International Journal of Mathematical Archive-6(6), 2015, 23-28
Available online through www.ijma.info ISSN 2229 – 5046

ESTIMATION OF PARAMETERS AND MISSING RESPONSES IN FIRST ORDER RESPONSE SURFACE DESIGN MODEL USING EM ALGORITHM

KUNCHAM SRINIVAS, N. CH. BHATRA CHARYULU

**Department of Statistics,
University College of Science, Osmania University, Hyderabad-7, India.**

(Received On: 05-06-15; Revised & Accepted On: 22-06-15)

SUMMARY

Expected Maximization algorithm maximizes pastiche estimates of parameters based on the observed sample in an iterative process. This paper attempts to estimate the parameters and missing responses using Expected Maximization algorithm if the design of the experiment satisfies the first order response surface design model.

Keywords: *EM algorithm, Missing responses, Response Surface Design Model.*

1. INTRODUCTION

Let there be \`v\` factors X_i (i =1, 2…v) each at \`s\` levels for experimentation and D denotes the design matrix with the combination of factor levels, given by

$$D = ((x_{u1}, x_{u2} \ldots x_{ui} \ldots x_{uv})) \tag{1.1}$$

where x_{ui} be the level of the i^{th} factor in the u^{th} treatment combination (i=1, 2 ... v; u =1, 2…N). Let Y_u denote the response at the u^{th} treatment combination. The factor-response relationship is given by

$$E(Y_u) = f(x_{u1}, x_{u2} \ \ldots \ x_{uv}) \tag{1.2}$$

is called the response surface. Design used for fitting the response surface model is termed as 'response surface design' and the model is called response surface design model.

Suppose it is required to fit a first order response surface design model expressed in the form

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon} \tag{1.3}$$

Where $\underline{Y} = (Y_1, Y_2 \ldots Y_N)'$ is the vector of responses,
$X_u = (1, x_{u1}, x_{u2} \ldots x_{uv})$ is the u^{th} row of X
$\underline{\beta} = (\beta_0, \beta_1, \beta_2 \ldots \beta_v)'$ is the vector of parameters
$\underline{\varepsilon} = (\varepsilon_1, \varepsilon_2 \ldots \varepsilon_N)$ is the vector of random errors and follows $N(0,\sigma^2 I)$.

The least square estimate of the parameter is given by

$$\hat{\beta} = (X'X)^{-1}X'Y \tag{1.4}$$

with $Var(\hat{\beta}) = (X'X)^{-1}\sigma^2$. Then the estimated responses can be obtained from the fitted model as

$$\hat{Y} = X\hat{\beta} \tag{1.5}$$

Even in some well-planned experiments and situations, the responses may not be available due to natural or manmade causes. If an observation lacks the resulting data, it is difficult to carry out the analysis as per the original plan of the experiment, it also may affect the orthogonality in case of RBD and LSD etc. Allan and Wishart (1930) initially made an attempt to estimate the missing response value in case of Randomized Block and Latin Square design using the least squares method. In case of more than one missing values, Yates (1933) developed an iterative process starting with some initial guess values. Healy and Westmacott (1956) described a more general iterative method for estimating the missing values. Later several authors made attempts on the estimation of missing values in design and analysis of experiments.

Corresponding Author: Kuncham Srinivas, Department of Statistics, University College of Science, Osmania University, Hyderabad-7, India.

2. EXPECTED MAXIMIZATION ALGORITHM

Expected Maximization algorithm maximizes pastiche estimates of parameters based on the observed sample in an iterative process. It is numerically stable and at each iteration increases the likelihood and, linearly converging.

Let $\underline{y} = [y_1, y_2, y_3, \ldots, y_n]'$ be the vector of observed sample of size ‘n’ correspondingly at design points X. Assume the sample drawn from a population is Normal with density function ***f(y)***. Assume the data may contain some unobserved or latent variable and unknown parameters. Let **L(*y*)** be the likelihood function and **log L(*y*)** be the log of likelihood function of the sample. First estimate the values of the parameters by maximizing the likelihood function based on the known observed sample. Then evaluate the expected value of log of likelihood function. The improved version of the parameter that maximizes the expected value of Log of Likelihood function can be evaluated by repeating the above two steps of evaluations until two successive iterations will results same value or with negligible difference.

No significant work is found on the estimation of missing values using expected maximization algorithm in the experimental design directly. Therefore, this paper attempts to develop the procedure to estimate the parameters and missing responses in case of first order response surface design model using expected maximization algorithm.

3. ESTIMATION OF PARAMETERS AND MISSING OBSERVATIONS IN RESPONSE SURFACE DESIGN MODEL USING EM ALGORITHM

Let $\underline{y} = [y_1, y_2 \ldots y_n]'$ be the vector of responses correspondingly at the design matrix $X_{nx(v+1)}$. Assume the factor-response relationship is linear model with first order response surface model in v factors, satisfying the model (1.3). Assume the response variable Y follows $N(X\beta, \sigma^2)$. Let us assume that at some design points the responses are missing. Then the model (1.3) can be expressed as

$$\begin{bmatrix} Y_1 \\ Y_m \end{bmatrix} = \begin{bmatrix} X_1 \\ X_m \end{bmatrix} \beta + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_m \end{bmatrix} \quad (3.1)$$

where $\underline{Y}_1$ is the vector of (n-m) known observations, $\underline{Y}_m$ is the vector of ‘m’ missing observations, X_1 is part of the design points corresponding to the known and X_m is corresponding to the missing observations design points. Let the error is also partitioned accordingly. The least square estimate of the parameters from the known observations is $\hat{\beta} = (X_1' X_1)^{-1} X'_1 Y_1$. Then the estimated missing observations can be obtained as $Y_m = X_m \hat{\beta}$.

Consider the problem of estimating the parameters and missing responses using expected maximization algorithm. The response variable Y follows $N(X\beta, \sigma^2)$. If $\underline{y} = [y_1, y_2 \ldots y_n]'$ be the observed responses (including missing responses) then the log of the likelihood function is

$$L(\underline{y}) = (2\pi\sigma^2)^{-n/2} \times \exp\left\{ \frac{-1}{2\sigma^2} \left[\sum_{j=1}^{n-m} (y_j - x_j\beta)^2 + \sum_{j=n-m+1}^{n} (y_j - x_j\beta)^2 \right] \right\} \quad (3.2)$$

$$\text{Log } L(\underline{y}) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\left[\sum_{j=1}^{n-m} (y_j - x_j\beta)^2 + \sum_{j=n-m+1}^{n} (y_j - x_j\beta)^2 \right] \quad (3.3)$$

In the expectation step, expected value of log of likelihood is evaluated, it results the conditional expectation of missing response observation as

$$E[\log L(y)] \Rightarrow E[y_u / y, X) = x_u \hat{\beta}^{(k)} \text{ and } E[y_u^2 / y, X) = (x_u \hat{\beta}^{(k)})' (x_u \hat{\beta}^{(k)}) + \sigma^{2(k)} \quad (3.4)$$

In the implementation of the Maximization step, minimize the current residual sum of squares using least square method and estimate the parameters β and σ^2. The estimates of parameters β and σ^2 and missing responses (y_u ; u = n-m+1, …, n) can be expressed in recurrence terms is presented below.

$$\frac{\partial}{\partial\beta}\log L(y) = 0 \quad \Rightarrow \hat{\beta} = \left[\sum_{j=1}^{n-m} x_j' x_j + \sum_{j=n-m+1}^{n} x_j' x_j \right]^{-1} \left[\sum_{j=1}^{n-m} x_j' y_j + \sum_{j=n-m+1}^{n} x_j' y_j \right]$$

$$\Rightarrow \quad \hat{\beta}^{(k+1)} = [X'X]^{-1} \left[\sum_{j=1}^{n-m} x_j' y_j + \sum_{j=n-m+1}^{n} x_j' y_j^{(k)} \right] \quad (3.5)$$

$$\frac{\partial}{\partial\sigma^2}\log L(y)=0 \Rightarrow \hat{\sigma}^2=\frac{1}{n}\left[\sum_{j=1}^{n-m}(y_j-x_j\hat{\beta})^2+\sum_{j=n-m+1}^{n}(y_j-x_j\hat{\beta})^2\right]$$

$$\Rightarrow\ \hat{\sigma}^{2(k+1)}=\frac{1}{n}\left[\sum_{j=1}^{n-m}(y_j-x_j\hat{\beta}^{(k+1)})^2+\sum_{j=n-m+1}^{n}(y_j-x_j\hat{\beta}^{(k+1)})^2\right]$$

$$\Rightarrow\ \hat{\sigma}^{2(k+1)}=\frac{1}{n}\left[(n-m)\hat{\sigma}_1^{2\ (k+1)}+m\hat{\sigma}_2^{2\ (k+1)}\right] \tag{3.6}$$

$$\frac{\partial}{\partial y_u}\log L(y)=0 \Rightarrow \hat{y}_u=x_u\hat{\beta}$$

$$\Rightarrow\ \hat{y}_u^{(k+1)}=\ x_u\hat{\beta}^{(k+1)} \tag{3.7}$$

The method of estimating the parameters and missing values using Least square and EM algorithm (with initial guess values as means or zeros) are illustrated through suitable examples in case of first order response surface design under with and without restrictions on the moment matrix.

Example: 3.1: 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_{27x4}. The vector of responses $\underline{Y}$ corresponding at the design points are given below.

$$X'=\begin{bmatrix} 1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\ -1&0&1&-1&0&1&-1&0&1&-1&0&1&-1&0&1&-1&0&1&-1&0&1&-1&0&1&-1&0&1\\ -1&-1&-1&0&0&0&1&1&1&-1&-1&-1&0&0&0&1&1&1&-1&-1&-1&0&0&0&1&1&1\\ -1&-1&-1&-1&-1&-1&-1&-1&-1&0&0&0&0&0&0&0&0&0&1&1&1&1&1&1&1&1&1 \end{bmatrix}$$

$\underline{Y}$ = [159 395 149 25 255 251 184 363 378 260 454 $\mathbf{Y_{12}}$ 98 422 270 237 362 363 146 417 150 103 455 172 $\mathbf{Y_{25}}$ 492 278]′

The estimation of missing responses at the 12th and 25th design points are evaluated and presented below

Case-(i): The estimated values of parameters and missing observations at the design points are evaluated using Least square method and are presented below. For the above design matrix X,

$$(X_1'X_1)=\begin{bmatrix} 25&0&0&-1\\ 0&16&2&1\\ 0&2&16&-1\\ -1&1&-1&17 \end{bmatrix}$$

$$(X_1'X_1)^{-1}=\begin{bmatrix} 0.040095141012 & -0.00016989466 & 0.0001698946650 & 0.002378525314\\ -0.00016989466 & 0.06379544682 & -0.008239891267 & 0.004247366632\\ 0.000169894665 & -0.008239891267 & 0.063795446822 & 0.004247366632\\ 0.002378525314 & 0.004247366632 & 0.004247366632 & 0.059463132857 \end{bmatrix}$$

$$\hat{\beta}=(X_1'X_1)^{-1}X_1'Y_1 \Rightarrow [\ 274.2528033\quad 45.23904179\quad 28.42762487\quad 18.32008155\ ]'$$

$$\hat{Y}_u=X_u\hat{\beta} \Rightarrow \hat{Y}_{(1\ 1\ -1\ 0)}=291.0642202 \quad\text{and}\quad \hat{Y}_{(1\ -1\ 1\ 1)}=275.7614$$

Case-(ii): Consider the initial guess values for missing observations at the 12th and 25th design points be $Y_{12} = Y_{25} = 0$.

$$(X'X) = \begin{bmatrix} 27 & 0 & 0 & 0 \\ 0 & 18 & 0 & 0 \\ 0 & 0 & 18 & 0 \\ 0 & 0 & 0 & 18 \end{bmatrix}; \quad (X'X)^{-1} = \begin{bmatrix} 0.03703704 & 0.00000000 & 0.00000000 & 0.00000000 \\ 0.00000000 & 0.05555556 & 0.00000000 & 0.00000000 \\ 0.00000000 & 0.00000000 & 0.05555556 & 0.00000000 \\ 0.00000000 & 0.00000000 & 0.00000000 & 0.05555556 \end{bmatrix}$$

The estimated parameters and missing observations values can be evaluated using (3.5), (3.7) at each iteration are presented in table 3.1.

Iteration No.	**Estimated Parameters values**				**Estimated Missing Values**	
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{Y}_{12}$	$\hat{Y}_{25}$
0	-	-	-	-	0	0
1	272.13032	45.90124	27.76543	16.39712	268.37043	241.14812
2	274.02436	45.49303	28.17364	18.02174	290.26612	270.39164
3	274.22483	45.31206	28.35461	18.26259	291.34367	274.72668
4	274.24860	45.25846	28.40821	18.30722	291.18231	275.53000
5	274.25202	45.24407	28.42260	18.31698	291.09886	275.70564
6	274.25263	45.24033	28.42633	18.31931	291.07354	275.74751
7	274.25276	45.23937	28.42729	18.31989	291.06666	275.75791
8	274.25200	45.23912	28.42754	18.42754	291.0648	275.76063
9	274.25280	45.23280	28.42760	18.32007	291.06442	275.76123
10	274.25281	45.23905	28.42762	18.32008	291.06434	275.76142
11	274.25280	45.23904	28.42763	18.32008	291.06424	275.76152
12	274.25280	45.23904	28.42763	18.32008	291.06420	275.76151

Table 3.1

Case-(iii): Let us assume that the initial missing values at the 12th and 25th design points be $Y_{12} = Y_{25} = 273.52000$ (average of known responses). The estimated parameters and missing observations values in each iteration are tabulated in table 3.2.

Iteration No.	**Estimated Parameters values**				**Estimated Missing Values**	
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{Y}_{12}$	$\hat{Y}_{25}$
0	-	-	-	-	273.52000	273.52000
1	274.19391	45.05704	28.60963	18.36691	288.63111	276.60444
2	274.25018	45.19600	28.47067	18.33963	290.64138	276.11342
3	274.25334	45.22839	28.43828	18.32581	290.97554	275.86453
4	274.25305	45.23636	28.43031	18.32161	291.04342	275.78900
5	274.25288	45.23836	28.42831	18.32045	291.05912	275.76864
6	274.25282	45.23887	28.42780	18.32018	291.06292	275.76332
7	274.25281	45.23900	28.42767	18.32011	291.06391	275.76192
8	274.25280	45.23903	28.42764	18.32009	291.06411	275.76161
9	274.25280	45.23904	28.42763	18.32008	291.06421	275.76150
10	274.25280	45.23904	28.42763	18.32008	291.06421	275.76150

Table-3.2

Example 3.2: Let us consider a response surface design conducted with three factors each with three levels with 18 design points given in the matrix X. The vector of responses Y corresponding at the design points are given below.

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

$\underline{Y}$ = [2.83 3.25 3.56 2.53 $\mathbf{Y_5}$ 3.19 2.23 2.65 3.06 2.57 3.08 3.5 2.42 2.79 3.03 $\mathbf{Y_{16}}$ 2.85 3.12]′

The estimation of missing responses at the 5th and 16th design points can be evaluated using

Case-(i): The estimated values of parameters and missing responses at the design points are evaluated and presented below. For the above design matrix X, we can obtain

$$(X_1'X_1) = \begin{bmatrix} 16 & -2 & -3 & -5 \\ -2 & 12 & 0 & 1 \\ -3 & 0 & 11 & 6 \\ -5 & 1 & 6 & 13 \end{bmatrix}$$

$$(X_1'X_1)^{-1} = \begin{bmatrix} 0.072523215404 & 0.010083746368 & 0.006665527260 & 0.024041474391 \\ 0.010083746368 & 0.085455477696 & 0.005640061527 & -0.005298239617 \\ 0.006665527260 & 0.005640061527 & 0.122372244060 & 0.054349683814 \\ 0.024041474391 & -0.005298239617 & 0.054349683814 & 0.111661824189 \end{bmatrix}$$

Then $\hat{\beta} = \begin{bmatrix} 2.8846738449 & 0.3144676124 & -0.020951489 & -0.2142636654 \end{bmatrix}'$

The resulting missing values at the design points are $\hat{Y}_{(1\ 0\ 0\ 1)} = 2.671$, $\hat{Y}_{(1\ -1\ 1\ 1)} = 2.336$

Case-(ii): Let us assume that the initial missing values at the 5th and 16th design points be $Y_5 = Y_{16} = 2.91625$ (The average of known responses).

$$(X'X) = \begin{bmatrix} 18 & -3 & -2 & -3 \\ -3 & 13 & -1 & 0 \\ -2 & -1 & 12 & 7 \\ -3 & 0 & 7 & 15 \end{bmatrix}$$

$$(X'X)^{-1} = \begin{bmatrix} 0.060139661 & 0.014322360 & 0.005771697 & 0.009334473 \\ 0.014322360 & 0.081017529 & 0.010260793 & -0.001923899 \\ 0.005771697 & 0.010260795 & 0.116075246 & -0.053014109 \\ 0.009334473 & -0.001923899 & -0.053014109 & 0.093273479 \end{bmatrix}$$

The estimated parameters and missing observations values can be evaluated using (3.5), (3.7) at each iteration are presented in table 3.3.

Iteration No.	**Estimated Parameters Values**				**Estimated Missing Values**	
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{Y}_5$	$\hat{Y}_{16}$
0	-	-	-	-	2.916250	2.915250
1	2.90196014	0.30640878	-0.01660326	-0.19492412	2.778072	2.495963
2	2.89020526	0.31206061	-0.01980384	-0.20797982	2.707036	2.384024
3	2.88643067	0.31371751	-0.02060342	-0.21225987	2.682225	2.350361
4	2.88523076	0.31423105	-0.02083858	-0.21362778	2.674171	2.339850
5	2.88485027	0.31439279	-0.02091155	-0.21406216	2.671603	2.336533
6	2.88472974	0.31444390	-0.02093449	-0.21419983	2.670788	2.335484
7	2.88469159	0.31446008	-0.02094174	-0.21424340	2.670530	2.335152
8	2.88467943	0.31446525	-0.02094408	-0.21425728	2.670448	2.335046
9	2.88467561	0.31446685	-0.02094478	-0.21426164	2.670422	2.335013
10	2.88467439	0.31446739	-0.02094505	-0.21426303	2.670414	2.335002
11	2.88467400	0.31446753	-0.02094508	-0.21426349	2.670411	2.334999
12	2.88467394	0.31446759	-0.02094514	-0.21426355	2.670411	2.334998
13	2.88467390	0.31446760	-0.02094510	-0.21426360	2.670410	2.334998
14	2.88467390	0.31446760	-0.02094510	-0.21426360	2.670410	2.334998
15	2.88467390	0.31446760	-0.02094510	-0.21426360	2.670410	2.334998

Table-3.3

Case-(iii): Let us assume that the initial missing values at the 5th and 16th design points be $Y_5 = Y_{16} = 0$. The estimated values of the parameters and missing values at each iteration are presented table 3.4.

Iteration No.	Estimated Parameters values				Estimated Missing Values	
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{Y}_5$	$\hat{Y}_{16}$
0					0	0
1	2.55689326	0.41762505	-0.03155408	-0.60856278	1.948330	1.499151
2	2.85288508	0.32768841	-0.02661992	-0.25071662	2.452170	2.062086
3	2.87462406	0.31870670	-0.02282503	0.22575497	2.602168	2.247860
4	2.88149212	0.31581475	-.002154761	-0.21789896	2.648869	2.307337
5	2.88366615	0.31489468	-0.02113655	-0.21541478	2.663593	2.326231
6	2.88435463	0.31460293	-0.02100581	0-.21541478	2.668251	2.332220
7	2.88457273	0.31451045	-0.02096433	0-.21437917	2.669726	2.334118
8	2.88464186	0.31448118	-0.02095123	-0.21430019	2.670194	2.334719
9	2.8846637	0.31447190	-0.02094710	-0.2142752	2.670342	2.334909
10	2.8846706	0.31446900	-0.02094580	-0.2142673	2.670389	2.334969
11	2.884672883	0.31446802	-0.02094529	-0.21426483	2.670403	2.334989
12	2.88467283	0.31446802	-0.02094529	-0.21426483	2.670408	2.334995
13	2.88467354	0.31446773	-0.02094518	-0.21426401	2.670408	2.334995
14	2.88467381	0.31446764	-0.02094515	-0.21426370	2.670410	2.334997
15	2.88467381	0.31446764	-0.02094515	-0.21426370	2.670410	2.334997

Table-3.4

Note:

1. The estimated values for the parameters and missing responses using the least square method and EM algorithm are resulting to same.
2. The number of iterations depend on the chosen initial values for the estimated responses.

ACKNOWLEDGEMENTS

The authors are thankful to the UGC for providing the fellowship to carry out this research work under the BSR RFSMS scheme.

REFERENCES

1. Allan,F.E & Wishart J., A method of estimating yield of missing plot in field experimental work, Journal of Agricultural Science, 20 (1930), 399-406.
2. Bhatra Charyulu N.Ch., Some studies on the second order response surface designs, unpublished Ph.D. theses submitted to Osmania University, Hyderabad, (1994).
3. Box M.J, Draper N.R and Hunter W.G., Missing values in multiresponse non linear model fitting, Technometrics, 12(3) (1970), 613-621.
4. Dempster.A.P, Laird.N.M, and Rubin.D.B, Maximum Likelihood from Incomplete Data via the EM Algorithm, Journal of the Royal statistical society, Series B, 39 (1977),1-38.
5. Draper N.R., Missing values in Response surface designs, Technometrics, 3 (1961), 389-398.
6. Healy, M.J.R. and Westmacott, M., Missing values in experiments analyzed on automatic computers, Applied Statistics 5 (1956), 203-206.
7. McLachlan Geoffrey J and Krishnan T, The EM Algorithms and Extensions, Wiley Inter science, 2nd edition, (2008).
8. Rubin, D.B. A Non-iterative Algorithm for least squares estimation of missing values in any analysis of variance design", Applied Statistics, 21 (1972), 136-141.
9. Subramani J and Ponnuswamy K.N. A non iterative least squares estimation of missing values in experimental designs, Journal of Applied Statistics, 16(1) (1989), 77-86.
10. Yates F.Y. The analysis of replicated experiments when the field results are incomplete, Empire Journal of Experimental Agriculture, 1 (1933), 129-142.

Source of support: UGC, BSR RFSMS scheme, India. Conflict of interest: None Declared