SuDoKu as an Experimental Design - Beyond the Traditional Latin Square Design I

Michael. F.A. Saba
Field Crops Research Institute, Agriculture research Center, Egypt
And
Sinha, Bikas K
Ex-Faculty, Applied Statistics Division, Indian Statistical Institute, Kolkata, India

Abstract

SuDoKu is an interesting combinatorial structure embedded within a Latin Square. It has been gaining popularity as combinatorial puzzle. Website provides plenty of examples of such SuDoKus. It is well known that Latin Square Designs [LSDs] go beyond CRDs and RCBDs in eliminating three external sources of variation in the experimental units [eu's] in an ANOVA set-up. We examine the possibility of viewing SuDoKus as experimental designs, going one step beyond LSDs. It is contemplated that one additional [fourth] component of variation can be suitably accommodated in a SuDoKu, though by sacrificing orthogonality. A detailed statistical analysis is provided along with the underlying ANOVA Table for such designs based on SuDoKus.. It is envisaged that the SuDoKUs will provide extra dimension of utility as experimental designs.

Key word: ***ANOVA Designs, CRD, RCBD, LSD, SuDoKu, Latin Square Designs with Internal Blocking***

1. Introduction

In combinatorics and in ANOVA-oriented experimental designs, a Latin Square is an n × n array filled with n different symbols, each occurring exactly once in each row and exactly once in each column. Latin Square Designs [LSDs] and their generalizations, known as Graeco Latin Square Designs, are among the basic experimental designs we are familiar with while dealing with ANOVA set-up. Several studies were done to investigate its properties (**Higham JeBey 1966; Preece and Freeman, 1983; Bailey, 1990; Chigbu, 2001 and 2003; Li *et al* 2005 and Rau, 2009).**

If each entry of an n × n Latin square is written as a triplet (r, c, s), where r is the row, c is the column, and s is the symbol, we obtain a set of n^2 triplets called the orthogonal array representation of the square. The definition of a Latin Square can be equivalently written in terms of an orthogonal array: a Latin Square is the set of all triples (r, c, s), where $1 \le r, c, s \le n$, such that all ordered pairs (r, c) are distinct, all ordered pairs (r, s) are distinct, and all ordered pairs (c, s) are distinct. For any Latin Square, there are n^2 triplets since choosing any two uniquely determines the third

Three very basic experimental designs (CRD, RCBD and LSD) have been developed in order to obtain valid and reliable conclusion from ANOVA-based field experiments. Overlooking the balanced nature, treatment number and randomization, the essential difference between these designs is in controlling the number of extraneous sources of variations due to environmental factors such as fertility variation of the soil in agricultural experiments. Identifying such 'assignable' source(s) of variation and eliminating their effects from the main analysis usually leads to more reliable understanding of the nature of pure errors and consequently improves the efficiency of the experimental designs. **Bailey, 1990 and Bolboaca *et al* 2009** revealed that CRD error > RCBD error > LSD error and the precisions of the underlying experimental designs are in reverse orientation LS > RCBD > CRD.

Designs, such as LSDs, eliminating effects of two factors, are commonly termed as row-column designs and a systematic study of such designs provides more general and comprehensive understanding of such effects than the standard LSDs **(Shah and sinha, 1996)**. In a broad perspective, notions of estimability, connectedness, efficiency and optimality have been discussed in the above-cited paper. Nevertheless, even within the framework of a standard LSD, there is a possibility for identifying an altogether different source of variation than the two extraneous sources represented by rows and columns.

Speculative sight in 9 x 9 SuDoKu puzzle, for example, suggests the possibility of exploiting its internal structure as an experimental design, yet identifying an additional source of variation, apart from those usually attributed to rows/columns/treatments. Although the 9×9 grid with 3×3 'inner regions' is by far the most common, many variations do exist. Some such SuDoKus deal with (i) 6 x 6 grids with 3 x 2 inner regions,, (ii) 16×16 grids with 4 x 4 inner regions, (iii) 25 x 25 grids with 5x5 inner regions and so on. We do not venture into the constructional aspects of such families of SuDoKus. **Free, Wikipedia, 2012** Our interest lies in their potential use as experimental designs, going one-step beyond the traditional LSDs. We will explain this aspect in the next section. Nevertheless, we have constructed and presented more variants, 8x8 grids with 2x4 inner regions, 10x10 grids with 2x5 inner regions, 12x12 grids with 3x4 inner regions, 18x18 grids with 3x6 inner regions and 20x20 grids with 4x5 inner regions (figure 2- 5).

2. SuDoKu as a Postulated Experimental Design

In the context of a SuDoKu, we will refer to the 'inner regions' as 'internal blocks' having potential advantage for accommodating an additional source of variation in the eu's. For example, in Figure 1, we display a SuDoKu of order 6 x 6 having 3x2 internal blocks. There are altogether six internal blocks in this SuDoKu and each internal block [having an allocation of all the six treatment symbols] might represent one specific 'differential fertility situation' within the set-up of LSD in an agricultural situation. We will refer to this feature as 'LSDs with internal blocking' and these designs originate from traditional LSDs wherein we already have the three sources of variation, viz., row-to-row variation, column-to-column variation and treatment-to-treatment variation. Therefore, SuDoKus, as experimental designs, build upon the LSDs and accommodate one more component of variation through the concept and formation of 'internal blocks'. It would be interesting and instructional to examine the contribution of this new source of variation in the context of the ANOVA Table derived thereupon. This is elaborated in the next section.

1	6	3	5	4	2
3	2	4	1	6	5
5	4	6	2	1	3
4	1	5	3	2	6
2	5	1	6	3	4
6	3	2	4	5	1

Figure 1: SuDoKu of order 6 with 6 internal blocks each of order 3x2

3. ANOVA for SuDoKu Designs

3.1. Linear Model for SuDoKu Designs

We start with a SuDoKu of order n consisting of n internal blocks each of order p x q. Each of these internal blocks corresponds to a group of p distinct rows and q distinct columns and there are p x q [=n] such internal blocks. For example, in Figure 1, let us label the internal blocks as B1, B2, B3 [upper set] and B4, B5 and B6 [lower set]. Then B1 corresponds to [R1, R2, R3] and [C1, C2]. Likewise, B2 corresponds to [R1, R2, R3] and [C3, C4], etc. It is understood that R_i denotes the row labeled 'i' and C_j denotes the column labeled 'j'. Regarded as an experimental design, we will designate the above as a SuDoKu design with parameters [n, p, q].

The postulated linear model has the obvious representation:

$$Y_{ijkl} = \mu + t_k + r_i + c_j + b_l + e_{ijkl} \qquad (1)$$

where Y_{ijkl} = observation recorded for treatment *k* associated with the combination (i, j), i = designated row involving treatment k, j = designated column involving treatment k and l designated block involving treatment k;

μ = general mean,

t_k= effect for treatment *k* (*k*=1 ... n) and $\sum t_k$ from 1 to n = zero,

r_i = effect for row *i* (*i* =1 ... n) and $\sum r_i$ from 1 to n=zero,

c_j = effect for column j (j = 1 ... n) and $\sum c_j$ from 1 to n =zero,

b_l = effect of internal block l (l=1 ... n) and $\sum b_l$ from 1 to n = zero,

e_{ijkl} = random error associated with Y_{ijkl} .

We will use the obvious notations:

$Y_{i...}$ = i-th row total

$Y_{.j..}$ = j-th column total

$Y_{..k.}$= k-th treatment total

$Y_{...l}$ = l-th internal block total

$Y_{....}$ = grand total (2)

It may be noted that the row-column combination (i, j) determines the label of the internal block (l) in the above representation.

3.2. Orthogonal Decomposition of Total Sum of Squares [TSS]

In the absence of the internal block effects (the Latin square case) [b_l ; l=1, 2, ..., n], we have the standard decomposition :

$$TSS = SSR + SSC + SSTr + SSE \qquad (3)$$

Regarding the newly added component of variation, we observe the following:

The arrangement of the inner blocks inside a Latin square n x n (where n=p x q) classify the n rows into p groups of rows each group contains q rows and classify the n columns into q groups each group contains p columns. Therefore

(I) [RG1 = $Y_{1...}$ + $Y_{2...}$ + ...+ $Y_{q...}$; .....; RGp = $Y_{(n-q+1)...}$ + $Y_{(n-q+2)...}$ +.. + $Y_{n...}$] are confounded with (p -1) IBC contrasts [This explains (item IV)] where RG_i is the total of row group i

(II) [CT1 = $Y_{.1..}$ + $Y_{.2..}$ + ...+ $Y_{.p..}$; .....; CTp = $Y_{.(n-p+1)..}$ + $Y_{.(n-p+2)..}$ +.. + $Y_{.n..}$] are confounded with (q-1) IBC contrasts [This explains (V)]

(III) Internal Block Classification [IBC] is orthogonal to the treatments;

(IV) Part of IBC is orthogonal to the rows;

(V) Part of IBC is orthogonal to the columns.

At this stage, it is instructional to point out that the IBC contrasts are estimated only in terms of the 'tetra differences'. Vide Shah and Sinha (1996). There are p x q IB or Location-related parameters and consequently, only the (p-1)(q-1) linearly independent tetra-differences provide estimates of IBC contrasts. All others are confounded. Specifically, (q-1) IBC contrasts are confounded with the row contrasts and (p-1) IBC contrasts are confounded with the column contrasts. That explains the decomposition of (pq-1) IBC contrasts.

Naturally, error df under the new model = error df under the LSD model – df for orthogonal estimation of IBC contrasts free from row/column contrasts = $(n^2 - 1) - 3(n-1) - (p-1)(q-1) = (n^2 - 1) - 3(n-1) - [n-p-q+1] = n^2 - 4n + p+q + 1$.

This explains the nature of the decomposition of the TSS. We are now in a position to prepare the ANOVA Table.

3.3. ANOVA TABLE for SuDoKu Design with parameters [n, p, q]

Let us accept the group classifications of the rows and columns to groups compatible with the inner blocks as the base the partitioning degree of freedom (d.f) and sum of squares (SS) to their components. Therefore, both d.f and SS of either rows or columns have to components between groups and within groups. The between group component for rows has d.f = p-1 and SS= ∑1p(∑1n rgi)2/np – C.F, where rgi is the eu's of row group gi, and within rows component has d.f= p(q-1) and SS=∑1n r2/n - ∑1p(∑1n rl)2/np. The between group component for columns has d.f = q-1 and SS= ∑1q(∑1n cJ)2/nq – C.F, where cgi is eu's of column group gj and within columns component has d.f = q(p-1) and SS=∑1n C2/n - ∑1q(∑1n CJ)2/nq.

Similarly, the IB can be partitioned to rows, column components and the remainder. The rows component of IB has d.f = p-1 and SS= ∑1p(∑1n bgi)2/np – C.F; this is completely confound with the between row groups component so it gives the same numerical value for both d.f and SS. Also, the columns component of BI, which confound with the between column groups component, has d.f = q-1 and SS ∑1q(∑1n CJ)2/nq – C.F. The remainder (the interaction between rows and columns) component has d.f = (p-1)(q-1) and SS= ∑1n B2/n - ∑1q(∑1n RI)2/np - ∑1p(∑1n CJ)2/nq + C.F. These components are presented on Table 1.

Table 1: Analysis of Variance of a SuDoKu design with parameters [n, p, q]

Source of Variation	d.f.	SS
Treatment	n – 1	$\Sigma_1^n t^2/n - C.F$
IB confounded with rows	p – 1	$\Sigma_1^p(\Sigma_1^n R_I)^2/np - C.F$
IB confounded with columns	q – 1	$\Sigma_1^q(\Sigma_1^n C_J)^2/nq - C.F$
Orthogonal IBC	(p-1)(q-1)	$\Sigma_1^n B^2/n - \Sigma_1^q(\Sigma_1^n R_I)^2/np - \Sigma_1^p(\Sigma_1^n C_J)^2/nq + C.F$
Rows within each of q subgroups	p(q -1)	$\Sigma_1^n r^2/n - \Sigma_1^p(\Sigma_1^n R_I)^2/np$
Columns within each of p subgroups	q(p – 1)	$\Sigma_1^n c^2/n - \Sigma_1^p(\Sigma_1^n C_J)^2/nq$
Error	n^2 -4n + p + q +1	By subtraction
Total	n^2 – 1	$\Sigma_1^{n^2} y^2 - C.F$ (*)

(*) $C.F = (\Sigma_1^{n^2} y)^2/n$, n=pq

4. Example

To test and exploit the postulated designs before applying them on real experimental data, simulated data have been created and utilized for n=12, p = 3 and q = 4 and the allocation chart is given in the following :

The layout

1	9	7	4	5	6	10	3	2	12	8	11
5	2	12	9	11	10	1	8	4	3	7	6
11	8	4	12	1	3	5	6	7	10	9	2
3	6	10	7	2	8	9	11	12	1	4	5
2	12	6	8	3	4	11	5	9	7	1	10
4	5	8	11	6	1	2	7	10	9	12	3
7	1	9	2	10	12	6	4	3	5	11	8
10	3	11	5	9	7	8	12	1	2	6	4
12	10	3	1	7	11	4	2	8	6	5	9
8	7	6	6	12	9	3	10	11	4	2	1
6	4	1	3	8	2	7	9	5	11	10	12
9	11	2	10	4	5	12	1	6	8	3	7

The data shown below were arranged according to the above layout:

10.5	10.8	10	9.7	9.4	9	9.7	9.1	10.3	7.2	8.4	8.8
9.7	10.5	9.5	11.2	10.3	10.4	10.1	9.7	9.1	7	8.5	8.6
10.5	9.9	9.3	9.9	10.3	9.3	9.2	9.2	9.8	8.1	9.3	8.5
9.1	7.8	10.3	10.4	10.1	10	10.3	10.5	9.3	8.5	7.8	7.9
9.7	8.9	7.9	9.9	8.5	9	9.1	8.6	9.6	8.8	10.3	9.5
8.6	9.2	9.3	10.7	7.9	10.4	8.8	8.8	9	9.7	9.1	8.4
9.2	10.1	10.3	10.5	9.5	9.2	8.4	8.1	8	8.7	10.3	9.1
9.5	8.7	10.1	9.8	10.1	9.7	8.2	8.3	9.4	9.3	8	8.5
8.7	9.7	8.6	9.8	8.3	9.5	8.9	10.4	9.8	6.9	9.8	8.9
9.1	9.3	9.6	8.3	7.8	9.6	8.8	10.2	10.6	7.1	8.9	8.7
8.4	8.7	10.1	8.3	8.2	9.2	9.5	10.8	8.5	8.4	8.7	7.5
10	10.1	9.8	9.4	7.7	9	9	10.6	9.1	7.6	7.7	8

The required totals for statistical analysis

serial	1	2	3	4	5	6	7	8	9	10	11	12	total	squares/ denominator
treatment	118.8	116	101.5	102.5	99.8	109.1	110.3	109.2	120.6	114	118.9	104.4	1325	12242.37
row	112.9	114.6	113.3	112	109.8	109.9	111.4	109.6	109.3	108	106.3	108	1325	12199.23
Row groups		452.8				440.7				431.6			1325	12198.39
column	113	113.7	114.8	117.9	108.1	114.3	110	114.3	112.5	97.3	106.8	102.4	1325	12225.52
Column groups		341.5			340.3			336.8			306.5		1325	12216.75
block	117.9	120	116.3	98.6	111.5	115.2	104.3	109.7	112.1	105.1	116.2	98.2	1325	12243.50
IB row		452.8				440.7				431.6			1325	12198.39
IB cilumn	341.5	340.3	336.8	306.5									1325	12216.75

The correction factor (CF)= 12193.68

Total sums of squares=12310.23

The corrected sums of squares were calculated and presented on Table 2

Table 2: Analysis of Variance of a SuDoKu Design [12, 4, 3]

Sources of Variation	d.f	SS	MS	F	P
Treatment effects	11	48.69	4.426	143.2	<0.000
Row effects confound with internal block effects	2	4.789	2.395	77.49	<0.000
Column effects confound with internal block effects	3	17.78	5.928	191.8	<0.000
Orthogonal Internal block effects	6	27.25	4.542	147	<0.000
Row effects orthogonal to internal blocks	9	0.764	0.085	2.749	0.006
Column effects orthogonal to internal blocks	8	14.06	1.757	56.87	<0.000
Error	104	3.214	0.031		
Total	143	116.5			

Despite of the main hypothesis for any experiment is the null effect of imposed treatment; we supposed the null effect of inner blocks. Really, the most important component has orthogonal contrasts is the remainder SS of the inner blocks after excluding the confound effects with either row or column effects. This component was highly significant so the null hypothesis was rejected. Following **Bailey, 1990 and Bolboaca *et al* 2009** who stated that CRD error > RCBD error > LSD error and the precisions of the underlying experimental designs are in reverse orientation LS > RCBD > CRD. We can safely claim that the Efficiency of the underlying experimental designs are in orientation SuDoKu>LS > RCBD > CRD.

4	1	3	5	7	2	8	6
8	5	2	7	1	6	3	4
7	3	1	6	4	8	5	2
2	6	4	8	3	5	1	7
1	7	6	2	8	3	4	5
3	2	5	1	6	4	7	8
5	8	7	4	2	1	6	3
6	4	8	3	5	7	2	1

Figure 2: SuDoKu of order 8 x 8 internal blocks each of order 2x4

10	8	5	3	1	4	6	9	7	2
2	5	9	10	7	6	8	1	3	4
4	1	2	6	8	5	7	3	10	9
6	3	4	7	9	2	10	5	1	8
9	7	1	8	3	10	2	4	5	6
3	6	7	4	2	9	5	10	8	1
8	10	3	5	4	1	9	6	2	7
5	2	10	9	6	7	1	8	4	3
7	9	8	1	10	3	4	2	6	5
1	4	6	2	5	8	3	7	9	10

Figure 3: SuDoKu of order 10 x 10 internal blocks each of order 2x5

1	17	5	10	18	14	16	3	6	8	11	9	12	2	15	13	4	7
13	15	11	6	7	4	2	14	9	12	5	17	18	10	1	3	16	8
18	2	12	17	1	15	10	5	13	6	16	7	3	4	8	14	9	11
7	10	8	3	16	9	4	11	17	2	1	15	6	14	13	12	18	5
4	16	3	2	12	8	15	7	1	14	13	18	11	5	9	6	10	17
9	6	14	11	13	5	8	12	18	4	10	3	7	16	17	15	1	2
16	9	6	13	14	3	5	18	8	11	15	12	17	7	10	1	2	4
8	1	17	12	9	10	14	2	7	13	3	16	4	15	5	11	6	18
15	5	10	8	2	17	13	4	11	7	18	1	9	12	6	16	14	3
2	18	13	16	4	11	1	17	15	10	9	6	8	3	14	7	5	12
11	14	4	7	15	6	12	10	3	5	2	8	1	18	16	17	13	9
3	12	7	1	5	18	9	6	16	17	14	4	13	11	2	10	8	15
14	3	18	9	11	7	17	13	4	16	8	5	10	6	12	2	15	1
17	8	2	18	6	12	3	16	14	1	7	10	15	9	4	5	11	13
10	11	16	4	17	1	18	8	2	15	12	14	5	13	3	9	7	6
12	7	1	5	3	13	6	15	10	9	4	2	14	8	11	18	17	16
5	4	9	15	10	2	11	1	12	18	6	13	16	17	7	8	3	14
6	13	15	14	8	16	7	9	5	3	17	11	2	1	18	4	12	10

Figure 4: SuDoKu of order 18 x 18 internal blocks each of order 3x6

15	7	16	8	18	5	11	6	20	14	2	13	10	3	12	9	17	1	19	4
5	20	3	13	8	14	1	17	4	7	19	9	2	16	18	6	15	12	11	10
12	19	4	6	13	20	10	2	15	11	1	18	17	5	7	14	3	9	8	16
1	14	9	18	3	16	19	12	5	17	10	8	15	20	11	4	7	6	2	13
17	10	2	11	9	7	15	4	3	16	12	6	13	8	1	19	14	5	20	18
11	1	19	12	20	6	3	8	14	9	18	7	16	17	10	5	2	13	4	15
20	5	6	4	16	13	17	15	10	19	3	2	14	18	9	11	12	8	7	1
7	16	18	3	5	19	12	9	11	4	15	17	8	1	13	2	6	20	10	14
2	9	13	17	14	10	18	7	16	1	8	12	20	6	4	15	11	3	5	19
14	8	15	10	1	4	2	11	6	20	13	5	3	12	19	7	18	16	9	17
13	3	20	15	2	9	5	19	12	6	14	10	4	7	8	16	1	17	18	11
9	6	17	2	4	15	7	16	13	8	20	11	18	14	3	1	19	10	12	5
16	11	5	19	12	18	14	13	7	15	17	1	9	10	2	20	8	4	3	6
8	18	14	1	6	3	20	10	19	5	4	16	12	11	15	17	9	7	13	2
10	4	12	7	11	17	8	1	2	18	9	3	6	19	5	13	16	14	15	20
4	12	11	16	10	1	6	20	9	2	7	19	5	15	14	8	13	18	17	3
19	2	7	5	17	12	9	14	18	10	6	15	1	13	20	3	4	11	16	8
3	15	1	9	7	8	16	18	17	13	11	20	19	4	6	10	5	2	14	12
18	13	8	20	15	11	4	5	1	3	16	14	7	2	17	12	10	19	6	9
6	17	10	14	19	2	13	3	8	12	5	4	11	9	16	18	20	15	1	7

Figure 5: SuDoKu of order 20 x 20 internal blocks each of order 4x5

References

Bailey R. A. (1990). An efficient semi-latin square for twelve treatments in blocks of size two, Journal of Statistical Planning and Inference, 1990, 26(3), p. 263-266.

Bolboaca, Sorana D., jäntschi, L. and R. E. Sestras (2009). Statistical Approaches In Analysis Of Variance: From random arrangements to Latin square experimental design. Leonardo Journal of Sciences Issue 15, July-December 2009 p. 71-82

Chigbu, P.E. (2001). Graeco-Latin squares versus Trojan squares, Journ. Nig. Stat. Assoc., 14, 11 – 17.

Chigbu, P.E. (2003). The "best" of the three optimal (4 x 4)/4 semi-Latin squares, Sankhya: The Indian Journal of Statistics, 65(3), 641 – 648.

Free Wikipedia (2012). en.wikipedia.org/wiki/**Sudoku**

Higham, JeBey T.(1966). Construction Methods for Row-Complete Latin Squares. A thesis in Combinatorics and Optimization, Waterloo, Ontario, Canada, 1996

Li L.-F., Dai Y., Zhang G.-Q. (2005). Generating test-case based on orthogonal Latin squares, Journal of Soochow University Engineering Science Edition, 2006, 26(2), pp. 19.

Preece, D.A. and Freeman, G.H. (1983). Semi- Latin squares and related designs, J. Roy. Statist.Soc. Ser. B, 45, 267 – 277.

Rau, A. R. P. (2009). R. A. Fisher, design theory, and the Indian connection. Journal of Biosciences, 2009, 34, 353–363.

Snedecor, G.W (1956). Statistical Methods, applied experiments in agriculture and biology. 5th Ed. The Iowa State College Press, Ames,Iowa, USA

Shah, K.R. and B.K. Sinha (1996). Design and Analysis of Experiments, Handbook of statistics Vol. 13, P:903-937, North-Holland, Amesterdam, Holland

السودوكو كتصميم تجريبي تطويرا لتصميم المربع اللاتيني

ميشيل فخري ابادير

معهد بحوث المحاصيل الحقلية، مركز البحوث الزراعية بالجيزة

بيكاس سنها

قسم البحوث التطبيقية، معهد الاحصاء الهندي كلكوتا، الهند

السودوكو تركيب توافقي مطمور بداخل مربع لاتيني، اكتسب شهرة كفزورة. ومواقع الشبكة الالكترونية العنكبوتية توفر امثلة كثيرة لمثل هذه الفوازير. ومن المعروف ان تصميمات المربع اللاتيني اكثر دقة من القطاعات كاملة العشوائية والتصميمات الكاملة العشوائية لأ ظهاره ثلاثة اسباب لتباين الوحدات التجيريبية في جداول تحليل التباين. وقد درسنا امكانيات السدوكو كتصميمات تجارب تتقدم خطوة للأمام بعد المربع الاتيني. وبالتفكير انها تعرض مكون اضافي (رابع) للتباين في جدول تحليل التباين لتصميم السودكو بالرغم من التضحية بفقدان التعامد (orthogonalty). وامكن تقديم تحليل احصائي ومعادلات واسس جدول تحليل التباين للسودوكو. ونوقعنا ان انواع السودوكو تقدم بعدا اضافي للأستفادة بها كتصميمات تجريبية.