Journal of Statistical Planning and Inference 163 (2015) 43–47

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Effective designs based on individual word length patterns

William Li [a], Qi Zhou [b,*], Runchu Zhang [c,d,e]

[a] Supply Chain and Operations Department, University of Minnesota, Minneapolis, MN 55455, USA
[b] Department of Statistics, Tianjin University of Finance and Economics, Tianjin 300204, China
[c] KLAS, School of Mathematics and Statistics, Northeast Normal University, Changchun 130024, China
[d] Department of Statistics, University of British Columbia, Vancouver, BC V6T 1Z4, Canada
[e] LPMC, School of Mathematical Sciences, Nankai University, Tianjin 300071, China

ARTICLE INFO

Article history:
Received 30 November 2013
Received in revised form 10 February 2015
Accepted 11 February 2015
Available online 25 February 2015

Keywords:
Effect hierarchy principle
Fractional factorial designs
Word length pattern

ABSTRACT

Most existing criteria for selecting efficient factorial designs are based on effect hierarchy principle. It focuses on making best estimations for lower-order effect, with the underlying assumption that the effects of the same order are equally important. However, experimenters often have some prior knowledge on which factors are more likely to be significant than others, and, consequently, may prefer designs that minimize aliasing between those important factors and other effects. We propose a new criterion that is defined on a column of the design matrix to measure the aliasing of the effect assigned to this column and effects involving other factors. We then use this criterion to rank the columns of a design and hence develop an approach for searching a class of effective designs. These designs are demonstrated to be more appropriate than minimum aberration designs in some scenarios. The results are tabulated for all 16-run and selected 32-run regular two-level designs for possible use by practitioners.

1. Introduction

The problem of selecting efficient fractional factorial designs has received much attention in the literature. Consider a design d with n runs and k factors having two levels each. The design can be represented by an $n \times k$ *design matrix* $\mathbf{X}_d = [\mathbf{x}_1, \ldots, \mathbf{x}_\mathbf{n}]'$, where each row of $\mathbf{X}_d$ is a k-dimensional vector whose elements are $+1$ or -1. A linear model for representing the response is $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{\beta}$ is a vector of m unknown parameters and the error vector $\boldsymbol{\epsilon}$ has mean $\mathbf{0}$ and variance $\sigma^2\mathbf{I}_\mathbf{n}$. The *model matrix* is given by $\mathbf{X} = [\mathbf{f}(\mathbf{x}_1), \ldots, \mathbf{f}(\mathbf{x}_\mathbf{n})]'$, where the functional $\mathbf{f}$ indicates which effects are present in the model. The least squares estimate of $\boldsymbol{\beta}$ is obtained by $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, which has the variance–covariance matrix $(\mathbf{X}'\mathbf{X})^{-1}\sigma^2$. Many optimality criteria are based on the information matrix $\mathbf{X}'\mathbf{X}$. For example, a D-optimal design maximizes $|\mathbf{X}'\mathbf{X}|$, and an A-optimal design minimizes $trace(\mathbf{X}'\mathbf{X})^{-1}$. For $\mathbf{X}_d$, we focus on two-level regular orthogonal designs in this article. The underlying model $\mathbf{f}$ considered here contains mostly main effects and two-factor interactions.

Among the most commonly used criteria are maximum resolution (Box and Hunter, 1961), minimum aberration (MA) (Fries and Hunter, 1980), and clear effects (Wu and Chen, 2002). These criteria are based on the effect hierarchy principle (Wu and Hamada, 2000), according to which a lower-order effect is more likely to be important than a higher-order

* Corresponding author.
E-mail address: daniel_zhouqi@hotmail.com (Q. Zhou).

http://dx.doi.org/10.1016/j.jspi.2015.02.005
0378-3758/

Table 1
WLP and I-WLPs of the 2^{6-2} design in Section 1.

Generating relations	WLP ($A_3\,A_4\,A_5\,A_6$)	Columns γ	Individual WLP ($A_3(\gamma)\,A_4(\gamma)\,A_5(\gamma)\,A_6(\gamma)$)
$E = AB, F = ACD.$	(1 1 1 0)	A	(1 1 0 0)
		B	(1 0 1 0)
		C	(0 1 1 0)
		D	(0 1 1 0)
		E	(1 0 1 0)
		F	(0 1 1 0)

one, and effects of the same order are equally important. These criteria are geared to the situation where all factors in the experiment are of equal importance. In many practical situations, however, the experimenters may have prior knowledge about the relative importance of the factors. Consequently, the effects of the same order may *not* be of equal importance. An interesting research question then arises: How can we assign factors to the columns of a chosen design in an optimal way? Intuitively, it is preferable to allocate columns having better aliasing properties to the factors that are more likely to be important.

The following example motivates the idea. Suppose a practitioner has decided to use a 2^{6-2} design, defined by $E = AB$ and $F = ACD$, to study six 2-level factors. From the prior knowledge, he/she knew that one factor was particularly important, and wanted to assign this factor to a column such that its aliasing with effects involving other factors is minimized. It can be easily seen from the defining relation, $I = ABE = ACDF = BCDEF$, that the six columns are not created equal. Columns A, B, and E appear in a length-3 word (ABE), implying that the effect of the factor assigning to them will be aliased with a two-factor interaction. In comparison, the effect of the factor assigning to columns C, D or F will be aliased with a three-factor interaction. By the hierarchy principle, the important factor should be assigned to columns C, D, or F.

The rest of this paper is organized as follows. In Section 2 we develop a new criterion called the individual word length pattern (I-WLP) that measures the level of aliasing between an individual factor and effects involving other factors. In Section 3, we obtain and tabulate I-WLPs of columns of all commonly used 16-run and selected 32-run fractional factorial designs. We also propose effective designs when the focus is on one or two particularly important factors. Concluding remarks are given in Section 4.

2. Individual word-length pattern

Consider a 2^{k-p} regular fractional factorial design d defined by p generators. There are $2^p - 1$ words in the defining relation, and the number of letters of a word is called its word length. We consider designs with at least resolution III, whose word length pattern can be denoted as $W(d) = (A_3, A_4, \ldots, A_k)$, where A_i is the number of length-i words in the defining relation.

For a column $\gamma \in d$, we use $A_i(\gamma)$ to denote the number of length-i words involving γ in the defining relation of d. We define the vector

$$W(d, \gamma) = \big(A_3(\gamma), A_4(\gamma), \ldots, A_k(\gamma)\big), \quad \text{for } \gamma \in d$$

and call it the *individual word-length pattern* (I-WLP) of γ in the design. Two columns $\gamma_1, \gamma_2 \in d$ are said to have the same I-WLP if $A_t(\gamma_1) = A_t(\gamma_2)$ for all $t = 3, \ldots, k$. On the other hand, if t is the first number such that $A_t(\gamma_1) \neq A_t(\gamma_2)$ and $A_t(\gamma_1) < A_t(\gamma_2)$, then γ_1 is said to have better I-WLP than γ_2.

In Table 1 we show the WLP and I-WLPs of the 2^{6-2} design described in the introduction section. Clearly, columns C, D and F have the best I-WLP, while column A has the worst I-WLP in this design. So if one factor is considered particularly important, then it can be assigned to C, D or F.

3. Column ranking and effective designs based on I-WLP

We obtained and tabulated the I-WLPs of all non-isomorphic 16-run and selected 32-run regular orthogonal designs, using the catalogs of Sun et al. (2008). In Tables 2 and 3, the superscript "MA" indicates the minimum aberration design for given numbers of factors k and generators p; and superscripts "1" and "2" indicate the effective designs when one or two factors are of particular importance, respectively. These tables can serve two purposes: (1) For a given design, we can easily find the I-WLP of each column from the table, and thus can assign more important factors to columns having better I-WLPs; (2) If the focus is on one or two particularly important factors, the effective designs for this objective are noted in the table. Consider, for instance, a study in which a 16-run design is used for studying seven factors. The MA design $7 - 3.3$ has I-WLP $= (0, 4, 0, 0, 0)$ for all factors. This design is appropriate if all factors are of equal importance. However, if one factor was considered to be particularly important, then Table 2 shows that design 7–3.1 should be used, with the important factor being assigned to column D, which has I-WLP $= (0, 0, 0, 0, 0)$.