ELSEVIER

Statistics & Probability Letters 28 (1996) 203–209

STATISTICS & PROBABILITY LETTERS

A measurement of multi-factor orthogonality

Lih-Yuan Deng [a], Dennis K.J. Lin [b,*], Jiannong Wang [c]

[a] Department of Mathematical Sciences, University of Memphis, Memphis, TN 38152, USA
[b] Department of Management Science, Penn State University, PA 16802, USA
[c] American Express, New York, NY 10285, USA

Received April 1995

Abstract

Orthogonality has been considered as an important design priority in the design literature. This article provides a set of criteria to measure "how orthogonal" a design may be when it is not perfectly orthogonal. These criteria can be obviously applied to a wide variety of designs. Properties of these criteria are discussed, using some existing supersaturated designs as examples.

Keywords: A criterion; B criteria; D criterion; Supersaturated design

1. Introduction

Screening experiments typically contain a large number of potential factors. Among them, only a few are believed to be active. The goal here is to identify those (relatively) few dominant active factors. A first-order model is tentatively assumed. When a small number of runs is desirable, the abandonment of orthogonality is sometimes inevitable. Since a lack of orthogonality results in lower efficiency (Box, 1959), it is always desirable to make the design as nearly orthogonal as possible when perfect orthogonality is unattainable.

One simple way to measure the degree of orthogonality between any two given design columns $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ is to consider their inner product, $s_{ij} = \boldsymbol{x}_i'\boldsymbol{x}_j$; larger $|s_{ij}|$ implies less orthogonality ($s_{ij} = 0$ implies perfect orthogonality). The measurement of orthogonality for more than two design columns is less obvious. For a factorial design with n observations and k factors with $k > n-1$, Booth and Cox (1962) and Lin (1993a) used $E(s^2) = \sum s_{ij}^2/\binom{k}{2}$, the average of all s_{ij}^2 pairs, as a measurement of the design orthogonality.

Note that once the few dominant active factors are identified, the initial design is then projected into a much smaller dimension. The implicit assumption with $E(s^2)$ is that there are, at most, two active factors. If the number of active factors, c, is larger than 2, there is no guarantee that the projective (reduced) design will be of full rank, i.e., a main effect model consists only of those active factors that may not be estimable.

The simple truth, due to the fact that s_{ij} is not transitive, is that while any two columns are highly near orthogonal, several ($c \geqslant 3$) columns may result in low orthogonality. A measurement for multi-factor

* Corresponding author.

0167-7152/96/$12.00 © 1996 Elsevier Science B.V. All rights reserved
SSDI 0167-7152(95)00125-5

orthogonality is thus desired. As pointed out by Lin (1993b), the "optimality" for a screening design should focus on its projective design, not the initial design. Supplemented by such a projection property, a new class of B criteria are proposed here. We first show how the classical design optimality can be extended as a competing alternative.

2. Extension of classical criteria to measure orthogonality

Let X be the design matrix with entries ± 1, and let c be the number of active factors, i.e., the number of design columns of the projected design matrix. For a given $s = (l_1, \ldots, l_c)$, a set of size c from $(1, \ldots, k)$, we can construct a $n \times c$ sub-matrix X_s from X. Following the idea of $E(s^2)$ which gives an efficiency measurement in an average sense, we can measure the "orthogonality" of X as follows:

$$V_c(X) = \frac{1}{\binom{k}{c}} \cdot \sum v(X_s), \tag{2.1}$$

where $v(X_s)$ is a function to measure the "orthogonality" of X_s and the summation is taken over all possible choice of s.

As an extension of the classical design optimality, some natural choices of $v(X_s)$ are:

(1) $v(X_s) = \det(X_s'X_s)^{-1}$.

(2) $v(X_s) = \operatorname{trace}(X_s'X_s)^{-1}$.

(3) $v(X_s) = \lambda_{(c)}(X_s'X_s)^{-1}$, where $\lambda_{(c)}$ denotes the largest eigenvalue of the matrix $(X_s'X_s)^{-1}$.

Note. (1) When $c = k \leqslant n$, these criteria are corresponding to (1) D optimal, (2) A optimal, and (3) E optimal criteria, respectively. When $k > n$ (see Section 5), the value of c cannot be larger than k, and in fact, is normally much smaller than k.

(2) When $c = 2$, all criteria are reduced to being similar to that proposed by Booth and Cox (1962) and Lin (1993a). One should note that the first two criteria will optimize a design by minimizing

$$E(1/(n^2 - s^2)), \tag{2.2}$$

whereas the criterion considered by Booth and Cox (1962) and Lin (1993a) is

$$E(s^2). \tag{2.3}$$

Note also that

$$\frac{1}{n^2 - s^2} = \frac{1}{n^2}\frac{1}{1 - s^2/n^2} = \frac{1}{n^2}\left(1 + \frac{s^2}{n^2} + \cdots\right).$$

Therefore, two criteria in (2.2) and (2.3) should be approximately equivalent to each other because s^2 is normally much smaller than n^2.

3. A new class of B-optimal criteria

Clearly, if a vector y is orthogonal to a group of vectors $Z = (z_1, z_2, \ldots, z_p)$, then the regression sum squares must be null, when regressing y on $(z_1, z_2, \ldots, z_p)$, i.e., $y'Z(Z'Z)^{-1}Z'y = 0$. Thus the value of $y'Z(Z'Z)^{-1}Z'y$, or equivalently $b'(Z'Z)b$, where $b = (Z'Z)^{-1}Z'y$ is the regression coefficient, provides a good measurement on how orthogonal the vector y to Z. Motivated by this, the proposed criteria, called "B optimality," reflects the dependence of a column to all other $c - 1$ columns by computing the regression coefficients of one column in X_s, x_i, over the remaining columns X_{s-i}. For any specific projection design X_s with size $n \times c$, we average $x_i' X_{s-i}(X_{s-i}' X_{s-i})^{-1} X_{s-i}' x_i$ over all possible i ($i = 1, 2, \ldots, c$) as a measurement of the design orthogonality. Of course, the value of c is typically small (see, e.g., Lin, 1993b).

In general, consider a class of new functions $v_g(X_s)$ to measure the "orthogonality" of X_s for $V_c(X)$ in (2.1):

$$v_g(X_s) = \sum_{i \in s} \beta_{s-i}' (X_{s-i}' X_{s-i})^g \beta_{s-i}, \tag{3.1}$$

namely,

$$V_c(X) = \frac{1}{\binom{k}{c}} \cdot \sum_{|s|=c} \sum_{i \in s} \beta_{s-i}' (X_{s-i}' X_{s-i})^g \beta_{s-i},$$

where

(1) $\beta_{s-i} = (X_{s-i}' X_{s-i})^{-1} X_{s-i}' x_i$,

(2) x_i is the $n \times 1$ column corresponding to the ith unit in s,

(3) X_{s-i} is the $n \times (c-1)$ matrix corresponding to units in $s - \{i\}$,

and g can be any scalar value to present the degree of penalty to the near-singularity of the $X_{s-i}' X_{s-i}$ matrix. In principle, the B-criteria can apply to any design when the projection property is under concern, regardless the number of levels, the number of factors and the number of runs. Three cases $g = 2, 1, 0$ deserve special mentioning:

(1) When $g = 2$ (B_2 criterion), we have

$$v_2(X_s) = \sum_{i \in s} (x_i' X_{s-i})(X_{s-i}' x_i) = \sum_{i \in s} \sum_{j \neq i} s_{ij}^2 \tag{3.2}$$

where $s_{ij} = x_i' x_j$. Thus, for $g = 2$ and any value of c, $v_2(X_s)$ is closely related to the $E(s^2)$ criterion considered by Booth and Cox (1962) and Lin (1993a).

(2) When $g = 1$ (B_1 criterion), we can see that

$$v_1(X_s) = \sum_{i \in s} \beta_{s-i}' X_{s-i}' X_{s-i} \beta_{s-i} = \sum_{i \in s} \hat{x}_i' \hat{x}_i = \sum_{i \in s} x_i' P_{s-i} x_i, \tag{3.3}$$

where

$$P_{s-i} = X_{s-i}(X_{s-i}' X_{s-i})^{-1} X_{s-i}' \tag{3.4}$$

is the projection matrix and $\hat{x}_i = X_{s-i}\beta_{s-i} = P_{s-i} x_i$ is the predicted value of x_i with all remaining columns in the design matrix.

(3) When $g = 0$ (B_0 criterion), $v_0(X_s)$ is the unweighted sum of squares of the regression coefficient β_{s-i}. Intuitively, if the matrix $X_{s-i}' X_{s-i}$ is nearly singular, the magnitude of the regression coefficients should be large. In that case, putting $g = 1$ in $v_g(X_s)$ will reduce the "penalty" of near-singularity of $X_{s-i}' X_{s-i}$. In contrast, $v_0(X_s)$, which treats all regression coefficients with equal weights, will impose a greater penalty for the inclusion of near-singular sub-matrices. It is easy to see that

$$v_0(X_s) = \sum_{i \in s} x_i' Q_{s-i} x_i,$$

where

$$Q_{s-i} = X_{s-i}(X_{s-i}' X_{s-i})^{-2} X_{s-i}'.$$

Another advantage of $v_0(X_s)$ is that it is scale invariant.

(4) For a $g < 0$, we will penalize even more for including a nearly singular sub-matrix $X'_{s-i}X_{s-i}$. The general formula is given as

$$v_g(X_s) = \sum_{i \in s} (x_i' X_{s-i})(X'_{s-i}X_{s-i})^{g-2}(X'_{s-i}x_i).$$

In general, the experimenter can use any value of g to compare designs. Of course, the smaller the g value is, the more penalty to the near-singularity of projective designs. The near-singularity (multicollinearity) problem has also received a great deal of attention in linear regression models. One popular measurement for multicollinearity is *variance inflation factors* (VIF) criterion introduced by Marquardt (1970).

The VIF are the main diagonal elements of the $X'X$ matrix in correlation form. They can also be defined as

$$\mathrm{VIF}_j = 1/(1 - R_j^2)$$

where R_j^2 is the coefficient of multiple determination obtained from regressing x_j, on the other regressor variables, when the X matrix under study is of full rank. As pointed out by one referee, this VIF criterion is closely related to the B_1 ($g = 1$) criterion for certain small values of c. Certainly, the VIF was obtained via the normalized $X'X$ matrix (i.e., in the correlation form). This can also be applied to all B criteria in general. For our major interest in two-level designs, the diference is only marginal, however. On the other hand, the B criteria may be useful measurements for multicollinearity. Further research in this direction is encouraging.

4. Some properties of the B-criteria

To compute the value of $v_g(X_s)$ given in (3.1), we need the value of $(X'_{s-i}X_{s-i})^{-1}$ for each $i \in s$. Now, $(X'_{s-i}X_{s-i})^{-1}$ easily can be computed from $(X_s'X_s)^{-1}$. Theorem 1 below gives the formula for $v_1(X_s)$ and $v_0(X_s)$ in terms of the elements in $(X_s'X_s)^{-1}$. Throughout this section, we will consider, without loss of generality, the sample $s = (l_1, \ldots, l_c)$ as $s = (1, \ldots, c)$.

Theorem 1. *Let $X_s = (x_1, \ldots, x_c)$ be a matrix with full rank of dimension $n \times c$ and let*

$$W = (X_s'X_s)^{-1} = (w_{ij}).$$

Then (1) $v_1(X_s) = nc - \sum_{i=1}^{c} 1/w_{ii}$ *and* (2) $v_0(X_s) = \sum_{i \neq j}^{c} w_{ij}^2/w_{ii}^2$.

Proof. Let S_i be a $c \times c$ matrix representing a shift operation that will move the ith column to the last column of the matrix X_s. That is, $X_sS_i = (X_{s-i}, x_i)$. Clearly, S_i is an orthogonal matrix, i.e. $S_i'S_i = I$.

The inverse of

$$M = \begin{pmatrix} X'_{s-i}X_{s-i} & X'_{s-i}x_i \\ x_i'X_{s-i} & x_i'x_i \end{pmatrix} = \begin{pmatrix} X'_{s-i} \\ x_i' \end{pmatrix} (X_{s-i}, x_i) = S_i'X_s'X_sS_i$$

is

$$M^{-1} = S_i'(X_s'X_s)^{-1}S_i = S_i'WS_i = \begin{pmatrix} m_{11} & m_{12} \\ m_{12}' & w_{ii} \end{pmatrix}, \tag{4.1}$$

where

$$\begin{aligned} m_{12} &= (w_{i1}, w_{i2}, \ldots, w_{i,i-1}, w_{i,i+1}, \ldots, w_{ic})' \\ &= Q^{-1}(X'_{s-i}X_{s-i})^{-1}X'_{s-i}x_i. \end{aligned} \tag{4.2}$$

Applying Eq. (7) in Draper and Smith (1981, p. 127) with $\boldsymbol{A} = \boldsymbol{X}'_{s-i}\boldsymbol{X}_{s-i}$, $\boldsymbol{B} = \boldsymbol{C}' = \boldsymbol{X}'_{s-i}\boldsymbol{x}_i$ and $\boldsymbol{D} = \boldsymbol{x}'_i\boldsymbol{x}_i$, we can see the (c,c) element of $\boldsymbol{M}^{-1}$ is

$$\boldsymbol{Q}^{-1} = 1/(\boldsymbol{x}'_i\boldsymbol{x}_i - \boldsymbol{x}'_i\boldsymbol{X}_{s-i}(\boldsymbol{X}'_{s-i}\boldsymbol{X}_{s-i})^{-1}\boldsymbol{X}'_{s-i}\boldsymbol{x}_i).$$

Part (1) now follows because

$$\begin{aligned}\frac{1}{w_{ii}} = \boldsymbol{Q} &= \boldsymbol{x}'_i\boldsymbol{x}_i - \boldsymbol{x}'_i\boldsymbol{X}_{s-i}(\boldsymbol{X}'_{s-i}\boldsymbol{X}_{s-i})^{-1}\boldsymbol{X}'_{s-i}\boldsymbol{x}_i \\ &= n - \beta'_{s-i}\boldsymbol{X}'_{s-i}\boldsymbol{X}_{s-i}\beta_{s-i} \end{aligned} \tag{4.3}$$

and

$$v_1(\boldsymbol{X}_s) = \sum_{i=1}^{c} \beta'_{s-i}\boldsymbol{X}'_{s-i}\boldsymbol{X}_{s-i}\beta_{s-i} = \sum_{i=1}^{c}(n - w_{ii}^{-1}).$$

To prove Part (2), we rewrite (4.2) as

$$\beta_{s-i} = \boldsymbol{m}_{12}\boldsymbol{Q} = \boldsymbol{m}_{12}(1/w_{ii}).$$

Therefore,

$$v_0(\boldsymbol{X}_s) = \sum_{i=1}^{c}\beta'_{s-i}\beta_{s-i} = \sum_{i=1}^{c}\frac{1}{w_{ii}^2}\boldsymbol{m}'_{12}\boldsymbol{m}_{12} = \sum_{i=1}^{c}\frac{1}{w_{ii}^2}\sum_{j\neq i} w_{ij}^2.$$

This proves Part (2). □

As we can see in Theorem 1, $v_1(\boldsymbol{X}_s) = nc - \sum_{i=1}^{c}(1/w_{ii})$ is closely related to the classical A-optimal criterion where $v(\boldsymbol{X}_s) = \mathrm{trace}(\boldsymbol{X}'_s\boldsymbol{X}_s)^{-1} = \sum_{i=1}^{c} w_{ii}$. Using Theorem 1, we can also express the classical A-optimal and D-optimal criteria in another forms as stated in Theorem 2.

Theorem 2. *Let $\boldsymbol{X}_s = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_c)$ be a matrix with full rank of dimension $n \times c$.*

(1) *The* A-*optimal criterion can be written as*

$$trace(\boldsymbol{X}'_s\boldsymbol{X}_s)^{-1} = \sum_{i=1}^{c}\frac{1}{\boldsymbol{e}'_i\boldsymbol{e}_i},$$

where

$$\boldsymbol{e}_i = \boldsymbol{x}_i - \hat{\boldsymbol{x}}_i = (\boldsymbol{I} - \boldsymbol{P}_{s-i})\boldsymbol{x}_i$$

is the residual vector of the ith column vector $\boldsymbol{x}_i$ over the remaining column vectors in $\boldsymbol{X}_s$.

(2) *The* D-*optimal criterion can be written as*

$$\det(\boldsymbol{X}'_s\boldsymbol{X}_s)^{-1} = \prod_{i=1}^{c}\frac{1}{\boldsymbol{d}'_i\boldsymbol{d}_i},$$

where

$$\boldsymbol{d}_i = \boldsymbol{d}(\boldsymbol{x}_i|\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{i-1}) = (\boldsymbol{I} - \boldsymbol{X}_{i-1}(\boldsymbol{X}'_{i-1}\boldsymbol{x}_{i-1})^{-1}\boldsymbol{X}'_{i-1})\boldsymbol{x}_i$$

Table 1
Comparisons on $(n,k) = (12,20)$ supersaturated design

c	Criterion	Booth and Cox	Lin
2	D	1.0757	1.0533
	A	0.1793	0.1755
	B_2	19.3684	13.6421
	B_1	1.6140	1.1368
	B_0	0.1345	0.0947
3	D	0.1050	0.0983
	A	0.2927	0.2801
	B_2	58.1053	40.9263
	B_1	4.9567	3.5801
	B_0	0.4556	0.3301
4	D	—	0.0099
	A	—	0.4025
	B_2	116.2105	81.8526
	B_1	10.1541	7.5520
	B_0	1.0433	0.7791
5	D	—	0.0011
	A	—	0.5523
	B_2	193.6842	136.4211
	B_1	—	13.2535
	B_0	—	1.5631

is the residual vector of the ith column vector $\boldsymbol{x}_i$ *over the column vectors in* $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{i-1}$, *for* $i \geqslant 2$; $\boldsymbol{d}(\boldsymbol{x}_i|\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{i-1}) = \boldsymbol{x}_i$, *for* $i = 1$ *and the matrix* $\boldsymbol{X}_{i-1}$ *is*

$$\boldsymbol{X}_{i-1} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{i-1}).$$

Proof. Part (1) follows easily from (4.3) derived in Theorem 1 and the fact that $\boldsymbol{P}_{s-i}$ is a projection matrix. Part (2) can be easily proved by using Eq. (9) in Draper and Smith (1981, p. 127). □

If the column vectors are orthogonal to each other, then the residual vector should remain "large". Thus, we offer another interpretation of the classical A criterion using the residual vectors $\boldsymbol{e}_i$. Similarly, we also present an interpretation of the classical D criterion using the residual vectors $\boldsymbol{d}(\boldsymbol{x}_i|\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{i-1})$.

5. Application to existing supersaturated designs

Recently, much attention has been focused on constructing systematic supersaturated designs. When such a design is used, the abandonment of orthogonality is inevitable. Thus the degree of orthogonality is clearly a major issue to be addressed. The classical optimal design criteria can not be used because the full design matrix is not of full rank here. We use the case $(n,k) = (12,20)$ as an example to illustrate and compare all the criteria discussed above (D, A and various B criteria). Supersaturated designs proposed by Booth and Cox (1962) and Lin (1993a) are considered here.

Table 1 summarizes the results. Whenever certain projective designs result in a singular matrix, this is indicated by "—" in the table. Of course, this is not desirable. As we can see, Lin (1993a) designs are clearly superior to the Booth and Cox (1962) design, judging by all criteria.

Note that the D criterion, which averages determinants of all possible projective design matrices, becomes smaller and smaller (i.e., higher orthogonality), as c increases. This contradicts to the fact that adding columns will only reduce the orthogonality, when the run size n is fixed. Such an unappealing property, however, does not appear in other criteria. The B_2 and B_1 criteria, on the other hand, are in a much larger scale to distinguish two designs and they are consistent with B_0 criterion.

Acknowledgements

Dennis Lin is partially supported by a Visiting Scientist Fellowship via IBM Watson Research Center, and National Science Foundation Grant DMS-9204007.

References

Box, G.E.P. (1959), Discussion on 'random balance designs', *Technometrics* **1**, 174–180.

Booth, K.H.V. and D.R Cox (1962), Some systematic supersaturated designs, *Technometrics* **4**, 489–495.

Draper N. and H. Smith (1981), *Applied Regression Analysis* (Wiley, New York, 2nd ed.).

Lin, D.K.J. (1993a), A new class of supersaturated designs, *Technometrics* **35**, 28–31.

Lin, D.K.J. (1993b), Another look at first-order saturated design: the p-efficient designs, *Technometrics* **35**, 284–292.

Marquardt, D.W. (1970), Generalized inversed, ridge regression, biased linear estimation and nonlinear estimation, *Technometrics* **12**, 591–612.