RASHI 2 (1) : 01-05 (2017)

Groups of Latin Square Designs in agricultural experiments with covariate

Bikas K Sinha, [1]Ganesh Dutta

ISI, Kolkata (Retired), [1]Basanti Devi Kollege, Kolkata

ABSTRACT

We plan to study the effectiveness of covariate analysis in the context of Latin Square Designs [LSDs] carried out in groups. We contemplate on three different experimental situations and in each case we examine the possibility of improving the performance of the estimates of the covariates' parameters. This is done in line with the study of optimal covariates' designs.

1. Introduction

In agricultural experiments, there are occasions when a number of covariates are in use and that too in the context of a Latin Square Design [LSD]. To fix ideas, we start with an LSD of order q conducted in a field with q rows, q columns and q treatments. Suppose there are plenty of resources available so that in each of the q^2 'cells', there are r otherwise identical plots available. In effect, we are referring to a single LSD with r independent observations in each cell. Thus altogether we are talking about rq^2 observations.

Since it is a replicated experiment, there are $(r–1)q^2$ df available for generating what we may term as 'pure experimental error'. Further to this, we have additional $(q – 1)(q – 2)$ df arising out of the model-based errors in the q-factor LSD. We refer to any standard text book on experimental designs dealing with such basic designs as CRD, RBD and LSD. Naturally, also we have q –1 df for each of the three orthogonal factors : Rows, Columns and Treatments.

In total, all the $rq^2 – 1$ df are taken care of. This is all without any involvement of covariates.

Now we wish to introduce a few controllable quantitative covariates Z_1, Z_2, ... - each one assuming values in the closed interval [–1, 1]. We specialize to the case of $q = 4$ *i.e.*, an LSD of order 4. Further, we take $r = 3$. Taking clue from Example 8.2.3 of Das *et al.* (Chapter 8, 2015), we find that it is possible to accommodate 6 covariates and provide most efficient orthogonal estimates of all the 6 γ-parameters attached to these covariates in the mean model underlying an LSD. Explicitly written, the 'optimal covariates design' is given in the table below in the form of an Orthogonal Array. The first two rows of the array represent rows and columns : 1 to 4. The third row of the array represents treatment allocations : 1 to 4. The underlying LSD is also shown below. The subsequent 6 rows of the array represent covariates allocations in the form of +1's and –1's. It must be noted that each covariate value in each row-column location is to be utilized 3 times in the 3 plots therein. Note that 'pure' errors absorb a total of 32 df.

The specic chosen field layout in the form of an LSD of order 4:

$$\begin{bmatrix} Rows & Column\ 1 & Column\ 2 & Column\ 3 & Column\ 4 \\ 1 & T1 & T2 & T3 & T4 \\ 2 & T2 & T1 & T4 & T3 \\ 3 & T3 & T4 & T1 & T2 \\ 4 & T4 & T3 & T2 & T1 \end{bmatrix} \tag{3.1}$$

The Orthogonal Array : (Given in next page)

As a matter of fact, there are 48 observations and hence 38 error df in this thrice replicated LSD of order 4. Thus one would wonder if we can accommodate more than 6 covariates optimally. A little reection shows that it is not possible to do so.

LSDs of Order 4 in Different Seasons with the Same Layout

We now consider an experimental situation wherein the same LSD of order 4 is replicated in 3 seasons with identical squares and without any covariates to start with.

E-mail : duttaganesh78@gmail.com

$$\begin{bmatrix} Rows & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 3 & 3 & 3 & 3 & 4 & 4 & 4 & 4 \\ Columns & 1 & 2 & 3 & 4 & 1 & 2 & 3 & 4 & 1 & 2 & 3 & 4 & 1 & 2 & 3 & 4 \\ Treatments & 1 & 2 & 3 & 4 & 2 & 1 & 4 & 3 & 3 & 4 & 1 & 2 & 4 & 3 & 2 & 1 \\ Cov\ 1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\ Cov\ 2 & 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \\ Cov\ 3 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 \\ Cov\ 4 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 \\ Cov\ 5 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 \\ Cov\ 6 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \end{bmatrix} \tag{3.2}$$

$$\begin{bmatrix} & & S1,\ S2,\ S3 & & \\ Rows & Column\ 1 & Column\ 2 & Column\ 3 & Column\ 4 \\ 1 & T1 & T2 & T3 & T4 \\ 2 & T2 & T1 & T4 & T3 \\ 3 & T3 & T4 & T1 & T2 \\ 4 & T4 & T3 & T2 & T1 \end{bmatrix} \tag{3.3}$$

Layout of 3 identical squares S1, S2 and S3 is given above.

It may be assumed that the seasons are separated enough to justify independence of successive observations from the same plot and this holds for all 16 plots. A simple model would suggest:

$$\textit{Model I : } y_{ijks} = \mu_0 + \theta_s + \rho_i + \nu_j + \tau_k + e_{ijks}$$

where 's' represents season, 'i' stands for row, 'j' stands for column and 'k' stands for treatment in the LSDs in different seasons. Routine computations yield : SS due to Seasons (2 df), SS due to Rows - SS due to Columns - SS due to Treatments - each carries 3 df. These leave 36 df for error - 6 df within each season [giving a total of 18 df] and 2 × 3 = 6 df for each Season × Eect [Row/Column/Treatment] interaction - once the model is extended by incorporating the interaction terms.

Now let us introduce a few covariates and discuss about the possibility of optimal estimation of the covariate parameters. Note that the only hurdle corresponds to the interaction terms involving the three seasons. Fortunately, for each combination of Season vs Eect Interaction [arising due to Rows / Column / Treatments], there are 4 observations. Therefore, the same covariate design as described above can optimally accommodate 6 covariates. We refer to Das *et al.* (2015) for necessary theory. Here optimality has to be understood in proper contexts. We will consider different situations :

(a) There are altogether 6 covariates and these are available in all the three seasons. Naturally all the covariates parameters can be optimally estimated and we attain 'global' optimality [minimum variance of $\sigma^2 / 48$].

(b) There are 6 covariates in each season and these are totally dierent from season to season. Here again all these can be optimality estimated and this time it is 'local' optimality with minimum variance of $\sigma^2 / 16$ for each.

(c) Altogether there are 6 covariates [C1 to C6] and these are grouped as : Seasons I /II : C1 and C2; Seasons I / III : C3 and C4; Seasons II / III : C5 and C6. A little reflection shows that this time we attain 'local' optimality with minimum variance of $\sigma^2 / 32$.

(d) Another interesting feature would be like : Season I : C1 and C4-C6; Season II : C2 and C4-C6; Season III : C3 - C6. The results are similar.

LSDs of Order 4 in Different Seasons with Different Design Layouts

Lastly, we consider an experimental situation wherein 3 different LSDs of order 4 are utilized in 3 seasons with non-isomorphic squares and without any covariates to start with.

Layout of 3 non-isomorphic LSDs in the 3 seasons S1, S2 and S3 is given below. Note that S1 is already introduced earlier.

$$\begin{bmatrix} & & S1 & & \\ Rows & Column\ 1 & Column\ 2 & Column\ 3 & Column\ 4 \\ 1 & T1 & T2 & T3 & T4 \\ 2 & T2 & T1 & T4 & T3 \\ 3 & T3 & T4 & T1 & T2 \\ 4 & T4 & T3 & T2 & T1 \end{bmatrix} \tag{3.4}$$

$$\begin{bmatrix} & & S2 & & \\ Rows & Column\ 1 & Column\ 2 & Column\ 3 & Column\ 4 \\ 1 & T4 & T3 & T2 & T1 \\ 2 & T3 & T2 & T1 & T4 \\ 3 & T2 & T1 & T4 & T3 \\ 4 & T1 & T4 & T3 & T2 \end{bmatrix} \tag{3.5}$$

$$\begin{bmatrix} & & S3 & & \\ Rows & Column\ 1 & Column\ 2 & Column\ 3 & Column\ 4 \\ 1 & T1 & T2 & T3 & T4 \\ 2 & T3 & T4 & T1 & T2 \\ 3 & T2 & T1 & T4 & T3 \\ 4 & T4 & T3 & T2 & T1 \end{bmatrix} \tag{3.6}$$

In each season, there are 16 observations and, as usual, we have 6 df for error. As is known, these errors are formed of observational contrasts and in case of LSDs, these arise out of study of 'tetra differences'. Vide Shah and Sinha (1996). Unless these errors are formed of +1's and –1's, covariates cannot be optimally estimated. Vide Das *et al*. (2015). For S1, we already have from the OA a set of 6 observational contrasts for optimal estimation of the covariates parameters. For the other two, we display below a set of 3 observational contrasts for each.

Just as in the above, we may consider the following cases towards optimal estimation of covariates parameters. Note that the seasons are well-separated and independence is assumed.

$$\begin{array}{l} Rows \\ Columns \\ Treatments \\ Cov\ 1 \\ Cov\ 2 \\ Cov\ 3 \end{array} \begin{bmatrix} 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 3 & 3 & 3 & 3 & 4 & 4 & 4 & 4 \\ 1 & 2 & 3 & 4 & 1 & 2 & 3 & 4 & 1 & 2 & 3 & 4 & 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 & 3 & 2 & 1 & 4 & 2 & 1 & 4 & 3 & 1 & 4 & 3 & 2 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{bmatrix} \quad (3.7)$$

$$\begin{array}{l} Rows \\ Columns \\ Treatments \\ Cov\ 1 \\ Cov\ 2 \\ Cov\ 3 \end{array} \begin{bmatrix} 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 3 & 3 & 3 & 3 & 4 & 4 & 4 & 4 \\ 1 & 2 & 3 & 4 & 1 & 2 & 3 & 4 & 1 & 2 & 3 & 4 & 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 & 3 & 4 & 1 & 2 & 2 & 1 & 4 & 3 & 4 & 3 & 2 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 \end{bmatrix} \quad (3.8)$$

(a) There are altogether 3 covariates and these are available in all the three seasons. Naturally all the covariates parameters can be optimally estimated and we attain 'global' optimality [minimum variance of $\sigma^2/48$]. From S1 we may choose any three of the covariates, followed by all three from each of S2 and S3. Note that the covariates values are cell-based and not treatment-based.

(b) There are covariates in each season and these are totally different from season to season. Here again all these covariates parameters [6 + 3 + 3 = 12] can be optimality estimated and this time it is 'local' optimality with minimum variance of $\sigma^2/16$ for each.

(c) Altogether there are 6 covariates [C1 to C6] and these are grouped as : Seasons I / II : C1 - C3; Seasons I / III : C4 - C6. A little reflection shows that this time we attain 'local' optimality with minimum variance of $\sigma^2/32$ for each parameter estimate.

(d) Altogether there are six covariates and these are common to all the three seasons. Since the seasons are well-separated and independence is assumed, can we optimally estimate all the six covariates parameters - with minimum variance of $\sigma^2/48$? By using the following allocation layout we can provide an affirmative answer :

[S1C1; S2C1; S3C1]; [S1C2; S2C1; –S3C1]; [S1C3; S2C2; S3C2];

[S1C4; S2C2; – S3C2]; [S1C5; S2C3; S3C3]; [S1C6; S2C3; – S3C3]:

Optimal choices of covariates designs/allocations is a very fascinating area for research and applications in different fields. We close this paper with yet another non-trivial example of optimal estimation of covariate effect parameter.

LSD of Order 5 with Optimal Allocation for a Single Covariate

Consider 5 × 5 Latin square:

$$L_1 = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 4 & 5 & 1 \\ 3 & 4 & 5 & 1 & 2 \\ 4 & 5 & 1 & 2 & 3 \\ 5 & 1 & 2 & 3 & 4 \end{bmatrix}$$

Model based on L_1 using only one covariate:

$$Model\ II : y_{ijk} = \mu + r_i + c_j + \tau_k + \gamma_{ijk} + e_{ijk};$$

$i, j, k \in D = \{(i, j, k) : k(i, j)\ j \neq k(i, j'), j \neq j' = 1(1)5, k(i, j), i \neq i' = 1(1)5\}$

In the absence of the covariate parameter , this is an orthogonal design. Routine computations suggest :

$$\mathrm{Var}(\hat{\gamma}) = \left(\sum_{(i,j,k)\in D} (z_{ijk} - \bar{z}_{i00} - \bar{z}_{0j0} - \bar{z}_{00k} + 2\bar{z}_{000})^2 \right)^{-1} \sigma^2$$

Now consider $\sum_{(i,j,k)\in D} (z_{ijk} - \bar{z}_{i00} - \bar{z}_{0j0} - \bar{z}_{00k} + 2\bar{z}_{000})^2 = \sum_{(i,j,k)\in D} z_{ijk}^2 - \sum_{i=1}^{5} \frac{R_i^2}{5} - \sum_{i=1}^{5} \frac{C_i^2}{5} -$

$\sum_{i=1}^{5} \frac{T_i^2}{5} + 2\frac{G^2}{25}$ and this is maximum when $z_{ijk} = \pm 1$ since this is a convex function. Consider another Latin square

$$L_2 = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 1 & 2 & 3 & 4 \\ 4 & 5 & 1 & 2 & 3 \\ 3 & 4 & 5 & 1 & 2 \\ 2 & 3 & 4 & 5 & 1 \end{bmatrix}$$

which is orthogonal with L_1. Now replace 1, 2, 3, 4, 5 of L_2 by +1, –1, +1, –1, +1 respectively and get following matrix version of the Latin square :

$$\begin{bmatrix} +1 & -1 & +1 & -1 & +1 \\ +1 & +1 & -1 & +1 & -1 \\ -1 & +1 & +1 & -1 & +1 \\ +1 & -1 & +1 & +1 & -1 \\ -1 & +1 & -1 & +1 & +1 \end{bmatrix}$$

Accordingly, with the above choice of the z-values *i.e.,*
z – vector = (+1,–1, +1,–1, +1, +1, +1,–1, +1,–1,–1, +1; +1,–1, +1, +1,–1, +1, +1,–1,–1, +1,–1, +1, +1) 0, it turns out that

$$\sum_{(i,j,k)\in D} (z_{ijk} - \bar{z}_{i00} - \bar{z}_{0j0} - \bar{z}_{00k} + 2\bar{z}_{000})^2 = \sum_{(i,j,k)\in D} z_{ijk}^2 - \sum_{i=1}^{5} \frac{R_i^2}{5} - \sum_{i=1}^{5} \frac{C_i^2}{5} - \sum_{i=1}^{5} \frac{T_i^2}{5} + 2\frac{G^2}{25} =$$

25 – 1 – 1 – 1 + 2 = 24 since $R_i = i^{th}$ row sum = 1 for all i, $C_j = j^{th}$ column sum =1 for all j and $T_k = k^{th}$ treatment sum = 1 for all k and G = Grand total = 5.

Thus we are in a position to introduce one covariate optimally, given the above structure of an LSD of order 5. The resulting variance of $\hat{\gamma}$ is $\sigma^2/24$ as against the non-attainable lower bound of $\sigma^2/25$.

REFERENCES

Shah, K. R. and Sinha, Bikas K. 1996. Row-column designs. In Handbook of Statistics 13: Design and Analysis of Experiments. Ed. S. Ghosh and C. R. Rao. North Holland, 903938.

Das, P., Dutta, G., Mandal, N. K. and Sinha, B. K. 2015. Optimal Covariates Designs and Their Applications. Springer, New York.