Estimation of Missing Value in Sudoku Square Design

Abstract

Missing values or missing data occur in experiments as a result of several reasons, these reasons could be natural or it happened due to failure on the part of experimenter. When missing value occurred it causes biasness to the analysis and failure in the efficiency. The study considered the Sudoku square design of order m^2 where row-blocks and column-blocks are equal, rows and columns are equal with one missing value. The missing value is estimated by comparing the missing value in respect to Latin square of order m^2 and also in respect to randomized block design, the estimator for the missing value is derived and numerical illustration is given to show how the estimator is used to obtain the estimate of a missing value when $k = 1$ and $m = 2$ in a squared Sudoku design.

Keywords: missing value, Sudoku square design, estimate, estimator, rowblock, columnblock

1.0 Introduction

Many years ago, attempts have been made on the conduct and analysis of experimental design in areas of agriculture, industry and so on. As observed by several authors such as Subramani and Ponnuswamy[1], Cochran and Cox[2], one cannot do without the case of missing values. No matter how well- planned, adequate and control an experiment is carried out, the situation like this could be caused by any form of natural disaster such as flood, fire outbreak, diseases and pests just to mention a few. The analysis of the resulting data that contains missing values that would still retain the original design was suggested by authors like Allan and Wishart[3], Subramanian [4], Subramani and Ponnuswamy [1], and Subramani and Aggarwal [5]. Though several authors such as Rubin [6], Little [7], Allision [8] and others have suggested methods to be used in case of missing values but such methods de-face the design of the experiment and also many important information got lost.

Yates [9] stated that the estimate of the missing observation can be obtained by minimizing the residual sum of squares. If assuming x is the unknown value of the missing observation by minimizing the residual error sum of squares given in the analysis of variance, one may obtain the estimate of the missing value as

$$\hat{X} = \frac{r\left(R' + C' + \sum^{k} T_i{}'\right) - (k+1)G'}{(r-1)(r-(k+1))} \qquad (1)$$

Where R', C', $T_i{}'$ are respectively the row, column and treatment (type$i$) total corresponding to the missing value and G' is the ground total of all known observation r is the number of rows and k is the order of the design.

Subramani [10] used method of non-iterative least square to estimate missing values in hyper-Graeco latin square design of kth order with r treatment in each of k types in r-rows and r-columns. Subramani [10] in his paper made an attempt to obtain explicit expressions for the estimators of the several missing values in Hyper-Graeco-Latin square designs. He further showed that the estimates of the missing values in Latin square designs and graeco-latin square designs are obtained as a particular case of the estimates of the missing values in hyper-graeco-latin square designs.

Estimation of one missing observation in randomized block designs using least square method was given by Allan and Wishart [3]. Yates [9] showed that through the values that minimize the residual sum of squares, one can obtain the correct least squares estimates of all estimable parameters as well as the correct residual sum of squares.

Bhatra and Daharamyadav [11] used least square method to estimate the values of missing observations in a random block design they used two approaches which are as follows. First approach was that the positions of the missing values are considered, the second approach used the RBD with the same number of missing observations with the assumption that the position of missing observations are not known at the early stage, some denotations were made. Then the residual sum of squares is obtained and the estimate of the missing observations were for the two approaches. The two approaches were compared and noted that two methods are the same.

Subramani and Ponnuswamy [1] observed that one has to face problem of missing values while conducting agricultural field experiments due to several reasons, such as the plant may be eaten away by animals or washed away by floods etc. Similarly, the agricultural plot yeilds may be mixed up during the time of transportation or due to handling at different stages. In all such situations, the resulting data become incomplete and are referred to as non-orthogonal data.

Bhatra and Daharamyadav [11] estimated two missing observations by specifying positions and use the method of principle of least squares. Collin et al. [12] identify that missing values may negatively impacted the analyses, interpretation and conclusion. It also posed bias estimation of parameters, inflate error rates (type I and II) as well as the decrease of statistical power.

Many methods have been adopted for handling missing values in statistical analysis. Some of the methods show that the approaches are default of statistical software packages which eventually lead to loss of efficiency owing to the fact that some observations have been discarded on the basis that some rows or columns contain missing values that must be deleted and biasedness in estimates.

Many authors like Subramani [13], Subramani and Ponnuswamy [14], Hui-Dong and Ru-Gen [15], Dauran et al. [16], Shehu and Dauran [17], Danbaba and Shehu [18], Danbaba [19], Danbaba [20], Danbaba [21] and many others. These authors have written on the areas of analysis of variance (ANOVA), construction of models, Construction of graeco Sudoku square design, multivariate analysis and variance components of Sudoku square models.

However, the area of estimation of missing value or missing data for Sudoku square models has not been given attention. It is on this note, that this study proposed a method of estimation of missing values for Sudoku square design that will produce correct estimate that is free of bias, no loss of vital information and no loss of efficiency.

2.0 Methodology

The method, this study proposed to use for the estimation of the missing values in Sudoku square design is to compare the missing observation in a small square and in a bigger Latin square. That is, to say the value of missing observation in a small square, assuming the small square is a ramdomized block design is equal in value with the missing value in whole Latin square. This method will only be used for estimation when one observation is missing, while some several missing values will be estimated with the help of some Lemma, ANOVA model and the estimator when one observation is missing.

Considering the missing value in a small square of a Sudoku square model as randomized block as

$$x = \frac{cR_s + rC_s - g}{(c-1)(r-1)} \qquad (2)$$

R_s is sum of row entries corresponding to the missing value in a small square

C_s is sum of column entries corresponding to the missing value in a small square

g is total sum of known values in the small square

r number of rows

c number of columns

The estimation of missing value in respect of the whole plot is considered as Graeco-Latin square

$$z = \frac{m\left(R_i + C_j + \sum_{i=1}^{k} T_k\right) - (k+1)G}{(m-1)(m-[(K+1)]} \quad (3)$$

Subramani [10]

R_i is sum of row entries corresponding to the missing value in a large square

C_j is sum of column entries corresponding to the missing value in a large square

T_k is sum of treatment entries corresponding to the missing value in a large square

G is the total sum of existing values in the large square

3.0 Estimation of Missing Values of Sudoku Square Design

For the missing value if we consider the ANOVA model below

$$Y_{ij\,(k,lp,q)} = x\beta + e_{i,j\,(k,l,p,q)} \quad (4)$$

		$CB1$				$CB2$					m	
		1			m	$m+1$	$m+2$		2m		$m-1)m+1$		m^2
$RB1$	1	1	2		m	$m+1$	$m+2$		$2m$		$m-1)m+1$		m^2
	2	$m+1$	$m+2$		$2m$	$2m+1$	$2m+2$		$3m$		1		m
	⋮	⋮	⋮		⋮	⋮	⋮		⋮		⋮		⋮
	m	$(m-1)m+1$	$m-1)m+2$		m^2	1	2		m		$m-2)m+1$		$m-1)m$
$RB2$	$m+1$	2	3		$m+1)$	$m+2$	$m+3$		$2m+1$		$(m-1)m+2$		1
	$m+2$	$m+2$	x		$2m+1$	$2m+2$	$2m+3$		$3m+1$		2		$m+1$
	⋮	⋮	⋮		⋮	⋮	⋮		⋮		⋮		⋮
	$2m$	$(m-1)m+2$	$(m-1)m+3$		1	2	3		$m+1$		$(m-1)m+2$		$(m-1)m+1$
⋮	⋮	⋮	⋮		⋮	⋮	⋮		⋮		⋮		⋮
	$(m-1)m+1$	m	$m+1$		$2m-1$	$2m$	$2m+1$		$3m-1$		m^2		$m-1$
	$(m-1)m+2$	$2m$	$2m+1$		$3m-1$	$3m$	$3m+1$		$4m-1$		m		$2m-1$
RBm	⋮	⋮	⋮		⋮	⋮	⋮		⋮		⋮		⋮
	m^2	m^2	1		$m-1$	m	$m+1$		$2m-1$		$(m-1)m$		m^2-1

Fig 1. Sudoku square design of order m^2containing a missing value

x is the missing value in the large square, since the latin square of order m^2, therefore the missing value in Hyper-graeco latin square from equation (3)

$$x = \frac{m(R_i+C_j+\sum_{i=1}^{k} T_k)-(k+1)G}{(m-1)(m-[(K+1)]}$$

$m+1$	2	3
$m+2$	$m+2$	x
⋮	⋮	⋮

$2m$	$(m-1)m+2$	$(m-1)m+3$

Fig 2. Sudoku design with missing value in a sub-square.

Assuming the small square is a ramdomized block design, if the order of the small square is $r \times c$ and let z be the missing value in the small square from equation (2)

$$x = \frac{R_s + x}{r} + \frac{C_s + x}{c} - \frac{g + x}{rc}$$

The value of z in the small square is equal to the value of x in the large square

$$\frac{R_s + z}{r} + \frac{C_s + z}{c} - \frac{g + z}{rc} = \frac{m\left(R_i + C_j + T_k\right) - 2(k+1)G}{(m-1)(m-[(K+1)]}$$

$$\frac{c(R_m + z) + r(C_m + z) - (g + z)}{rc} = \frac{m\left(R_i + C_j + T_k\right) - (k+1)G}{(m-1)(m-[(K+1)]}$$

$$z = \frac{cr\left(m\left(R_i + C_j + T_k\right) - (k+1)G\right)}{(m-1)(m-[(K+1)](r+c-1)} - \frac{cR_s + rC_s - g}{r+c-1}$$

$$z = \frac{cr\left[m\left(R_i + C_j + T_k\right) - (k+1)G\right] - (cR_s + rC_s - g)(m-1)[(m-[(k+1)]}{(m-1)[(m-[(k+1)](r+c-1)} \qquad (5)$$

When several observations are missing, one may decided by the use of above estimator repeatedly or the use of non-iterative procedure given by Subramani and Ponnuswamy[1] to estimate the missing values. It has its merit, when using non-iterate method, it gives direct estimates for the missing values.

Numerical Illustration

	CB1		CB2		
$RB1$	5113 (A)	5398 (B)	5307 (C)	4678 (D)	
	5346 (C)	4719 (D)	X (A)	5272 (B)	15337
$RB2$	5272 (B)	4986 (C)	4410 (D)	5254 (A)	
	4748 (D)	4919 (A)	4986 (B)	4542 (C)	
Total			14703		74950

Fig 3. Numerical Illustration

If we put $k = 1$ into equation (5)

$$z = \frac{cr\left[M\left(R_i + C_j + T_k\right) - (k+1)G\right] - (cR_s + rC_s - g)(m-1)[(m - [(k+1)]}{(m-1)[(m - [(k+1)](r+c-1)}$$

$$z = \frac{cr\left[M\left(R_i + C_j + T_k\right) - 2G\right] - (cR_s + rC_s - g)(m-1)[(m-2]}{(m-1)[(m-2](r+c-1)}$$

$c = r = 2 \quad M = 4 \quad R_m = 5272 \;\; C_m = 5307 \quad g = 15257$

$R_i = 15337 \qquad C_j = 14703 \quad T_k = 15286 \;\; G = 74950$

$$x = \frac{4 \times (4(15337 + 14703 + 15286) - 2 \times 74950) - 3(2 \times 5272 + 2 \times 5307 - 15257)}{(4^2 - 3 \times 4 + 2)(2 + 2 - 1)}$$

$$= 5011.667$$

$\cong 5012$

4.0 Conclusion

In this study, the estimator for missing value in Sudoku square design, when the number of rowblocks and columnblocks are equal as well as number of rows and columns are equal is obtained. The estimator is devoid of loss of some vital information, there was no deletion of any row or column as in case of complete linkage method of estimation of missing value, Numerical example was given to illustrate how it works.

Reference

[1] Subramani, J. & Ponnuswamy, K.N., (1989). A non-iterative least square estimation of missing values in experimental designs *Journal of Application Statistics*. 16,77-86.

[2] Cochran, W. G., & Cox, G. M., (1957): *Experimental Designs***,** 2nd edn. New York. Wiley.

[3] Allan, F. & Wishart, J. (1930.) A method of estimating the yield of a missing plot in field experimental work. *Journal of Agricultural Science,***20**(3), 399-406.

[4] Subramani, J. (1989). Estimation of several missing values in experimental designs. *Journal of Application Statistics*. **16**, 77-86.

[5] Subramani, J. & Aggarwal M.L. (1993). Estimation of several missing values in F- Square Designs designs. *Biometrical Journal*, **35**: 455-463

[6] Rubin D.B. (1972). A non-iterative algorithm for least squares estimation of missing values in any analysis of variance design. *Appl. Statist* **21**, 136-14.

[7] Little, R. J.A. (1986). Survey nonresponse adjustments. *Int. Statis. Rev* **54,** 134-157

[8] Allision, P.D. (2003). Missing data techniques for structural equation modelling. *Journal of Abnormal Psychology*, **112**(4),545-557.

[9] Yates, F. (1933).The Analysis of replicated experiments when the field results are incomplete. *Journal of Experimental Agriculture*, 129 -142.

[10] Subramani, J. (1993). Non-iterative least square estimation of missing values in Hyper-Graeco Latin square designs. *Biometika Journal* **35** (A) 465-470.

[11] Bhatra charyulu N. Ch & Dharamyadav T.(2013). Estimation of observation in ramdomized block design. *International Journal of Technology and Engineering Science.* **6**, 618-621.

[12] Collins, L. M., Schafer, J. L., & Kam, C. M (2001). A comparison of inclusive and restrictive strategies inmodern missing data procedures. *Psychological Methods*, *6*, 330-351.

[13] Subramani, J. (2012). Construction of Graeco Sudoku Square Designs of Odd Orders, *Bonfring International Journal of Data Mining*, 2 **(2)**, 37 – 41.

[14] Subramani, J. & Ponnuswamy, K.N. (2009). Construction and analysis of Sudoku designs, *Model Assisted Statistics and Applications*, 4**(4)**, 287-301.

[15]Hui-Dong, M. & Ru-Gen, X. (2008). Sudoku Square — a New Design in Field *Experiment, Acta Agron Sin*, 34(9), 1489–1493.

[16] Dauran, N. S., 2Odeyale, A. B. and *3Shehu, A. (2020) Construction and Analysis of Balanced Incomplete Sudoku Square Design *FUDMA Journal of Sciences (FJS iSSN online: 2616- 1370 ISSN print: 2645 – 2944 Vol. 4 No. 2, June, 2020, pp 290 – 299*

[17] Shehu, A. and Dauran, N. S. (2020) Manova: Power Analysis of Models of Sudoku Square Designs *FUDMA Journal of Sciences (FJS) SSN online: 2616-1370 ISSN print: 2645 – 2944 Vol. 4 No. 2, June, 2020, pp 350 – 364*

[18] Danbaba A. and Shehu A. (2016).On the Combined Analysis of Sudoku Square Designs with Some Common Treatments International Journal of Statistics and Applications 2016, 6(6): 347-351

[19] Danbaba, A (2016). Construction and Analysis of Samurai Sudoku. International *Journal Mathematical, Computational, Physical,Electrical and Computer Engineering* Vol.**10** No.4, 126-131.

[20] Danbaba, A. (2016a).Combined Analysis of Sudoku Square Designs with same treatments World Academy of Science and Technology. *International Journal of Mathematical,Computational, Physical,Electrical and Computer Engineering* Vol.**10** No. 4 155-159.

[21] Danbaba, A. (2016b). Construction and Analysis of Partially Balanced Sudoku Design of Prime Order Conference Proceedings Dubai UAE,April 08-09,2016, **18**(4) Part iv. pp 519- 520.