https://doi.org/10.56093/JISAS.V78I3.2
Available online at http://isas.org.in/isa/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 78(3) 2024 193–200

Expected Maximization Algorithm for the Estimation of Missing Responses in Experimental Designs

Kuncham Srinivas and N. Ch. Bhatra Charyulu

University College of Science, Osmania University, Hyderabad

Received 06 September 2024; Revised 21 October 2024; Accepted 23 October 2024

SUMMARY

This paper presents the estimation of missing responses in Randomized block Design (RBD), Latin Square Design (LSD) and Response Surface Design (RSD) using least squares method and expected maximization algorithm. The parameter relations also presented when initial values are taken as zero, mean of known responses and arbitrary. The methods are illustrated with suitable examples.

Keywords: RBD, LSD, RSD.

1. INTRODUCTION

Experimenter is often encountered with a situation in which observation(s) of particular plot(s) may be lost or much affected by some extraneous causes that it would not be desirable to regard these observations as normal experimental observations and the omissions naturally affect the method of analysis. The data with missing observations are generally analyzed through missing plot techniques like estimating the missing value(s) by minimizing the residual sum of squares iteratively or non iteratively or by the method of fitting constants or by using the covariates.

Expected Maximization algorithm is one of the popular methods used to estimate missing information. Further it is an ad-hoc approach for the estimation of missing data by choosing arbitrarily initial values for missing data. The estimated missing values and parameters were iteratively updated until they converge. The name Expected Maximization algorithm was given by Dempster, Laird and Rubin (1977).

DEFINITION 1.1: The experimental material is partitioned into 'b' homogeneous groups of experimental units called blocks such that each block contains 'v' experimental units. Assign the 'v' treatments to the experimental units with a separate randomization for each block, such a design is said to be Randomized Block Design (RBD).

Let Y_{ij} be the response correspond to i^{th} treatment in j^{th} block (i =1, 2, … v; j=1, 2, … b) (let N=vb, p=v+b+1). Assume that the responses are independent and follow Normal distribution with mean μ and variance σ^2. The general linear model for a randomized block design is

$$Y = X\beta + \varepsilon \qquad (1.1)$$

where Y_{Nx1} be the vector of observations Y_{ij}, X_{Nxp} is the design matrix and β_{px1} is the vector of parameters μ, α_i, β_j, where μ is the general mean, α_i is the effect due to the i^{th} of the treatment, β_j is the effect due to the j^{th} block, and $\underline{\varepsilon} \sim N(0, \sigma^2 I_n)$.

DEFINITION 1.2: The experimental material is partitioned into v^2 experimental units in such a way that the experimental units can be grouped (blocking) in two ways, row wise and column wise, such that each contains 'v' experimental units. The 'v' treatments are allocated to the experimental units in such a way that each treatment occurs once in each row, and in each

Corresponding author: Kuncham Srinivas
E-mail address: kunchamsrinu86@gmail.com

column such a design is said to be Latin Square Design (LSD).

Let Y_{ijk} be the response corresponding to the k^{th} treatment belongs to i^{th} row, j^{th} column and (i = 1, 2, … v; j = 1, 2, … v, k = 1, 2, … v), N= v^2 and p=3v+1 Assume $Y_{ijk} \sim N(\mu, \sigma^2)$. The general linear model for a Latin square design is

$$Y = X\beta + \varepsilon \tag{1.2}$$

where Y_{Nx1} be the vector of observations Y_{ij}, X_{Nxp} is the design matrix and β_{pxl} is the vector of parameters μ, α_i, β_j γ_k and $\underline{\varepsilon} \sim N(0, \sigma^2 I_n)$. where μ is the general mean, α_i is the effect due to the i^{th} of the row, β_j is the effect due to the j^{th} column, γ_k is the effect due to the k^{th} treatment (i=1, 2, … v; j=1, 2, … v, k=1, 2, … v).

DEFINITION 1.3: Let $X_1, X_2, \ldots X_v$ be the set of 'v' factors, each has 's' levels used for the experimentation and let $D = ((\boldsymbol{x}_{u1}, \boldsymbol{x}_{u2}, \ldots, \boldsymbol{x}_{uv}))$ be the design matrix, where $\boldsymbol{x}_{ui}$ be the level of the i^{th} factor in the u^{th} design point. Let Y_u denote the response at the u^{th} design point. The functional relationship between the response and factor combination is $E(Y_u) = \boldsymbol{f}(\boldsymbol{x}_{u1}, \boldsymbol{x}_{u2}, \ldots, \boldsymbol{x}_{uv})$ called 'Response Surface', the design used for fitting the response surface is called a 'Response Surface Design', and the model fitted to the design and responses is called 'Response Surface Design Model'.

2. EXPECTED MAXIMIZATION ALGORITHM ESTIMATION FOR MISSING RESPONSES FOR AN EXPERIMENTAL DESIGN MODEL

The Expected Maximization algorithm detailed step by step procedure for implementation to any experimental design model is presented below.

Step 1: Let $\underline{x} = (x_1, x_2, \ldots, x_n)$ be the observed sample drawn from a population with probability density function P(x, θ), where θ is unknown.

Step 2: Evaluate the likelihood function of the observed sample $L(\underline{x}, \theta)$ and log of the $L(\underline{x}, \theta)$. Choose some arbitrary values initially for the missing information.

Step 3: Evaluate the expected value of log likelihood function $E[\log L(\underline{x}, \theta)]$ and evaluate the improved version of the parameter that maximizes the expected value of log of Likelihood function.

Step 4**:** Repeat the step 3 and step 4 until two successive iterations difference of the estimated values for the parameter is negligible.

Now, consider the general linear model for an experimental design model

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon} \tag{2.1}$$

where, $\underline{Y}_{Nx1}$ be the vector of responses, $\underline{\beta}_{k\times 1}$ be the vector of parameters, $X_{N\times k}$ be the design matrix of an experiment, $\underline{\varepsilon}$ be the vector of random error. Assume $Y \sim N(\mu_y, \sigma_y^2)$ and $\underline{\varepsilon} \sim NI(0, \sigma_y^2)$.

Let $\underline{Y} = [Y_1, Y_2, \ldots Y_{N-m}, | Y_{N-m+1}, \ldots Y_N]'$ be the vector of 'N' responses with (N-m) known and 'm' missing responses. Assume $E(Y) = \mu_y = X\beta$. Then the partitioned model of (2.1) is

$$\begin{bmatrix} Y_{N-m} \\ Y_m \end{bmatrix} = \begin{bmatrix} X_{N-m} \\ X_m \end{bmatrix}\beta + \begin{bmatrix} \varepsilon_{N-m} \\ \varepsilon_m \end{bmatrix} \tag{2.2}$$

where $\underline{Y}_{N-m}$ is the vector of (N-m) known response values and $\underline{Y}_m$ is the vector of 'm' missing response values, X_{N-m} and X_m are the partitioned design matrices corresponding to the known and missing response vectors. Assume ε is also partitioned accordingly.

The likelihood function of the 'N' observed responses $\underline{Y}$ including missing is

$$L(\underline{Y}/\mu_y, \sigma_y^2) = (2\pi\sigma_y^2)^{-N/2} \times \exp\left\{\frac{-1}{2\sigma_y^2}\left[\sum_{j=1}^{N-m}(Y_j - \mu_y)^2 + \sum_{j=N-m+1}^{N}(Y_j - \mu_y)^2\right]\right\} \tag{2.3}$$

The log of the likelihood function is

$$\log L(\underline{Y}/\mu_y, \sigma_y^2) = -\frac{N}{2}\log 2\pi - \frac{N}{2}\log\sigma_y^2 - \frac{1}{2\sigma_y^2}\left[\sum_{j=1}^{N-m}(Y_j - \mu_y)^2 + \sum_{j=N-m+1}^{N}(Y_j - \mu_y)^2\right] \tag{2.4}$$

The maximum likelihood estimates for the μ and σ^2 are

$$\frac{\partial}{\partial\mu_y}\log L(Y/\mu_y, \sigma_y^2) = 0 \Rightarrow \sum_{j=1}^{N}(Y_j - \mu_y) = 0$$

$$\Rightarrow \hat{\mu}_y = \frac{1}{N}\left[\sum_{j=1}^{N-m} Y_j + \sum_{j=N-m+1}^{N} Y_j\right] \tag{2.5}$$

$$\frac{\partial}{\partial \sigma_y^2}\log L(Y/\mu_y,{\sigma_y}^2)=0$$

$$\Rightarrow \hat{\sigma}_y^2=\frac{1}{N}\left[\sum_{j=1}^{N-m}(Y_j-\hat{\mu}_y)^2+\sum_{j=N-m+1}^{N}(Y_j-\hat{\mu}_y)^2\right] \quad (2.6)$$

The expected value of log of likelihood

$$E\left[\log L(\underline{Y}/\mu_y,\sigma_y^2)\right]=\int Y\left\{-\frac{N}{2}\log 2\pi-\frac{N}{2}\log\sigma_y^2-\frac{1}{2\sigma_y^2}\left[\sum_{j=1}^{N-m}(Y_j-\mu_y)^2+\sum_{j=N-m+1}^{N}(Y_j-\mu_y)^2\right]\right\}dY \quad (2.7)$$

The estimate of missing response can be obtained from E [log L(Y)] as

$$E\,[Y_m / Y, X) = X_m\hat{\beta}^{(k)} \quad (2.8)$$

The conditional the expectation for missing response is

$$E\,[{Y_m}^2 / Y, X) = [\,X_m\hat{\beta}^{(k)}\,]'\,[\,X_m\hat{\beta}^{(k)}\,]+\sigma^{2\,(k)} \quad (2.9)$$

From (2.4) and E(Y) = Xβ, we have

$$\log L(\underline{Y}/\mu_y,\sigma_y^2)=\mathrm{Log}(\underline{Y})-\frac{N}{2}\log 2\pi-\frac{N}{2}\log\sigma_y^2-\frac{1}{2\sigma_y^2}\left[\sum_{j=1}^{N-m}(Y_j-X_j\beta)^2+\sum_{j=N-m+1}^{N}(Y_j-X_j\beta)^2\right]$$

The pastiche estimates $\hat{\beta}$, $\hat{\sigma}^2$ and $\hat{Y}_m$ can be obtained by minimizing the residual sum of squares as

$$\frac{\partial}{\partial\beta}\log L(Y)=0$$

$$\Rightarrow\sum_{j=1}^{N-m}\left(X_j'Y_j-X_j'X_j\beta\right)+\sum_{j=N-m+1}^{N}\left(X_j'Y_j-X_j'X_j\beta\right)=0$$

$$\Rightarrow\left[\sum_{j=1}^{N-m}X_j'X_j+\sum_{j=N-m+1}^{N}X_j'X_j\right]\beta=\left[\sum_{j=1}^{N-m}X_j'Y_j+\sum_{j=N-m+1}^{N}X_j'Y_j\right]$$

$$\hat{\beta}=\left[\sum_{j=1}^{N-m}X_j'X_j+\sum_{j=N-m+1}^{N}X_j'X_j\right]^{-1}\left[\sum_{j=1}^{N-m}X_j'Y_j+\sum_{j=N-m+1}^{N}X_j'Y_j\right]$$

$$\Rightarrow\hat{\beta}=\left[X'X\right]^{-1}\left[\sum_{j=1}^{N-m}X_j'Y_j+\sum_{j=N-m+1}^{N}X_j'Y_j\right]$$

$$\Rightarrow\hat{\beta}^{(k+1)}=\left[X'X\right]^{-1}\left[\sum_{j=1}^{N-m}X_j'Y_j+\sum_{j=N-m+1}^{N}X_j'Y_j^{(k)}\right] \quad (2.10)$$

$$\frac{\partial}{\partial\sigma^2}\log L(Y)=0$$

$$\Rightarrow\hat{\sigma}^2=\frac{1}{N}\left[\sum_{j=1}^{N-m}(Y_j-X_j\hat{\beta})^2+\sum_{j=N-m+1}^{N}(Y_j-X_j\hat{\beta})^2\right]$$

$$\Rightarrow\hat{\sigma}^{2(k+1)}=\frac{1}{N}\left[\sum_{j=1}^{N-m}(Y_j-X_j\hat{\beta}^{(k+1)})^2+\sum_{j=N-m+1}^{N}(Y_j-X_j\hat{\beta}^{(k+1)})^2\right] \quad (2.11)$$

$$\Rightarrow\hat{\sigma}^{2(k+1)}=\frac{1}{N}\left[(N-m)\hat{\sigma}_1^{2\,(k+1)}+m\hat{\sigma}_2^{2\,(k+1)}\right]$$

where

$$(N-m)\hat{\sigma}_1^{2(k+1)}=\sum_{j=1}^{N-m}(Y_j-X_j\hat{\beta}^{(k+1)})^2;$$

$$m\hat{\sigma}_2^{2(k+1)}=\sum_{j=N-m+1}^{N}(Y_j-X_j\hat{\beta}^{(k+1)})^2$$

$$\frac{\partial}{\partial Y_m}\log L(Y)=0\Rightarrow\hat{Y}_m=X_m\hat{\beta}$$

where m = N-m+1, … , N.

$$\Rightarrow\hat{Y}_m^{(k+1)}=X_m\hat{\beta}^{(k+1)} \quad (2.12)$$

Set the values for missing responses be either zero or mean of known responses or any arbitrary value. The parameters $\hat{\sigma}^2$, $\hat{\beta}$ and missing responses $\hat{Y}_m$ are valuated iteratively using the equations (2.10), (2.11) and (2.12) till the values for parameters and missing responses are stabilized.

3. EXPECTED MAXIMIZATION ESTIMATION WITH DIFFERENT INITIAL SEEDS

The missing responses are estimated in an iterative approach using Expected maximization algorithm with different initial values for missing responses as either zeroes or mean of known response or arbitrary values. In each of the case the estimated parameter and its successive parameters are derived and presented below.

3.1 Estimation with Seed Value Zero

Assume $Y_m = 0_m$ (zero vector), then the model (2.1) becomes

$$\begin{bmatrix}Y_{N-m}\\0_m\end{bmatrix}=\begin{bmatrix}X_{N-m}\\X_m\end{bmatrix}\beta+\begin{bmatrix}\varepsilon_{N-m}\\\varepsilon_m\end{bmatrix} \quad (3.1)$$

Let $\tilde{\beta}^{(1)}$ denotes the first estimate of β with the initial value set to zero. The estimate of vector of parameters and missing responses from (2.10) and (2.12) as

$$\tilde{\beta}^{(1)} = (X'X)^{-1}(X'_{N-m}Y_{N-m}) \tag{3.2}$$

$$\tilde{Y}_m^{(1)} = X_m\tilde{\beta}^{(1)} \tag{3.3}$$

The second estimate of the parameter is $\tilde{\beta}^{(2)}$ is

$$\tilde{\beta}^{(2)} = (X'X)^{-1}(X'_m Y_m^{(1)} + X'_{N-m}Y_{N-m})$$

$$= (X'X)^{-1}\left[X'_m X_m (X'X)^{-1}X'_{N-m}Y_{N-m} + X'_{N-m}Y_{N-m}\right]$$

$$= (X'X)^{-1}(X'X - X'_{N-m}X_{N-m})(X'X)^{-1}X'_{N-m}Y_{N-m} + (X'X)^{-1}X'_{N-m}Y_{N-m}$$

$$\tilde{\beta}^{(2)} = \left[2I - (X'X)^{-1}(X'_{N-m}X_{N-m})\right]\tilde{\beta}^{(1)} \tag{3.4}$$

The modified estimate of response is given by

$$\tilde{Y}_m^{(2)} = X_m\left[2I - (X'X)^{-1}(X'_{N-m}X_{N-m})\right]\tilde{\beta}^{(2)} \tag{3.5}$$

The iterative procedure is continued until the two successive responses difference is negligible.

3.2 Estimation with Seed Value Mean of Known Responses

Assume $Y_m = \bar{Y}_{N-m}$ i.e. Mean of the known (N-m) responses. Then the model (2.1) becomes

$$\begin{bmatrix} Y_{N-m} \\ \bar{Y}_{N-m} \end{bmatrix} = \begin{bmatrix} X_{N-m} \\ X_m \end{bmatrix}\beta + \begin{bmatrix} \varepsilon_{N-m} \\ \varepsilon_m \end{bmatrix} \tag{3.6}$$

Let $\bar{\beta}^{(1)}$ denotes the first estimate of β with the initial values set to mean of the known responses. The estimate of vector of parameter and missing responses from (2.10) and (2.12) as

$$\bar{\beta}^{(1)} = \left(X'X\right)^{-1}\left(X'_m\bar{Y}_{N-m} + X'_{N-m}Y_{N-m}\right) \tag{3.7}$$

$$\bar{Y}_m^{(1)} = X_m\,\bar{\beta}^{(1)} \tag{3.8}$$

The modified estimates for the parameter and missing response are

$$\bar{\beta}^{(2)} = (X'X)^{-1}(X'_m\bar{Y}_{N-m}^{(1)} + X'_{N-m}Y_{N-m})$$

$$= (X'X)^{-1}(X'_m\bar{Y}_{N-m}^{(1)}) + (X'X)^{-1}(X'_{N-m}Y_{N-m})$$

$$= (X'X)^{-1}\left[(X'_mX_m)(X'X)^{-1}(X'_m\bar{Y}_{N-m}) + X'_{N-m}Y_{N-m}\right] + (X'X)^{-1}X'_{N-m}Y_{N-m}$$

$$= (X'X)^{-1}(X'X - X'_{N-m}X_{N-m})\left[(X'X)^{-1}X'_m\bar{Y}_{N-m} + (X'X)^{-1}X'_{N-m}Y_{N-m}\right] + \tilde{\beta}$$

$$= (X'X)^{-1}(X'X - X'_{N-m}X_{N-m})\left[(X'X)^{-1}\left\{X'_m\bar{Y}_{N-m} + X'_{N-m}Y_{N-m}\right\}\right] + \tilde{\beta}$$

$$= \left[I - (X'X)^{-1}(X'_{N-m}X_{N-m})\right]\bar{\beta}^{(1)} + \tilde{\beta}$$

$$\bar{\beta}^{(2)} = (\text{I-B})\,\bar{\beta}^{(1)} + \tilde{\beta}\text{ , where } B = (X'X)^{-1}(X'_{N-m}X_{N-m}) \tag{3.9}$$

The modified estimate of response is given by

$$\bar{Y}_m^{(2)} = X_m\left[(X'X)^{-1}\left\{(X'_m\bar{Y}_{N-m}^{(1)}) + (X'_{N-m}Y_{N-m})\right\}\right]$$

$$\bar{Y}_m^{(2)} = X_m\left[(I-B)\bar{\beta}^{(1)} + \tilde{\beta}\right] \tag{3.10}$$

This iterative procedure is continued until the two successive responses difference is negligible.

3.3 Estimation with Seed Value is an Arbitrary Value

Assume $Y_m = A_m$ i.e. arbitrary value. Then the model (2.1) becomes

$$\begin{bmatrix} Y_{N-m} \\ A_m \end{bmatrix} = \begin{bmatrix} X_{N-m} \\ X_m \end{bmatrix}\beta + \begin{bmatrix} \varepsilon_{N-m} \\ \varepsilon_m \end{bmatrix} \tag{3.11}$$

Let $\ddot{\beta}_A^{(1)}$ denote the first estimate of β with the initial values set to arbitrary value. The estimate of vector of parameter and missing response from (2.10) and (2.12) as

$$\ddot{\beta}_A^{(1)} = (X'X)^{-1}(X'_mA + X'_{N-m}Y_{N-m}) \tag{3.12}$$

$$\ddot{Y}_A^{(1)} = X_m\,\ddot{\beta}_A^{(1)} \tag{3.13}$$

The modified estimates for parameters and missing responses are

$$\ddot{\beta}_A^{(2)} = (X'X)^{-1}(X'_m\ddot{Y}_A^{(1)} + X'_{N-m}Y_{N-m})$$

$$= (X'X)^{-1}\left[X'_mX_m\left(X'X\right)^{-1}\left(X'_mA_m + X'_{N-m}Y_{N-m}\right) + X'_{N-m}Y_{N-m})\right]$$

$$= (X'X)^{-1}\left[\left(X'_mX_m\left(X'X\right)^{-1}\right)\left(X'_mA_m + X'_{N-m}Y_{N-m}\right)\right] + \tilde{\beta}$$

$$= (X'X)^{-1}(X'X - X'_{N-m}X_{N-m})\left[(X'X)^{-1}X'_mA_m + (X'X)^{-1}X'_{N-m}Y_{N-m}\right] + \tilde{\beta}$$

$$= (X'X)^{-1}(X'X - X'_{N-m}X_{N-m})\left[(X'X)^{-1}\left\{X'_mA_m + X'_{N-m}Y_{N-m}\right\}\right] + \tilde{\beta}$$

$$= \left[I - (X'X)^{-1}(X'_{N-m}X_{N-m})\right]\ddot{\beta}^{(1)} + \tilde{\beta}$$

$$\ddot{\beta}^{(2)} = (\text{I-B})\,\ddot{\beta}^{(1)} + \tilde{\beta}\text{, where } B = (X'X)^{-1}(X'_{N-m}X_{N-m}) \quad (3.14)$$

The modified estimate of response is given by

$$\ddot{Y}_m^{(2)} = X_m\left[(I-B)\ddot{\beta}^{(1)} + \tilde{\beta}\right] \quad (3.15)$$

This iterative procedure is continued until the two successive responses difference is negligible.

3.4 Inter Relationships Between $\hat{\beta}$, $\tilde{\beta}$, $\overline{\beta}$ & $\ddot{\beta}$

The properties of the estimated parameters such as interrelationship between the parameter, unbiasedness variances and covariance of the parameters and responses presented below.

3.4.1. Parametric Relationship Between $\hat{\beta}$ and $\tilde{\beta}$

The relationship between estimated of parameters $\hat{\beta}$ (least square) and $\tilde{\beta}$ (zero as initial guess) is established and is presented below.

$$\hat{\beta} = (X'_{N-m}X_{N-m})^{-1}X'_{N-m}Y_{N-m}$$

$$= \left[X'X - X'_mX_m\right]^{-1}X'_{N-m}Y_{N-m}$$

$$= \left[X'X - X'_mX_m(X'X)^{-1}X'X\right]^{-1}X'_{N-m}Y_{N-m}$$

$$= (X'X)^{-1}\left[I - X'_mX_m(X'X)^{-1}\right]^{-1}X'_{N-m}Y_{N-m}$$

$$= (X'X)^{-1}\left[I + X'_m\left\{I - X_m(X'X)^{-1}X'_m\right\}^{-1}X_m(X'X)^{-1}\right]X'_{N-m}Y_{N-m}$$

$$= \left[I + (X'X)^{-1}X'_m\left\{I - X_m(X'X)^{-1}X'_m\right\}^{-1}X_m\right](X'X)^{-1}X'_{N-m}Y_{N-m}$$

$$\hat{\beta} = \left[I + (X'X)^{-1}X'_mMX_m\right]\tilde{\beta}^{(1)};$$

where $M = \left\{I - X_m(X'X)^{-1}X'_m\right\}^{-1}$.

3.4.2 Relationship Between $\tilde{\beta}$ and $\overline{\beta}$

The relationship between the estimated parameters $\tilde{\beta}$ (zero as the initial values) and $\overline{\beta}$ (mean of known responses as initial value) is established and presented below.

$$\overline{\beta}^{(1)} = (X'X)^{-1}X'_m\overline{Y}_{N-m} + (X'X)^{-1}X'_{N-m}Y_{N-m}$$

$$\overline{\beta}^{(1)} = (X'X)^{-1}X'_m\overline{Y}_{N-m} + \left[I + (X'X)^{-1}X'_mMX_m\right]^{-1}\tilde{\beta}^{(1)}$$

3.4.3 Parametric Relations Between $\tilde{\beta}$ and $\ddot{\beta}$

The relationship between the estimated parameters $\tilde{\beta}$ (zero as initial values) and initial values $\ddot{\beta}$ ('arbitrary value' as initial value) is established and presented below.

$$\ddot{\beta}^{(1)} = (X'X)^{-1}(X'_mA_m + X'_{N-m}Y_{N-m})$$

$$\ddot{\beta}^{(1)} = (X'X)^{-1}(X'_mA_m) + \tilde{\beta}$$

EXAMPLE 3.1: Nigam and Gupta (1979) conducted a manorial trail with six of farm Yard Manus with four replications in a random block design layout to study the rate of decomposition of organic matters in the soil and its synthetic capacity in soil on cotton crop. Treatments: six levels of farm yard manus are 0, 12.4, 24.7, 61.18, 123.6, 247.2 as treatments and with four Replications each of Plot size gross:27.42 m×20.12m, Net: 8.23m ×7.32 m., the yield per plot in kg for different levels of farm Yard Manus is given the Table 1.

Table 1.

Levels of farm yard manus	Replications			
	I	II	III	IV
1	6.9	4.6	6.10	4.81
2	6.48	5.57	4.28	4.45
3	6.52	7.6	5.3	5.3
4	Y_{41}	6.65	6.75	7.75
5	6	6.18	5.5	5.5
6	7.9	7.57	6.8	7.45

Assume Y_{41} = 0 (zero), 6.17217(Mean of known responses) and 5 (arbitrary). The stabilized estimated values are $\hat{\hat{a}}$ = [4.41086, 0.08892, -0.31858, 0.66642, 1.77626, 0.28142, 1.91642, 1.82222, 1.21566, 0.64233, 0.73066]′ and = 8.00933.

EXAMPLE 3.2: The following are the field layout and yields in bushes per acre of an experiment on dusting wheat with Sulphur to control stem rust. The treatments are A-dusted before rains, B- dusted after rains, C- dusted once each week, D- drifting once each week, E- control or check. Analyze the data presented in the Table 2.

Table 2.

B 4.9	D 6.4	E 3.3	A 9.5	C X
C 9.3	A 4.0	B 6.2	E 5.1	D 5.4
D 7.6	C 15.4	A 6.5	B 6.0	E 4.6
E 6.3	B 7.6	C 13.2	D 8.6	A 4.9
A 9.3	E 6.3	D Y	C 15.9	B 7.6

Assume Y_{41} = 0 (zero), The stabilized estimated values are $\hat{\beta}$ = [5.21875, -0.46625, -0.84625, 6.14375, 2.57375, -2.18625, 2.20375, -1.30625, 0.71375, 0.81375, 2.79375, 0.17375, 2.63375, 0.81375, 1.71375, -0.11625]′ and = 8.009333 and estimated missing responses are 13.45, 11.4

EXAMPLE 3.3: Let us consider a response surface design conducted with three factors each with three levels with 27 design points given in the design matrix X_{27x10}. Assume the response vector Y satisfying second order response surface model. The design points of matrix are (-1 -1 0 0) (1 -1 0 0) (-1 1 0 0) (1 1 0 0) (0 0 -1 -1) (0 0 1 -1) (0 0 -1 1) (0 0 1 1) (0 0 0 0) (-1 0 0 -1) (1 0 0 -1) (-1 0 0 1) (1 0 0 1) (0 -1 -1 0) (0 1 -1 0) (0 -1 1 0) (0 1 1 0) (0 0 0 0) (0 -1 0 -1) (0 1 0 -1) (0 -1 0 1) (0 1 0 1) (-1 0 -1 0) (1 0 -1 0) (-1 0 1 0) (1 0 1 0) (0 0 0 0). The vector of responses Y corresponding at the design points are given below and found that the response values missing at 3rd, 12th and 19th and 25th design points.

Y = [11.28 8.44 Y_3 7.71 8.94 10.9 11.85 11.03 8.26 7.87 12.08 Y_{12} 7.98 10.48 10.14 10.22 10.53 9.5 Y_{19} 11.02 10.98 9.56 8.78 9.02 Y_{25} 8.24 9.79]′. (Fig. 1)

The parameter β and missing responses are estimated. The resulting estimated values when the initial values as taken $Y_3 = Y_{12} = Y_{19} = Y_{25} = 0$, 9.76522 (Mean of Known responses) and 10 (arbitrary), the estimated values are stabilizes various iterations and =[9.93257 -0.14697 -0.39313 0.14316 -1.07302 0.24101 0.35626 0.5652 -0.50198 0.1625] and $\hat{Y} = [8.2892, 9.00651, 10.5667, 10.00791]\hat{\beta}$

4. COMPARISON OF ESTIMATED PARAMETERS AND RESPONSES

A table of comparison of mean and variances of responses before and after estimation of missing responses and estimated parameters and responses using the two approaches, mean square error and confidence interval for the parameters with 95% confidence are is presented in Table 3.

The $(X'X)$, $(X'X)^{-1}$, can be obtained respectively as

$$\begin{bmatrix} 27 & 0 & 0 & 0 & 12 & 12 & 12 & 0 & 0 & 0 \\ 0 & 12 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 12 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 12 & 0 & 0 & 0 & 0 & 0 & 0 \\ 12 & 0 & 0 & 0 & 12 & 4 & 4 & 0 & 0 & 0 \\ 12 & 0 & 0 & 0 & 4 & 12 & 4 & 0 & 0 & 0 \\ 12 & 0 & 0 & 0 & 4 & 4 & 12 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 \end{bmatrix}$$ and

$$\begin{bmatrix} 0.185185 & 0 & 0 & 0 & -0.111111 & -0.111111 & -0.11111 & 0 & 0 & 0 \\ 0 & 0.083333 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0.083333 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -0.166667 & 0 & 0 & 0 & 0 & 0 \\ -0.111111 & 0 & 0 & 0 & 0.166667 & 0.041667 & 0.041667 & 0 & 0 & 0 \\ -0.111111 & 0 & 0 & 0 & 0.041667 & 0.166667 & 0.041667 & 0 & 0 & 0 \\ -0.111111 & 0 & 0 & 0 & 0.041667 & 0.041667 & 0.166667 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.25 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.25 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.25 \end{bmatrix}$$

Fig. 1

Table 3. Comparison Table

	Example 3.1 (Single Missing)	Example 3.2 (Two Missing)	Example 3.3 (Four Missing)
Mean (known responses)	6.17217	7.56087	9.76522
Variance (known responses)	1.20446	11.27794	1.77504
Least Square Estimates (Missing values)	8.00934	13.45 11.4	8.2892, 9.00651 10.5667, 10.00791
Expected Maximization Estimates (Missing values)	8.00933	13.45 11.4	8.2892, 9.00651 10.5667, 10.00791
Mean of Responses (including estimated)	6.24872	7.95	9.72112
Variance of responses (including estimated)	1.29273	12.23958	1.63284
Estimated Parameters	4.41086, 0.08892, -0.31858, 0.66642, 1.77626, 0.28142, 1.91642, 1.82222, 1.21566, 0.64233, 0.73066	5.21875, -0.46625, -0.84625, 6.14375, 2.57375, -2.18625, 2.20375, -1.30625, 0.71375, 0.81375, 2.79375, 0.17375, 2.63375, 0.81375, 1.71375, -0.11625	9.93257, -0.14697 -0.39313, 0.14316 -1.07302, 0.24101 0.35626, 0.5652 -0.50198, 0.1625
Mean Square Error	0.50509	2.527	—
Lower Limits	4.11749, -0.20446 -0.61196, 0.37304 1.48288, -0.01196 1.62304, 1.52884 0.92228, 0.34895 0.43728	4.5638, -1.12119, -1.50119, 5.48881, 1.91881, -2.84119, 1.54881, -1.96119, 0.05881, 0.15881, 2.13881, -0.48119, 1.97881, 0.15881, 1.05881, -0.77119	8.78174, -0.91897 -1.16513, -0.62884 -2.1648, -0.85076 -0.73551, -0.77194 -1.83912, -1.17464
Upper limits	4.70424, 0.3823 -0.02521, 0.9598 2.06964, 0.5748 2.2098, 2.1156 1.50904, 0.93571 1.02404	5.87369, 0.18869, -0.19131, 6.79869 3.22869, -1.5313 2.85869, -0.65131 1.36869, 1.46869 3.44869, 0.82869 3.28869, 1.46869 2.36869, 0.53869	11.0834, 0.62503 0.37888, 0.91516 0.01875, 1.33279 1.44804, 1.90234 0.83517, 1.4996

5. CONCLUSIONS

1. Expected Maximization algorithm is maximizing pastiche estimates of parameters based on the observed sample and is an alternative approach to the least square method for the estimation of missing responses in Design and analysis of experiments
2. Expected Maximization estimate is derived from the distribution of responses whereas least square estimate is obtained from the model and both are equally efficient.
3. The E-step of each iteration only involves taking expectations over complete data conditional distributions and the M-step of each iteration requires complete data ML-estimation, which is often in simple closed form with increasing of likelihood in each iteration and linearly converges.
4. If the number of missing responses increases, the difficulty level for estimating them is increase in both the approaches.
5. The number of iterations is depending on the initial value chosen, variance of the known responses, number of missing responses and on the design matrix.
6. The iterative procedures of Yates (1933) and Healy and Westmascat (1956) are also providing the same results as that of Expected Maximization approaches.
7. The relationship between the parameters for different assumptions for missing values are developed.

ACKNOWLEDGEMENTS

The authors are thankful to the UGC for providing financial assistance through UGC-BSR-RFSMS to carry out present research. The authors are also thankful to the referee for the valuable comments for improvisation of this manuscript.

REFERENCES

Dempster A.P, Laird N.M, and Rubin D.B (1977): Maximum likelihood from incomplete data via the EM algorithm, Journal of the Royal Statistical Society, Series B, vol 39, pp 1-38.

Draper N.R (1961): "Missing values in Response Surface Designs", Technometrics, vol. 3(3), pp 389-398.

Healy M.J.R. and Westmacott M. (1956): "Missing values in experiments analyzed on automatic computers", Applied Statistics, vol 5, pp 203-206.

Nigam and Gupta (1979): "Handbook on Analysis of Agricultural Experiments", IASRI publications, New Delhi.

Srinivas Kuncham (2018): "Some contributions to estimation of missing values in design and analysis of experiments using expected maximization algorithm", an unpublished Ph.D. thesis submitted to Osmania University.

Yates F.Y. (1933): "The analysis of replicated experiments when the field results are incomplete", Empire Journal of Experimental Agriculture, vol 1, pp 129-142.