SCOPUA Journal of

Applied Statistical Research

Vol. 1, Issue 2, August 2025

DOI: https://doi.org/10.64060/JASR.v1.i2.4

The Exact Analysis of Augmented Incomplete Latin Square Design with One Missing Observation

Raneem Khalil[1], Abdulrahman AlAita [2*], Yasser Al Zaim [1]

[1]Department of Statistics, Faculty of Science, Damascus University, Syria

[2]Department of Agricultural Economy, Faculty of Agricultural Engineering, Damascus University, Syria

* Corresponding Email: abdulrahman22.alaita@damascusuniversity.edu.sy

Received: 10 March 2025 / Revised: 28 July 2025 / Accepted: 29 July 2025 / Published online: 16 August 2025

ABSTRACT

As Federer's augmented Latin square design (ALSD) is one of the most important augmented designs used in plant breeding programs, this paper aims to introduce an exact analysis for an ALSD with a missing check treatment. Namely, the novelty of this present work is evident in proposing and defining the augmented incomplete Latin square design (AILSD). In this light, the effects of rows, columns, checks, and new treatments were evaluated. All previous known studies were limited by focusing on dealing with complete datasets, paying less attention to the problem of the presence of a missing value, and this is the gap we are trying to fill here. Moreover, computing the regression sum of squares (RSS) for both full and reduced models was essential to figure out the other needed sum of squares. A numerical example and R simulation study were carried out to assess the performance of the proposed design. The importance of employing R comes from filling the calculation gaps involved in analysing AILSD.

Keywords*:* Augmented Design; Experimental Design; Latin Square; Incomplete Latin Square; One Missing; Empirical Type I Error

1. INTRODUCTION

In modern world experiments, the existence of one or more missing observations is familiar and common, specifically when the experimental units are related to biological varieties. It is well-known that many statistical approaches fail in the existence of missing values. This major fail is seen in producing bias estimation of treatment effects, resulting in an inflation in the estimate variability. Moreover, the improper handling of missings may reduce the power of testing treatment effects. In Latin square design (LSD), dealing with one missing value was accomplished by applying an approximate method, and it would be improved by applying a biased adjustment in two steps, dealing with missing value analysis. As blocking structure of designs becomes increasingly complex and unbalanced, the missing problem gets more and more complicated. We may refer to Gokul and Pachamuthu (2025), Ezievuo et al. (2025), Brenda and Oladugba (2024) and Sirikasemsuk et al. (2024) as previous existing studies dealing with handling missing observations in different designs.

On the other hand, the ALSD is an unbalanced extension of LSD, where many new treatments are added to both rows and columns, leading to adding more complexity to the addressed problem.

The LSD was originally defined by Fisher (1925), and one may find a wide review and many examples of LSD in Montgomery (2017) and Ott et al. (2015). In fact, the exact analysis of LSD with one missing is studied by Sirikasemsuk (2017). In addition, the construction and analysis of a balanced incomplete Latin square design can be found in Ai et al. (2013) and Mandal and Dash (2017). After introducing the augmented design by Federer (1956), many researches focused on developing and expanding this concept since then. In this context, Federer (1961) proposed the augmented incomplete block design. Also, the augmented row-column design is discussed by Federer and Raghavarao (1975) and Federer et al. (1975). Moreover, Federer (2002) suggested an augmented resolvable row-column design. Furthermore, Piepho and Williams (2016) defined the augmented row-column to include a small number of checks and discussed many cases. Vo-Thanh and Piepho (2020) presented a general approach for searching for augmented quasi-sudoku designs using three blocking factors for any number of unreplicated varieties and replicated varieties in field trials. Likewise, Federer (2002) discussed ALSD, and further, the augmented split plot and augmented split block designs have been presented by Federer (2005a) and Federer (2005b). Federer and Wolfinger (2003), Federe (2003), and Wolfinger et al. (1997) set forth SAS code for analysing data that originally come from augmented designs. We may refer the interested reader to Federer and Crossa (2012), Federer and King (2007) and Müller et al. (2010) as good reviews for augmented designs. Other recent references can be found in Zhou et al. (2025), Anchang et al. (2025), Williams and Piepho (2025), Sandborn et al. (2024).

Confronting a missing observation in an ALSD leads to dealing with an unbalanced design and hence non-orthogonality of the treatments, rows, and columns. In such situations, the experimenter can estimate the missing and impute it as a true value, then conduct an approximate analysis. But this approach may lead to misspecification and ambiguities, for example, there is a higher chance of rejecting the null hypothesis for a new treatment when it is true, that is, a higher type I error.

Formally speaking, an exact analysis approach to handle missing values based on a design with complete data has been less heeded; therefore, we are motivated to fill that gap and develop such an approach to facilitate computations as well as to promote a more precise analysis.

For all the reasons mentioned above, we seek a more precise methodology to analyse AILSD by adopting an exact approach. To be more precise, by performing an exact analysis approach, we focus on dealing with missing control treatments in ALSD. As stated before, the existence of one missing observation leads to non-orthogonality of treatments, rows, and columns, which in turn makes the known previous estimates invalid. So, we are handling this issue and computing unbiased estimates for the treatments. The new estimates can also be utilised to compute the adjusted sums of squares, relying on the well-known full-reduced model approach.

Again, we assure that the novelty of this research comes from presenting and analysing AILSD, which was not handled by any previous studies, and this is a crucial gap to fill.

In the next section, the numerical data of a Package *plant breeding* is presented. The proposed method in this study is described in Section 3. In Section 4, the model is identified and the notations are introduced. The analytical results are presented in Section 5. The numerical data and simulation examples are coming in section 6. The paper is concluded in section 7.

2. Data

This data was taken from the Package *plantbreeding* in R (Rosyara and Rosyara, 2012). In this experiment, seventy-five cultivars of genotypes were given using ALSD with five blocks (rows and columns). Each row and column contains five control (check) treatments and ten genotypes (new treatments). The data was displayed in Table 2. By missing at least one observation at random (MAR) related to a control treatment, the corresponding Latin square design part gets unbalanced. Thus, the pre-obtained parameter estimation and the sum of squares are not valid anymore. Usually, the estimated missing value is considered the actual value, and the data is analysed as a complete set. However, this is an approximate analysis approach. This is a motivation to develop an exact approach to resolve the missing value problem and compare the inferential results of the two approaches.

3. Methodology

We focus on exact analysis with a missing value of control treatments because a new genotype among an enormous number of these treatments can be discarded at the current stage of investigation. Regardless of the new treatments, the design with control treatments is an LSD with missing values; so, the approximation approach for analysing the data in ALSD is the same as LSD with a missing value because both are established based on the SS_E obtained from replicated check treatments in the experiment through an LSD.

By this approach, the estimated value $y_{qumr} = \frac{b(y_{qu..}+y_{q.m.}+y_{q..r})-2y_{q...}}{(b-1)(b-2)}$ is imputed for the true missing observation for the r^{th} check treatment in u^{th} row and m^{th} column. This statistic is a newly suggested quantity in the paper, and we evaluate its performance in what follows.

To assess such an approximation method and as an illustration for motivation, a simple simulation study, which we will go over in detail subsequently, has been carried out using the approximate approach and the empirical type I error of testing treatment effects for the current approximate analysis approach was calculated, given the significance level, α.

The results for data with one and two missing observations, along with the calculated empirical type I error for the case of a full data set, are presented in Table 1. The coefficient factor (CF) is the rate of empirical to nominal type I error. The empirical type I error is deviated upward from the nominal α in the approximate approach. That is, utilising the approximate approach in LSD with many new genotypes may lead to detecting non-significant new treatments as significant. This is a motivation to seek a more precise approach.

Table 1:
Empirical type I error rate for testing treatment effects in ALSD with one missing observation data

Design	Test	α	Empirical type I error rate	CF for empirical type I error rate
ALSD with parameters ($r = c = ch = 5, n = 50, N = 75$)	Complete	0.01	0.009819	0.9819
		0.05	0.049723	0.99446
	Approximate	0.01	0.010626	1.0626
		0.05	0.052467	1.04934

Following, we suggest the exact analysis approach to deal with missing values. In this approach, we focused on the missing observation of one control treatment and considered the control design part as an incomplete Latin square (ILSD). So, we face a disarrangement and incompleteness in design and hence non-orthogonality of treatments in blocks (rows and columns) due to not having all checks in each row and column.

Obviously, this influences both control and new treatments estimation and sums of squares. That is, the exact analysis is accomplished based on the complete case analysis of the observations in an ILSD. For the case of one missing, there is an empty cell in an LSD corresponding to the missing control treatment observation in a block.

4. Model and Notations

Following Federer (1956), we present the models for ALSD in this section and use them throughout the paper. An ALSD can be formed for b checks and n new treatments in b rows and b columns. To construct such an augmented design, a Latin square design with b rows and b columns is used for the b check treatments; then, the blocks are expanded to include n new units in addition to the b replicated checks so that each row and column contains n/b new genotypes. an analysis standpoint, the aim in an augmented design is to assess the effects of new treatments in comparison with the checks.

Evidently, a common measure of experimental error is obtained based on the replication of checks in blocks as an LSD in Federer (1956).

For a $b \times b$ Latin square, consider the statistical model of the ALSD as (Federer (1956)),

$$y_{hijkg} = \mu + \alpha_i + \beta_j + \tau_{qk} + \tau_{lijg} + \varepsilon_{hijkg}, \begin{cases} i = 1,2,\dots,b \\ j = 1,2,\dots,b \\ k = 1,2,\dots,b \\ g = 1,2,\dots,n_{lij} \end{cases} \quad (1)$$

Where $h = l\ or\ q$ stands for the effects associated with new treatments or checks, respectively, μ is an overall mean, α_i is the effect of the i^{th} row, β_j is the effect of the j^{th} column, τ_{bk} is the effect of the k^{th} check, τ_{lijg} is the effect of the g^{th} new treatment in i^{th} row and j^{th} column, and ε_{hijkg} is the random error assumed to have a mean of zero and variance σ_ε^2. The total number of all plots in the blocks (rows and columns) is $N = \text{v} + b^2$, where $\text{v} = \sum_{i=1}^{b}\sum_{j=1}^{b} n_{lij}$ is the number of new treatments, b is replications, i.e., check treatments, rows and columns, therefore $e = \text{v} + b$ is the total number of new and check treatments. The estimate of parameters was obtained by Federer (1956).

The ANOVA Table of the ALSD is given by Federer and Crossa (2012).

5. Analytical Exact Analysis: Estimation and Sum of Squares

In this section, the estimation of the parameters in model (1) and corresponding sums of squares with one and two missing values are developed for ARCBD through the exact analysis approach as suggested above. To estimate the parameters, the normal equations are obtained and resolved for each case. The estimates will be used in deriving the adjusted sum of squares for control and new treatments through the regression sums of squares (RSS) for full and reduced models.

5.1 Missing a check

Estimation: To estimate the parameters where r^{th} check treatment in the u^{th} row and m^{th} column is missed, we obtain the least squares normal equations (NE), using the constraints $\sum_{i=1}^{b} \hat{\alpha}_i = 0$ $\sum_{j=1}^{b} \hat{\beta}_j = 0, \sum_{k=1}^{b} \hat{\tau}_{qk} + \sum_{i=1}^{b} \sum_{j=1}^{b} \sum_{g=1}^{n_{lij}} \hat{\tau}_{lijg} = 0$, as

$$\mu: (\mathrm{v} + b^2 - 1)\hat{\mu} + (b-1)\sum_{\substack{k=1 \\ k\neq r}}^{b} \hat{\tau}_{qk} + (b-2)\hat{\tau}_{qr} + \sum_{i=1}^{b}\sum_{j=1}^{b} n_{lij}\, \hat{\alpha}_i - \hat{\alpha}_u + \sum_{j=1}^{b}\sum_{i=1}^{b} n_{lij}\, \hat{\beta}_j - \hat{\beta}_m$$

$$= y_{....},$$

$$\alpha_u: (b + \sum_{j=1}^{b} n_{luj} - 1)(\hat{\mu} + \hat{\alpha}_u) + \sum_{\substack{k=1 \\ k\neq r}}^{b} \hat{\tau}_{qk} + \sum_{j=1}^{b}\sum_{g=1}^{n_{luj}} \hat{\tau}_{lujg} + \sum_{j=1}^{b} n_{luj}\hat{\beta}_j - \hat{\beta}_m = y_{.u..},$$

$$\beta_m: (b + \sum_{i=1}^{b} n_{lim} - 1)(\hat{\mu} + \hat{\beta}_m) + \sum_{\substack{k=1 \\ k\neq r}}^{b} \hat{\tau}_{qk} + \sum_{i=1}^{b}\sum_{g=1}^{n_{lim}} \hat{\tau}_{limg} + \sum_{i=1}^{b} n_{lim}\hat{\alpha}_i - \hat{\alpha}_u = y_{..m.},$$

$$\tau_{br}: (b-1)(\hat{\mu} + \hat{\tau}_{qr}) + \sum_{\substack{i=1 \\ i\neq u}}^{b} \hat{\alpha}_i + \sum_{\substack{j=1 \\ j\neq m}}^{b} \hat{\beta}_j = y_{q..r},$$

$$\tau_{lumg}: \hat{\mu} + \hat{\alpha}_u + \hat{\beta}_m + \hat{\tau}_{lumg} = y_{lumg},$$

By solving the above NE, the estimates of the parameters in model (1) are

$$\hat{\mu} = \frac{1}{(v+b)}\left(y_{....} - \frac{(b-1)}{b}\sum_{\substack{k=1 \\ k\neq r}}^{b} y_{b..k} - \frac{(b-2)}{(b-1)} y_{b..r} + \frac{\hat{\alpha}_u}{b-1} + \frac{\hat{\beta}_m}{b-1}\right),$$

$$\hat{\alpha}_u = \frac{1}{b(b-2)}\left((b-1)y_{bu..} + y_{b.m.} - \sum_{\substack{k=1 \\ k\neq r}}^{b} y_{b..k}\right), \hat{\alpha}_i = \frac{1}{b}\left(y_{bi..} - \sum_{k=1}^{b} \bar{y}_{b..k}\right) = \frac{1}{b} y_{bi..} - \hat{\mu}_b, i \neq u,$$

$$\hat{\beta}_m = \frac{1}{b(b-2)}\left((b-1)y_{b.m.} + y_{bu..} - \sum_{\substack{k=1 \\ k\neq r}}^{b} y_{b..k}\right), \hat{\beta}_j = \frac{1}{b}\left(y_{b.j.} - \sum_{k=1}^{b} \bar{y}_{b..k}\right) = \frac{1}{b} y_{b.j.} - \hat{\mu}_b, j \neq m,$$

$$\hat{\tau}_{br} = \frac{y_{b..r} + \hat{\alpha}_u + \hat{\beta}_m}{b-1} - \hat{\mu}, \hat{\tau}_{bk} = \frac{y_{b..k}}{b} - \hat{\mu}, k = 1,2,\ldots,b, k \neq r,$$

$$\hat{\tau}_{lijg} = y_{lijg} - \hat{\alpha}_i - \hat{\beta}_j - \hat{\mu}, i,j = 1,2,\ldots,b, g = 1,2,\ldots,n_{lij}, i \neq u, j \neq m,$$

Where $\hat{\mu}_b = \frac{(b-3)y_{b...} + y_{bu..} + y_{b.m.} + y_{b..r}}{b(b-1)(b-2)}$.

Here, we point out that in Federer (1956), all estimations were made for a model with a full dataset, whereas the above equations are all computed in the presence of a missing value, and these are new estimates we derived.

In the same manner as above, we obtain the estimates of the parameters of the reduced treatment- and block-models, obtained by deleting the block and treatment effects from model (1), respectively.

Regression Sum of Squares: Now, in this section, the RSSs in models (1) are calculated based on the exact estimates of parameters with a missing check value obtained as in the above section. The RSS of full model (1), say $R(\mu,\alpha,\beta,\tau)$ is

$$R(\mu,\alpha,\beta,\tau) = SS_{Tr(\text{unadj.})} + \frac{y_{....}^2}{N} + \frac{1}{b}\left(\sum_{i=1,i\neq u}^{b} y_{qi..}^2 + \sum_{j=1,j\neq m}^{b} y_{q.j.}^2\right) - \hat{\mu}_b\left(2y_{q...} - y_{qu..} - y_{q.m.}\right) + P \qquad (2)$$

Where $P = \frac{\left((b-1)y_{qu..}+y_{q.m.}+y_{q..r}-y_{q...}\right)\left(y_{qu..}+\frac{y_{q..r}}{b-1}\right)+\left((b-1)y_{q.m.}+y_{qu..}+y_{q..r}-y_{q...}\right)\left(y_{q.m.}+\frac{y_{q..r}}{b-1}\right)}{b(b-2)}$.

Similarly, the RSSs of the reduced row, $R(\mu,\beta,\tau)$, -column, $R(\mu,\alpha,\tau)$, and -treatment, $R(\mu,\alpha,\beta)$, effect models were also derived.

The following theorem establishes the RSSs for ALSD with a missing observation of the control treatment.

The simple form of the $SS_{Tr(\text{adjusted})}$ in (5) is $R(\tau|\mu,\alpha,\beta) = R(\mu,\alpha,\beta,\tau) - R(\mu,\alpha,\beta)$ obtained by some simplification of subtraction. Totally, the sums of squares in ARCBD with one missing control value are

$$SS_{Row(\text{adj.})} = \frac{1}{b}\left(\sum_{i=1,i\neq u}^{b} y_{qi..}^2\right) - \hat{\mu}_q\left(y_{q...} - y_{qu..}\right) + \frac{\left((b-1)y_{qu..}+y_{q.m.}+y_{q..r}-y_{q...}\right)\left((b-1)y_{qu..}+y_{q.m.}+\frac{b}{b-1}y_{q..r}\right)}{b(b-1)(b-2)}, \qquad (3)$$

$$SS_{Col(\text{adj.})} = \frac{1}{b}\left(\sum_{j=1,j\neq u}^{b} y_{q.j.}^2\right) - \hat{\mu}_q\left(y_{q...} - y_{q.m.}\right) + \frac{\left((b-1)y_{q.m.}+y_{qu..}+y_{q..r}-y_{q...}\right)\left((b-1)y_{q.m.}+y_{qu..}+\frac{b}{b-1}y_{q..r}\right)}{b(b-1)(b-2)} \qquad (4)$$

$$SS_{Tr(\text{adj.})} = SS_{Tr(\text{unadj.})} - SS_{Row(\text{unadj.})} - SS_{Col(\text{unadj.})} - \frac{y_{....}^2}{N} + \frac{1}{b}\left(\sum_{i=1,i\neq u}^{b} y_{qi..}^2 + \sum_{j=1,j\neq m}^{b} y_{q.j.}^2\right) \qquad (5)$$

where $P = \frac{\left((b-1)y_{qu..}+y_{q.m.}+y_{q..r}-y_{q...}\right)\left(y_{qu..}+\frac{y_{q..r}}{b-1}\right)+\left((b-1)y_{q.m.}+y_{qu..}+y_{q..r}-y_{q...}\right)\left(y_{q.m.}+\frac{y_{q..r}}{b-1}\right)}{b(b-2)}$,

$$Q = \frac{(y_{.u..}+y_{..m.}+(b+n-2)y_{....})\hat{\mu}^{NT}}{b+n-2} - \frac{(y_{.u..}^2+y_{..m.}^2)+2(b+n-1)y_{.u..}y_{..m.}}{(b+n)(b+n-1)(b+n-2)},$$

$SS_{Row(\text{unadj.})}$, $SS_{Col(\text{unadj.})}$ and $SS_{Tr(\text{unadj.})}$ are given in Federer (1956).

6. Numerical Data and Simulation Illustration

In this section, we numerically assess the performance of the proposed exact analysis approach by a numerical data example obtained from an ALSD. A simulation study is also accomplished for evaluation. The MSE is used to assess the efficiency of the approach.

The data for this example were taken from Package *plantbreeding* in R by Rosyara and Rosyara (2012). Five hypothetical check treatments named $A, B, C, D,$ and E, and 50 hypothetical new treatments, denoted by $1,2,\ldots,50$, are arranged in an ALSD with 5 rows and 5 columns. The data is given in Table 2. We deleted a check, B, in the third row and fourth column at random.

Table 2:
Data for ALSD (observation values in rows and columns)

Rows	**Columns**														
		1			**2**			**3**			**4**			**5**	
1	C 2.10	19 2.69	37 4.56	44 7.36	8 4.03	A 2.88	25 3.67	B 3.89	35 6.61	E 4.30	2 3.80	26 3.20	10 5.45	D 4.45	24 4.20
2	46 4.26	12 3.39	E 1.02	39 4.10	C 2.95	4 3.18	A 4.42	16 3.68	11 5.22	27 8.08	D 9.95	28 3.23	5 6.70	45 4.65	B 5.90
3	34 2.17	D 9.06	21 2.49	15 3.24	1 3.29	E 6.69	C 2.57	42 1.87	36 3.40	7 5.30	~~B 5.52~~	32 6.94	20 6.00	38 7.15	A 4.60
4	17 4.28	B 5.38	6 3.60	D 8.53	13 2.86	31 2.96	E 7.84	30 3.90	9 3.98	22 5.13	A 5.82	41 4.77	C 3.50	49 7.20	33 4.05
5	A 2.70	23 4.66	43 6.65	18 2.84	48 5.47	B 5.16	47 6.68	D 7.84	29 2.57	3 9.56	50 5.57	C 3.20	40 6.55	E 4.50	14 6.45

The parameters have been estimated and the sums of squares were calculated based on the exact missing method and the corresponding ANOVA is presented in Table 3. The treatment effect is not significant, which is consistent with the complete data in ALSD given in ANOVA Table 4, both with the same $p - value = 0.13$. Moreover, the MSE relative efficiency of the exact missing method in comparison to the MSE of the complete data is 0.97, which demonstrates the perfect performance of the proposed method in estimating the error variance, unbiasedly.

Table 3:
ANOVA (A) for the AILSD experiment

Sources of variation	df	SS	MS	F-value	p-value
Rows (unadjusted)	4	10.55	2.64	--	--
Columns (unadjusted)	4	30.48	7.17		
Treatments (adjusted)	54	209.12	3.87	1.89	0.13
Checks (adjusted)	4	72.44	18.11	8.85	0.00
New and new vs. check	50	136.68	2.73	1.34	0.31
Error	11	22.51	2.05	--	--
Total	73	272.66	--	--	--

Table 4:
ANOVA for the ALSD experiment with full data set

Sources of variation	df	SS	MS	F-value	p-value
Rows (unadjusted)	4	10.36	2.59	--	--
Columns (unadjusted)	4	31.02	7.76	--	--
Treatments (adjusted)	54	206.54	3.82	1.81	0.13
Error	12	25.30	2.11	--	--

7. Comparison illustration

The performance of the proposed exact approach in analysing data is evaluated by a simulation study. In this study, we calculate the empirical type I error and the power of the test for treatment effect at a given nominal significance level. These are calculated for the exact, approximate missing analysis approaches in ALSD data and compared with that calculated for the full data set. The simulation study was carried out for the given number of treatments and blocks, which are already presented in various existing examples (e.g., Federer and Crossa (2012); Rosyara and Rosyara (2012)). The data are first generated from the corresponding normal distribution. Then, by assuming nominal significant levels of 0.05 and 0.01, we used the MC method to calculate the empirical type I error and the power of the test for treatment effects. The simulations for the exact, approximate and full data analysis were replicated 10000 times.

The simulated type I error for approaches as well as full data analysis issummarised in Table 5. The calculated empirical type I error differs from the nominal significance level for the approaches; however, the deviance for full data analysis is negligible and is almost consistent with the exact approach. The approximate approach exposes the highest deviation, where the empirical type I error is inflated. This means that the treatments may erroneously be considered significant. To see the deviation more precisely, the coefficient factor (CF) was defined as $CF = Empirical\ type\ I\ error / nominal\ \alpha - level$ and are given in the last columns of Table 5.

Table 5:
Empirical type I error rate for ARCBD with parameters (r, c, ch, n, N)

(r,c,ch,n,N)		**(5, 5, 5, 50, 75)**	
Approach	α	**Empirical α**	**CF**
Exact	0.01	0.009962	1.003814495
	0.05	0.049784	1.004338743
Full	0.01	0.009819	1.018433649
	0.05	0.049723	1.005570863
Approximate	0.01	0.010626	0.941087898
	0.05	0.052467	0.952979968

In this study, the power of the test for testing the treatment effects through the ANOVA was calculated for exact and approximate missing analyses as well as for full data. The results are given in Table 5 for ALSD with $(r,c,ch,n,N) = (5,5,5,50,75)$, for a different set of check means, $(\mu_1,\mu_2,\mu_3,\mu_4,\mu_5)$. In order to compare the exact and approximate approaches' performance, the power of the test in Table 6 is displayed in Figure 1, with that of full data, without loss of generality, the powers were plotted in an ascending order. Evidently, the power of the test with Full data is higher than the power with a missing value; however, the exact approach performs better than the approximate in testing the treatment effects.

Table 6:
Simulation results for ALSD (one missing value) with parameters (r,c,ch,n,N) for Check treatment means $(\mu_1,\mu_2,\mu_3,\mu_4,\mu_5)$ and different values of new treatments

$(\mu_1,\mu_2,\mu_3,\mu_4)$ $\boldsymbol{\gamma}$	**Approach**	α	**Empirical power** (2,3,3,3,4) 1	(1,3,3,3,4) 2	(3,3,3,4,4) 3	(4,3,3,4,4) 4	(2,3,4,4,5) 5	(3,4,4,5,5) 6	(4,4,5,5,6) 7	(0,2,4,6,8) 8
$(r,c,ch,n,N) = (5,5,5,50,75)$	**Exact**	0.01	0.4115	0.4442	0.4726	0.5194	0.5818	0.6944	0.8177	0.9586
		0.05	0.7621	0.7957	0.8161	0.8461	0.8935	0.9413	0.9809	1.0000
New treatments: $(\mu_i = 0, \mu_j = 1, \mu_k = 2, \mu_u = 3, \mu_s = 4)$ $i,j,k,u,s = 1,\ldots,10$	**Complete**	0.01	0.4831	0.5136	0.5440	0.5885	0.6745	0.7833	0.8966	1.0000
		0.05	0.8133	0.8405	0.8572	0.8848	0.9290	0.9670	0.9962	1.0000
	Approximate	0.01	0.4014	0.4310	0.4564	0.4958	0.5727	0.6697	0.7880	0.9158
		0.05	0.7354	0.7671	0.7855	0.8092	0.8618	0.9003	0.9373	0.9520

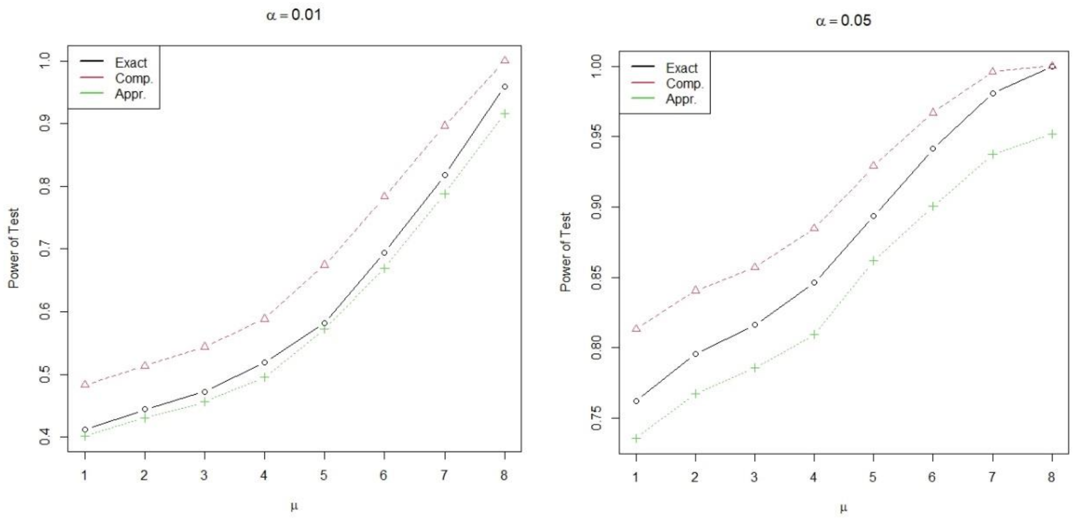

Figure 1: Power of test for ARCBD with one missing value of checks (Row=5, Column=5, Check= 5, New Treatment=50, N=75)

8. Conclusion

This study was conducted in order to improve the analysis of data in augmented Latin square designs with a missing value. A missing value might be dealt with using an existing approximation method; however, this would result in precision loss, as well as a higher probability of rejecting non-significant treatments. Therefore, we sought to find an exact approach to dealing with the missing value in augmented Latin square designs, because there are no treatment effect estimates or sums of squares for the checks and new treatments available. The exact approach to dealing with missing values in the study proposed revealed that the details in the ANOVA table for check and new treatments are more precise than what one could obtain using the approximate method or even using the augmented incomplete Latin square designs analysis obtained by ignoring the missing values, which has been programmed in R. In fact, using the R program to analyse the set of data in ALSD with missing values as an AILSD considers the check with missing value as a new treatment; this may increase the type I error, which misleads in determining the null new treatments as significant. The exact approach provides more powerful results than the approximate method in testing treatment effects. In addition, in the case of a large sample size, the power of the test is similar to the power of the test without missing data.

Finally, we mention that the proposed methodology in this paper has limitations considering dealing with small sample sizes, and this fact is directly linked to the significant impact of the sample size on experimental designs in general. So, as future research, we consider improving our method to deal with small sample sizes, which may be solved by assuming that the sample is large but involves many missing. Another limitation is related to the deviation from the normality assumption of the underlying model, and future research is needed to see how this specific deviation might have affected our results.

Declaration

Conflict of Study: The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

Author Contribution Statement: RK, AA and YAZ wrote the paper and did the statistics and the relevant table. RK critically and linguistically revised and reviewed the whole literature. All authors read and approved the final manuscript.

Funding Statement: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Availability of Data and Material: Data will be made available by the corresponding author on reasonable request.

Consent to Publish: All authors have agreed to publish this manuscript in the SCOPUA Journal of Applied Statistical Research (JASR)-ISSN(e): 3104-4794.

Ethical Approval: Not Applicable

Consent to Participate: Not Applicable

Acknowledgment: The authors are thankful to the editor and reviewers for their valuable suggestions to improve the quality and presentation of the paper.

References

[1] Ai, M., Li, K., Liu. S. and Lin, D. (2013). "Balanced incomplete Latin square designs", *Journal of Statistical Planning and Inference*, **143(9),** 1575-1582.

[2] Anchang, M. M., Berebon, D. P. and Okafor, G. I. (2025). "Production and optimization of probiotic-enriched breakfast cereals: A mixture augmented design approach", *Food and Humanity*, **4**, 100508.

[3] Brenda, Y. M. and Oladugba, A. V. (2024). "Robustness of Augmented Third-order Response Surfaces Designs to Missing Observation", *Journal of the Iranian Statistical Society*, **23(1)**, 33-49.

[4] Burgueño, J., Crossa, J., Rodríguez, F. and Yeater, K. M. (2018). "Augmented designs-experimental designs in which all treatments are not replicated", *Applied Statistics in Agricultural, Biological, and Environmental Sciences*, Barry Glaz and Kathleen M. Yeater: editors, Chapter **13**, 345-370. https://doi.org/10.2134/appliedstatistics.2016.0005.c13

[5] Ezievuo, C. S., Oladugba, A. V. and Babatunde, O. T. (2025). "Evaluation of orthogonal composite designs for second-order model in presence of missing observation", *Quality and Reliability Engineering International*, **41(1)**, 204-234.

[6] Federer, W.T. (1956). "Augmented (or hoonuiaku) designs", *Biometrics Unit Technical Reports*, **55,** 191–208. https://doi.org/10.2307/2529707

[7] Federer, W., Nair, R. and Raghavarao, D. (1975). "Some augmented row-column designs", *Biometrics*, **31**, 361-373. https://doi.org/10.2307/2529426

[8] Federer, W. T. and Raghavarao, D. (1975). "On augmented designs", *Biometrics*, **31**, 29- 35. https://doi.org/10.2307/2529707

[9] Federer, W. T. and Wolfinger, R. D. (2003). "*Augmented row–column design and trend analyses*", Handbook of formulas and software for plant geneticists and breeders, 291- 295.

[10] Federer, W.T. (1961). "Augmented designs with one-way elimination of heterogeneity", *Biometrics*, **17(3)**, 447-473. https://doi.org/10.2307/2527837

[11] Federer, W.T. (1998a). "Recovery of interblock, intergradient, and intervariety information in incomplete block and lattice rectangle designed experiments", *Biometrics*, **54**, 471-481. https://doi.org/10.2307/3109756

[12] Federer, W.T. (1998b). "*A simple procedure for constructing resolvable row-column experiment designs",* BU-1438-M in the Technical Report Series of the Department of Biometrics, Cornell University, Ithaca, NY. https://ecommons.cornell.edu/handle/1813/32036

[13] Federer, W.T. (2002). "Construction and analysis of an augmented lattice square design", *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, **44(2)**, 251-257. https://doi.org/10.1002/1521-4036(200203)44:2%3C251::AID-BIMJ251%3E3.0.CO;2-N

[14] Federer, W.T. (2003). "*Analysis for an experiment designed as augmented lattice square design*", Handbook of Formulas and Software for Plant Geneticists and Breeders, 283-289.
[15] Federer, W.T. and Crossa, J. (2012). "I.4 screening experimental designs for quantitative trait loci, association mapping, genotype-by environment interaction, and other investigations", *Frontiers in Physiology*, ***3***, 1–9. https://doi.org/10.3389/fphys.2012.00156
[16] Federer, W. T. and King, F. (2007). "*Variations on split plot and split block experiment designs"*, (Volume 654), John Wiley & Sons.
[17] Fisher, R.A. (1925). "*Statistical methods for research workers",* Edinburg, Oliver and Boyd, Auflage.
[18] Gokul, A. R. and Pachamuthu, M. (2025). "Assessing the predictive resilience of composite experimental designs under missing observations: the influence of α alpha value variations", *Reliability: Theory and Applications*, **20(2(84))**, 518-534.
[19] Mandal, B. and S. Dash (2017). "On balanced incomplete Latin square designs", *Communications in Statistics-Theory and Methods*, **46(13)**, 6258-6263.
[20] Montgomery, D.C. (2017). "*Design and analysis of experiments*", John Wiley & sons.
[21] Müller, B. U., Schützenmeister, A. and Piepho, H. P. (2010). "Arrangement of check plots in augmented block designs when spatial analysis is used", *Plant breeding*, ***129*****(6)**, 581-589. https://doi.org/10.1111/j.1439-0523.2010.01803.x
[22] Ott, N.c. and Longnecker, M.T. (2015). "*An introduction to statistical methods and data analysis"*, 7th Edition, Cengage Learning.
[23] Piepho, H.P. and Williams, E.R. (2016). "Augmented row–column designs for a small number of checks", *Agronomy Journal*, **108(6)**, 2256-2262. https://doi.org/10.2134/agronj2016.06.0325
[24] Sandborn, M., Olea, C., Said, A., Shabir, M., Volgyesi, P., Koutsoukos, X. and White, J. (2024). "Towards AI-augmented design space exploration pipelines for UAVs", In book: Artificial Intelligence: Machine Learning, Convolutional Neural Networks and Large Language Models, 1, 309-332.
[25] Sirikasemsuk, K. and K. Leerojanaprapa (2017). "One missing value problem in Latin square design of any order: Exact analysis of variance", *Cogent Engineering*, ***4*****(1)**, https://doi.org/10.1080/23311916.2017.1411222
[26] Sirikasemsuk, K., Wongsriya, S. and Leerojanaprapa, K. (2024). "Solving the incomplete data problem in Greco-Latin square experimental design by exact-scheme analysis of variance without data imputation", *AIMS mathematics*, **9(12)**, 33551-33571.
[27] Vo-Thanh, N. and Piepho, H. P. (2020). "Augmented quasi-sudoku designs in field trials", *Computational Statistics and Data Analysis*, ***150,*** https://doi.org/10.1016/j.csda.2020.106988
[28] Williams, E. R. and Piepho, H. P. (2025). "A note on the construction of augmented designs in square arrays", *arXiv (preprint)*, 2501.08448.
[29] Wolfinger, R., Federer, W.T. and Cordero-Brana, O. (1997). "Recovering information in augmented designs, using SAS PROC GLM and PROC MIXED", *Agronomy Journal*, ***89*****(6)**, 856-859. https://doi.org/10.2134/agronj1997.00021962008900060002x
[30] Zhou, C., Liu, X., Yu, C., Tao, Y. and Shao, Y. (2024). "Trust in AI-augmented design: Applying structural equation modeling to AI-augmented design acceptance", *Heliyon*, **10(1)**.

Author(s) Bio / Authors' Note

Raneem Khalil:

R.K *from Department of Statistics, Faculty of Science, Damascus University, Syria.*

Email: raneem.khallil92@damascusuniversity.edu.sy

Abdulrahman AlAita:

A.A *is from Department of Agricultural Economy, Faculty of Agricultural Engineering, Damascus University, Syria.* Email: abdulrahman22.alaita@damascusuniversity.edu.sy

Yasser Al Zaim:

Y.A.Z *from Department of Statistics, Faculty of Science, Damascus University, Syria.*

Email: y.alzaim@damascusuniversity.edu.sy