American Society for Quality

Missing Values in Response Surface Designs
Author(s): Norman R. Draper
Source: *Technometrics*, Vol. 3, No. 3 (Aug., 1961), pp. 389-398
Published by: American Statistical Association and American Society for Quality
Stable URL: http://www.jstor.org/stable/1266729
Accessed: 09/03/2010 10:52

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/action/showPublisher?publisherCode=astata.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

American Statistical Association and *American Society for Quality* are collaborating with JSTOR to digitize, preserve and extend access to *Technometrics.*

http://www.jstor.org

Missing Values in Response Surface Designs[1]

NORMAN R. DRAPER[2]

Mathematics Research Center, United States Army, Madison, Wisconsin

The estimation of missing values for a general design is described and discussed. Formulae are provided for the estimation of missing values for two well-known, three factor, second order rotatable designs, with zero to six center points. A worked example illustrates the use of the formulae in the case of the cube plus octahedron plus one center point design.

1. INTRODUCTION

Suppose we fit by least squares, to the results $\mathbf{y}' = (y_1, y_2, \cdots, y_N)$ of an experimental investigation, a regression equation of the form $\hat{\mathbf{y}} = \mathbf{Xb}$, i.e., the model considered is $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, where $\mathbf{e} \sim N(\mathbf{0}, \mathbf{I}\sigma^2)$. Then $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. Usually when an observation is missing, we simply drop out the corresponding row in the $\mathbf{X}$ matrix. However, in the case of a rotatable design and certain other response surface designs, where the $\mathbf{X}'\mathbf{X}$ matrix and its inverse are already known (e.g. designs given by Cochran and Cox [10]), we might prefer to estimate the missing value(s) and then proceed with the analysis as originally planned, except for adjustments due to the loss of degrees of freedom. We shall now see what this involves.

Suppose that the matrix $\mathbf{X}'$ is divided into $[\mathbf{X}_1', \mathbf{X}_2']$ in such a way that $\mathbf{X}_1$ is associated with yield values $\mathbf{y}_1$ that are observed, and $\mathbf{X}_2$ is associated with yield values $\mathbf{y}_2$ that are missing. This is easily effected by rearranging the order of the symbols $y_1, y_2, \cdots, y_N$ so that the f (say) missing values occupy the last f places $y_{(N-f+1)}, \cdots, y_N$ and re-arranging the rows of $\mathbf{X}$ to correspond so that the model remains the same.

2A. METHOD (A)

Tocher [1] gives the following:

$$E(\mathbf{y}) = E\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix} = \begin{bmatrix}\mathbf{X}_1\\ \mathbf{X}_2\end{bmatrix}\boldsymbol{\beta}.$$

Thus the estimates from the observed responses would be

$$\begin{aligned}\mathbf{b} &= (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y}_1\\ &= (\mathbf{X}'\mathbf{X} - \mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_1'\mathbf{y}_1\\ &= (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{I} - \mathbf{X}_2'\mathbf{X}_2(\mathbf{X}'\mathbf{X})^{-1})^{-1}\mathbf{X}_1'\mathbf{y}_1\\ &= (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{I} + \mathbf{X}_2'\mathbf{M}\mathbf{X}_2(\mathbf{X}'\mathbf{X})^{-1})\mathbf{X}_1'\mathbf{y}_1\end{aligned}$$

---

[1] Sponsored by the United States Army under Contract No. DA-11-022-ORD-2059, Mathematics Research Center, United States Army, Madison, Wisconsin.

[2] Now at the Department of Statistics, University of Wisconsin.

where $\mathbf{M} = (\mathbf{I} - \mathbf{X}_2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_2')^{-1}$ and the identity $(\mathbf{I} + \mathbf{AB})^{-1} = \mathbf{I} - \mathbf{A}(\mathbf{I} + \mathbf{BA})^{-1}\mathbf{B}$, [1], has been employed. After some algebra, the expected values for the missing observations are seen to be

$$\hat{\mathbf{y}}_2 = \mathbf{X}_2\mathbf{b} = \mathbf{M}\mathbf{X}_2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_1'\mathbf{y}_1 .$$

Since

$$\mathbf{X}_1'\mathbf{y}_1 = (\mathbf{X}_1' , \mathbf{X}_2')\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{0} \end{bmatrix} = \mathbf{X}'\mathbf{y}^{(0)}, \quad \text{say, it follows that} \quad \hat{\mathbf{y}}_2 = \mathbf{M}\mathbf{X}_2\mathbf{b}_0$$

where $\mathbf{b}_0$ is the estimate of $\boldsymbol{\beta}$ obtained from the data assuming the missing observations have zero values. The estimate of $\boldsymbol{\beta}$ using the data $\mathbf{Z} = [\mathbf{y}_1 , \hat{\mathbf{y}}_2]'$ is thus

$$\begin{aligned}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z} &= (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}_1'\mathbf{y}_1 + \mathbf{X}_2'\hat{\mathbf{y}}_2) \\ &= (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}_1'\mathbf{y}_1 + \mathbf{X}_2'\mathbf{M}\mathbf{X}_2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_1'\mathbf{y}_1) \\ &= (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{I} + \mathbf{X}_2'\mathbf{M}\mathbf{X}_2(\mathbf{X}'\mathbf{X})^{-1})\mathbf{X}_1'\mathbf{y}_1 \\ &= \mathbf{b}.\end{aligned}$$

It follows that the results of an experiment with missing values can be analyzed in the following way. Perform a standard analysis with all missing observations given zero values to obtain $\mathbf{b}_0$. Evaluate $\hat{\mathbf{y}}_2 = \mathbf{M}\mathbf{X}_2\mathbf{b}_0$. Perform the standard analysis using $\hat{\mathbf{y}}_2$ for the missing observations. The final coefficients $\mathbf{b}$ obtained will be the same as those that would have been obtained if $(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y}_1$ had been evaluated

2B. Method (B)

We estimate the missing values $\mathbf{y}_2$ by choosing them in such a way that the residual sum of squares is minimized with respect to those values. Now:

$$\begin{aligned}\text{Residual sum of squares } S^2 &= \mathbf{y}'\mathbf{y} - \mathbf{b}'\mathbf{X}'\mathbf{y} \\ &= \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= \mathbf{y}'\mathbf{H}\mathbf{y}\end{aligned}$$

where

$$\mathbf{H} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

(Note that $\mathbf{X}'\mathbf{H} = \mathbf{X}' - \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{0}$. This provides a useful check on $\mathbf{H}$.) Let $\mathbf{H} = (\mathbf{h}_1 , \mathbf{h}_2 , \cdots , \mathbf{h}_N) = (\mathbf{h}_1' , \mathbf{h}_2' , \cdots , \mathbf{h}_N')'$, since $\mathbf{H}$ is symmetric. Differentiating S^2 partially with respect to y_i and setting the result equal to zero to satisfy the condition for a minimum we find that $\mathbf{h}_i'\mathbf{y} = 0$, or $h_{i1}y_1 + h_{i2}y_2 + \cdots + h_{iN}y_N = 0$. Since $(\partial/\partial y_i)(\mathbf{h}_i'\mathbf{y}) = h_{ii} > 0$, because $\mathbf{H}$ is positive definite when $S^2 > 0$, this equation does give a minimum.

Solution of the equation for y_i provides us with the estimate $\hat{y}_i$ of a single missing value y_i . If two values y_i and y_j are missing, we must solve the simultaneous equations $\mathbf{h}_i'\mathbf{y} = \mathbf{h}_j'\mathbf{y} = 0$ for $\hat{y}_i$ and $\hat{y}_j$. The obvious extension applies for more missing values.

It is easy to see that this method is equivalent to the one given by Tocher. For, since $\mathbf{X}' = [\mathbf{X}_1' , \mathbf{X}_2']$, it follows that $\mathbf{X}'\mathbf{X} = \mathbf{X}_1'\mathbf{X}_1 + \mathbf{X}_2'\mathbf{X}_2$ and

$$\mathbf{H} = \left[\begin{array}{c|c} \mathbf{I} - \mathbf{X}_1(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_1' & -\mathbf{X}_1(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_2' \\ \hline -\mathbf{X}_2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_1' & \mathbf{I} - \mathbf{X}_2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_2' \end{array}\right]$$

Thus if $\mathbf{y}_1' = (y_1 , y_2 , \cdots , y_{N-f})$ is the vector of observed values and

$\mathbf{y}_2' = (y_{N-f+1}, \cdots, y_N)$ is the vector of missing values, we obtain the estimates $\hat{\mathbf{y}}_2$ from

$$-\mathbf{X}_2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_1'\mathbf{y}_1 + (\mathbf{I} - \mathbf{X}_2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_2')\hat{\mathbf{y}}_2 = \mathbf{0}$$

which implies that

$$\hat{\mathbf{y}}_2 = \mathbf{M}\mathbf{X}_2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_1'\mathbf{y}$$

as before. Thus the two methods (A) and (B) are equivalent and lead to the same estimates for $\hat{\mathbf{y}}_2$ and $\mathbf{b}$.

3. Analysis of Variance

The correct analysis of variance table is

Source	S. S.	d. f.	M. S.
Coefficients $\mathbf{b}$	$\mathbf{b}'\mathbf{X}_1'\mathbf{y}_1$	j	
Residual	by difference	$N - f - j$	s^2
Total	$\mathbf{y}_1'\mathbf{y}_1$	$N - f$	

in which only the observations $\mathbf{y}_1$ appear and where j is the number of coefficients estimated. The correct variance-covariance matrix of the $\mathbf{b}$ coefficients is, [1], estimated by s^2 times

$$(\mathbf{X}_1'\mathbf{X}_1)^{-1} = (\mathbf{X}'\mathbf{X})^{-1} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_2'\mathbf{M}\mathbf{X}_2(\mathbf{X}'\mathbf{X})^{-1},$$

as can easily be verified.

4. Discussion

In general it would be preferable to use method (A), especially for a design used only once, since this would involve the evaluation of only that portion of the matrix $\mathbf{H} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ which was required, and thus considerable calculation would be avoided. Note particularly that $\mathbf{M}$ is an f by f matrix and so in the case of one or two missing values (i.e., $f = 1$ or 2), the calculations would not be difficult. However when a design is frequently used, it is better to evaluate $\mathbf{H}$ once and for all and then make use of the appropriate portion of it when required. This speeds up the estimation part of the work and also reduces the chance of error in the computation. Furthermore, it is particularly easy to explain the method to computing assistants. In cases where the statistician decides either that he is prepared to use $(\mathbf{X}'\mathbf{X})^{-1}s^2$ instead of the correct $(\mathbf{X}_1'\mathbf{X}_1)^{-1}s^2$ for the estimated variance-covariance matrix of the regression coefficients, or that he does not wish to examine the standard errors of individual coefficients at all, only the missing values values are required and, when $\mathbf{H}$ is available, this involves very little work indeed. (The effect on the standard errors of using the incorrect variance-covariance matrix is illustrated in a worked example later in the text.) In Section 6, we shall give the matrix $\mathbf{H}$ for the following well-known [2, 3, 4, 5, 6, 10], three-factor, second order rotatable designs:

(1) cube plus doubled octahedron plus n center points, $n \leq 6$.

(2) cube plus octahedron plus n center points, $n \leq 6$.

Because these designs have some levels which are neither integral nor rational, some or all of the numbers which occur in their $\mathbf{H}$ matrices cannot be expressed in rational form (see Tables 1 and 2). When designs with integral or rational levels are considered, however, this difficulty does not arise, and the estimation of missing values for such designs is extremely simple, once $\mathbf{H}$ is available. These remarks would apply, for example, to the three factor, three level response surface designs discussed by De Baun [7] and the three level designs of Box and Behnken [8]. Appropriate calculations for five designs given by De Baun appear in an earlier version of this paper [9].

5. Notation

To the results of a group of experiments on three coded factors x_1, x_2, x_3, it is desired to fit the second order model

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2 + \beta_{12} x_1 x_2 + \beta_{23} x_2 x_3 + \beta_{31} x_3 x_1 .$$

Thus each row of $\mathbf{X}$ consists of the values of

$$1,\ x_1,\ x_2,\ x_3,\ x_1^2,\ x_2^2,\ x_3^2,\ x_1x_2,\ x_2x_3,\ x_3x_1,$$

at one point of the experimental design and the vector of coefficients to be estimated is

$$\boldsymbol{\beta}' = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_{11}, \beta_{22}, \beta_{33}, \beta_{12}, \beta_{23}, \beta_{31}).$$

The two designs mentioned above both have $\mathbf{X}'\mathbf{X}$ of the same form, namely

$$\mathbf{X}'\mathbf{X} = \left[\begin{array}{cccccccccc}
N & 0 & 0 & 0 & C & C & C & 0 & 0 & 0 \\
 & C' & 0 & 0 & & & & & & \\
 & & C' & 0 & & \mathbf{0} & & & \mathbf{0} & \\
 & & & C' & & & & & & \\
 & & & & D & B & B & & & \\
 & & & & & D & B & & \mathbf{0} & \\
 & & & & & & D & & & \\
 & & & & & & & B' & 0 & 0 \\
 & & & & & & & & B' & 0 \\
\text{Symmetric} & & & & & & & & & B'
\end{array}\right]
\begin{array}{l}
1 \\ x_1 \\ x_2 \\ x_3 \\ x_1^2 \\ x_2^2 \\ x_3^2 \\ x_1x_2 \\ x_2x_3 \\ x_3x_1
\end{array}$$

where the design consists of the N points (x_{1u}, x_{2u}, x_{3u}), $u = 1, 2, \cdots, N$, and where

$$C = C' = \sum_u x_{iu}^2, \quad D = \sum_u x_{iu}^4, \quad B = B' = \sum_u x_{iu}^2 x_{ju}^2, \quad i,\ j = 1, 2 \text{ or } 3,$$

$i \neq j$. Also, since the designs are rotatable, $D = 3B$. (The column to the right of $\mathbf{X}'\mathbf{X}$ indicates the order in which the rows and columns of $\mathbf{X}'\mathbf{X}$ have been arranged and thus the order in which the estimated coefficients of the second order model will emerge.) The inverse $(\mathbf{X}'\mathbf{X})^{-1}$ is of the same form, with

$$
\begin{aligned}
P &= 10B^2A && \text{in place of } N,\\
Q &= -2BCA && \text{in place of } C,\\
R &= \{4NB - 2C^2\}A && \text{in place of } D,\\
S &= (C^2 - NB)A && \text{in place of } B,\\
&1/B' && \text{in place of } B',\\
&1/C' && \text{in place of } C',
\end{aligned}
$$

where $1/A = 2B(5NB - 3C^2)$, see [2].

6. The Missing Value Formulae for the Chosen Designs

6.1. *Cube plus doubled octahedron plus n center points, rotatable.*

$$
\begin{array}{c}
\begin{array}{cccccccccc}
\underline{1} & \underline{x_1} & \underline{x_2} & \underline{x_3} & \underline{x_1^2} & \underline{x_2^2} & \underline{x_3^2} & \underline{x_1x_2} & \underline{x_2x_3} & \underline{x_3x_1}
\end{array}\\
\mathbf{X} = \left[\begin{array}{cccccccccc}
1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1\\
1 & 1 & -1 & -1 & 1 & 1 & 1 & -1 & 1 & -1\\
1 & -1 & 1 & -1 & 1 & 1 & 1 & -1 & -1 & 1\\
1 & 1 & 1 & -1 & 1 & 1 & 1 & 1 & -1 & -1\\
1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & -1 & -1\\
1 & 1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & 1\\
1 & -1 & 1 & 1 & 1 & 1 & 1 & -1 & 1 & -1\\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1\\
\hdashline
1 & -a & 0 & 0 & a^2 & 0 & 0 & 0 & 0 & 0\\
1 & -a & 0 & 0 & a^2 & 0 & 0 & 0 & 0 & 0\\
1 & a & 0 & 0 & a^2 & 0 & 0 & 0 & 0 & 0\\
1 & a & 0 & 0 & a^2 & 0 & 0 & 0 & 0 & 0\\
1 & 0 & -a & 0 & 0 & a^2 & 0 & 0 & 0 & 0\\
1 & 0 & -a & 0 & 0 & a^2 & 0 & 0 & 0 & 0\\
1 & 0 & a & 0 & 0 & a^2 & 0 & 0 & 0 & 0\\
1 & 0 & a & 0 & 0 & a^2 & 0 & 0 & 0 & 0\\
1 & 0 & 0 & -a & 0 & 0 & a^2 & 0 & 0 & 0\\
1 & 0 & 0 & -a & 0 & 0 & a^2 & 0 & 0 & 0\\
1 & 0 & 0 & a & 0 & 0 & a^2 & 0 & 0 & 0\\
1 & 0 & 0 & a & 0 & 0 & a^2 & 0 & 0 & 0\\
\hdashline
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right]
\end{array}
$$

where $a = \sqrt{2}$. The matric $\mathbf{H}$ is size $(20 + n)$ by $(20 + n)$ and has the form

$$
\begin{array}{c}
\begin{array}{cccccccccccccccccccccccc}
1&2&3&4&5&6&7&8&9&10&11&12&13&14&15&16&17&18&19&20&21&&&(20+n)
\end{array}\\
\mathbf{H} = \left[\begin{array}{cccccccc|cccccccccccc|cccc}
p&r&r&s&r&s&s&q&t&t&u&u&t&t&u&u&t&t&u&u&v&\cdot&\cdot&v\\
&p&s&r&s&r&q&s&u&u&t&t&t&t&u&u&t&t&u&u&v&\cdot&\cdot&v\\
&&p&r&s&q&r&s&t&t&u&u&u&u&t&t&t&t&u&u&v&\cdot&\cdot&v\\
&&&p&q&s&s&r&u&u&t&t&u&u&t&t&t&t&u&u&v&\cdot&\cdot&v\\
&&&&p&r&r&s&t&t&u&u&t&t&u&u&u&u&t&t&v&\cdot&\cdot&v\\
&&&&&p&s&r&u&u&t&t&t&t&u&u&u&u&t&t&v&\cdot&\cdot&v\\
&&&&&&p&r&t&t&u&u&u&u&t&t&u&u&t&t&v&\cdot&\cdot&v\\
&&&&&&&p&u&u&t&t&u&u&t&t&u&u&t&t&v&\cdot&\cdot&v\\
\hline
&&&&&&&&w&w'&x&x&y&y&y&y&y&y&y&y&z&\cdot&\cdot&z\\
&&&&&&&&&w&x&x&y&y&y&y&y&y&y&y&z&\cdot&\cdot&z\\
&&&&&&&&&&w&w'&y&y&y&y&y&y&y&y&z&\cdot&\cdot&z\\
&&&&&&&&&&&w&y&y&y&y&y&y&y&y&z&\cdot&\cdot&z\\
&&&&&&&&&&&&w&w'&x&x&y&y&y&y&z&\cdot&\cdot&z\\
&&&&&&&&&&&&&w&x&x&y&y&y&y&z&\cdot&\cdot&z\\
&&&&&&&&&&&&&&w&w'&y&y&y&y&z&\cdot&\cdot&z\\
&&&&&&&&&&&&&&&w&y&y&y&y&z&\cdot&\cdot&z\\
&&&&&&&&&&&&&&&&w&w'&x&x&z&\cdot&\cdot&z\\
&&&&&&&&&&&&&&&&&w&x&x&z&\cdot&\cdot&z\\
&&&&&&&&&&&&&&&&&&w&w'&z&\cdot&\cdot&z\\
&&&&&&&&&&&&&&&&&&&w&z&\cdot&\cdot&z\\
\hline
&&&&&&&&&&&&&&&&&&&&z'&y'&\cdot&y'\\
&&&&&&&&&&&&&&&&&&&&&z'&\cdot&\cdot\\
&&&&&&&&&&&&&&&&&&&&&&\cdot&\cdot\\
\text{Symmetric}&&&&&&&&&&&&&&&&&&&&&&&z'
\end{array}\right]
\end{array}
$$

The numerical values of the symbols in the body of the matrix $\mathbf{H}$ can be found by referring to Table 1, using the column with the appropriate number of center points n indicated at the head of the column. We also note that

$$
\begin{aligned}
N &= 20 + n, & P &= 5/(4 + 5n),\\
C &= 16, & Q &= -2/(4 + 5n),\\
D &= 24, & R &= (4 + n)/4(4 + 5n),\\
B &= 8, & S &= (12 - n)/16(4 + 5n).
\end{aligned}
$$

TABLE 1
Cube plus doubled octahedron plus n center points, rotatable

n	0	1	2	3	4	5	6
p	5	49	39	107	17	165	97
q	-5	-41	-31	-83	-13	-125	-73
r	-1	-5	-3	-7	-1	-9	-5
s	1	13	11	31	5	49	29
t	$-\sqrt{2}$	$-4-9\sqrt{2}$	$-4-7\sqrt{2}$	$-12-19\sqrt{2}$	$-2-3\sqrt{2}$	$-20-29\sqrt{2}$	$-12-17\sqrt{2}$
u	$\sqrt{2}$	$9\sqrt{2}-4$	$7\sqrt{2}-4$	$19\sqrt{2}-12$	$3\sqrt{2}-2$	$29\sqrt{2}-20$	$17\sqrt{2}-12$
v	*	16	8	16	2	16	8
w	10	94	74	202	32	310	182
w'	-6	-50	-38	-102	-16	-154	-90
x	-2	-14	-10	-26	-4	-38	-22
y	0	4	4	12	2	20	12
y'	*	-80	-40	-80	-10	-80	-40
z	*	-16	-8	-16	-2	-16	-8
z'	*	64	72	224	38	384	232
Divisor	16	144	112	304	48	464	272

Note: Each entry in a column must be divided by the divisor of that column to give the correct value. However this is necessary only when obtaining **M** for the analysis of variance. In the estimation equations the divisor is common and cancels.
* does not occur.

6.2. *Cube plus octahedron plus n center points, rotatable.*

For **X**, refer to Section 6.1, set $a = 1.681793$, $a^2 = 2\sqrt{2}$, and delete the tenth, twelfth, fourteenth, sixteenth, eighteenth and twentieth rows of the **X** matrix given there; alternatively, see [10], noting that our x_1x_3 and x_2x_3 columns are interchanged. The matrix **H** is size $(14 + n)$ by $(14 + n)$ and has the same form as the **H** matrix in Section 6.1 but with the tenth, twelfth, fourteenth, sixteenth, eighteenth and twentieth rows and columns deleted.

The numerical values of the symbols in the body of the matrix **H** must now be taken from Table 2, using the column with the appropriate number of center points n indicated at the head of the column. We also note that

$$N = 14 + n, \qquad P = 5/(34 - 24\sqrt{2} + 5n),$$

$$C = 8 + 4\sqrt{2} = 13.656856, \qquad Q = -(4 + 2\sqrt{2})/4(34 - 24\sqrt{2} + 5n),$$

$$D = 24, \qquad R = (8 - 4\sqrt{2} + n)/4(34 - 24\sqrt{2} + 5n),$$

$$B = 8, \qquad S = (8\sqrt{2} - 2 - n)/16(34 - 24\sqrt{2} + 5n).$$

TABLE 2

Cube plus octahedron plus n center points, rotatable

n	0	1	2	3	4	5	6
p	0.280331	0.329749	0.330038	0.330136	0.330184	0.330214	0.330233
q	-0.280331	-0.230913	-0.230624	-0.230526	-0.230478	-0.230448	-0.230429
r	-0.073223	-0.023805	-0.023516	-0.023418	-0.023370	-0.023340	-0.023321
s	0.073223	0.122641	0.122930	0.123028	0.123076	0.123106	0.123125
t	-0.123146	-0.193034	-0.193443	-0.193580	-0.193649	-0.193691	-0.193718
u	0.123146	0.053258	0.052849	0.052712	0.052643	0.052601	0.052574
v	*	0.023982	0.012061	0.008056	0.006048	0.004842	0.004036
w	0.292894	0.391730	0.392308	0.392503	0.392600	0.392659	0.392698
x	-0.292894	-0.194058	-0.193479	-0.193285	-0.193188	-0.193129	-0.193090
y	0	0.098836	0.099415	0.099609	0.099706	0.099765	0.099804
y'	*	-0.988362	-0.497074	-0.332030	-0.249266	-0.199530	-0.166340
z	*	-0.033915	-0.017057	0.011393	0.008553	0.006847	0.005708
z'	*	0.011638	0.502926	0.667970	0.750734	0.800470	0.833660

* does not occur.

It will be noticed in Table 2 that, as n increases from 1 to 6, the values of p, q, r, s, t, u, w, x and y change very little, while v, y', z and z' change very nearly in proportion to $1/n$. Thus the estimates of any missing values obtained by using the columns for $n = 2, 3, 4, 5$ or 6 would differ very little from estimates obtained by first finding the mean yield from the n center points when $n \geq 1$, and then using this mean as a single center point observation with the $n = 1$ column. The estimates of missing values in other designs are more sensitive to additional center point observations, however.

7. Illustrative Example

(This example is constructed. The data were obtained by rounding off selected observations in a numerical example of response surface analysis given by Cochran and Cox [10, Table 8A.8].)

The cube plus octahedron plus one center point design was used for a group of experiments to the results of which it was intended to fit a second order response surface of the form shown in Section 5. The results $\mathbf{y}$, in the order defined by the order of the rows of matrix $\mathbf{X}$ in section 6.2, were

$$\mathbf{y}' = (16, c, 16, 7, 15, 8, 20, 5; d, 0, 25, 18, 7, 12; 24).$$

The letters c and d represent missing yields which are to be estimated. Missing are the second and ninth results. Using the notation of section 2B and referring to the matrix **H** of section 6.2 (with $n = 1$), we see that

$$\mathbf{h}_2' = (r, p, s, r, s, r, q, s; u, t, t, u, t, u; v),$$

$$\mathbf{h}_9' = (t, u, t, u, t, u, t, u; w, x, y, y, y, y; z),$$

where the values of the symbols are as given in the $n = 1$ column of Table 2. The estimation equations $\mathbf{h}_2'\mathbf{y} = \mathbf{h}_9'\mathbf{y} = 0$ are thus

$$p\hat{c} + u\hat{d} = -20q - 31r - 36s - 32t - 30u - 24v = 4.944919,$$

$$u\hat{c} + w\hat{d} = -62y - 67t - 20u - 24z = 6.554246.$$

Now

$$\mathbf{M} = \begin{bmatrix} p & u \\ u & w \end{bmatrix}^{-1} = \begin{bmatrix} 3.100700 & -0.421558 \\ -0.421558 & 2.610095 \end{bmatrix}$$

so that the required estimates are $\hat{c} = 12.570$, $\hat{d} = 15.023$.

From the formulae given in section 6.2, $P = 0.988364$, $Q = -0.337448$ $R = 0.165212$ and $S = 0.102712$. The estimates of the coefficients of the second order response surface are found in the usual way and are shown in Table 3. The analysis of variance table, calculated as shown in section 3, follows.

Source	Sum of Squares	d. f.	Mean Square
b	2971.15	10	
Residual	21.85	3	$7.28 = s^2$
Total	2993.00	13	

Using the value of s^2 just obtained, we can find two different sets of values for the standard errors of the estimated coefficients **b**, incorrectly, by using the diagonal terms of $(\mathbf{X}'\mathbf{X})^{-1}s^2$, or correctly, by using the diagonal terms of $(\mathbf{X}_1'\mathbf{X}_1)^{-1}s^2$ as given in section 3. In the latter case it is not necessary to calculate the whole of the matrix, only the ten diagonal terms. Table 3 shows, in addition to the values of the **b** coefficients, the standard errors which would be obtained in the two cases. Columns labelled (1) are the values obtained when $(\mathbf{X}_1'\mathbf{X}_1)^{-1}s^2$ is used; columns labelled (2) are values obtained when $(\mathbf{X}'\mathbf{X})^{-1}s^2$ is used.

The final column of table 3 shows the ratio of the incorrect standard error (2) to the correct standard error (1) for each coefficient. This ratio will always be less than unity, since use of the incorrect variance-covariance matrix will give standard errors that are smaller than the actual ones. It follows that confidence statements for the values of the regression coefficients based on the incorrect standard errors will be attributed a higher probability than is actually the case. While care must be exercised, use of the incorrect variance-covariance matrix will not, in most practical situations, appreciably affect the statistician's conclusions.

TABLE 3
The b *coefficients and their standard errors*

Coefficients	Value	s.e.(1)	s.e.(2)	(2)/(1)
b_0	23.957	2.690	2.682	0.997
b_1	-4.371	0.999	0.730	0.731
b_2	-1.124	0.809	0.730	0.903
b_3	0.354	0.809	0.730	0.903
b_{11}	-5.757	1.234	1.097	0.889
b_{22}	-0.812	1.131	1.097	0.970
b_{33}	-5.054	1.131	1.097	0.970
b_{12}	-1.696	1.124	0.953	0.848
b_{23}	0.946	1.124	0.953	0.848
b_{31}	-1.196	1.124	0.953	0.848

If it is not desired to examine the standard errors of the individual coefficients then, after the estimation of the missing yields, the analysis can be continued in the way indicated by Cochran and Cox [10].

REFERENCES

[1] Tocher, K. D., (1952), "The design and analysis of block experiments," *Journal of the Royal Statistical Society*, Series B, Vol. 14, pp. 45–100.

[2] Box, G. E. P. and Hunter, J. S., (1957), "Multi-factor experimental designs," *Annals of Mathematical Statistics*, Vol. 28, pp. 195–241.

[3] De Baun, Robert M., (1956), "Block effects in the determination of optimum conditions," *Biometrics*, Vol. 12, pp. 20–22.

[4] Gardiner, D. A., Grandage, A. H. E., and Hader, R. J., (1959), "Third order rotatable designs for exploring response surfaces," *Annals of Mathematical Statistics*, Vol. 30, pp. 1082–1096.

[5] Bose, R. C. and Draper, Norman R., (1959), "Second order rotatable designs in three dimensions," *Annals of Mathematical Statistics*, Vol. 30, pp. 1097–1112.

[6] Dykstra, O., (1960), "Partial duplication of response surface designs," *Technometrics*, Vol. 2, pp. 185–195.

[7] De Baun, Robert M., (1959), "Response surface designs for three factors at three levels," *Technometrics*, Vol. 1, pp. 1–8, corrections p. 419.

[8] Box, G. E. P. and Behnken, D. W., (1960), "Some new three level designs for the study of quantitative variables," *Technometrics*, Vol. 2, pp. 455–475.

[9] Draper, Norman R., "Missing value formulae for certain three factor, second order response surface designs," Report No. 201, Mathematics Research Center, U. S. Army, Madison, Wisconsin.

[10] Cochran, W. G. and Cox, G. M., *Experimental Designs*, second edition (1957), John Wiley and Sons.