ON ESTIMATION AND CONSTRUCTION IN FRACTIONAL REPLICATION[1]

BY K. S. BANERJEE AND W. T. FEDERER

Kansas State University and Cornell University

1. Summary and introduction. It was shown by the authors [1], [2] how to adjust the treatment design matrix X to furnish estimates of effects as orthogonal linear functions of observations for any irregular fractional replicate from an N treatment factorial. The fractional replicate considered earlier was such that the design matrix X was of dimensions $p \times p$ implying that p effect parameters be estimated from p observations. The method consisted in finding a matrix λ such that the design matrix X and the observation vector Y were augmented to become $X_1 = [X' \vdots X'\lambda]'$ of dimensions $(p + m) \times p$ and $Y_1 = [Y' \vdots Y'\lambda]'$ of dimensions $(p + m) \times 1$ with $p + m = N$ in such a way that $[X_1'X_1]$ reduced to a diagonal matrix. In the present note, the earlier results have been generalized in the sense that the design matrix X need not be square, that is, of dimensions $p \times p$, but is of dimensions $(p + m_1) \times p$, $p + m_1 < N$, implying that p effect parameters be estimated from $(p + m_1)$ observations. Besides this generalization the following additional results were obtained: (i) the structural relationship between the effect parameters retained and the observations omitted was derived, (ii) a working rule was developed for constituting the irregular fractional replicate with observations that are internally consistent making it possible to estimate the effect parameters, and (iii) a desirable procedure of designing the fractional replicate to obtain maximum efficiency was set forth.

2. Notation and the background material. A set of $\nu = p + m_1$ observational equations is denoted by $Y = X\mathbf{B} + e$, where Y is a $\nu \times 1$ random vector of observations with elements y_i, X is the $\nu \times p$ treatment design matrix with rank $p \leqq \nu$, $\mathbf{B}$ is the $p \times 1$ vector of effect parameters and e is a $\nu \times 1$ random vector of errors with $E(ee') = \sigma^2 I_\nu$. The least squares estimates of $\mathbf{B}$ are given by $\mathbf{B}^+ = [X'X]^{-1}X'Y$ with the covariance matrix as cov $(\mathbf{B}^+) = [X'X]^{-1}\sigma^2$.

With the augmentation as referred to in Section 1, the observational equations take the form, $Y_1 = X_1\mathbf{B}_1 + e_1$, where the dimensions of X_1 are $N \times p$ and those of Y_1 and e_1 are $N \times 1$ and the p columns of X_1 are orthogonal. With this adjustment and with $\nu = p$ it was demonstrated [1], [2] that the least squares estimates $\mathbf{B}_1^+$ are the same as $\mathbf{B}^+$ and that cov $(\mathbf{B}_1^+) =$ cov $(\mathbf{B}^+) = [X'X]^{-1}\sigma^2$.

Partitioning the design matrix X (dimensions $N \times N$) pertaining to a full replicate as

$$(2.1) \qquad X = \begin{bmatrix} X_{11} & \vdots & X_{12} \\ X_{21} & \vdots & X_{22} \end{bmatrix} = [X_1 \vdots X_2],$$

Received 18 October 1965.

[1] Supported in part by NSF Grant No. GP-4618 and by NIH Research Grant PHS-GM-#5900.

we note that X_{11} (dimensions $p \times p$) would correspond to the design matrix X for the fractional replicate dealt with earlier and that X_{21} (dimensions $m \times p$) to $\lambda' X$ (denoted earlier by X_m) with which the design matrix X was augmented. Denoting the augmented part $\lambda' Y$ of the observation vector Y_1 by Y_m, it was proved [2] that if solutions to the equations $X'_{22} Y_m{}^+ = -X'_{12} Y$ exist, it is always possible to find the matrix $\lambda = -X_{12}[X_{22}]^{-1}$. The procedure amounted to finding the values of m omitted observations from m equations obtained by equating to zero the m effects which were taken as aliases of the estimated effect parameters. Such a method of estimation is similar in principle to estimating "missing values," as has been done by Tocher [3].

3. The generalized situation. In the full factorial partitioned into N orthogonal contrasts of the N observations $X'X$ is a diagonal matrix D, say, with diagonal elements d_i, $i = 1, \cdots, N$. In the 2^n case $X'X/2^n = I_N =$ the identity matrix with dimensions $N \times N$. From this we note that $X'X = 2^n I_N = XX'$. This is not true for other factorial design matrices X since they are only columnwise orthogonal. However, we may transform the design matrix and the parameters in the following manner: let $X\mathbf{B} = (XD^{-\frac{1}{2}})(D^{\frac{1}{2}}\mathbf{B}) = W\mathbf{C}$, where $W = XD^{-\frac{1}{2}}$, $\mathbf{C} = D^{\frac{1}{2}}\mathbf{B}$, and $D^{\frac{1}{2}}$ is an $N \times N$ diagonal matrix with diagonal elements $d_i^{\frac{1}{2}}$ and similarly for $D^{-\frac{1}{2}}$. Then, since $X'X = D$, when we pre- and post-multiply both sides by $D^{-\frac{1}{2}}$ we obtain $D^{-\frac{1}{2}}X'XD^{-\frac{1}{2}} = D^{-\frac{1}{2}}DD^{-\frac{1}{2}} = I_N = W'W$. Likewise when we pre-multiply $W'W = I_N$ by W and then post-multiply both sides by W^{-1} we obtain $WW'WW^{-1} = WW^{-1} = I_N = WW'$. Partitioning the X matrix as in (2.1) we note that the following relations hold:

$$X'_{11}X_{11} + X'_{21}X_{21} = D_p \quad \text{(a } p \times p \text{ diagonal matrix);} \tag{3.1}$$

$$X'_{12}X_{12} + X'_{22}X_{22} = D_m \quad \text{(an } m \times m \text{ diagonal matrix);} \tag{3.2}$$

$$X'_{11}X_{12} + X'_{21}X_{22} = 0 \quad \text{(a } p \times m \text{ null matrix);} \tag{3.3}$$

$$X'_{12}X_{11} + X'_{22}X_{21} = 0 \quad \text{(an } m \times p \text{ null matrix).} \tag{3.4}$$

Likewise, in the 2^n case and for the transformed design matrix we note the following relations:

$$\begin{aligned} &\begin{bmatrix} W'_{11} & W'_{21} \\ W'_{12} & W'_{22} \end{bmatrix} \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix} \\ &\quad = \begin{bmatrix} W'_{11} W_{11} + W'_{21} W_{21} & W'_{11} W_{12} + W'_{21} W_{22} \\ W'_{12} W_{11} + W'_{22} W_{21} & W'_{12} W_{12} + W'_{22} W_{22} \end{bmatrix} = \begin{bmatrix} I_p & 0 \\ 0 & I_m \end{bmatrix}; \end{aligned} \tag{3.5}$$

$$\begin{aligned} &\begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix} \begin{bmatrix} W'_{11} & W'_{21} \\ W'_{12} & W'_{22} \end{bmatrix} \\ &\quad = \begin{bmatrix} W_{11} W'_{11} + W_{12} W'_{12} & W_{11} W'_{21} + W_{12} W'_{22} \\ W_{21} W'_{11} + W_{22} W'_{12} & W_{21} W'_{21} + W_{22} W'_{22} \end{bmatrix} = \begin{bmatrix} I_{p+m_1} & 0 \\ 0 & I_{m_2} \end{bmatrix}. \end{aligned} \tag{3.6}$$

I_p, I_m, I_{p+m_1}, and I_{m_2} are identity matrices with dimensions $p \times p$, $m \times m$, $(p + m_1) \times (p + m_1)$, and $m_2 \times m_2$, respectively.

Let the observational equations in the generalized situation be denoted by:

$$Y_{p+m_1} = X_{11}\mathbf{B}_p + e_{p+m_1}, \tag{3.7}$$

where the dimensions of X_{11}, $\mathbf{B}_p$, Y_{p+m_1} and e_{p+m_1} are $(p + m_1) \times p$, $p \times 1$, $(p + m_1) \times 1$ and $(p + m_1) \times 1$ respectively with $E(e_{p+m_1}e'_{p+m_1}) = \sigma^2 I_{p+m_1}$. The dimensions of the observational vector Y_{m_2} will be taken as $m_2 \times 1$. Using this notation the following two theorems and the derived variance formulas are presented to generalize previous results [1], [2].

THEOREM 1. *If X'_{22} is of rank m_2, then the least squares estimates $Y^+_{m_2}$ obtained from the observational equations $X'_{22}Y^+_{m_2} = -X'_{12}Y_{p+m_1}$ (the error part not indicated) and expressed as $\Lambda' Y_{p+m_1}$, are such that $X_{21} = \Lambda' X_{11}$.*

PROOF. The least squares estimates $Y^+_{m_2}$ are given by

$$Y^+_{m_2} = -[X_{22}X'_{22}]^{-1}X_{22}X'_{12}Y_{p+m_1}.$$

Thus, $\Lambda' = -[X_{22}X'_{22}]^{-1}X_{22}X'_{12}$, and

$$\Lambda' X_{11} = -[X_{22}X'_{22}]^{-1}X_{22}X'_{12}X_{11} = [X_{22}X'_{22}]^{-1}[X_{22}X'_{22}]X_{21} = X_{21}$$

by (3.4).

THEOREM 2. *If X_{11}, Y_{p+m_1} and e_{p+m_1} are augmented respectively by $\Lambda' X_{11} = X_{21}$, $\Lambda' Y_{p+m_1}$ and $\Lambda' e_{p+m_1}$ to become X_1, Y_1 and e_1, then the least squares estimates, $\mathbf{B}_p{}^*$, obtained from the observational equations $Y_1 = X_1\mathbf{B}_p + e_1$, are algebraically the same as the least squares estimates, $\mathbf{B}_p{}^+$, obtained from the observational equations $Y_{p+m_1} = X_{11}\mathbf{B}_p + e_{p+m_1}$.*

PROOF. The least squares estimates $\mathbf{B}_p{}^*$ obtained from the observational equations $Y_1 = X_1\mathbf{B}_p + e_1$ are given by

$$\begin{aligned}
\mathbf{B}_p{}^* &= [X_1{}'X_1]^{-1}X_1{}'Y_1 \\
&= [X'_{11}X_{11} + X'_{21}X_{21}]^{-1}[X'_{11} \vdots X'_{11}\Lambda] \begin{bmatrix} Y_{p+m_1} \\ \Lambda' Y_{p+m_1} \end{bmatrix} \qquad (3.8) \\
&= [X'_{11}X_{11} + X'_{21}X_{21}]^{-1}[X'_{11} + X'_{11}\Lambda\Lambda']Y_{p+m1}.
\end{aligned}$$

Now transform the design matrix and the parameters such that $X_{11}\mathbf{B}_p{}^+ = W_{11}\mathbf{C}_p{}^+$ and $X_1\mathbf{B}_p{}^* = W_1\mathbf{C}_p{}^*$ as described above. The least squares estimates become:

$$\mathbf{C}_p{}^* = [W_1{}'W_1]^{-1}W_1{}'Y_1 = [W'_{11}W_{11} + W'_{21}W_{21}]^{-1}[W'_{11} + W'_{11}\Delta\Delta']Y_{p+m_1}, \tag{3.9}$$

where $\Delta' = -[W_{22}W'_{22}]^{-1}W_{22}W'_{12}$ in the same manner as for Λ' when using the X matrix. Using Equations (3.5) and (3.6), Equation (3.9) may be manipulated to obtain the least squares equation for $\mathbf{C}_p{}^*$ as follows:

$$
\begin{aligned}
&[W_{11}'W_{11} + W_{21}'W_{21}]^{-1}[W_{11}' - W_{21}'(W_{22}W_{22}')^{-1}W_{22}W_{12}']Y_{p+m_1} \\
&\quad = [W_{11}'W_{11} + W_{21}'W_{21}]^{-1}[I_p + W_{21}'(W_{22}W_{22}')^{-1}W_{21}] \\
&\quad\quad \cdot (W_{11}'W_{11})(W_{11}'W_{11})^{-1}W_{11}'Y_{p+m_1} \\
&\quad = [W_{11}'W_{11} + W_{21}'W_{21}]^{-1}[W_{11}'W_{11} + W_{21}'W_{21}](W_{11}'W_{11})^{-1}W_{11}'Y_{p+m_1} \\
&\quad = (W_{11}'W_{11})^{-1}W_{11}'Y_{p+m_1} = \mathbf{C}_p^{+}.
\end{aligned} \tag{3.10}
$$

Premultiplying $\mathbf{C}_p^{+}$ and $\mathbf{C}_p^{*}$ by $D^{-\frac{1}{2}}$, we find that $\mathbf{B}_p^{+} = \mathbf{B}_p^{*}$.

4. Variance of the estimated effects under the above adjustment. We have $e_1 = [e_{p+m_1}' \vdots e_{p+m_1}'\Lambda]'$ corresponding to $Y_1 = [Y_{p+m_1}' \vdots Y_{p+m_1}'\Lambda]'$. Thus

$$E(e_1 e_1') = E\begin{bmatrix} ee' & ee'\Lambda \\ \Lambda'ee' & \Lambda'ee'\Lambda \end{bmatrix} = \begin{bmatrix} I_{p+m_1} & \Lambda \\ \Lambda' & \Lambda'\Lambda \end{bmatrix}\sigma^2.$$

The covariance matrix of $\mathbf{B}_p^{*}$ obtained from $Y_1 = X_1\mathbf{B}_p + e_1$ is given by

$$\operatorname{cov}(\mathbf{B}_p^{*}) = S_1^{-1} X_1' \begin{bmatrix} I_{p+m_1} & \Lambda \\ \Lambda' & \Lambda'\Lambda \end{bmatrix} X_1 S_1^{-1} \sigma^2, \tag{4.1}$$

where $S_1 = [X_1'X_1] = D_p$. It can further be shown [2] that

$$X_1' \begin{bmatrix} I_{p+m_1} & \Lambda \\ \Lambda' & \Lambda'\Lambda \end{bmatrix} X_1 = [X_{11}' \vdots X_{21}'] \begin{bmatrix} I_{p+m_1} & \Lambda \\ \Lambda' & \Lambda'\Lambda \end{bmatrix} \begin{bmatrix} X_{11} \\ \cdots \\ X_{21} \end{bmatrix} \tag{4.2}$$

$$= X_{11}'(1 + \Lambda\Lambda')^2 X_{11} = X_{11}'(1 + 2\Lambda\Lambda' + \Lambda\Lambda'\Lambda\Lambda')X_{11}.$$

After some matrix manipulations and simplifications using the relations in (3.1) to (3.6) we find that the covariance matrix reduces to $[X_{11}'X_{11}]^{-1}\sigma^2$ which is the same as cov $(\mathbf{B}_p^{+})$, where $\mathbf{B}_p^{+}$ is obtained from observational Equations (3.7). Thus, for finding the covariance matrix, expression (4.1) may be preferred, as it would involve inversion of a matrix of smaller dimensions in that the inverse of $[X_{22}X_{22}']$, or $[W_{22}W_{22}']$, needs to be obtained.

5. Estimability of the effect parameters that are retained and their connection with the observations that are omitted. The methodology given above consists in getting the least squares estimates of the omitted observations from equations obtained by equating the "negligible" effects to zero. The question of estimability of the omitted observations is thus connected with the question of what effect parameters are being ignored and consequently with the question of what effects are being retained. In other words, given that a certain set of effects is neglected (or equivalently, given that a given set of effects may be retained), it would not be possible to cut out the observations arbitrarily to obtain a fractional replicate. We now prove a theorem showing that the question of being able to estimate (in the least squares sense) the retained effect parameters is inseparably connected with the question of being able to estimate (in the least squares sense) the omitted observations. The retained effect parameters are p

in number, which have to be estimated from $(p + m_1)$ observations. We recall that the dimensions of X_{11} and X_{22} are $(p + m_1) \times p$ and $m_2 \times m$ $(p + m = N, m_1 + m_2 = m, p > N/2)$ respectively. In order that it be possible to have least squares estimates for each of the p effect parameters and also for each of the m_2 omitted observations, $X'_{11}X_{11}$ and $X_{22}X'_{22}$ should be of rank p and m_2 respectively.

THEOREM 3. *Each of the set of p effect parameters (as denoted by* $\mathbf{B}_p$*) is estimable (in the least squares sense), if and only if each of the set of* m_2 *omitted observations is estimable.*

PROOF. As in Theorem 2, we shall transform the X matrix to the W matrix. This does not alter the rank of the matrix. Since Rank (W) is $p + m_1 + m_2 = N$, Rank $(W_{11} \vdots W_{12})$ is $p + m_1$ and Rank $(W_{21} \vdots W_{22})$ is m_2. Now suppose that Rank (W_{11}) is p. This implies that $(W'_{11}W_{11})^{-1}$ exists. Then from Equation (3.6) we write $W_{21}W'_{11} = -W_{22}W'_{12}$. Post-multiplying both sides by W_{11} we obtain $W_{21}W'_{11}W_{11} = -W_{22}W'_{12}W_{11}$. Then $W_{21} = W_{22}F$ where $F = -W'_{12}W_{11}(W'_{11}W_{11})^{-1}$. Since $(W_{21} \vdots W_{22}) = W_{22}(F \vdots I_m)$ and since Rank $(W_{21} \vdots W_{22})$ is m_2, Rank (W_{22}) is m_2. Hence $(W_{22}W'_{22})^{-1}$ exists.

Now suppose that W_{22} has rank m_2. From Equation (3.5) we have $W'_{21}W_{22} = -W'_{11}W_{12}$. Post-multiplying both sides by W'_{22} we obtain $W'_{21}W_{22}W'_{22} = -W'_{11}W_{12}W'_{22}$. Hence $W'_{21} = W'_{11}G$, where $G = -W_{12}W'_{22}(W_{22}W'_{22})^{-1}$. Since Rank $(W'_{11} \vdots W'_{21})$ is p and since $(W'_{11} \vdots W'_{21}) = (W'_{11} \vdots W'_{11}G)$, Rank (W_{11}) is p. Thus, if each of the m_2 omitted observations is estimable then each of the p effect parameters is estimable, and vice versa.

6. A working rule for omitting observations against a given set of negligible effects. From what has been indicated in Section 5 it will be clear that it is permissible to omit only those observations for which it is possible to have least squares estimates from the omitted effect parameters, as this permissibility implies the estimability of the effect parameters that are retained. If it is possible to have least squares estimates for each in the set of omitted observations, we shall call this set *consistent*, otherwise, *inconsistent*. We give below a working rule for omitting observations against a given set of omitted effects:

1. Write the full set of observational equations as

$$X\mathbf{B} = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix} \begin{bmatrix} \mathbf{B}_p \\ \mathbf{B}_m \end{bmatrix} = \begin{bmatrix} Y_{p+m_1} \\ Y_{m_2} \end{bmatrix}$$

where Y_{m_2} are the $m_2 \leqq m$ observations to be omitted and $\mathbf{B}_m$ are the m parameters not to be estimated.

2. Specify two observations to be omitted; if the rows in X_{22} for the two observations are not multiples of each other, the two observations are consistent, and, hence, may be omitted.

3. If the row in X_{22} for the third observation to be omitted is not a multiple of the rows of X_{22} for the previous two omitted observations, it is consistent and, therefore, can be omitted.

4. Continue this process until m_2 observations are omitted. If the rows of

X_{22} are independent and if there are at least m_2 independent columns, then $X_{22}X'_{22}$ has an inverse and the omitted set of m_2 observations is consistent.

7. The most efficient way of omitting observations given a choice. Efficiency of a design may be judged by more than one criterion. The design given by X_{11} may be considered to have maximum efficiency if the value of the det $|X'_{11}X_{11}|$ is maximized, or that of $|(X'_{11}X_{11})^{-1}|$ is minimized. Since the dimensions of X_{11} may be different in different designs, for the purpose of comparing two different designs, the criterion of average trace of $[X_{11}\vdots X_{11}]^{-1}$ may also be adopted a sa measure of efficiency, a design being considered to have maximum efficiency when the average trace is the minimum. There are situations, however, when any one of these two criteria would imply the other.

We consider below the question, given a choice, of omitting observations suitably so as to maximize the efficiency of the design for the 2^n series. (The same arguments may be used for the general case by transforming the X matrix to the W matrix.) We have

$$\begin{aligned} \text{Tr }(X'_{11}X_{11})^{-1} &= (1/N^2)\text{ Tr }[X'_{11}(1+\Lambda\Lambda')^2X_{11}] \\ &= (1/N^2)[\text{Tr }(X'_{11}X_{11}) + 2\text{ Tr }(X_{11}X'_{11}\Lambda\Lambda') \\ &\quad + \text{Tr }(X_{11}X'_{11}\Lambda\Lambda'\Lambda\Lambda')]. \end{aligned} \tag{7.1}$$

The second term of (7.1) may be shown to be equal to 2 Tr $(NI_{m_2} - X_{22}X'_{22})$ and the third term equal to Tr $[N(X_{22}X'_{22})^{-1} - 2NI_{m_2} + (X_{22}X'_{22})]$ by relations (3.1) to (3.6) for the 2^n case. On substitution of these expressions in (7.1) we get Tr $(X'_{11}X_{11})^{-1}$ reduced to

$$\text{Tr }(X'_{11}X_{11})^{-1} = (1/N^2)[\text{Tr }(X'_{11}X_{11}) + N^2\text{ Tr }(X_{22}X'_{22})^{-1} - \text{Tr }(X_{22}X'_{22})]. \tag{7.2}$$

Now (7.2) will be minimized if Tr $(X_{22}X'_{22})$ is maximized and Tr $(X_{22}X'_{22})^{-1}$ is minimized.

It is known that if the elements of X (square) are restricted to lie between ± 1, then the maximum possible value of the det $|X|$ is given when X is orthogonal in the sense that $X'X$ is diagonal. In such a situation, det $|X'X|$ is maximum, det $|(X'X)^{-1}|$ is minimum, Tr $(X'X)$ is maximum and Tr $(X'X)^{-1}$ is minimum. [These aspects are connected in a way with the problem of Hadamard matrices and Hotelling's weighing problem.] Hence, it would be clear that if the rows of X_{22} are chosen to be orthogonal to each other Tr $(X_{22}X'_{22})^{-1}$ is minimized and Tr $(X_{22}X'_{22})$ is maximized simultaneously. In other words, if it is possible to estimate the omitted observations with maximum efficiency, the effect parameters are also estimated with maximum efficiency. When, however, it is not possible to have the rows of X_{22} orthogonal, we have to cut out observations, if there is a choice, in such a way as to minimize the trace of $(X_{22}X'_{22})^{-1}$.

Acknowledgment. Appreciation is expressed for the helpful comments of K. R. Shah and B. L. Raktoe on earlier versions of the manuscript.

REFERENCES

[1] Banerjee, K. S. and Federer, W. T. (1963). On estimates for fractions of a complete factorial experiment as orthogonal linear combinations of the observations. *Ann. Math. Statist.* **34** 1068–1078.

[2] Banerjee, K. S. and Federer, W. T. (1964). Estimates of effects for fractional replicates. *Ann. Math. Statist.* **35** 711–715.

[3] Tocher, K. D. (1952). The design and analysis of block experiments. *J. Roy. Stat. Soc. Ser. B* **14** 45–100.