

Sudoku Squares as Experimental Designs

Jyotirmoy Sarkar and Bikas K Sinha



(left) Jyotirmoy Sarkar is a Professor and Statistics Consultant at Indiana University-Purdue University, Indianapolis. His research includes applied probability, mathematical statistics and reliability theory.

(right) Bikas K Sinha retired as Professor of Statistics from Indian Statistical Institute, Kolkata. He has done theoretical and applied research in many topics in statistics.

Sudoku is a popular combinatorial puzzle. We give a brief overview of some mathematical features of a Sudoku square. Then we focus on interpreting Sudoku squares as experimental designs in order to meet a practical need.

1. Sudoku Squares and Sudoku Puzzles

The (standard) Sudoku square of order 9 entails a 9×9 array of numbers 1 through 9 such that every row, every column and every 3×3 internal block contains each number exactly once. Sudoku squares of any other order $n = pq$ are also similarly defined as an $n \times n$ array of numbers 1 through n such that every row, every column and every $p \times q$ internal block contains each number exactly once. The internal blocks are generally shown as ‘bordered’ squares/rectangles of specified order(s).

A well-designed Sudoku puzzle consists of a partially filled array (the filled values are called clues) which can be extended to a complete Sudoku square in a unique way using pure logic and requiring no guesswork. To solve the puzzle, the player must find this complete Sudoku square solution (as quickly as possible).

We sketch a broad-stroke history of the Sudoku puzzle. In the late nineteenth century, a version of the puzzle appeared in French newspapers such as *L’Echo de Paris*, but it had disappeared by the time of World War I. The modern Sudoku puzzle first appeared under the name ‘Number Place’ in the New York puzzle magazine, *Dell Math Puzzles and Logic Problems*. In 1984, a Japanese puzzle company Nikoli introduced the puzzle under the name ‘*Suji wa dokushin ni kagiru*’ (meaning ‘the digits must be single’), which was shortened to Sudoku.

Keywords

Adjacency matrix, internal block, Latin square, orthogonal Latin squares, orthogonal Sudoku squares, Sudoku graph, Sudoku puzzle.



In 1997, Wayne Gould, a retired Chief District Judge in Hong Kong discovered a puzzle while visiting Japan. He began writing a computer program to produce more puzzles. He sent a puzzle to *The Times* in London, where it was first published in November 2004, and continues to be published ever since. By May 2005, the Sudoku craze had spread around the world and puzzles had begun to appear in newspapers in Australia, Canada, Israel, Eastern Europe, India, and eventually the United States. Today, the Sudoku puzzle has become a regular feature in many newspapers and magazines all over the world, inviting readers from all walks of life to solve this entertaining and addictive combinatorial puzzle. Many websites and computer applications offer the puzzle electronically on hand-held devices, making it accessible at all times in all places.

Sudoku puzzles are classified in increasing order of difficulty as easy, medium, hard, very hard and evil. Alternatively, the Sudoku puzzlers may be categorized as novice, player, expert and master. The likelihood of solving a puzzle and the time to solve it depends on the expertise and diligence of the player. Some websites give hints based on the next logical step a computer program would take in solving the puzzle.

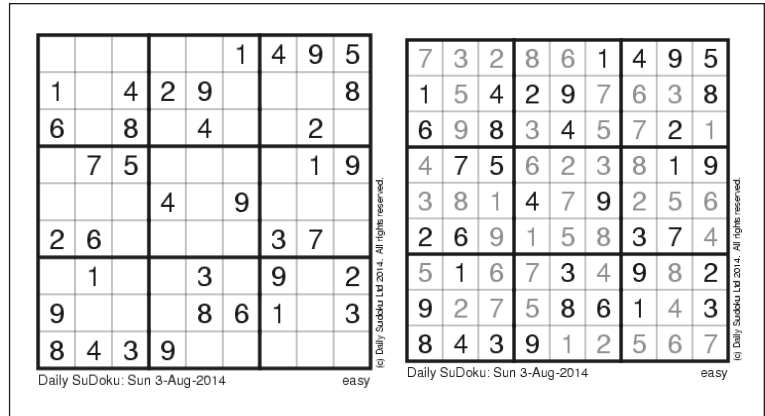
Figure 1 shows an example of a Sudoku puzzle that appeared in [1].



Figure 1. The daily Sudoku on the day this article was written.



Figure 2. The previous day's puzzle and its solution.



Typically, Sudoku solutions are available the next day or in the next issue of the magazine. For instance, we found the previous day's puzzle and its solution in [2] (*Figure 2*).

Although not required, in most Sudoku puzzles, the cells in which values are already given exhibit a half-turn (180° rotation) symmetry! See *Figures 1* and *2*.

It should be pointed out that the Sudoku solution only shows that the puzzle is a legitimate one. It helps neither the solver nor the player who failed to solve it, unless the solution includes the order in which the numbers are filled in. Such information seems to be missing in all published solutions the authors have come across.

Drawing a parallel to Magic Squares¹ (where numbers 1 through n are arranged in an $n \times n$ array in such a way that the sum in each row, in each column and along each diagonal is a constant $n(n^2+1)/2$), one may inquire whether the numbers along the diagonals of a Sudoku square are all distinct. In general, most Sudoku squares do not have distinct elements along its main diagonal or alternative diagonal. See *Figure 2*, where the main diagonal has (7, 5, 8, 6, 7, 8, 9, 4, 7) and the alternative diagonal has (5, 3, 7, 3, 7, 1, 6, 2, 8). Also, in the Sudoku puzzle in *Figure 1*, the alternative diagonal already has two 8's.

¹ For more information on Magic Squares, see articles at *Resonance*, Vol.11, No.9, pp.76–89, 2006, Vol.12, No.10, p.79, 2007 and Vol.15, No.8, pp.733–736, 2010.



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

Figure 3. A diagonal Sudoku square of order 9.

Needless to say that it requires a special care to construct Sudoku squares in which the main or the alternative diagonal contains distinct elements. In the rare cases when both the main diagonal and the alternative diagonal contain distinct elements, we call the Sudoku square a diagonal Sudoku square. *Figure 3* exhibits a diagonal Sudoku square.

A $p^2 \times p^2$ Sudoku board can be depicted as the Sudoku graph of rank p by letting each of the p^4 cells (numbered 1 through p^4 row-by-row) correspond to a vertex, and by connecting two vertices by an edge if and only if the cells that they correspond to are in the same row, or column, or internal block. The adjacency matrix of the Sudoku graph of rank p is a $p^4 \times p^4$ matrix $A = ((a_{ij}))$, with $a_{i,j} = 1$ if vertices i and j are connected, and 0 otherwise. It can be shown that for all $p \geq 2$, the eigenvalues of the adjacency matrix of the Sudoku graph of rank p are integers [3].

Furthermore, one can assign a distinct color to each number 1 through p^2 , and color all vertices of the Sudoku graph of order p that correspond to cells containing a given number with the color assigned to that number. Then, a Sudoku square becomes equivalent to a coloring of the corresponding Sudoku graph such that no two adjacent vertices have the same color, and solving a Sudoku puzzle becomes equivalent to completely coloring a partially-colored Sudoku graph.



A typical Sudoku puzzler need not worry about the following mathematical questions which arise naturally:

- How many Sudoku squares are there? (6670903752021072936960 $\approx 6.67 \times 10^{21}$ of order 9 [4]. Also, if the symmetries due to the permutation of numbers, bands (top, middle and bottom sets of three rows), rows within a band, stacks (left, middle and right sets of three columns), columns within a stack, and symmetries due to reflections and rotations of the entire square are not differentiated, then there are essentially 5,472,730,538 *distinct* Sudoku squares of order 9 [5]).
- How many solutions does a given Sudoku puzzle have? (As we mentioned before, a well-designed Sudoku puzzle has a unique solution. But, in general, it is very hard to determine (without the aid of a computer) if an arbitrary Sudoku puzzle has a solution; and if it does, whether the solution is unique or there are several solutions. However, the problem was completely solved – a method was developed to calculate the number of ways in which one can color the vertices of a partially colored graph in such a way that no two adjacent vertices have the same color [6]).
- What is the fewest number of entries that must be given so that the completed Sudoku square is unique? (The answer is 17 [7]; or, ≤ 18 , if we further require that the cells with given clues exhibit a half-turn symmetry. Also, see [8] for almost 50,000 *distinct* uniquely solvable Sudoku puzzles of order 9 with 17 clues).
- How is the level of difficulty of a Sudoku puzzle determined? (This can be quite subjective. Oftentimes, the difficulty rating depends on the relevance and the location of the clues rather than the number of clues. A computer program can determine the difficulty for a human player based on the complexity of the required solving techniques [9]).



However, the puzzler can benefit from answers to the question: What are some strategies for solving a Sudoku puzzle of a given difficulty level? We will not address this important question in this article, except to advise the interested puzzlers to search the Internet using key phrases such as ‘How to Solve a Sudoku?’ and ‘Math Behind Sudoku’. Hopefully, the puzzlers will arrive at helpful sites such as [10] and [11].

In this article, we will focus on using Sudoku squares as statistical experimental designs in order to meet some practical constraints. But no knowledge of statistics is needed to understand this article.

2. Sudoku Square as a Specialized Latin Square

Every Sudoku square is a special kind of a Latin square² (where numbers 1 through n are arranged in an $n \times n$ array such that every row and every column has each number exactly once). In fact, Sudoku squares form a tiny proportion of Latin squares of the same order. For example, there are about 5.525×10^{27} Latin squares of order 9, which is about 828,000-fold bigger than the number of Sudoku squares of order 9 [12].

² Also see *Resonance*, Vol.17, No.9, p.895–902, 2012.

Latin squares are used as statistical experimental designs that study the effect of three explanatory variables, each at n levels, on a response variable, using only n^2 experimental units. In a Latin square design (LSD), the experimental units are laid out in an $n \times n$ square array. The Row (R) represents one explanatory variable, the Column (C) represents a second explanatory variable, and the Treatment (T), shown as a number in the (i, j) -th cell of the square, represents a third explanatory variable.

Note that, in an LSD, the roles of R , C and T can be interchanged. Hence, we can say that every row and every treatment has each column exactly once; or that every treatment and every column has each row exactly once.



(i) T within $(R \times C)$, R <table style="display: inline-table; border-collapse: collapse;"> <tr><td style="border-right: 1px solid black; padding: 5px;">1</td><td style="padding: 5px;">2</td><td style="padding: 5px;">3</td><td style="padding: 5px;">4</td><td style="padding: 5px;">5</td></tr> <tr><td style="border-right: 1px solid black; padding: 5px;">2</td><td style="padding: 5px;">3</td><td style="padding: 5px;">4</td><td style="padding: 5px;">5</td><td style="padding: 5px;">1</td></tr> <tr><td style="border-right: 1px solid black; padding: 5px;">3</td><td style="padding: 5px;">4</td><td style="padding: 5px;">5</td><td style="padding: 5px;">1</td><td style="padding: 5px;">2</td></tr> <tr><td style="border-right: 1px solid black; padding: 5px;">4</td><td style="padding: 5px;">5</td><td style="padding: 5px;">1</td><td style="padding: 5px;">2</td><td style="padding: 5px;">3</td></tr> <tr><td style="border-right: 1px solid black; padding: 5px;">5</td><td style="padding: 5px;">1</td><td style="padding: 5px;">2</td><td style="padding: 5px;">3</td><td style="padding: 5px;">4</td></tr> </table>	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	(ii) C within $(R \times T)$, R <table style="display: inline-table; border-collapse: collapse;"> <tr><td style="border-right: 1px solid black; padding: 5px;">1</td><td style="padding: 5px;">2</td><td style="padding: 5px;">3</td><td style="padding: 5px;">4</td><td style="padding: 5px;">5</td></tr> <tr><td style="border-right: 1px solid black; padding: 5px;">5</td><td style="padding: 5px;">1</td><td style="padding: 5px;">2</td><td style="padding: 5px;">3</td><td style="padding: 5px;">4</td></tr> <tr><td style="border-right: 1px solid black; padding: 5px;">4</td><td style="padding: 5px;">5</td><td style="padding: 5px;">1</td><td style="padding: 5px;">2</td><td style="padding: 5px;">3</td></tr> <tr><td style="border-right: 1px solid black; padding: 5px;">3</td><td style="padding: 5px;">4</td><td style="padding: 5px;">5</td><td style="padding: 5px;">1</td><td style="padding: 5px;">2</td></tr> <tr><td style="border-right: 1px solid black; padding: 5px;">2</td><td style="padding: 5px;">3</td><td style="padding: 5px;">4</td><td style="padding: 5px;">5</td><td style="padding: 5px;">1</td></tr> </table>	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	(iii) R within $(T \times C)$, T <table style="display: inline-table; border-collapse: collapse;"> <tr><td style="border-right: 1px solid black; padding: 5px;">1</td><td style="padding: 5px;">5</td><td style="padding: 5px;">4</td><td style="padding: 5px;">3</td><td style="padding: 5px;">2</td></tr> <tr><td style="border-right: 1px solid black; padding: 5px;">2</td><td style="padding: 5px;">1</td><td style="padding: 5px;">5</td><td style="padding: 5px;">4</td><td style="padding: 5px;">3</td></tr> <tr><td style="border-right: 1px solid black; padding: 5px;">3</td><td style="padding: 5px;">2</td><td style="padding: 5px;">1</td><td style="padding: 5px;">5</td><td style="padding: 5px;">4</td></tr> <tr><td style="border-right: 1px solid black; padding: 5px;">4</td><td style="padding: 5px;">3</td><td style="padding: 5px;">2</td><td style="padding: 5px;">1</td><td style="padding: 5px;">5</td></tr> <tr><td style="border-right: 1px solid black; padding: 5px;">5</td><td style="padding: 5px;">4</td><td style="padding: 5px;">3</td><td style="padding: 5px;">2</td><td style="padding: 5px;">1</td></tr> </table>	1	5	4	3	2	2	1	5	4	3	3	2	1	5	4	4	3	2	1	5	5	4	3	2	1
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																																																																									
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																																																																									
1	5	4	3	2																																																																									
2	1	5	4	3																																																																									
3	2	1	5	4																																																																									
4	3	2	1	5																																																																									
5	4	3	2	1																																																																									

Figure 4. Various representations of the same Latin square design of order 5:

- (i) T within $(R \times C)$,
- (ii) C within $(R \times T)$, and
- (iii) R within $(T \times C)$.

See *Figure 4* for three representations of the same LSD, even though the Latin squares appear to be different.

Whenever a Latin square is usable as an experimental design, one can instead choose a Sudoku square of the same order as the experimental design. This will allow incorporating a fourth explanatory variable (internal block) *without increasing the number of experimental units*, and thereby provide an added practical benefit. For example, a standard Sudoku square of order $n = 9$ also forces every 3×3 internal block to contain each number exactly once (just as every row and every column does). Which additional variable can we identify with internal block?

Following [13] and [14], we discuss the use of a Sudoku as an experimental design in the context of a manufacturing example in which the Row (R) represents the suppliers (S) of raw material, the Column (C) represents the days (D) of operation, the Treatment (T) is identified with Machines (M) used in the operation, represented by the numbers 1 through 9, and the Internal Block (B) is identified with Operators (O), also represented by the

$S \backslash D$	$D1$	$D2$	$D3$	$D4$	$D5$	$D6$	$D7$	$D8$	$D9$
$S1$	$M1$	$M4$	$M7$	$M6$	$M2$	$M5$	$M9$	$M8$	$M3$
$S2$	$M9$	$M5$	$M8$	$M1$	$M7$	$M3$	$M4$	$M6$	$M2$
$S3$	$M2$	$M6$	$M3$	$M8$	$M9$	$M4$	$M5$	$M7$	$M1$
$S4$	$M6$	$M9$	$M5$	$M2$	$M4$	$M8$	$M3$	$M1$	$M7$
$S5$	$M7$	$M1$	$M4$	$M9$	$M3$	$M6$	$M2$	$M5$	$M8$
$S6$	$M3$	$M8$	$M2$	$M5$	$M1$	$M7$	$M6$	$M4$	$M9$
$S7$	$M5$	$M3$	$M1$	$M7$	$M6$	$M9$	$M8$	$M2$	$M4$
$S8$	$M4$	$M2$	$M6$	$M3$	$M8$	$M1$	$M7$	$M9$	$M5$
$S9$	$M8$	$M7$	$M9$	$M4$	$M5$	$M2$	$M1$	$M3$	$M6$
	$O 1$	$O 2$	$O 3$	$O 4$	$O 5$	$O 6$	$O 7$	$O 8$	$O 9$

Figure 5. A Sudoku of order 9 with one operator (O) assigned to each internal block.



numbers 1 through 9. See *Figure 5*. The identification of an internal block as a fourth explanatory variable was a simple yet profound achievement that gave a new life to the otherwise artificial requirements of a Sudoku square. A Sudoku-based statistical experimental design was born, which we now call a Sudoku design. With reference to the above example, the Sudoku of order 9 allows some flexibility on the availability of the 9 operators, who are available as teams of three for 3 days only: Operators $O1$, $O2$ and $O3$ will work on Days 1-2-3, Operators $O4$, $O5$ and $O6$ will work on Days 4-5-6, and Operators $O7$, $O8$ and $O9$ will work on Days 7-8-9.

3. Balanced Incidence Matrix and Orthogonality of Variables in a Sudoku Design

Consider the four explanatory variables in a Sudoku design: Row (R), Column (C), Internal Block (B) and Treatment (T). The incidence matrix for any two of these four variables is a two-way display of the number of units that receive each level combination of these two variables. Since there are exactly n^2 units in total, the two variables each with n options are said to be *perfectly balanced* if their incidence matrix is an $n \times n$ matrix with all entries 1; that is, the incidence matrix is $\mathbf{J}_n = \mathbf{1}_n \mathbf{1}'_n$, where $\mathbf{1}'_n = (1, 1, \dots, 1)$ is an n -component vector of unity; that is, every level combination of the two variables occurs in exactly one experimental unit.

The combinatorial requirement of perfect balance between two variables translates into these two variables becoming orthogonal to each other. Orthogonality of two variables means that the effect of one variable can be measured without it being contaminated by the effect of the other variable. Without orthogonality, measuring effects of variables may be impossible, or difficult and less precise. So, orthogonality is a very desirable statistical property. For further details, we refer the readers to any standard statistics textbook such as [15]. For the



Figure 6. Operator assignment in the Sudoku of *Figure 5*.

block	C1	C2	C3	C4	C5	C6	C7	C8	C9
R1 R2 R3		1			4			7	
R4 R5 R6		2			5			8	
R7 R8 R9		3			6			9	

purpose of this article, it suffices to interpret orthogonality between two variables as the combinatorial property of perfect balance between them.

We have already mentioned (in Section 2, paragraph 1) that a Latin square design exhibits row-column, row-treatment and treatment-column orthogonality. What about the orthogonality between any two variables in a Sudoku design, where we also have internal blocking as a fourth explanatory variable, besides the three variables already in a Latin square design? To answer this question, let us first number the internal blocks of the usual Sudoku square of order $n = 9$ column-by-column, as shown in *Figure 6* (and in *Figure 5*).

The most striking feature of a Sudoku design follows from its definition: Treatment (T) is orthogonal to all three variables: Row (R), Column (C) and Internal Block (B); that is, the incidence matrices for $T \times R$, $T \times C$, $T \times B$ are all \mathbf{J}_9 . Of course, by the definition of a square, the incidence matrix for $R \times C$ is already \mathbf{J}_9 , yielding row-column orthogonality. The incidence matrix for $R \times B$ and $B \times C$ are shown in *Figure 7*.

Figure 7. Incidence matrix for $R \times B$ (left) and $B \times C$ (right).

3	0	0	3	0	0	3	0	0	3	3	3	0	0	0	0	0	0
3	0	0	3	0	0	3	0	0	3	3	3	0	0	0	0	0	0
3	0	0	3	0	0	3	0	0	3	3	3	0	0	0	0	0	0
0	3	0	0	3	0	0	3	0	0	0	0	3	3	3	0	0	0
0	3	0	0	3	0	0	3	0	0	0	0	3	3	3	0	0	0
0	3	0	0	3	0	0	3	0	0	0	0	3	3	3	0	0	0
0	0	3	0	0	3	0	0	3	0	0	0	0	0	0	3	3	3
0	0	3	0	0	3	0	0	3	0	0	0	0	0	0	3	3	3
0	0	3	0	0	3	0	0	3	0	0	0	0	0	0	3	3	3



Hence, row and internal block variables are non-orthogonal; and column and internal block variables, too are non-orthogonal. Thus, of the $\binom{4}{2} = 6$ possible pairs of variables, 4 pairs are orthogonal and 2 are not.

Returning to the example in *Figure 5*, we note that Operators 1, 2, 3 are available on Days 1, 2, 3; Operators 4, 5, 6 on Days 4, 5, 6; and Operators 7, 8, 9 on Days 7, 8, 9. In spite of this restricted availability of the 9 operators, Machines are orthogonal to Suppliers, Days and Operators. Why is orthogonality so desirable? Because when orthogonality is achieved, it is known that comparison among the nine Machine Effects can be made with maximum accuracy! This precisely is the statistical utility resulting from the combinatorial beauty of a Sudoku design!

However, this statistical analysis of a Sudoku design [13] has the following shortcomings: No operator worked with raw materials from all 9 suppliers, and no operator worked on all 9 days. Consequently, operator–supplier and operator–day pairs of variables were non-orthogonal. This went unnoticed in the analysis provided in [13]. Subsequently, this was properly accounted for in [4]. Their analysis is quite intricate, but is not essential for this article.

In fact, if orthogonality between pairs of variables was the only concern, the literature already contains a perfect solution to ensure such orthogonality: One can simply choose a pair of orthogonal Latin squares (OLS) (or a Graeco–Latin squares or Euler squares), one representing Machine allocation and the other Operator assignment. Recall that two Latin squares are said to be orthogonal if, when superimposed, they exhibit every ordered pair of numbers exactly once. There is a well-developed mathematical theory for constructing a complete set of OLS. In fact, a pair of OLS exist for all orders $n \geq 3$, but $n \neq 6$ [16]. *Figure 8* displays a pair of



$S \setminus D$	D1	D2	D3	D4	D5	D6	D7	D8	D9
S1	M1, O1	M2, O2	M3, O3	M4, O4	M5, O5	M6, O6	M7, O7	M8, O8	M9, O9
S2	M2, O7	M3, O8	M1, O9	M5, O1	M6, O2	M4, O3	M8, O4	M9, O5	M7, O6
S3	M3, O4	M1, O5	M2, O6	M6, O7	M4, O8	M5, O9	M9, O1	M7, O2	M8, O3
S4	M4, O2	M5, O3	M6, O1	M7, O5	M8, O6	M9, O4	M1, O8	M2, O9	M3, O7
S5	M5, O8	M6, O9	M4, O7	M8, O2	M9, O3	M7, O1	M2, O5	M3, O6	M1, O4
S6	M6, O5	M4, O6	M5, O4	M9, O8	M7, O9	M8, O7	M3, O2	M1, O3	M2, O1
S7	M7, O3	M8, O1	M9, O2	M1, O6	M2, O4	M3, O5	M4, O9	M5, O7	M6, O8
S8	M8, O9	M9, O7	M7, O8	M2, O3	M3, O1	M1, O2	M5, O6	M6, O4	M4, O5
S9	M9, O6	M7, O4	M8, O5	M3, O9	M1, O7	M2, O8	M6, O3	M4, O1	M5, O2

Figure 8. An Orthogonal Latin Square Design (OLSD) of order 9.

9×9 OLS, forming an orthogonal Latin square design (OLSD) of order 9, wherein the first entry represents Machine and the second entry represents Operator.

Do mutually orthogonal Sudoku squares (OSS) (or Graeco Sudoku squares) exist? Indeed, they do [17]. Also, see [18] for the construction of OSS of odd order. As an illustration, the example in *Figure 9* superimposes an orthogonal (diagonal) Sudoku square to the orthogonal (diagonal) Sudoku square of *Figure 3*.

Note, however, that in any OLSD or in any OSSD, every operator must be available to work on every day. In contrast, for the design shown in *Figure 5*, each operator is required to work on only 3 days, which results in a significant operational efficiency. Thus, when operator availability is restricted, one *cannot* invoke an OLSD or an OSSD. On the other hand, the Sudoku design of *Figure 5* accommodates restrictions on the operator availability only at the cost of imposing non-orthogonality for both operator–supplier and operator–day. But, as already reasoned above, it is highly desirable that each

Figure 9. An Orthogonal (diagonal) Sudoku Square Design (OSSD) of order 9.

$S \setminus D$	D1	D2	D3	D4	D5	D6	D7	D8	D9
S1	M1, O1	M2, O2	M3, O3	M4, O4	M5, O5	M6, O6	M7, O7	M8, O8	M9, O9
S2	M7, O4	M8, O5	M9, O6	M1, O7	M2, O8	M3, O9	M4, O1	M5, O2	M6, O3
S3	M4, O7	M5, O8	M6, O9	M7, O1	M8, O2	M9, O3	M1, O4	M2, O5	M3, O6
S4	M6, O8	M4, O9	M5, O7	M9, O2	M7, O3	M8, O1	M3, O5	M1, O6	M2, O4
S5	M3, O2	M1, O3	M2, O1	M6, O5	M4, O6	M5, O4	M9, O8	M7, O9	M8, O7
S6	M9, O5	M7, O6	M8, O4	M3, O8	M1, O9	M2, O7	M6, O2	M4, O3	M5, O1
S7	M8, O6	M9, O4	M7, O5	M2, O9	M3, O7	M1, O8	M5, O3	M6, O1	M4, O2
S8	M5, O9	M6, O7	M4, O8	M8, O3	M9, O1	M7, O2	M2, O6	M3, O4	M1, O5
S9	M2, O3	M3, O1	M1, O2	M5, O6	M6, O4	M4, O5	M8, O9	M9, O7	M7, O8



$S \setminus D$	$D1$	$D2$	$D3$	$D4$	$D5$	$D6$	$D7$	$D8$	$D9$
$S1$	$M1, O1$	$M2, O2$	$M3, O3$	$M4, O4$	$M5, O5$	$M6, O6$	$M7, O7$	$M8, O8$	$M9, O9$
$S2$	$M4, O1$	$M5, O2$	$M6, O3$	$M7, O4$	$M8, O5$	$M9, O6$	$M1, O7$	$M2, O8$	$M3, O9$
$S3$	$M7, O1$	$M8, O2$	$M9, O3$	$M1, O4$	$M2, O5$	$M3, O6$	$M4, O7$	$M5, O8$	$M6, O9$
$S4$	$M2, O3$	$M3, O1$	$M1, O2$	$M5, O6$	$M6, O4$	$M4, O5$	$M8, O9$	$M9, O7$	$M7, O8$
$S5$	$M5, O3$	$M6, O1$	$M4, O2$	$M8, O6$	$M9, O4$	$M7, O5$	$M2, O9$	$M3, O7$	$M1, O8$
$S6$	$M8, O3$	$M9, O1$	$M7, O2$	$M2, O6$	$M3, O4$	$M1, O5$	$M5, O9$	$M6, O7$	$M4, O8$
$S7$	$M3, O2$	$M1, O3$	$M2, O1$	$M6, O5$	$M4, O6$	$M5, O4$	$M9, O8$	$M7, O9$	$M8, O7$
$S8$	$M6, O2$	$M4, O3$	$M5, O1$	$M9, O5$	$M7, O6$	$M8, O4$	$M3, O8$	$M1, O9$	$M2, O7$
$S9$	$M9, O2$	$M7, O3$	$M8, O1$	$M3, O5$	$M1, O6$	$M2, O4$	$M6, O8$	$M4, O9$	$M5, O7$

operator should work with raw materials from all 9 suppliers. Thus, even though we are willing to sacrifice operator–day orthogonality in order to accommodate restricted availability of operators, we would like to preserve operator–supplier orthogonality. Is that possible? Indeed, it is! We display one such Sudoku design in *Figure 10*.

Figure 10 . A Sudoku design with orthogonal incidence structure for operator–supplier, when operators are available as a team of three in sets of three days.

In *Figure 10*, as in *Figure 5*, the incidence matrices for $S \times D$, $S \times M$, $D \times M$, $O \times M$ are all \mathbf{J}_9 . Additionally, in *Figure 10*, the incidence matrix for $S \times O$ is also \mathbf{J}_9 , unlike that for $R \times B$ in *Figure 5*. But the incidence matrix for $O \times D$ in *Figure 10*, as for $B \times C$ in *Figure 5*, is **not** \mathbf{J}_9 . Thus, of the 6 possible pairs of variables only one pair exhibits non-orthogonality.

The design in *Figure 10* is an illustration of a special sub-class of Sudoku designs called the cylindrical-shift Sudoku designs. We do not discuss any statistical properties of such Sudoku designs in this article. Nor do we discuss how to incorporate additional explanatory variables orthogonal to the existing variables, or how to allow other scenarios of operator availability such as Operator i is available on days $i, i + 1, i + 2$ (modulo 9), or on days $i, i + 2, i + 4$ (modulo 9), and still maintain operator–supplier orthogonality. For such matters, we refer interested readers to [19].

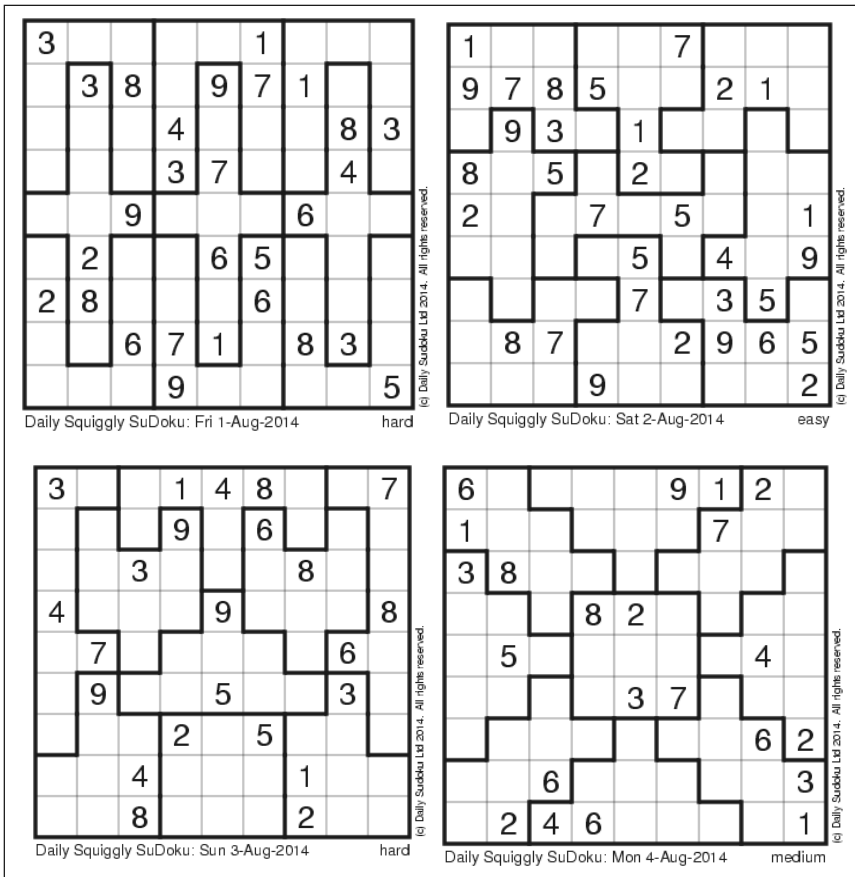


4. Irregular Sudoku Puzzles

We conclude this article by displaying some irregular Sudoku puzzles that appeared in [20] (*Figure 11*). Here, ‘irregular’ means the bordered internal blocks are allowed to be of any shape (not necessarily squares/rectangles) formed by contiguous cells. Although not absolutely necessary, the irregular-shaped (or squiggly) internal blocks usually exhibit, perhaps for aesthetic reasons, some form of symmetry – lateral reflection, vertical reflection, quarter-turn (90° rotation), or half-turn (180° rotation).

Furthermore, although not required, in all of these irregular Sudoku puzzles, the cells in which values are already given exhibit either a reflection or a rotation symmetry!

Figure 11. Some irregular Sudoku puzzles.



Glossary of Some Terms

Units: Single entities on which experimental conditions are applied and on which measurements are taken; as for example, plots in an agricultural experiment.

Response variable: The main variable of primary interest in a statistical study; as for example, yield of a variety of wheat grown in a plot.

Explanatory variables: Other variables that help us learn about the response variable once we can figure out how these variables relate to one another; as for example, plot size, quantity of rainfall, amount of manure, etc.

Treatment: An experimental condition assigned to the units by the experimenter; as for example, wheat varieties applied in experimental plots.

Levels: Distinct values of a variable; as for example, small, medium, large quantities of manure applied per plot.

Statistical experimental design: Assignment of treatments to units in a planned manner accounting for various constraints; as for example, planning an agricultural experiment involving experimental plots, varieties of wheat, different levels of soil fertility, different amount of irrigated water etc. When planned well, an experimental design can extract maximum information using as few units as are necessary.

Latin square: A $p \times p$ array with entries taken from $\{1, 2, \dots, p\}$ in such a way that every row (column) contains each symbol exactly once. A Latin square experimental design uses the rows of a Latin square as the p classes of explanatory variable A , its columns as the p classes of explanatory variable B , and the p symbols as the levels of a treatment variable. The total number of units is only p^2 (and not p^3 as a complete experiment would require).

Orthogonal Latin squares, or Graeco–Latin squares, or Euler squares: Two Latin squares which when superimposed, exhibit every pair of ordered symbols formed from $\{1, 2, \dots, p\}$ exactly once. An OLS experimental design assigns levels of two treatment variables to each of the p^2 experimental units.

Acknowledgment

The authors thank the associate editors and two anonymous referees for their guidance in determining the right focus for this article, and in suggesting several mathematical properties of Sudoku squares.

Suggested Reading

- [1] *Daily Sudoku Puzzle*, 4 Aug 2014. <http://dailysudoku.com/sudoku/>
- [2] *Daily Sudoku Puzzle*, 3 Aug 2014. http://dailysudoku.com/sudoku/archive/2014/08/2014-08-3_solution.shtml



- [3] T Sander, Sudoku graphs are integral, *The Electronic Journal of Combinatorics*, Vol.16, No.25, 2009.
- [4] B Felgenhauer and F Jarvis, Enumerating possible Sudoku grids, Unpublished, <http://www.afjarvis.staff.shef.ac.uk/sudoku/sudoku.pdf>
- [5] E Russell and F Jarvis, Mathematics of Sudoku II. Unpublished, http://www.afjarvis.staff.shef.ac.uk/sudoku/russell\jarvis_spec2.pdf
- [6] A M Herzberg and M R Murty, Sudoku squares and chromatic polynomials, *Notices of the American Mathematical Society*, Vol.54, No.6, pp.708–717, 2007.
- [7] G McGuire, B Tugemann and G Civario, There is no 16-clue Sudoku: Solving the Sudoku minimum number of clues problem, Unpublished, <http://arxiv.org/pdf/1201.0749v1.pdf>
- [8] Gordon Royle, Minimum Sudoku, <http://staffhome.ecm.uwa.edu.au/~00013890/sudokumin.php>
- [9] T Davis, The Mathematics of Sudoku, Unpublished, <http://www.geometer.org/mathcircles/sudoku.pdf>
- [10] How to Solve Sudoku Puzzles, About.com Puzzles, http://puzzles.about.com/library/sudoku/blsudoku_tutorial01.htm
- [11] The Math Behind Sudoku, Cornell University Department of Mathematics, <http://www.math.cornell.edu/~mec/Summer2009/Mahmood/Intro.html>
- [12] Sammel and J Rothstein, The number of 9×9 Latin squares, *Discrete Math*, Vol.11, pp.93–95, 1975.
- [13] J Subramani, and K N Ponnuswamy, Construction and analysis of Sudoku designs, *Model Assisted Statistics and Applications*, Vol.4, pp.287–301, 2009.
- [14] F A Saba and B K Sinha, SuDoKu as an experimental design – Beyond the traditional Latin square design, *Statistics and Applications*, Vol.12, (1 & 2), pp.15–20, 2014.
- [15] C R Rao, *Linear statistical inference and its applications*, Wiley, New York, 1973.
- [16] Orthogonal Latin Squares, *Encyclopedia of Mathematics*, Springer and the European Mathematical Society, http://www.encyclopedia-ofmath.org/index.php/Orthogonal_Latin_squares
- [17] S Golomb, Proposed Problem, *American Mathematical Monthly*, Vol.25, No.3, 268, 2006.
- [18] J Subramani, Construction of Graeco Sudoku square designs of odd order, *Bonfring International Journal of Data Mining*, Vol.2, pp.37–41, 2012.
- [19] J Sarkar and B K Sinha, Variants of SuDoKu as useful experimental designs, *Statistics and Applications*, Vol.12, (1 & 2), pp.35–60, 2014.
- [20] Daily Sudoku Puzzle, 1–4 Aug 2014, <http://dailysudoku.com/sudoku/archive/index.html>

*Address for Correspondence*Jyotirmoy Sarkar¹Bikas K Sinha²¹ Indiana University-Purdue
University

Indianapolis, USA

Email: jsarkar@iupui.edu² Indian Statistical Institute

Kolkata

Email:

bikassinha1946@gmail.com