One Missing Observation in Graeco Latin Square Design: An Approximate Analysis of Variance

Author's Details:

[1]Joseph A Kupolusi [2]Oluwadare O Ojo

[1,2] Department of Statistics-Federal University of Technology Akure, Nigeria

Corresponding author's email: **jakupolusi@futa.edu.ng**

Abstract

Background: *Experimental results can seriously be affected by different degrees of variation that arise from unknown or uncontrollable design factors which is not of interested to the experimenter that may probably has effect on the response. Blocking is an extremely important design technique that can be used to systematically eliminate the effect of the uncontrollable design factors. Graeco-Latin Square design is used to eliminate three sources of variability; that is, it systematically allows blocking in three directions. Thus, rows, columns and Greek letters actually represent three restrictions on randomization. There are situations whereby one observation is occasionally missed in a Graeco Latin Square design of order P x P. In this paper, we proposed an approximate method for one missing observation in Graeco Latin Square Design of any order.*

Results: *The proposed approximate method was applied to a data set of Graeco Latin Square Design of order 4 and the results are presented in ANOVA tables. Based on this result, Mean Sum of Squares (MSE) reduced drastically compared to that of complete data set. Reduction in MSE is a clear indication that the proposed method can be used to obtain a better result with a minimum variance and unbiased estimate.*

Conclusion: *The result of the analysis indicated that the proposed approximate method for Graeco Latin Square is appropriate for estimation of missing observation through a simulation study of 1000 experimental runs. The result converged to the real data set. Hence, the method derived in this research is capable to handling the problem of missing observation in Graeco-Latin Square design.*

Keywords: *Latin Square, Graeco Latin Square, ANOVA, missing data, experimental error*

1. Background

The result of an experiment can seriously be affected by different degrees of variation that arise from nuisance factors. Kittiwat, S. & Kanogkan, L. (2017). A nuisance factor can be defined as unknown or uncontrollable design factor that is not of interested to the experimenter which may probably has effect on the response. Sometimes, the existence of this factor is latent and it can even be changing levels while we are performing the experiment. Randomization technique is used to guard such nuisance factors. Blocking is an extremely important design technique that can be used to systematically eliminate the effect of the nuisance factor from statistical comparison among treatments. Graeco-Latin Square design is used to eliminate three sources of variability; that is, it systematically allows blocking in three directions. Thus, rows, columns and Greek letters actually represent three restrictions on randomization. Two Latin squares are said to be orthogonal if each pair of letters appears exactly once in the superimposed squares. The superimposed square is called Graeco-Latin square design. There are situations under certain condition that experimenters might have the problem of a set of incomplete experimental observations. This can be categorized into two situations: (1) the deliberate plan to feature the incomplete observations due to limited number of experimental units. This can be existence of unbalanced arrangement or balance characteristic like Youden square design, the balanced incomplete block design (BIBD), and the balanced incomplete Latin square design (BILSD) proposed respectively by Youden (1937), Yates (1936), and Ai, Li, Liu, and Lin (2013), (2) situations whereby the incomplete observation occurs accidentally which might be as a result of bad control of some variables, the reading values from experiment are abnormal or not observed. Hence, their values might be cut from a set of observations, leading to the unbalanced or asymmetrical arrangement.

The analysis of incomplete data problem was first considered by Allan and Wishart (1930) in their paper based on differentiation of overall mean. Yates (1933) and Sirikasemsuk (2016a) proposed non-iterative and iterative missing plot techniques whereby the differential calculus was utilized to determine the missing experimental data with minimal error sum of squares. The estimates of the missing experimental data increases bias of the treatment sum of squares. Thus, the bias is determined and subtracted from the initial treatment sum of squares (Little & Rubin, 2002).

Existing methods to solve the incomplete-data experimental problems are tabulated in Table 1. Many recent research studies considered aspects of combinatorics, examples of which were the studies on the construction of the orthogonal Latin squares by Zhang (2013) and Donovan and Şule Yazıcı (2014); and the studies on the completability of the incomplete Latin squares from the partial Latin squares by Euler (2010) and Casselgren and Häggkvist (2013).

Table 1. Existing methods to solve the incomplete-data experimental problems as reported by Kittiwat, S. & Kanogkan, L. (2017)

	Method	Description	Author
1	Missing plot technique by minimizing the error sum of squares with non-iterative method	Differentiating the estimated parameter of the overall mean with respect to each missing value	Allan and Wishart (1930)
		General method for estimating several missing values in Latin square design	Kramer and Glass (1960)
		Differentiating the error sum of squares to each missing value (when only one observation is missing)	Yates (1933)
		Non-iterative Rubin method	Rubin (1972)
2	Missing plot technique with iterative method Iterative	Yates method (based on the work of Allan and Wishart (1930) when more than one observations are missing)	Yates (1933)
		Healy-Westmacott method based on regression imputation	Healy and Westmacott (1956)
3	Exact approach with general linear model	General regression significance test	Montgomery (2008)
4	Multiple imputation (MI) method	A combination of raw maximum likelihood and EM method	Rubin (1987)
5	Expectation maximization algorithm (EM Algorithm)	Iterative method with maximum likelihood estimation	Dempster, Laird, and Rubin (1977)
6	Analysis of covariance (ANCOVA) technique	A combination of regression analysis and ANOVA consisting of the covariate	Coons (1957), Cochran (1957), and Wilkinson (1958)
7	One missing value problem in Latin square design of any order	Exact analysis of variance	Kittiwat S. & Kanogkan L. (2017)

There are situations whereby one observation is occasionally missed in a Graeco Latin Square design of order P x P, this can pose a serious threat to the analysis of variance if not treated with care. Various methods to handle missing data have been discussed in the literature which include but not limited to multiple imputation method by (Rubin, 1987), Missing plot technique with iterative method by (Yate, 1933), Analysis of covariance (ANCOVA) technique by (Coons, 1957), (Cochran, 1957), and (Wilkinson, 1958) and Expectation maximization algorithm (EM Algorithm) by (Dempster, Laird, and Rubin, 1977). All these methods have successfully estimated missing observation in one way or the other. This work focused at obtaining an approximate missing data method for Graeco Latin Square design.

The rest of the paper is organized as follows. In section 2, we provide the Graeco-Latin Square layout and its Analysis of Variance (ANOVA) table while section 3 discusses derivation of the proposed method and research methodology. Application of the method in ANOVA table are provided in section 4. The discussion of the analysis are presented in section 5 while section 6 handled conclusion. Abbreviations used in the research are listed in section 7

2. **Methods**

The proposed statistical method and the underlined model is presented in this section.

The research is aimed at developing an approximate method to determining missing observation in Graeco Latin Square Design

The layout of Graeco-Latin Square design is given in Table 2

Table 2: Layout of Graeco-Latin Square design

	Column			
Row	$A\alpha$	$B\beta$	$C\gamma$	$D\delta$
	$B\delta$	$A\gamma$	$D\beta$	$C\alpha$
	$C\beta$	$D\alpha$	$A\delta$	$B\gamma$
	$D\gamma$	$C\delta$	$B\alpha$	$A\beta$

The null hypothesis of equal row, column, Latin letter and Greek letter treatments would be tested by dividing the corresponding mean square by mean square error. The rejection region is the upper tail point of the $F_{p-1,(p-3)(p-1)}$ distribution.

Table 3: Analysis of Variance (ANOVA) for Graeco Latin Square Design

Source of variation	Sum of Square	Degree of freedom
Latin letter treatments	$SSl = \frac{1}{p}\sum_{j=1}^{p} y_{.j..}^2 - \frac{y_{....}^2}{N}$	P-1
Greek letter treatments	$SSG = \frac{1}{p}\sum_{k=1}^{p} y_{..k.}^2 - \frac{y_{....}^2}{N}$	P-1
Row	$SSRow = \frac{1}{p}\sum_{i=1}^{p} y_{i...}^2 - \frac{y_{....}^2}{N}$	P-1
Column	$SSCol. = \frac{1}{p}\sum_{l=1}^{p} y_{...l}^2 - \frac{y_{....}^2}{N}$	P-1
Error	SSE(by Subtraction)	(P-3)(P-1)
Total	$SST = \sum\sum\sum\sum y_{....}^2 - \frac{y_{....}^2}{N}$	P^2-1

Where ANOVA represent Analysis of Variance; SSl is Latin Letters Sum of Square; SSG is Greek Letters Sum of Square; SSR is Rows Sum of Square; SSCol is Columns Sum of Square; SSE is Error Sum of Square and SST is Total Sum of Square.

The factor represented by the Greek letters is orthogonal to rows, columns and Latin letters because the Greek letters appear exactly once in each row and column and exactly once with each Latin letter. Therefore a sum of squares due to the Greek letter factor may be computed from the Greek letter totals and the experimental error is further reduced by this amount.

The statistical model for Graeco-Latin square design is

$$y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \omega_l + \varepsilon_{ijkl} \begin{cases} i = 1,2,\ldots, p \\ j = 1,2,\ldots, p \\ k = 1,2,\ldots, p \\ l = 1,2,\ldots, p \end{cases} \qquad (1)$$

Where

μ: grand mean

α_i :ith block one effect (row effect)

β_j : jth treatment effect

γ_k : kth block two effect (column effect)

ω_l : lth block three effect (Greek letter)

and $\varepsilon_{ijkl} \sim N(0, \sigma^2)$

The Error Sum of Square (SSE) for the model (1) is obtained as equation (2)

$$\text{SSE=} \left\lfloor y_{ijkl} - \left(\bar{y}_{i\ldots} + \bar{y}_{.j..} + \bar{y}_{..k.} + \bar{y}_{\ldots l}\right) + 3\bar{y}_{\ldots} \right\rfloor \qquad (2)$$

There are situations whereby an observation in one of the block is missing due to carelessness of error beyond the control of the experimenter such as unavoidable damage to an experimental unit. Due to non-orthogonality of treatment to block, thereby another problem to analyzing the data arises. Hence, we propose an approximate approach to estimate the missing data.

Suppose 'w' is missed in Graeco Latin Square then, its SSE is given as obtained in (3)

$$\text{SSE } = w^2 - \frac{1}{p}(y'_{i\ldots} + w)^2 - \frac{1}{p}(y'_{.j..} + w)^2 - \frac{1}{p}(y'_{..k.} + w)^2 - \frac{1}{p}(y'_{\ldots l} + w)^2 + \frac{1}{p^2}(3y'_{\ldots\ldots} + 3w)^2 + R \quad 3$$

Where the primes indicate total for row, column, treatment and Greek letter with missing value while $y'_{\ldots\ldots}$ represent grand total for the missing value and R include all terms not involving w.

Differentiating (3) with respect to w and equate the solution to zero gives us (4)

$$w = \frac{p\left(y'_{i\ldots} + y'_{.j..} + y'_{..k.} + y'_{\ldots l}\right) - 3y'_{\ldots\ldots}}{(p-1)(p-3)} \qquad (4)$$

3. **Results**

In this section, results obtained based on the methods highlighted under section 2 are presented and also discussed with the use of real data application. Data in Table 4 is example of an experiment conducted and reported by Zhu (Purdue University 2005) to compare four gasoline additives by testing them on four cars with four drivers over four days. Only four runs can be conducted in each day. The response of the experiment is the amount of automobile emission. The treatment factor is gasoline additive denoted by A, B, C and D, the Block factor1 is driver denoted by 1,2,3,4; Block factor2 is day denoted by 1,2,3,4 and the Block factor 3 is car which is denoted by Greek letters α, β, γ, δ.

Table 4; Automobile emission experiment

Driver/Days	1	2	3	4
1	Aα=32	Bβ=25	Cγ=31	Dδ=27
2	Bδ=24	Aγ=36	Dβ=20	Cα=25
3	Cβ=28	Dα=30	Aδ=23	Bγ=31
4	Dγ=34	Cδ=35	Bα=29	Aβ=33

Table 5: Analysis of variance table for Automobile emission experiment

Source of Variation	Degree of freedom	Sum of squares	Mean sum of square	F-calculated	F-tabulated
Driver	3	310.94	103.65	1.51	**3.29**
Day	3	68.19	22.73	0.33	
Additive	3	36.69	12.23	0.18	
Car	3	101.19	33.73	0.49	
Error	3	205.94	68.65		
Total	15	722.94			

Suppose one observation is completely missed at random, say Dβ=20 in row two column three. Calculation of the missed observation is carried out by equation (4), the estimated observation is obtained to be Dβ=17 by using the proposed method (equation 4). Then, the new ANOVA Table for the missing data is presented in Table 6.

Table 6: Analysis of variance table for Automobile emission experiment with one missing observation

Source of Variation	Degree of freedom	Sum of squares	Mean sum of square	F-calculated	F-tabulated
Driver	3	108.50	36.16	4.43	**3.29**
Day	3	89.00	29.66	3.42	
Additive	3	45.50	15.16	1.86	
Car	3	117.50	39.16	4.80	
Error	3	24.50	8.16		
Total	15	385.00			

Model Adequacy checking

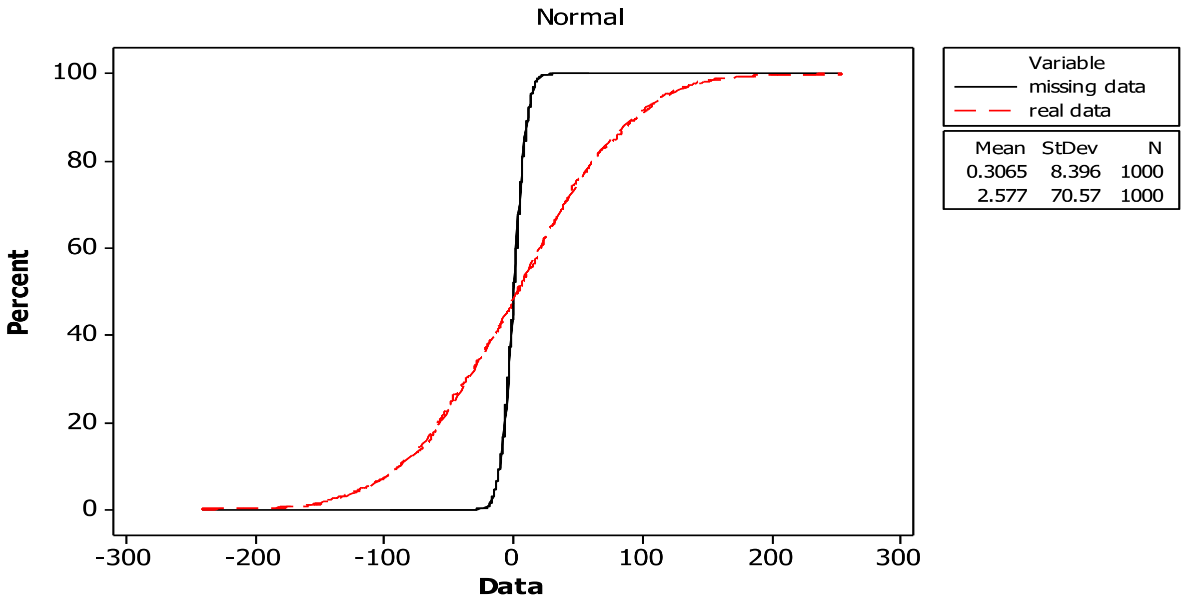

Figure I: CDF of both missing data experiment and real experiment

4. Discussion

The result of analysis for the proposed method is presented in Table 5. Based on this result, Mean Sum of Squares (MSE) reduced drastically compared to that of Table 6 for complete data set. Reduction in MSE is a clear indication that with an application of approximate method derived for Graeco-Latin Square Design, we can obtain a better result with a minimum variance unbiased estimate. Despite the sharp reduction of MSE, the conclusion and interpretation of the result is consistent with the result obtained in table 6 which is demonstrated in figure1 for model adequacy checking.

5. Conclusion

The research had discussed the existence of one missing observation in Graeco-Latin Square design. Approximate method to address the problem of missing data was derived for the design. The result of the analysis revealed a significant impact of the proposed method which has not altered the conclusion and interpretation of the treatment effect in the analysis. We were able to check for model adequacy of the proposed method by comparing the result with an experiment without missing data through a simulation study of 1000 experimental runs. The result converged with real data set as shown in figure 1. Hence, we recommend that the method derived in this research is capable to handling the problem of missing observation in Graeco-Latin Square design.

Abbreviations

BIBD: Balanced Incomplete Block Design; BILSD: Balanced Incomplete Latin Square Design; EM: Expectation Maximization; ANCOVA: Analysis of Covariance; ANOVA: Analysis of Variance; SSl : Latin Letters Sum of Square; SSG : Greek Letters Sum of Square; SSR: Rows Sum of Square; SSCol :Columns Sum of Square; SSE: Error Sum of Square; SST: Total Sum of Square; CDF: Cumulative Distribution Function; MSE: Mean Square Error

Declarations

Ethics approval and consent to participate: Not applicable.
Consent for publication: Not applicable.

Availability of data and material: The datasets used and/ or analyzed during the current study are available from the corresponding author on reasonable request.
Competing interests: The authors declare that they have no competing interest.
Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors**,** and funded by the authors
Authors' Contributions: All authors share in every step of this work and all of them contribute to writing of the manuscript. The authors read and approved the manuscript.
Acknowledgements: Not applicable.

References

i *Ai, M., Li, K., Liu, S., & Lin, D. K. (2013). Balanced incomplete Latin square designs. Journal of Statistical Planning and Inference, 143(9), 1575–1582. https://doi.org/10.1016/j.jspi.2013.05.001*

ii *Allan, F. E., & Wishart, J. (1930). A method of estimating the yield of a missing plot in field experimental work. The Journal of Agricultural Science, 20(3), 399–406. https://doi.org/10.1017/S0021859600006912*

iii *Bartlett, M. S. (1937). Some examples of statistical methods of research in agriculture and applied biology. Supplement to the Journal of the Royal Statistical Society, 4, 137–183. https://doi.org/10.2307/2983644*

iv *Casselgren, C. J., & Häggkvist, R. (2013). Completing partial Latin squares with one filled row. Column and Symbol. Discrete Mathematics, 313, 1011–1017. https://doi.org/10.1016/j.disc.2013.01.019*

v *Cochran, W. G. (1957). Analysis of covariance: Its nature and uses. Biometrics, 13(3), 261–281. https://doi.org/10.2307/2527916*

vi *Coons, I. (1957). The analysis of covariance as a missing plot technique. Biometrics, 13(3), 387–405. https://doi.org/10.2307/2527922*

vii *Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society Series B (Methodological), 39, 1–38.*

viii *Donovan, D. M., & Şule Yazıcı, E. (2014). A polynomial embedding of pairs of orthogonal partial Latin squares. Journal of Combinatorial Theory, Series A, 126, 24–34. https://doi.org/10.1016/j.jcta.2014.04.003*

ix *Euler, R. (2010). On the completability of incomplete Latin squares. European Journal of Combinatorics, 31, 535–552. https://doi.org/10.1016/j.ejc.2009.03.036*

x *Healy, M., & Westmacott, M. (1956). Missing values in experiments analysed on automatic computers. Applied Statistics, 5, 203–206. https://doi.org/10.2307/2985421*

xi *Kramer, C. Y., & Glass, S. (1960). Analysis of variance of a Latin square design with missing observations. Applied Statistics, 9, 43–50. https://doi.org/10.2307/2985758*

xii *Kittiwat, S. & Kanogkan, L. | (2017) One missing value problem in Latin square design of any order: Exact analysis of variance, Cogent Engineering, 4:1, 1411222*

xiii *Little, R. J. A., & Rubin, D. B. (2002). Statistical analysis with missing data (2nd ed.). Hoboken: John Wiley & Sons. https://doi.org/10.1002/9781119013563*

xiv *Mead, R., Gilmour, S. G., & Mead, A. (2012). Statistical principles for the design of experiments: Applications to real experiments, Vol. 36. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139020879*

xv *Montgomery, D. C. (2008). Design and analysis of experiments (7th ed.). New York, NY: John Wiley & Sons.*

xvi *Ott, R. L., & Longnecker, M. (2010). An introduction to statistical methods and data analysis (6th ed.). Boston: Cengage Learning.*

xvii *Rubin, D. B. (1972). A non-iterative algorithm for least squares estimation of missing values in any analysis of variance design. Applied Statistics, 136–141. https://doi.org/10.2307/2346485* xviii *Rubin, D. B. (1987). Multiple imputation for non-response in surveys. Hoboken: John Wiley & Sons. https://doi. org/10.1002/SERIES1345*

xix *Sirikasemsuk, K. (2016a). A review on incomplete latin square design of any order. In N. Rusli, W. M. Zaimi, K. A. Khazali, M. J. Masnan, W. S. Daud, N. Abdullah, ..., Y. N. Yusuf (Eds.),*

xx *AIP Conference Proceedings 2016, (vol. 1775, pp. 030022). Melville: AIP Publishing.*

xxi *Sirikasemsuk, K. (2016b). One missing value problem in Latin square design of any order: Regression sum of squares. In 2016 Joint 8th International Conference on Soft computing and Intelligent Systems (SCIS) and 17th International Symposium on Advanced Intelligent Systems (pp. 142–147). Japan: IEEE Press.*

xxii *Speed, T. P. (1992). Introduction to fisher (1926) the arrangement of field experiments. In S. Kotz & N. L. Johnson (Eds.), Breakthroughs in statistics, Vol. 2 (pp. 71–81). New York, NY: Springer. https://doi. org/10.1007/978-1-4612-4380-9*

xxiii *Wilkinson, G. N. (1958). Estimation of missing values for the analysis of incomplete data. Biometrics, 14(2), 257–286. https://doi.org/10.2307/2527789*

xxiv *Yates, F. (1933). The analysis of replicated experiments when the field results are incomplete. Empire Journal of, Experimental Agriculture, 2(1), 129–142.*

xxv *Yates, F. (1936). Incomplete randomized blocks. Annals of eugenics, 7, 121–140. https://doi. org/10.1111/j.1469-1809.1936.tb02134.x*

xxvi *Youden, W. J. (1937). Use of incomplete block replications in estimating tobacco-mosaic virus. Contributions from Boyce Thompson Institute, 9(1), 41–48.*

xxvii *Zhang, H. (2013). 25 new r -self-orthogonal Latin squares. Discrete Mathematics, 313, 1746–1753. https://doi.org/10.1016/j.disc.2013.04.021*