of squares	Mean squares	F
Age (A)	1	21.125	21.125	56.333
Weight (B)	1	28.125	28.125	75.000
Physical fitness (C)	1	34.125	34.125	96.333
Error	4	1.500	0.375	
Total	7	86.875		

tions assumption, let the data set related to this design be revised as in Table 7.

The analysis of variance table for the data set in Table 7 is given in Table 8.

Since, $F_{1,4,0.05}$=7.71, age, weight and physical fitness are found to be significant factors in explaining the time for male swimmers to swim 100m.

As a result, the outcome obtained using smallest set of treatment combinations is the same with the outcome obtained using all treatment combinations in 2^3 factorial design. For this reason, using smallest set of treatment combinations is more advantageous then using all treatment combinations in terms of time and cost. In light of this result, this method can be preferred over other methods due to providing practical solutions to most of the problems.

CONCLUSION

In general, when only S main and interaction effects are non-zero, all possible sets of S treatment combinations are not sufficient to estimate these effects in a 2^k factorial design. For this reason, a method is introduced to obtain the smallest set of S treatment combinations. Two smallest sets are given for all possible scenarios of interest for 2^3 factorial designs using this method. The Phase I simplex method used for solving linear programming models is used for obtaining these sets. WINQSB is used for solving this method.

The method introduced in this article can be used for any 2^k factorial designs. Another method for obtaining the smallest set of treatment combinations is fractional factorial designs. The method introduced in this article is a better method than fractional factorial design method. Because fractional factorial designs method deals with the sum of treatment combinations. However, the method introduced gives the smallest set and the optimum number of treatment combinations.

In addition, the method can produce multiple minimal sets of treatment combinations by continuing to perform pivoting after having obtained the first optimal solution to the linear programming. As a conclusion, with this method the researcher obtains the smallest set of treatment combinations and tests the necessary main and interaction effects using this set. In this study, a 2^3 factorial design is considered and shown to be very advantageous in some situations. The practical use of this method given by Tsao ve Wibowo can be extended to 2^k factorial designs.

REFERENCES

Cochran, W.G., Cox, G.M., 1992. Experimental designs. 2nd edition, JohnWiley&Sons, 148-181.

Erbaş, S.O., Olmuş, H., 2005. Deney düzenleri. 1st edition, Gazi Kitabevi, 236, 371-379.

Jacob Tsao, H.S., Wibowo, I., 2005. A method for identifying a minimal set of test conditions in 2^k experimental design", Computers & Industrial Engineering 48, 141-151.

Kirk, R.E., 1982. Experimental Design: Procedures for the Behavioral Sciences, 2nd edition, Wadsworth, Inc., Belmont, California, 570-591.

Montgomery, D.C., 1984. Design and analysis of experiments. 6th edition, JohnWiley&Sons, 203-254, 373-390.

Taha, H.A., 1982. Operations research. 6th edition., New York: MacMillan Publishing Co., 67-108.

Wang, Z., 2005. Experimental Design and Analysis 2^k Factorial Design, Lecture Notes, http://www.csl.mtu.edu/cs4481/www/lectures/ experimentaldesign.pdf.

Winston, W.L., 2004. Operations research: applications and algorithms. 4th edition, Belmont, CA, 127-202.