International Journal of Mathematics and Statistics Invention (IJMSI)
E-ISSN: 2321 – 4767 P-ISSN: 2321 - 4759
www.ijmsi.org Volume 3 Issue 7 || November. 2015 || PP-06-08

Bayesian Estimation for Missing Values in Latin Square Design

R. Ajantha and N.Ch. Bhatra Charyulu
Department of Statistics, University College of Science, Osmania University, Hyderabad-7.

ABSTRACT: *In this paper, an attempt is made to estimate the missing values in Latin Square Design in Bayesian approach. The method is illustrated for one, two and m-missing values using R and WinBUG software's. Some remarks on the method are presented.*

Keywords: *Missing values, LSD, Bayesian estimation.*

I. INTRODUCTION

Consider the statistical linear model for a Latin Square Design with 'v' rows, 'v' columns and for 'v' treatments as

$$\underline{\mathbf{Y}} = \mathbf{X}\,\underline{\boldsymbol{\beta}} + \underline{\boldsymbol{\varepsilon}}$$

where $\underline{\mathbf{Y}} = [y_{111} \ldots y_{11k} \ldots y_{11v} \mid \ldots \mid y_{ij1} \ldots y_{ijk} \ldots y_{ijv} \mid \ldots \mid y_{vv1} \ldots y_{vvk} \ldots y_{vvv}]'$ is the vector of observations where y_{ijk} is the observation belongs to i^{th} row, j^{th} column and k^{th} treatment. $\underline{\boldsymbol{\beta}} = [\mu \mid \alpha_1 \ldots \alpha_i \ldots \alpha_v \mid \beta_1 \ldots \beta_j \ldots \beta_v \mid \gamma_1 \ldots \gamma_k \ldots \gamma_v]'$ vector of parameters, μ is the mean, α_i, β_j , γ_k are the effects due to the i^{th} row, j^{th} column and k^{th} treatment respectively. $\underline{\boldsymbol{\varepsilon}} = [\varepsilon_{11\,1} \ldots \varepsilon_{1jk} \ldots \varepsilon_{1vv} \mid \ldots \mid \varepsilon_{i11} \ldots \varepsilon_{ijk} \ldots \varepsilon_{ivv} \mid \ldots \mid \varepsilon_{v11} \ldots \varepsilon_{vjk} \ldots \varepsilon_{vvv}]'$ vector of random error, ε_{ijk} is random error corresponding to y_{ijk} and X is the design matrix.

$$X = \begin{bmatrix}
1 & 1 & 0 & . & 0 & 1 & 0 & . & 0\,1 & 0 & 0 & . & 0 \\
1 & 1 & 0 & . & 0 & 0 & 1 & . & 0\,0 & 1 & 0 & . & 0 \\
. & . & . & . & . & . & . & . & . . & . & . & . & . \\
1 & 1 & 0 & . & 0 & 0 & 0 & . & 1\,0 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & . & 0 & 1 & 0 & . & 0\,0 & 1 & 0 & . & 0 \\
1 & 0 & 1 & . & 0 & 0 & 1 & . & 0\,0 & 0 & 1 & . & 0 \\
. & . & . & . & . & . & . & . & . . & . & . & . & . \\
1 & 0 & 1 & . & 0 & 0 & 0 & . & 1\,1 & 0 & 0 & . & 0 \\
. & . & . & . & . & . & . & . & . . & . & . & . & . \\
1 & 0 & 0 & . & 1 & 1 & 0 & . & 0\,0 & 0 & 0 & . & 1 \\
1 & 0 & 0 & . & 1 & 0 & 1 & . & 0\,0 & 0 & 0 & . & 0 \\
. & . & . & . & . & . & . & . & . . & . & . & . & . \\
1 & 0 & 0 & . & 1 & 0 & 0 & . & 1\,1 & 0 & 0 & . & 0
\end{bmatrix}$$

The estimated response is $\hat{Y} = X\hat{\beta}$, where $\hat{\beta} = (X'X)^{-1}X'Y$

If an observation is missing the resulting data is incomplete to carry out the analysis as per the original plan of the experiment and also affecting the orthogonality. So, it is necessary to estimate the missing values to carry out the analysis as per the original plan of experiment. Several authors made attempts since 1930. An attempt is made to estimate the missing values in Latin Square Design (LSD) in Bayesian approach is presented in section 2 and is illustrated with suitable examples.

II. ESTIMATION OF MISSING ALUES IN BAYESIAN APPROACH

Consider the general linear model and partition the vector of responses into known (Y_1) and missing (Y_m) response vectors and accordingly model can be expressed as

$$\begin{bmatrix} Y_1 \\ Y_m \end{bmatrix} = \begin{bmatrix} X_1 \\ X_m \end{bmatrix} \beta + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_m \end{bmatrix}$$

Let $\underline{\mathbf{y}} = (y_1, y_2, \ldots, y_n)'$ be the vector of observed sample $\underline{\mathbf{y}} \in R^n$. Assume the sample drawn from normal population with mean $X\underline{\beta}$ and variance σ^2 i.e. follows $N(X\underline{\beta}, \sigma^2)$, where the parameters $\underline{\beta}$ and σ^2 are unknown and $X \in R^{n\times p}$ be the design matrix. The likelihood function of the observed sample with the estimated parameters is $P(\underline{\mathbf{y}} / \underline{\beta}, \sigma^2)$. The distribution of the parameter $\underline{\beta}$ is follows normal with mean $\hat{\beta}$ and variance $\sigma^2/\sum x_i^2$. The sample mean follows $N(\bar{y}, \sigma^2/n)$ and precision follows Gamma distribution with parameters (a, b), where a=(n-k) and b= σ^2/(n-k) The posterior probability of $\underline{\beta}$ given **y** can be evaluated by generating a sequence of sample values in such a way that, as more and more sample values as possible, such that the distribution of sample values more closely approximates the desired distribution and is used to evaluate the normalized constant $P(\underline{\mathbf{y}}) = \int P(\underline{\beta}).\ P(\underline{\mathbf{y}}/\underline{\beta})\ d\beta$. Generate a large sample (n is large) from a normal using Gibbs simulation technique by setting initial parameters using WinBUG software and estimate the vector of parameters $\hat{\beta}$ to evaluate the estimated responses using the normal equation $\hat{Y}_m = X_m \hat{\beta}$. The estimation of parameters and missing values using Bayesian approach is illustrated for one, two and m missing values with suitable examples in example 2.1, 2.2 and in 2.3.

EXAMPLE 2.1: Consider the Latin square design experimental data with 3 treatments presented in Table 2.3, related to bioequivalence study with an observation is missing.

	Period		
Subject	**1**	**2**	**3**
1	(A) 1186	(B) 642	(C) 1183
2	(B) 984	(C) x_1	(A) 1305
3	(C) 1426	(A) 1540	(B) 873

Table 2.3

The estimated parameters of normal populated observations are 1142.375 and 88081.41. The estimated vector of parameters β mean and variances are [(590, 11010.18), (20.3333, 29360.47), (273.3333, 44040.71), (296.3333, 29360.47), (215.3333, 29360.47), (237.6666, 44040.71), (137, 29360.47), (360.3333, 29360.47), (-150.3333, 29360.47), (380, 44040.71)]. The precision follows Gamma distribution with parameters 3.5 and 0.000000324. Estimate parameter β using Win BUG software, by generating a large sample from normal as [590, 20.33, 273.3, 296.3, 215.3, 237.7, 137, 360.3, -150.3, 380]′. The estimated missing response is 1481.

EXAMPLE 2.2: Consider the problem of manufacturer of disk drives, interested in studying the effect of four substrates (Aluminium, Nickel-plated and two types of Glass) on the amplitude of the signal that is received. There were four machines, four operators, four days of production that were to be involved with machines, operators and days to save as blocking variables. It can be noted that two observations are missing and the data presented in Table 2.2.

8 (A)	11 (C)	2 (D)	8 (B)
7 (C)	x_1 (A)	2 (B)	4 (D)
3 (D)	9 (B)	7 (A)	x_2 (C)
4 (B)	5 (D)	9 (C)	3 (A)

Table 2.2

The estimated parameters of normal populated observations are 5.8571 and 8.5934. The estimated values of vector β of parameters mean and variances are [(3.5535, 0.6138), (1.9196, 2.1483), (-0.5178, 2.8644), (2.2321, 2.1483), (-0.0803, 2.1483), (0.1696, 2.8644), (2.4821, 2.1483), (-0.3303, 2.1483), (1.2321, 2.8644), (0.7321, 2.1483), (0.4196, 2.1483), (4.2321, 2.8644), (-1.8303, 2.1483)]. The precision follows Gamma distribution with parameters 6.5 and 0.0179. Estimate parameter β using Win BUG software, by generating a large sample from normal, as [2.453, 1.606, -0.2539, 1.702, -0.0956, 0.0775, 2.121, 0.1747, 1.046, 0.5138, 0.475, 3.728, -1.146]′. The estimated values for missing responses are 6.0046, 8.929.

EXAMPLE 2.3: Consider the data presented in Table 2.1 obtained through 25 plots can be arranged in two way blocking with five rows and five columns for testing five treatments with seven missing values.

(D) 376	(E) 371	(C) x_1	(B) 356	(A) 335
(B) 316	(D) 338	(E) 336	(A) 356	(C) x_2
(C) x_3	(A) 326	(B) 335	(D) x_4	(E) 330
(E) 317	(B) x_5	(A) 330	(C) 327	(D) 336
(A) 321	(C) 332	(D) x_6	(E) x_7	(B) 306

Table 2.1

The estimated parameters of normal populated observations are 335.7778 and 340.3007. The mean and variances of estimated vector β of parameters values are: [(228.1523,18.90559), (67.3771,85.07518), (39.3704, 85.07518), (47.0638, 113.4336), (38.8171, 85.07518), (40.5238, 113.4336), (39.2438, 85.07518), (40.683, 113.4336), (51.5971, 113.4336), (65.0571, 85.07518), (31.5704, 85.07518), (14.1866, 113.4336), (4.6466, 85.07518), (-0.266, 340.3007), (37.8266, 170.1504), (22.6666, 113.4336)]. The precision follows Gamma distribution with parameters 8.5 and 0.00034. Estimate parameter β using WinBUG software, by generating a large sample from normal: (228.1, 67.37, 39.37, 47.06, 38.82, 40.52, 39.24, 40.68, 51.6, 65.06, 31.57, 14.19, 4.645, -0.2663, 37.83, 22.67)′. The estimated missing responses obtained using R software are (346.8037, 298.7737, 314.1337, 378.05, 312.245, 309.0337, 358.05)′.

III. REMARKS ON ESTIMATED PARAMETERS AND MISSING VALUES

The Bayes procedure for estimating missing values in LSD is compared with least square approach and made some remarks and are presented below.

1. It can be observed that the Least square and Bayes estimated values for the parameters are nearly same due to the normality. It can be noted that the least squares and Maximum likelihood estimates for missing values are same.
2. It does not provide any formulae to estimate missing value even in case of single missing observation and it is ideal for many missing values when data is small or large.
3. Bayes approach is complicated when compared with least squares approach because it is a simulating procedure and involves distribution for generation of samples and is difficult to solve manually.
4. The limit for number of missing values belong to a treatment is (0, v-1) and the limit for number of missing values belong to each row or column is also same.
5. As the number of missing values is increasing the efficiency will decrease. The error sum of squares in case of missing observations is least when compared with original data.

ACKNOWLEDGEMENTS: The authors are grateful to UGC-BSR RFMS for providing financial assistance to carry out this work.

REFERENCES

[1] Box M.J, Draper N.R and Hunter W.G (1970): "Missing values in multi response non linear model fitting", Technometrics, Vol 12(3), pp 613-621.
[2] Box M.J (1971) "A parameter estimation criterion for multi response models applicable when some observations are missing", Journal of applied statistics, Vol 20(1), pp.1-7.
[3] Draper N.R (1961): "Missing values in Response surface designs" Technometrics, Vol 3(3), pp 389-398.
[4] Peter M. Lee (2012): "Bayesian Statistics- An Introduction", John-Wiley & sons Ltd, 4th Edition.
[5] Rubin D.B. (1972). "A Non-iterative Algorithm for least squares estimation of missing values in any analysis of variance design" , Applied Statistics, Vol 21, pp 136-141.
[6] Subramani J and Ponnuswamy K.N. (1989): "A non iterative least squares estimation of missing values in experimental designs", Journal of Applied Statistics, Vol. 16(1), pp 19-24.
[7] Subramani J and Balamurali S (2012): "Loss of efficiency in randomized block designs with two missing values", Elixir Statistics, Vol. 49, pp 10042-44
[8] Vanlier J, Tiemann C.A., Hilbers P.A.J., Van Riel N.A.W. (2012): "A Bayesian approach to targeted experimental design" Bioinformatics, Vol. 28(8), pp 1136-42.
[9] Yates, F. (1933): "The Analysis of replicated experiments when the field results are incomplete", Empire Journal of Experimental Agriculture, Vol. 1, pp. 129-142.