International Journal of Statistical Sciences *ISSN 1683–5603*
Vol. 24(1) March, 2024, pp 65-73
DOI: https://doi.org/10.3329/ijss.v24i1.72020

Study of Covariates' Effects in the Presence of Neighbor Effects : An Informative Review

Sobita Sapam[1*], Bikas K. Sinha[2] and KK Singh Meitei[3]
[1,3]Department of Statistics, Manipur University,
Imphal, Manipur, India
[2]Retd. Professor, ISI, Kolkata, India
*Correspondence should be addressed to Sobita Sapam
(Email: sobita1@yahoo.com)

[Received December 29, 2023; Accepted February 20, 2024]

Abstract

With reference to a Gauss-Markov Model, Analysis of Covariance (ANCOVA) is a standard exercise in the study of differential treatment effects in the presence of covariates. Again in the presence of 'Neighbor Effects', we carry out necessary data analysis in a routine manner. In this paper we present a review of this area of research, encompassing both covariates' effects and neighbor effects.

Keywords and Phrases: ANCOVA, Neighbor effects, Optimal covariate designs.

AMS Classification: 62K10.

1 Introduction

During our undergraduate/post-graduate studies, in a course in Linear Models, we are told about ANOVA in considerable details, with illustrative examples encompassing One-way and Two-way classified data. As an extension, we are also told about ANCOVA which deals with non-stochastic covariates, also known as regressors. This is a routine extension of ANOVA.

Neighbor effects, on the other hand, is a concept/term/key word less discussed in text books and it is more confined to research-level expositions. The concept of neighbor designs was introduced by Rees (1967) in the experiment of serology when the blocks are circular in nature, such that the first plot and the last plot in a block are appearing as neighbors. Azais et al (1993) constructed a series of neighbor-balanced designs incomplete blocks and also mentioned about the analysis of different models with one-sided, two-sided neighbor effects. Meitei (1996) gives a method of construction of incomplete block neighbor design in which number of blocks is not a multiple of number of treatments. Jaggi et al (2018) described some construction methods of circular neighbor balanced and also circular partially neighbor balanced block designs for estimation of direct and neighbor effects of the treatments. Kunert (1984), Monod (1992), Bailey (2003), Jaggi et al (2007), Varghese et al (2014), Sapam et al (2019a, 2019b), Sapam and Sinha (2020) are also some related references of neighbor designs.

Troya (1982a, 1982b) introduced the concept of Optimal Covariates Designs (OCDs) and presented optimality results in the context of Completely Randimized Designs (CRDs). Optimal covariate designs are the designs which provide optimal or most efficient estimation of the covariates' effects in terms of the parameters in an assumed linear model. Inspired by Troya's formulation of optimality problems involving covariates effects, Das et al (2003) provided some combinatorial solutions which served as a modest beginning of the Monograph titled 'Optimal Covariate Designs' by Das et al (2015). Sapam et al (2021) worked on a linear ANCOVA model with the key reference of the mentioned Monograph to study its analysis - with special emphasis on the question of estimability of the regression coefficient(s) involving the covariates in the presence of neighbor effects under Randomized Block Designs (RBDs), balanced incomplete block design and latin square design set ups.

Sapam et al (2021) focused on OCDs incorporating the neighbor effects in four directions viz., left-sided, right-sided, top-sided and bottom-sided in the assumed linear model in different RBD set ups. Sinha and Dutta (2017) worked on three different seasons of Latin Square Designs (LSDs) of order four without any neighbor effects. Dutta et al (2014), Shah and Sinha (1989) are also related references on optimal designs.

The standard ANOVA-based analysis models of RBD without covariates as well as neighbor effects is given by Type 1 and by Type 2, RBD with covariates are shown below:
Type 1 model: $y_{ij} = \mu + \alpha_i + \tau_j + e_{ij}$
Type 2 model: $y_{ij} = \mu + \alpha_i + \tau_j + \beta x_{ij} + e_{ij}$

Next, by Type 3 model, RBD without covariates in the presence of neighbor effects [two-sided] and by Type 4 model, RBD with covariates and neighbor effects are as follows:
Type 3 model: $y_{ij} = \mu + \alpha_i + \tau_j + LN_{(i-1)} + RN_{(i+1)} + e_{ij}$
Type 4 model: $y_{ij} = \mu + \alpha_i + \tau_j + \beta x_{ij} + LN_{(i-1)} + RN_{(i+1)} + e_{ij}$
where $LN_{(i-1)} + RN_{(i+1)}$ are left and right-sided neighbor effects of the treatment i, assuming the blocks are circular.

1.1 Covariates' Designs

We start with a block design as an illustration. We assume the existence of a non-stochastic yet quantitative covariate (X) attached to each experimental unit. Specifically, the model assumes the form:

$$y_{ij} = \mu + \alpha_i + \tau_j + \beta x_{ij} + e_{ij} \tag{1}$$

where y_{ij} is the observation in the experimental unit corresponding to i-th block and j-th treatment; μ is the general mean effect; α_i is the i-th block effect; τ_j is the j-th treatment effect; β is the regression parameter and x_{ij} is the covariate value attached to the experimental unit labeled (i, j) associated with the linear effects parameter β and e_{ij} is the random error effect .
In general terms, for any number of covariates and any experimental design set- up, it transpires that $var(\hat{\beta}) \geq \sigma^2/\sum {x_{i,j}}^2$. Without any loss of generality, we can argue that $-1 \leq x_{i,j} \leq 1$, for all covariate values.

This takes the variance bound to σ^2/n, where n is the total number of observations. We need to examine the case of 'equality' and that too, for each of the covariates and there again, we need to attain 'equality' simultaneously for all the covariates parameters' estimates, in case there are more covariates in the experimental set-up.

The literature, henceforth, is vast and varied. Das et al (2015) gives an account of extensive studies in this area of research, prior to introduction of neighbor effects.

As and when the neighbor effects were introduced to account for the situations wherein the mean models are distorted by such effects, researchers got involved in the study of optimal estimation of covariates' parameters. It transpires that existential results are not necessarily available even in case of nicely behaving block design layouts.
Below we discuss some such available results in both senses (existential/non-existential).

1.2 Some review of literature in the context of OCDs

Optimal covariate designs are widely discussed in literature. Dutta et al (2009) constructed OCDs in different series of PBIBDs through various method of construction viz., singular group divisible design, semi regular group divisible design and regular group divisible design, latin square and triangular design. Further, the model (1) mentioned above may be rewritten as

$$(Y, \mu 1_N + X_1\alpha + X_2\tau + Z\beta, \sigma^2 I) \tag{2}$$

and the necessary and sufficient condition for optimality under the model is given by

$$Z^{'}X_1 = 0,\ Z^{'}X_2 = 0\ and\ Z^{'}Z = nI \tag{3}$$

The optimality conditions in (3) may be expressed, equivalently, in terms of W-matrices introduced in Das et al (2003). W-matrices connect the incident matrices of the treatments across the blocks, row/column effects etc with the covariates' values, assumed to be +/- 1. We omit the details which ensure the properties displayed below:

C_1: Each W -matrix has all treatment sums equal to zero.
C_2: Each W -matrix has all row/block sums equal to zero.
C_3: The grand total of all the entries in the Hadamard product of any two distinct W-matrices reduces to zero.

These three conditions play a very important role in the construction of optimal covariate designs under different complete/ incomplete block design set-ups without any neighbor effects. The monograph of Das et al (2015) shows in detail. Hore et al (2014) proposed a new algorithm for optimal or near optimal allocation where there is large number of experimental units and several covariates. Sinha et al (2014) discussed the formulation of problem of allocation of one controllable covariate in the context of an experiment involving several treatments. Sinha and Rao (2019) also initiated a study involving the issues of optimal allocation of covariate values in the context of 2^n-factorial experiments. Dutta and Sinha (2017a,b) provide in two parts, optimality results regarding nearly optimal covariate designs under an ANCOVA model with the examples of CRD, RBD.

OCDs incorporating the neighbor effects whether, one-sided, two-sided or four sided seems to be rare in the literature. We have studied the OCDs in the presence of neighbor effects in Sapam et al (2021). The monograph by Das et al (2015), have covered in vast under the Type 2 model. Sapam et al (2019a, 2019b, 2020) provided under Type 3 model with different design set-ups. We focus for current research is centered around Type 4.

Specifically, if we are dealing with an RBD involving b blocks and v treatments and if there are k covariates $(X_{(1)}, X_{(2)}, ..., X_{(k)})$, we will attain 'equality' in the variance bound simultaneously for all the covariates if and only if the following conditions are met :

(i)$\sum_j x_{u;(i,j)} = 0$, 1≤i≤b;
(ii) $\sum_i x_{u;(i,j)} = 0$, 1≤j≤v;
(iii) $\sum_{1\leq i\leq b}\sum_{1\leq j\leq v} x_{u;(i,j)}\ x_{u^*;(i,j)} = nI(u,u^*), 1 \leq u, u^* \leq k;$

where, in the above, I(.,.) is the usual indicator function and n = bv, $x_{u;(i,j)}$ is the covariate values attached to the u-th experimental unit labeled in the ith block and jth treatment associated with the linear effects parameter β.

The following two additional conditions (iv)-(v) are also necessary for optimality incorporating the neighbor effects [two-sided] in the model.
(iv) sum of x- values in the positions of all the left-neighbors of a given treatment is zero and that holds for each treatment;
(v) sum of x- values in the positions of all the right-neighbors of a given treatment is zero and that holds for each treatment.

Below we discuss some such available results from Sapam and Sinha (GJSDS, accepted paper) through examples in both senses (existential/non-existential).

In an RBD with v=b=4, there are $(4!)^4$ choices of design layouts - though plenty of them are 'permutation invariant' in subsets. We considered one subset of them viz., 4!=24 RBDs starting with the first three blocks in the natural order of the treatments i.e., treatment-levels 1, 2, 3, 4. Then only the last row [i.e., 4th row] is made to be composed of a permutation of the treatment levels 1, 2, 3, 4 resulting 4! = 24 distinct design layouts. We have checked the existence and non-existence of X-matrices satisfying the conditions (i)-(iv) laid down above. We claim that there are two distinct design layouts among these 24 choices whose X-matrices are in opposite directions. For one of them, we can construct an X-matrix satisfying all the conditions. For another, we argue that there does not exist any such X-matrix satisfying all the conditions laid down above.

Table 1: RBD (v=b=4)

1	2	3	4
1	2	3	4
1	2	3	4
1	2	3	4

For the choice of RBD (v=b=4) in Table 1 we argue that there exist X-matrix satisfying all the four conditions (i)-(iv) given above, one such X- matrix is shown in the Table 2.

Whereas for the choice of RBD (v=b=4) shown in Table 3, we argue that there does not exist any underlying X-matrix satisfying all the four conditions mentioned above. Having understood the existence/non-existence results in the particular frameworks of selected values of b and v, the general case: RBD (b = 2p, v = 2q, p and q being positive integers).

Table 2: RBD (v=b=4)

1	-1	1	-1
-1	1	-1	1
1	-1	1	-1
-1	1	-1	1

Table 3: RBD (v=b=4)

1	2	3	4
1	2	3	4
1	2	3	4
2	1	3	4

Towards the non-existence result, we start with the $RBD(b, v)$ design layout as

$$D_{b\times v} = \begin{bmatrix} 1 & 2 & 3 & \cdots & v \\ 1 & 2 & 3 & \cdots & v \\ \vdots & \vdots & \ddots & \vdots & \\ 1 & 2 & 3 & \cdots & v \\ 2 & 1 & 3 & \cdots & v \end{bmatrix}$$

Further, let

$$X_{b\times v} = [\ ((x_{ij}))\ ]$$

be the usual matrix of associated covariate-values.

And towards existence result, as in the particular cases, we may start with the standard RBD for treatment allocations in the natural order for each block and follow the allocations of the x-values as $+1$'s and -1's alternately as are shown in the particular cases.

In matrix notation, in this case, the solution matrix is represented as the matrix product $H_2 \otimes J_{p\times q}$ which is a succession of the Hadamard matrix of order 2 i.e., of H_2 matrix in every row and column covering the matrix J of order $p \times q$.

$$H_2 \otimes J_{p\times q} = \begin{bmatrix} H_2 & H_2 & \cdots & H_2 \\ H_2 & H_2 & \cdots & H_2 \\ \vdots & \vdots & \ddots & \vdots \\ H_2 & H_2 & \cdots & H_2 \\ H_2 & H_2 & \cdots & H_2 \end{bmatrix}$$

This study was meant to examine robustness [with respect to presence of neighbor effects] of optimal covariates designs in RBD set-ups.

Sapam et al (2021) examined that the existence of 'optimal covariates designs' in the presence of neighbor-effects under the design set -ups (i) RBD (b = v = 4), (ii) BIBD (b =v = 7, r = k = 4, λ= 2) and (iii) LSD of order 4. The model adopted is linear in the general mean, block - effects / row-column effects, treatment effects and circularly located neighbor- effects. The presence of covariates makes the analysis complicated unless their effects are optimally and orthogonally estimated. This study shows that at times we are in a position to achieve this by suitably allocating the covariates values in the experimental units. Even though the experimental set-ups are simple, the results are non-trivial and worth noting.

Further, Sapam and Sinha (accepted in SSCA) studied four non isomorphic LSDs viz., S_1, S_2, S_3, S_4 of order four with and without neighbor effects and summarized that in the absence of neighbor effects, for each of the designs S_1, S_2, S_3 and S_4 of LSD of order 4, we can find out all the possible (six) optimal X-matrices. On the other hand, in all the four LSDs, these optimal matrices fail to be optimal when we incorporate the four sided neighbor effects. Only for the designs S_1 and S_2 all the six optimal X-matrices continue to be so even in the presence of neighbor effects. The other two LSDs S_3 and S_4 has no X-matrix.

Acknowledgment

The first author thanks her Mentor Professor KK Singh Meitei for providing all facilities towards successfully pursuing her research in the broad area of DoE and also for arranging visits of Professor Sinha to the Manipur University for collaborative research. The first author also acknowledges financial support from **DST, Women Scientist Scheme-A, Project Sanction order No. DST/WOS-A/PM-27/2021**.
The authors profusely thank the reviewer for a very critical reading of the manuscript and for making a note of corrections for the manuscript to be clearly readable.

References

[1] Azais, J. M., Bailey, R. A. and Monod, H. (1993). A catalogue of efficient neighbor designs with border plots. Biometrics, 49(4):1252–61. doi:10.2307/2532269.

[2] Bailey, R. A. (2003). Designs for one sided neighbor effects. Journal of Indian Society of Agricultural Statistics, 56, 302-314.

[3] Das, K., Mandal, N. K. and Sinha, B. K. (2003). Optimal experimental designs for models with covariates. Journal of Statistical Planning and Inference, 115, 273-285.

[4] Das, P., Dutta, G., Mandal, N. K. and Sinha, B. K. (2015). Optimal Covariate Designs. Springer Verlag Text Book Series.

[5] Dutta, G., Das, P. and Mandal, N. K. (2009). Optimum covariate designs in partially balanced incomplete block (PBIB) design set-ups. Journal of Statistical Planning and Inference, 139:2823-2835.

[6] Dutta, G., Das, P. and Mandal, N. K. (2014). D-optimal designs for covariate models. Communication in Statistics-Theory and Methods, 43, 165-174.

[7] Dutta, G. and Sinha, B. K. (2017a). Nearly optimal covariate designs- Part I. Communication in Statistics-Theory and Methods, 46(19), 9691-9702.

[8] Dutta, G. and Sinha, B. K. (2017b). Nearly optimal covariate designs- Part II. Communication in Statistics-Theory and Methods, 46(19), 9703-9725.

[9] Jaggi, S., Varghese, C. and Gupta, V. K. (2007). Optimal circular block designs for competition effects. Journal of Applied Statistics, 34, 577-584.

[10] Jaggi, S., Pateria, D. K., Varghese, C., Varghese, E. and Bhowmik, A. (2018). A note on circular neighbor balanced designs. Communication in Statistics - Simulation and Computation, 47, 2896-2905.

[11] Hore, S., Dewanji, A. and Chatterjee, A. (2014). Design issues related to allocation of experimental units with known covariates into two treatment groups. Journal of Statistical Planning and Inference, 155, 117-126.

[12] Kunert, J. (1984). Designs balanced for circular residual effects. Communication in Statistics-Theory and Methods, 13, 2665-2671.

[13] Meitei, K. K. Singh (1996). A series of incomplete block neighbour designs. Sankhya Series B, 58(1), 145-147.

[14] Monod, H. (1992). Two- factor neighbour designs in incomplete blocks for intercropping experiments. The Statistician, 41(5), 487-497.

[15] Rees, D.H. (1967). Some designs of use in serology. Biometrics, 23, 779-791.

[16] Sapam, S., Mandal, N. K. and Sinha, B. K. (2019a). Latin square designs with neighbor effects. Journal of Indian Society of Agricultural Statistics, 73, 91-98.

[17] Sapam, S., Mandal, N. K. and Sinha, B. K. (2019b). Latin square designs with neighbor effects -part II. Communication in Statistics - Theory and Methods, https://doi.org/10.1080/03610926.2019.1702694.

[18] Sapam, S. and Sinha, Bikas K. (2020). On the status of variance balanced block designs in the presence of both-sided neighbor effects: two examples. Statistics and Applications (New Series), 18(2), 15-29.

[19] Sapam, S., Meitei, K. K. S. and Sinha, B. K. (2021). Randomized block designs, balanced incomplete block designs and latin square designs with neighbor effects in the presence of covariates. Statistics and Applications, (New Series), 19, 19-31.

[20] Sapam, S., and Sinha, B. K. On the robustness of LSD layouts in the presence of neighbor effects. Statistics and Applications. [Accepted for publication]

[21] Sapam, S. and Sinha, B. K. On the existence and non-existence of robust optimal covariate designs in the presence of neighbor effects: RBDs. Gujarat Journal of Statistics and Data Sciences. [Accepted for publication]

[22] Shah, K. R. and Sinha, B. K. (1989). Theory of optimal designs. Lecture notes in Statistics. Series 54, Springer, New York.

[23] Sinha, B. K., Mathew, T., Rao, P.S.S.N.V.P. and Rao, S.B. (2014). A new class of optimal designs in the presence of a quantitative covariate. International Journal of Statistical Sciences, 14(1-2), 01-16.

[24] Sinha, B. K. and Dutta, G. (2017). Groups of Latin square designs in agricultural experiments with covariate, RASHI, 2, 01-05.

[25] Sinha, B. K. and Rao, P.S.S.N.V.P. (2019). Some aspects of optimal covariate designs in factorial experiments. Statistics and Applications, 17(2), 01-05.

[26] Troya, L. J. (1982a). Optimal designs for covariates models. Journal of Statistical Planning and Inference, 6, 373-419.

[27] Troya, L. J (1982b). Cyclic designs for covariates models. Journal of Statistical Planning and Inference, 7, 49-75.

[28] Varghese, E., Jaggi, S. and Varghese, C. (2014). Neighbor balanced-row column designs. Communication in Statistics- Theory and Methods, 43, 1261-1276.