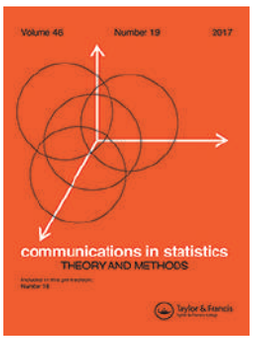

Communications in Statistics - Theory and Methods

ISSN: 0361-0926 (Print) 1532-415X (Online) Journal homepage: https://www.tandfonline.com/loi/lsta20

Nearly optimal covariate designs—Part I

Ganesh Dutta & Bikas Kumar Sinha

To cite this article: Ganesh Dutta & Bikas Kumar Sinha (2017) Nearly optimal covariate designs—Part I, Communications in Statistics - Theory and Methods, 46:19, 9691-9702, DOI: 10.1080/03610926.2016.1217020

To link to this article: https://doi.org/10.1080/03610926.2016.1217020

Accepted author version posted online: 20 Sep 2016.
Published online: 23 Jun 2017.

COMMUNICATIONS IN STATISTICS—THEORY AND METHODS
2017, VOL. 46, NO. 19, 9691–9702
https://doi.org/10.1080/03610926.2016.1217020

Taylor & Francis
Taylor & Francis Group

Nearly optimal covariate designs—Part I

Ganesh Dutta[a] and Bikas Kumar Sinha[b]

[a]Basanti Devi College, Rash Behari Avenue, Kolkata, West Bengal, India; [b]Indian Statistical Institute, Kolkata, India

ABSTRACT

We propose to discuss at length several examples from standard text books. All of these examples deal with analysis of covariance (ANCOVA) models and related analyses of data. We intend to capitalize on our understanding of optimal covariate designs (OCDs) in different ANCOVA models and re-visit these examples with a view to suggest optimal/nearly optimal designs for estimation of the covariate parameter(s). As we will see, for some examples our task is very much routine but for others, it is indeed a highly non trivial exercise.

We intent to cover a total of six examples—divided in two parts. This is Part I—dealing with two examples.

ARTICLE HISTORY
Received 1 March 2016
Accepted 21 July 2016

KEYWORDS AND PHRASES
Analysis of covariance (ANCOVA); analysis of variance (ANOVA); covariate parameter(s); Hadamard matrices; nearly optimal designs; optimal covariate designs (OCDs).

MATHEMATICS SUBJECT CLASSIFICATION
Primary 62K05; Secondary 62K10

1. Introduction and preliminaries

Most standard text books in the area of linear models and design of experiments provide discussions on what are known as analysis of covariance (ANCOVA) models applied to completely randomized designs (CRDs), randomized block designs (RBDs), and latin square designs (LSDs). It is well-accepted practice in experimental design contexts to use one or more available and meaningful covariates together with *local control* to reduce the experimental error. Such a model comprises of three components: local control parameter(s) (if any), "treatment" parameters, and the covariate parameter(s), apart from the error. This generates a family of "covariate models"—serving as a "blend" of "regression models" (in the absence of treatment parameters) and "varietal design models" (in the absence of covariates). These are the so-called ANCOVA models. Generally, for such models, emphasis is given on analysis of the data. Inference-related procedures are fairly routine exercises and are well discussed in the texts.

Let the following covariate model be considered:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\theta} + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{e} \tag{1}$$

where $\mathbf{Y}^{n\times 1}$ denotes the observation vector, $\mathbf{X}^{n\times p}$ denotes the coefficient matrix for the analysis of variance (ANOVA) effects parameters $\boldsymbol{\theta}' = (\theta_1, \theta_2, \ldots, \theta_p)$ and $\mathbf{Z}^{n\times c} = (z_{ij})$ denotes the matrix of the values given to c covariates viz., $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_c)$. In the above, $\mathbf{Z}$ is also called the covariate design matrix of the vector of covariate effects $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_c)'$. As usual, $\mathbf{e}$ is the random error component with $\mathrm{E}(\mathbf{e}) = \mathbf{0}$, $\mathrm{Disp}(\mathbf{e}) = \sigma^2\mathbf{I}_n$, where $\mathbf{I}_n$ is the identity

CONTACT Bikas Kumar Sinha bikassinha1946@gmail.com Retired Faculty, Indian Statistical Institute, Kolkata 700108, India.

matrix of order n. We represent the above set-up by the triplet:

$$\left(\mathbf{Y}, \mathbf{X}\boldsymbol{\theta} + \mathbf{Z}\boldsymbol{\gamma}, \sigma^2\mathbf{I}_n\right) \tag{2}$$

Here the observations are uncorrelated and variances of each of the observations are equal to σ^2. This is to note that the information matrix with respect to model (2) under design d is given by $\sigma^{-2}\mathbf{I}_d(\boldsymbol{\eta})$, where

$$\mathbf{I}_d(\boldsymbol{\eta}) = \begin{pmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} \end{pmatrix} \tag{3}$$

and $\boldsymbol{\eta}' = (\boldsymbol{\theta}', \boldsymbol{\gamma}')$. The information matrix of $\boldsymbol{\gamma}$ is given by

$$\sigma^{-2}I_d(\boldsymbol{\gamma}) = \mathbf{Z}'\mathbf{Z} - \mathbf{Z}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Z} \tag{4}$$

where $(\mathbf{X}'\mathbf{X})^{-}$ is a generalized inverse of $\mathbf{X}'\mathbf{X}$ satisfying

$$\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X} = \mathbf{X}'\mathbf{X}$$

(cf. Rao, 1973, p. 24). It is evident that $\mathbf{Z}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Z}$ is a positive semi-definite matrix.

Therefore, from (4), it follows that

$$\sigma^{-2}I_d(\boldsymbol{\gamma}) \leq \mathbf{Z}'\mathbf{Z} \tag{5}$$

in the Loewner order sense (vide Pukelsheim, 1993) where for two non negative definite matrices $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{A}$ is said to dominate $\mathbf{B}$ in the Lowener order sense if $\mathbf{A} - \mathbf{B}$ is a non negative definite matrix.

Equality in (5) is attained whenever

$$\mathbf{X}'\mathbf{Z} = \mathbf{0} \tag{6}$$

If $\mathbf{Z}$ satisfies (6), then ANOVA effects and covariate effects are orthogonally estimated. Again under condition (6), the information matrix $\mathbf{I}_d(\boldsymbol{\gamma})$ reduces to $\mathbf{I}_d(\boldsymbol{\gamma}) = \mathbf{Z}'\mathbf{Z}$. It is also tacitly assumed that all covariate effects parameters are estimable—though the parameters in the other part viz., varietal effects parameters, and block effects parameters are not necessarily all estimable.

For the covariates, it is assumed without loss of generality, the (location-scale)-transformed version: $z_{ij} \in [-1, 1]$; $\forall i,\ j$; (cf. Das et al., 2003).

Then the covariate effects are estimated with the maximum efficiency if and only if

$$\mathbf{Z}'\mathbf{Z} = n\mathbf{I}_c \tag{7}$$

along with (6).

The designs allowing the estimators with the minimum variance are called *globally optimal designs* (cf. Shah and Sinha, 1989, p. 143). Henceforth, we shall only be concerned with such optimal estimation of regression parameters and by optimal covariate design, *to be abbreviated as OCD hereafter*, we shall only mean *globally optimal design*, unless otherwise mentioned.

In this paper, we deal with a number of application areas wherein optimality study in the context of uses of covariates has a natural scope for enhancing the experimental results. In this part, we work on two motivating examples and provide details of the computations. While working on both the parts, we have derived inspiration from Sinha (2009).

Inference-related procedures are fairly routine exercises and are well discussed in the texts. We are all well aware of three basic considerations in the context of design of experiments: randomization, replication, and local control. Randomization is a technique used as much

Table 1. Original allocation of patients based on covariate values (patient serial number, covariate value).

1.	Treatment A:	(P1, 3), (P2, 5), (P3, 6), (P4, 6), (P5, 8), (P6, 10), (P7, 11), (P8, 11), (P9, 14), (P10, 19)
2.	Treatment D:	(P11, 5), (P12, 6), (P13, 6), (P14, 7), (P15, 8), (P16, 8), (P17, 8), (P18, 15), (P19, 18), (P20, 19)
3.	Control F:	(P21, 7), (P22, 9), (P23, 11), (P24, 12), (P25, 12), (P26, 12), (P27, 13), (P28, 16), (P29, 16), (P30, 21)

as feasible and possible to ensure the statistical validity of inference procedures. Unfortunately, however, in most settings, an experimenter is not in a position to apply randomization (and hence to derive benefits out of it) since a satisfactory and valid application of this tool becomes prohibitive. In the context of covariates models, it is well known that the same phenomenon prevails. We depend largely on distributional assumptions such as normality with homoscedastic errors. We will not reiterate this limitation of our study any further.

2. Examples: Eye-openers

At times there lies a (possibly huge) potential for improving the experimental results by suitably classifying/re-classifying the existing experimental units through a study of the associated covariate values or by first suitably choosing the covariate values from a larger lot and then, hopefully, identifying the associated experimental units from a larger pool.

In the following, we cite a motivating example from Snedecor and Cochran (1989, p. 377), suitably presented to explain our point.

Example 1. It relates to leprosy study mentioned in Snedecor and Cochran's book (1989, p. 377). The point we wish to make is that there may be ample scope of improvement in the efficiency of the estimates for the covariates' parameters if we have a "free" hand in the recruitment of the patients and if a "pool" is made available to us. There are thirty patients for the study and there are three drugs (two antibiotics A and D, and one control F) to be compared - each to be applied to ten patients. To explain our point, the basic design is taken to be a CRD. Optimality theory is well established. Vide Troya Lopes (1982) and Das et al. (2003). We display the original treatment-patient allocation design in Table 1. Also we display in Table 2 the optimal scheme of recruitment of the patients in terms of their possession of original pre-treatment score (count of bacilli)—under the supposition that we have a "free choice" of the patients from a conceivably larger pool. It is a routine task to assert that as against the given patients' ad-hoc recruitment scheme in Table 1, the above optimal scheme in Table 2 provides more than 300% gain in efficiency towards estimation of the covariate parameter. Even with the "given" pool of 30 patients as in Table 1, a suitable *re-allocation* of the patients across the three treatments, as indicated in Table 3, would have provided 12% gain in efficiency against the "adhoc" allocation. In the study of regression designs (or covariate designs) involving a quantitative regressor or quantitative covariate, say z^* with its

Table 2. Proposed optimal plan (d_{opt} (say)) for recruitment of patients based on pre-treatment score in actual units (patient serial number, covariate value).

1.	Treatment A:	(P1, 3), (P2, 3), (P3, 3), (P4, 3), (P5, 3), (P6, 21), (P7, 21), (P8, 21), (P9, 21), (P10, 21)
2.	Treatment D:	(P11, 3), (P12, 3), (P13, 3), (P14, 3), (P15, 3), (P16, 21), (P17, 21), (P18, 21), (P19, 21), (P20, 21)
3.	Control F:	(P21, 3), (P22, 3), (P23, 3), (P24, 3), (P25, 3), (P26, 21), (P27, 21), (P28, 21), (P29, 21), (P30, 21)

Table 3. "Improved" allocation of patients based on given covariate values.

1.	Treatment A :	(P1, 3), (P3, 6), (P4, 6), (P22, 9), (P6, 10), (P7, 11), (P24, 12), (P9, 14), (P19, 18), (P10, 19)
2.	Treatment D :	(P2, 5), (P12, 6), (P14, 7), (P5, 8), (P15, 8), (P8, 11), (P25, 12), (P18, 15),(P28,16), (P20, 19)
3.	Control F :	(P11, 5), (P13, 6), (P21, 7), (P16, 8), (P17, 8), (P23, 11), (P26, 12), (P27, 13), (P29, 16), (P30, 21)

natural range of variation as $(a,\ b)$ with $a < b$ or in a given context, data-driven range as $(z^*_{\min}\ ,\ z^*_{\max})$, we may make a linear transformation to $z = \frac{2(z^*-a)}{(b-a)} - 1$ or $z = \frac{2(z^*-z^*_{\min})}{(z^*_{\max}-z^*_{\min})} - 1$. This has two advantages. The range of variation of z is now $[-1, 1]$ and a linear relation in terms of z^* gets transferred to a linear relation in terms of z. The standardized covariate z is unit-free and it applies universally to all covariates with arbitrary units of measurements. It is evident that based on data-driven range, both the extremes "+/−1" are attained by the transformed covariate z. In applications, the code -1 (respectively, $+1$) is to be replaced by $z^*_{\min}$ (respectively, $z^*_{\max}$) which are "3" and "21" in the above example.

We will now carry out the non trivial exercise of arriving at the design indicated in Table 3 as obtained through adequate re-allocation of the covariate values of the given pool of 30 patients as in the given design in Table 1, to be denoted by d_0. For the sake of completeness, we display the allocation of covariate-values over the three treatments as in d_0.

$$\begin{array}{|l|}\hline \text{A}: 3,\ 5,\ 6,\ 6,\ 8,\ 10,\ 11,\ 11,\ 14,\ 19 \\ \hline \text{D}: 5,\ 6,\ 6,\ 7,\ 8,\ 8,\ 8,\ 15,\ 18,\ 19 \\ \hline \text{F}: 7,\ 9,\ 11,\ 12,\ 12,\ 12,\ 13,\ 16,\ 16,\ 21 \\ \hline \end{array} = d_0,\ \text{say}$$

and the Z-scores are:

A :	−1.00	−0.78	−0.67	−0.67	−0.44	−0.22	−0.11	−0.11	0.22	0.78
D :	−0.78	−0.67	−0.67	−0.56	−0.44	−0.44	−0.44	0.33	0.67	0.78
F :	−0.56	−0.33	−0.11	0.00	0.00	0.00	0.11	0.44	0.44	1.00

It follows that, in terms of the Z-scores (i.e., pre-treatment score (count of bacilli) ranging in $[-1, 1]$,

$$\mathbf{I}_{d_0}(\boldsymbol{\eta}) = \begin{pmatrix} 10 & 0 & 0 & -3.0000 \\ 0 & 10 & 0 & -2.2222 \\ 0 & 0 & 10 & 1.0000 \\ -3.0000 & -2.2222 & 1.0000 & 8.8148 \end{pmatrix}$$

Routine computation yields: Information for γ (i.e., covariate parameter), $I_{d_0}(\gamma) = 7.3210$.

Towards an 'improved' allocation, we arrange the data of pre-treatment scores of all the 30 patients in ascending order: 3, 5, 5, 6, 6, 6, 6, 7, 7, 8, 8, 8, 8, 9, 10, 11, 11, 11, 12, 12, 12, 13, 14, 15, 16, 16, 18, 19, 19, 21.

Since $I_{d_0}(\gamma) = 7.3210$ and $I_{d_{opt}}(\gamma) = 30.00$, the optimal scheme (d_0) in Table 2 provides more than 300% gain in efficiency $(\frac{I_{d_{opt}}(\gamma) - I_{d_0}(\gamma)}{I_{d_0}(\gamma)} \times 100\% > 300\%)$ towards estimation of the covariate parameter. However, the design d_{opt} "ignores" the given experimental units with given covariate values and instead embarks on a fresh choice of experimental units with

chosen z-values. This is a hypothetical situation to make us aware of improvements, if we can act well in advance during the time of choice of experimental units.

Keeping the given experimental units with stated z-values, we now want to explore the possibility of improving "information" on the covariate parameter by suitably relocating the experimental units across different treatments. Now by using following algorithms, we make an attempt to search for a design for which the information of γ is increased.

Algorithm 1

Step 1: We conveniently divide ordered Z-scores into three blocks. Block 1 consists of the first nine observations of arranged data, i.e. (3, 5, 5, 6, 6, 6, 6, 7, 7); Block 2 consists of the next 12 observations, i.e. (8, 8, 8, 8, 9, 10, 11, 11, 11, 12, 12, 12); Block 3 consists of the last nine observations (13, 14, 15, 16, 16, 18, 19, 19, 21).

Step 2: In block 1 we allocate the first three observations, i.e. (3, 5, 5) under Treatment A, the next three observations, i.e. (6, 6, 6) under Treatment D and the last three observations, i.e. (6, 7, 7) under Treatment F.

Step 3: In block 2 we allocate the first four observations, i.e. (8, 8, 8, 8) under Treatment D, the next four observations, i.e. (9, 10, 11, 11) under Treatment F and the last four observations, i.e. (11, 12, 12, 12) under Treatment A.

Step 4: In block 3 we allocate the first three observations, i.e. (13, 14, 15) under Treatment F, the next three observations, i.e. (16, 16, 18) under Treatment A and the last three observations, i.e. (19, 19, 21) under Treatment D.

Hence we get the following arrangement:

	Block 1	Block 2	Block 3
A	3, 5, 5	11, 12, 12, 12	16, 16, 18
D	6, 6, 6	8, 8, 8, 8	19, 19, 21
F	6, 7, 7	9, 10, 11, 11	13, 14, 15

$= d_1$, say.

The information of γ from d_1=$I_{d_1}(\gamma) = 8.1852$

Step 5: Start with d_1. Consider the left block, i.e., Block 1. Permute the rows and generate 3! = 6 options for this block, while keeping the middle and the right block intact. Work out $I_d(\gamma)$ for all the 6 options generated from the left block. Identify the best case scenario and hold this intact while passing into the middle block. Here the best design is found to be d_1.

Step 6: For the middle block, i.e., Block 2, we follow a similar step. Here the best design using Step 6 is

	Block 1	Block 2	Block 3
A	3, 5, 5	9, 10, 11, 11	16, 16, 18
D	6, 6, 6	8, 8, 8, 8	19, 19, 21
F	6, 7, 7	11, 12, 12, 12	13, 14, 15

$= d_2$, say.

The information of γ for d_2=$I_{d_2}(\gamma) = 8.2$.

Step 7: For the last block, i.e., Block 3, we again follow similar step. Ultimately we get d_2 as the best design.

We now consider other aspects towards improving d_2.

Algorithm 2

Here we consider three allocations:

(I) (ADF—DFA—FAD—ADF—DFA—FAD—ADF—DFA—FAD—ADF)
(II) (ADF—DFA—FAD—ADF—DFA—FAD—ADF—DFA—FAD—DFA)
(III) (ADF—DFA—FAD—ADF—DFA—FAD—ADF—DFA—FAD—FAD)

and the the designs are, respectively:

	Block 1	Block 2	Block 3
A	3, 6, 7	8, 10, 11, 12	15, 16, 19
D	5, 6, 7	8, 8, 11, 12	13, 18, 19
F	5, 6, 6	8, 9, 11, 12	14, 16, 21

$= d_{(I)}$, say;

	Block 1	Block 2	Block 3
A	3, 6, 7	8, 10, 11, 12	15, 16, 21
D	5, 6, 7	8, 8, 11, 12	13, 18, 19
F	5, 6, 6	8, 9, 11, 12	14, 16, 19

$= d_{(II)}$, say;

	Block 1	Block 2	Block 3
A	3, 6, 7	8, 10, 11, 12	15, 16, 19
D	5, 6, 7	8, 8, 11, 12	13, 18, 21
F	5, 6, 6	8, 9, 11, 12	14, 16, 19

$= d_{(III)}$, say.

For the above three designs, $\mathrm{I}_{d_{(I)}}(\gamma) = 8.2198$, $\mathrm{I}_{d_{(II)}}(\gamma) = 8.2148$ and $\mathrm{I}_{d_{(III)}}(\gamma) = 8.2148$.

Algorithm 3

We may consider another allocation:

(AFD—FDA—DAF—AFD—FDA—FDA—AFD—DAF—FDA—AFD)

and the corresponding design is

	Block 1	Block 2	Block 3
A	3, 6, 7	8, 10, 11, 12	14, 18, 19
D	5, 6, 6	8, 9, 11, 12	13, 16, 21
F	5, 6, 7	8, 8, 11, 12	15, 16, 19

$= d_3$, say.

Here also $I_{d_3}(\gamma) = 8.2198$.

Heuristic search:

	G_1	G_2	G_3
A	3, 6, 6	9, 10, 11, 12	14, 18, 19
D	5, 6, 7	8, 8, 11, 12	15, 16, 19
F	5, 6, 7	8, 8, 11, 12	13, 16, 21

$= d_4$, say.

This yields $I_{d_4}(\gamma) = 8.2198$ and d_3 is equivalent to d_4. Further, these are also equivalent to d_I in the sense of same information.

In the final analysis, we find that there is gain in efficiency (more than 12%) in the performance of the design d_I or d_3, as against the original design d_0. This is the design (d_I or d_3) displayed in Table 3.

The purpose of Example 2 is to indicate a step-by-step procedure towards improved allocation of a covariate design in the set-up of an RBD.

Example 2. We now elaborate on another example taken from Rao (1973, p. 291)—also to be found in Scheffé (1999, p. 217)—suitably modified to suit our discussion. This essentially

refers to an RBD with $b = 5, v = 3$. Here b stands for number of blocks and v stands for number of treatments. In the literature, we find results at length on constructions of optimal OCDs under RBD set-ups but mostly dealing with the "regular" cases viz., both b and v being multiples of 4 so that Hadamard matrices exist (cf. Das et al., 2003; Rao et al., 2003). Here is a notable deviation and we take this rare opportunity to discuss the example quite in details.

This study refers to 15 male and 15 female pigs with their initial weights used as values of a single covariate. The blocks correspond to five pens and treatments correspond to three levels of feeding with increasing proportions of protein. Here is the original allocation design in Table 4.

We consider an alternative representation of the above data to give an insight into the allocation problem (Table 5). Now one may consider the standard covariate model (ANCOVA) for two-way RBD Pen × Treatment layout with a single covariate (initial weights of pigs) separately for females and males.

Under RBD ANCOVA model with a single covariate, the standard expression for information on γ

$$\begin{aligned} I(\gamma) &= \sum_{i=1}^{5}\sum_{j=1}^{3} z_{ij}^2 - \frac{1}{3}\sum_{i=1}^{5} R_i^2 - \frac{1}{5}\sum_{j=1}^{3} C_j^2 + \frac{G^2}{15} \\ &= \sum_{i=1}^{5}\sum_{j=1}^{3} z_{ij}^2 - \frac{1}{3}\sum_{i=1}^{5} R_i^2 - 5\sum_{j=1}^{3} (\bar{z}_{0j} - \bar{z}_{00})^2 \end{aligned} \tag{8}$$

Table 4. Given allocation design.

Pen	Treatment	Sex	Initial weight
	A	F	48
	B	F	48
I	C	F	48
	C	M	48
	B	M	39
	A	M	38
	B	F	32
	C	F	28
II	A	F	32
	C	M	37
	A	M	35
	B	M	38
	C	F	33
	A	F	35
III	B	F	41
	B	M	46
	C	M	42
	A	M	41
	C	F	50
	A	M	48
IV	B	F	46
	A	F	46
	B	M	40
	C	M	42
	B	F	37
	A	F	32
V	C	F	30
	B	M	40
	C	M	40
	A	M	43

Data Source: Rao (1973), p. 291 and Scheffé (1999), p. 217.

Table 5. Initial weight distribution as per allocation of pigs.

	5(a) FEMALE Treatment					5(b) MALE Treatment			
Pen	A	B	C	Totals	Pen	A	B	C	Totals
1	48	48	48	144	1	38	39	48	125
2	32	32	28	92	2	35	38	37	110
3	35	41	33	109	3	41	46	42	129
4	46	46	50	142	4	48	40	42	130
5	32	37	30	99	5	43	40	40	123
Totals	193	204	189	586	Totals	205	203	209	617

Table 6. Female.

	Treatment			
Pen	A	B	C	Totals
1	48	48	48	144
2	32	32	28	92
4	46	46	50	142
3	35	41	33	109
5	32	37	30	99
Totals	193	204	189	586

where R_i is ith row total, C_j is jth column total, G is grand total, and $\bar{z}_{0j} = C_j/5$, $\bar{z}_{00} = G/15$.

The notations are standard and we use γ_F and γ_M to, respectively, denote the covariate effect for female and male pigs. These are routine computations and for the given allocation design in Table 5, to be denoted by d_0, $I_{d_0}(\gamma_F) = 57.8667$ and $I_{d_0}(\gamma_M) = 116.2667$.

Our aim is to maximize the information of γ_F as well as γ_M given in (8) by properly allocating the pigs in the two-way RBD layout for both female and male pigs. We will take up the study for female pigs only and thereby concentrate on data in the relevant table. Note that towards maximization of information, the row totals of the covariate values should be as close as possible and the same is to hold true of the column totals. We start with the 5×3 table of covariate values and proceed through the following steps:

Step 1: First we arrange the rows in three sets where the first set consists of the rows where all the covariate values are equal; in the second set, we consider those rows where two of the three covariate values are not equal and the last set consists of the rows where all the covariate values are unequal. The arrangement is shown in Table 6.

Step 2: We select the first row of second set (i.e., Pen No. 2) and permute the covariate values keeping the other rows fixed. Next we compute the information of γ_F under each permutation. Then we choose the design in which the information of γ_F will be a maximum. We

Table 7. d_{F1}.

	Treatment			
Pen	A	B	C	Totals
1	48	48	48	144
2	32	28	32	92
4	46	46	50	142
3	41	35	33	109
5	30	37	32	99
Totals	197	194	195	586

Table 8. A's in ascending and B's in descending orders.

Treatment			
A	B	C	Totals
30	48	*	78
32	46	*	78
41	37	*	78
46	35	*	81
48	28	*	76

do the same for the second row of the second set (i.e., Pen No. 4) keeping the other rows of the new design intact. Similarly, we do the same for third set also (Pen No. 3 and 5).

Step 3: We repeat Step 2 until all column totals (C_{Fj}'s) are as close as possible to $G_F/3$. Finally, we get the design where the information of γ_F is a maximum with C_{Fj}'s are as close as possible to $G_F/3$. We denote it by d_{F1} and $I_{d_{F1}}(\gamma_F) = 81.0667$

Step 4: We arrange the initial weights under Treatment A in ascending order and the initial weights under Treatment B in descending order. The arrangement is shown in Table 8. Since the sum of the two entries in each of 5 rows are 78, 78, 78, 81, 76, we fill the entries under Treatment C as 48, 33, 32, 32, 50. Then we get the design d_{F2} and here $I_{d_{F2}}(\gamma_F) = 782.4$.

For another option, we arrange the initial weights under Treatment A in ascending order and the initial weights under Treatment C in descending order. The arrangement is shown in Table 10.

Since the sum of the two entries in each of 5 rows are 80, 80, 74, 78, and 80, we fill the entries under Treatment B as 37, 35, 48, 46, and 28. Then we get the design d_{F3} and here $I_{d_{F3}}(\gamma_F) = 817.0667$.

Lastly we arrange the initial weights under Treatment B in ascending order and the initial weights under Treatment C in descending order. The arrangement is shown in Table 12.

Table 9. d_{F2}.

	Treatment			
	A	B	C	Totals
	30	48	48	126
	32	46	33	111
	41	37	32	110
	46	35	32	113
	48	28	50	126
Totals	197	194	195	586

Table 10. A's in ascending and C's in descending orders.

Treatment			
A	B	C	Totals
30	*	50	80
32	*	48	80
41	*	33	74
46	*	32	78
48	*	32	80

Table 11. d_{F3}.

	Treatment			
Pen	A	B	C	Totals
	30	37	50	117
	32	35	48	115
	41	48	33	122
	46	46	32	124
	48	28	32	108
Totals	197	194	195	586

Table 12. B's in ascending and C's in descending orders.

	Treatment			
Pen	A	B	C	Totals
1	*	28	50	78
2	*	35	48	83
4	*	37	33	70
3	*	46	32	78
5	*	48	32	80

Table 13. d_{F4}.

	Treatment			
Pen	A	B	C	Totals
	41	28	50	119
	30	35	48	113
	48	37	33	118
	46	46	32	124
	32	48	32	112
Totals	197	194	195	586

Table 14. d_{F5}.

	Treatment			
Pen	A	B	C	Totals
	41	28	48	117
	30	35	50	115
	48	37	33	118
	46	46	32	124
	32	48	32	112
Totals	197	194	195	586

Table 15. d_{F6}.

	Treatment			
Pen	A	B	C	Totals
	41	28	48	117
	30	37	50	117
	48	35	33	116
	46	46	32	124
	32	48	32	112
Totals	197	194	195	586

Table 16. d_{F7}.

	Treatment			
Pen	A	B	C	Totals
	46	28	48	122
	30	37	50	117
	48	35	33	116
	41	46	32	119
	32	48	32	112
Totals	197	194	195	586

Table 17. d_{F8}.

	Treatment			
Pen	A	B	C	Totals
	46	28	48	122
	30	37	50	117
	48	35	32	115
	41	46	32	119
	32	48	33	113
Totals	197	194	195	586

Since the sum of the two entries in each of 5 rows are 78, 83, 70, 78, and 80, we fill the entries under Treatment A as 41, 30, 48, 46, and 32. Then we get the design d_{F4} and here $I_{d_{F4}}(\gamma_G) = 838.4$.

Now we start with d_{F4} and proceed with Step 1 and Step 2. Then we observe that d_{F4} is a better design. Next we can improve over d_{F4} by interchanging the first element and the second element under Treatment C and denote the design by d_{F5}. Here $I_{d_{F5}}(\gamma_F) = 843.7333$.

Again we can improve d_{F5} by interchanging the second element and the third element under Treatment B and we denote the design by d_{F6}. Here $I_{d_{F6}}(\gamma_F) = 845.0667$.

We can further improve d_{F6} by interchanging the first element and the forth element under Treatment A and denote the design by d_{F7}. Here $I_{d_{F7}}(\gamma_F) = 851.7333$.

Lastly we improve d_{F7} by interchanging the third element and the forth element under Treatment C and denote the design by d_{F8}. Here $I_{d_{F8}}(\gamma_F) = 853.7333$.

Now we construct design d_{F9} by interchanging the third element under Treatment A and the forth element under Treatment B of d_{F8}. Here $I_{d_{F9}}(\gamma_G) = 846.5333$ which is less than $I_{d_{F8}}(\gamma_F)$ even though column sums are more or less equal. We stop here and recommend the design d_{F8} for use. Now we consider the "ideal" situation for female data where the row sums are 117, 117, 117, 117, and 118 and column sums are 195, 195, 196, and $G_F = 586$. In this situation, $I(\gamma_F) = 870.5333$. Therefore, the relative efficiency of $d_{F8} = 98.07\%$ whereas

Table 18 d_{F9}.

	Treatment			
Pen	A	B	C	Totals
	46	28	48	122
	30	37	50	117
	46	35	32	113
	41	48	32	121
	32	48	33	113
Totals	195	196	195	586

the relative efficiency of γ_F under design d_0 is 6.6473%. The gain in efficiency of d_{F8} against d_0 for female pigs is 1375.35%.

The designs [d_{F1}, d_{F2}, d_{F3}, d_{F4}, d_{F5}, d_{F6}, d_{F7}, d_{F8} and d_{F9}] are shown respectively in Tables 7, 9, 11, 13, 14, 15, 16, 17 and 18.

For male pigs, the procedure is similar. Interested readers may work out the details.

3. Concluding remark

In Example 1, we undertook a study of a CRD model involving three treatments and with 15 observations under each. This represents a non regular case and we elaborated on the procedures to enhance the efficiency of covariates designs. In Example 2, we considered 5 blocks and three treatments and this is not amenable to usual study of optimal RBDs as is considered in Troya Lopes (1982) or in Das et al. (2003). We have elaborated the steps involved and settled for nearly optimal covariate designs. We believe these exercises will be instructional to the experimenters in exploring the possibility of improved experimental designs.

Acknowledgments

The authors acknowledge the anonymous referees for their suggestions and valuable comments.

References

Das, K., Mandal, N.K., Sinha, B.K. (2003). Optimal experimental designs with covariates. *J. Stat. Plann. Inference* 115:273–285.

Pukelsheim, F. (1993). *Optimal Design of Experiments*. New York, NY: Wiley.

Rao, C.R. (1973). *Linear Statistical Inference and Its Applications*. Second Edition. USA: John Wiley & Sons, Inc.

Rao, P.S.S.N.V.P., Rao, S.B., Saha, G.M., Sinha, B.K. (2003). Optimal designs for covariates' models and mixed orthogonal arrays. *Electr. Notes Discrete Math.* 15:157–160.

Scheffé, H. (1999). *The Analysis of Variance*. New York: John Wiley & Sons, Inc.

Shah, K.R., Sinha, B.K. (1989). *Theory of Optimal Designs*. Lecture notes in Statistics. Ser. 54. New York, Inc.: Springer-Verlag.

Sinha, B. K. (2009). A reflection on the choice of covariates in the planning of experimental designs. *J. Indian Soc. Agric. Stat.* 63:219–225.

Snedecor, George W., Cochran, William G. (1989). *Statistical Methods*. Eighth Edition. Ames, Iowa: Iowa State University Press.

Troya Lopes, J. (1982). Optimal designs for covariate models. *J. Stat. Plann. Inference* 6:373–419.