FILE

DEPARTMENT OF STATISTICS
University of Wisconsin
Madison, Wisconsin

Technical Report No. 24
January, 1964

A Note on Partial Duplication of Designs

by
George E. P. Box

This research was supported by the United States Navy through the Office of Naval Research, under Contract Nonr - 1202(17), Project NR 042 222.

A NOTE ON PARTIAL DUPLICATION OF DESIGNS

George E. P. Box
University of Wisconsin

The two-level factorials and fractionals are often carried out in circumstances where replication of the whole design to provide an estimate of real experimental error is not economically justifiable. In these circumstances it is useful to duplicate only a component sub-fraction of the original design. Typical plans selected by Daniel, 1957 and Dykstra, 1959 are that the original design leaves main effects and two-factor interactions unconfounded one with another while the component sub-fraction confounds main effects with two-factor interactions but not with other main effects. It has recently been shown (Box, 1959), that partial duplication of this kind allows an unbiased estimate of the variances of the effects to be obtained even when errors in setting the factor levels are allowed for.

In an article in this journal referred to above, Dykstra showed how to estimate the effects and their variances and covariances for the augmented arrangements resulting from particular schemes of partial duplication. The object of the present note is to supply a simple and general procedure.

Suppose that there are n_1 observations in the original design from which, on the usual assumptions, unbiased estimates T_1, T_2, etc. of individual "effects" θ_1, θ_2, etc. may be obtained. Suppose that n_2 further observations are now added at conditions which duplicate those of the original design and provide on their own a smaller fractional factorial design. Suppose finally that from this second set of n_2 observations alone an unbiased estimate L_2 of the certain linear function of effects $L(\theta_i)=\alpha_1\theta_1 + \ldots +\alpha_i\theta_i +\ldots +\alpha_p\theta_p$ may be obtained. In the present application the α's are + 1's and -1's. And $\theta_1, \theta_2, \ldots, \theta_p$ are those effects which are aliases one of another. We may ask the question: How should our estimate T_i of θ_i obtained for the original * design be modified in the light of the extra information supplied by the later experiments?

Now information about θ_i from the second set of runs enters only through the estimate L_2 of $L(\theta_i)$, and an alternative and independent estimate of $L(\theta_i)$ is available from the <u>original</u> set of n_1 experiments. This is

$$L_1 = \alpha_1 T_1 + \ldots + \alpha_i T_i + \ldots + \alpha_p T_p.$$

As might be expected, the modifying factor depends on the difference between the two estimates L_2 and L_1. In fact the least squares estimate $\hat{\theta}_i$ of θ_i for the complete set of $n_1 + n_2$ observations is simply

$$\hat{\theta}_i = T_i + \frac{\alpha_i n_2}{n_1 + n_2 p} (L_2 - L_1) \tag{1}$$

With the usual definition of main effects and interactions (Yates, 1937) and with σ^2 the experimental error variance, the variance of the estimate is

$$V(\hat{\theta}_i) = \frac{4\sigma^2}{n_1} \left[\frac{n_1 + n_2(p-1)}{n_1 + n_2 p}\right] \tag{2}$$

where σ^2 is the variance of a single observation.

The factor in the square brackets indicates the amount by which the variance is reduced by the additional runs. The covariance between estimates of effects θ_i and θ_j both of which are in the linear combination $L(\theta_i)$ is

$$\mathrm{Cov}(\hat{\theta}_i \hat{\theta}_j) = \pm \frac{\sigma^2}{n_1} \left[\frac{n_2}{n_1 + n_2 p}\right] \tag{3}$$

The coefficient of correlation between the estimates $\hat{\theta}_i$ and $\hat{\theta}_j$ is

$$\rho_{ij} = \pm \frac{n_2}{4(n_1 + n_2(p-1))} \qquad (4)$$

The sign to be attached to this covariance and the coefficient of correlation is negative if α_i and α_j are of like sign and positive otherwise.

Example:

For illustration, let us employ a situation considered by Dykstra. Suppose that the $n_1 = 64$ runs required by a suitable 2^{8-2} fractional factorial in 8 factors, A to H, has been completed and, on the usual assumptions, unbiased estimates of all main effects and two-factor interactions have been calculated. Suppose now that a suitably selected set of $n_2 = 16$ runs is duplicated and from these runs alone estimates of various linear combinations of main effects and two-factor interactions obtained. Suppose in particular that from these later runs we have an estimate L_2 of A - CD - GH. If {A}, {CD}, and {GH} are estimates obtained from the original 2^{8-2} design then an independent estimate of this quantity is $L_1 = \{A\} - \{CD\} - \{GH\}$. Using equation (1) the modifying factor is $\pm \frac{1}{7}(L_2 - L_1)$.

The least squares estimates for A, CD, and GH from the complete set of $64 + 16 = 80$ observations are $\widehat{A} = \{A\} + \frac{1}{7}(L_2 - L_1)$, $\widehat{CD} = \{CD\} - \frac{1}{7}(L_2 - L_1)$ $\widehat{GH} = \{GH\} - \frac{1}{7}(L_2 - L_1)$.

Using equation (2) the variance of each of these estimates is $\left\{\frac{1}{16}\sigma^2\right\}\frac{6}{7}$ as compared with $\frac{1}{16}\sigma^2$ for the estimates from the original design. Using equation (4) the coefficient of correlation between any two estimates is the group $\pm \frac{1}{6}$. The sign is positive for the associations (A, CD) and (A, GH) and negative for the association (CD, GH).

APPENDIX

Suppose that for p parameters $\theta_1, \theta_2, \ldots, \theta_p$ we have unbiased uncorrelated estimates $T_1, T_2, \ldots, T_p$ respectively, each with variance $1/w_1$, and that for the linear combination $L(\underline{\theta}) = \sum_{i=1}^{p} \alpha_i \theta_i$ we have an unbiased estimate L_2 uncorrelated with the previous estimate and with variance $1/w_2$. Let $\sum_{i=1}^{p} \alpha_i^2 = q$, $\frac{w_1}{w_2} = w$ and $\sum_{i=1}^{p} \alpha_i T_i = L_1$. Then the least squares estimate $\hat{\theta}_i$ of θ_i is given by

$$\hat{\theta}_i = T_i + \frac{\alpha_i}{w+q}\left\{L_2 - L_1\right\} \qquad (5)$$

and

$$V(\hat{\theta}_i) = \frac{1}{w_1}\left\{\frac{w+q-\alpha_i^2}{w+q}\right\} \qquad (6)$$

$$\mathrm{Cov}(\hat{\theta}_i \hat{\theta}_j) = \frac{-1}{w_1}\left\{\frac{\alpha_i \alpha_j}{w+q}\right\} \qquad (7)$$

whence the coefficient of correlation between $\hat{\theta}_i$ and $\hat{\theta}_j$ is

$$-\alpha_i \alpha_j \Big/ \sqrt{\left\{w+q-\alpha_i^2\right\}} \sqrt{(w+q-\alpha_j^2)} \qquad (8)$$

The results directly applicable to fractional factorials are obtained by substituting $\frac{1}{w_1} = \frac{4\sigma^2}{n_1}$, $\frac{1}{w_2} = \frac{4\sigma^2}{n_2}$, $w = \frac{n_1}{n_2}$ and $q = \sum_{i=1}^{p} \alpha_i^2 = p$.

Proof:

The least squares estimates of the θ's using all the data are the values $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_p$ which render a minimum the quadratic form

$$S = w_1 \sum_{i=1}^{p} \left\{T_i - \theta_i\right\}^2 + w_2 \left\{L_2 - L(\underline{\theta})\right\}^2$$

The partial derivative $S_i = \left(\frac{\partial s}{\partial \theta_i}\right)$ is such that

$$-\tfrac{1}{2} S_i = w_1 \left\{T_i - \theta_i\right\} + w_2 \alpha_i \left\{L_2 - L(\underline{\theta})\right\} \qquad (9)$$

and on equating S_i to zero we obtain

$$\hat{\theta}_i = T_i + \frac{\alpha_i}{w}\left\{L_2 - L(\hat{\underline{\theta}})\right\} \qquad (10)$$

where $w = w_1/w_2$.

Now, from (9)

$$-\tfrac{1}{2} L(\underline{S}) = w_1\left\{L_1 - L(\underline{\theta})\right\} + w_2 q\left\{L_2 - L(\underline{\theta})\right\} \qquad (11)$$

where $q = \sum_{i=1}^{p} \alpha_i^2$ and $L_1 = L(\underline{T}) = \sum_{i=1}^{p} \alpha_i T_i$.

On equating $L(\underline{S})$ to zero we obtain

$$L(\hat{\underline{\theta}}) = \left\{w L_1 + q L_2\right\} / \left\{w + q\right\} . \qquad (12)$$

By substituting (12) in (10) we obtain the least squares estimate given in equation (5).

Since $\hat{\theta}_i$ is a linear combination of the uncorrelated estimates $T_1, T_2, \ldots, T_p$ and L_2 its variance and its covariance with any other estimate $\hat{\theta}_j$ is easily obtained directly. Alternatively as is sometimes convenient in linear least squares problems we may use the fact that the variance-covariance matrix of the least squares estimates is given by $\left\{\tfrac{1}{2} S_{ij}\right\}^{-1}$ where $\left\{S_{ij}\right\}$ is a p x p symmetric matrix with $S_{ij} = \partial^2 S/\partial\theta_i\partial\theta_j$ the element in the i^{th} row and j^{th} column. Differentiating both sides of equation (9) with respect to θ_i and then with respect to θ_j we obtain

$$\tfrac{1}{2} S_{ii} = w_1 + w_2\alpha_i^2, \quad \tfrac{1}{2} S_{ij} = w_2\alpha_i\alpha_j$$

Let $\underline{I}$ be a p x p unit matrix and $\underline{\alpha}$ be a p x 1 vector with elements $\alpha_1, \alpha_2, \ldots, \alpha_p$ so that $\alpha'\alpha = q$. Then

$$\left\{\tfrac{1}{2} S_{ij}\right\} = w_1 \underline{I} + w_2 \underline{\alpha\alpha'}$$

A matrix form with the well known inverse

$$\left\{\frac{1}{2} S_{ij}\right\}^{-1} = \frac{1}{w_1} \left\{ \underline{I} - \frac{1}{w + q} \underline{\alpha}\, \underline{\alpha}' \right\}$$

Equations (6) and (7) for the variances and covariances of the least squares estimates follow at once.

General Note

Industrial experience, see for example Davies, 1954, has repeatedly shown the value of the sequential use of two-level fractionals used as building blocks. An infinite variety of situations can occur and it would seem pointless to try to discuss each case separately. Although minimum variance estimates can always be obtained by a proper application of least squares theory, the procedure for finding the estimates is not unique and it is easy to introduce unnecessary complexities. Greatest simplicity is usually obtained by making maximum use of the properties of orthogonality and balance enjoyed by the individual fractionals. Uncorrelated estimates of effects and of linear combinations of effects may be very simply obtained from these individual designs. Those estimates for the various building blocks where expected values contain common effects may then be combined by simple application of weighted least squares theory with those individual estimates playing the part of observations.

In some cases the estimates to be combined have known correlation. It is perhaps worth reminding the reader, therefore, of the general results due to Aitken, 1935, which allow us to apply least squares not only to unequally weighted quantities but also to correlated quantities. Suppose we have unbiased estimates $T_1, T_2, \ldots, T_s, \ldots, T_u, \ldots, T_n$ of

$L_1(\theta)$, $L_2(\theta)$, ..., $L_s(\theta)$, ..., $L_u(\theta)$, ..., $L_n(\theta)$ then the L's are linear functions of $\theta_1, \theta_2, \ldots, \theta_i, \ldots, \theta_p$. Suppose also that the n estimate have variance-covariance matrix $\{v_{su}\} = \{c_{su}\}\sigma^2$ where the c_{su} are known but the constant σ^2 in general is not known. Then the unbiased linear estimates of the θ's with smallest possible variance are those which minimize the quadratic form

$$S = \sum_{s=1}^{n} \sum_{u=1}^{n} w_{su}(T_s - L_s(\theta))\,(T_u - L_u(\theta))$$

where the weights w_{su} to be applied to the squares and cross product terms are simply the elements of the inverse matrix $\{v_{su}\}^{-1} = \{w_{su}\}$.

This quantity S can be manipulated in the same way as before and in particular

$$\left\{\tfrac{1}{2} S_{ij}\right\}^{-1} \quad \text{where} \quad S_{ij} = \partial^2 S/\partial\theta_i \partial\theta_j$$

provides the variance-covariance matrix of the estimates.

NOTES

* It is convenient to talk of the "original" design of n_1 experiments and the "later" set of n_2 experiments. In fact the original design would often be carried out in blocks with the component sub-fraction a duplicated block, randomization being carried out within blocks. If there were no blocking the complete set of $n_1 + n_2$ terms would be randomized. So far as the final estimates are concerned it is immaterial which one of the duplicated blocks or, for the fully randomized design, which set of n_2 runs is regarded as falling into the "original" design and which into the component sub-fraction.

REFERENCES

1. Aitken, A. C. (1935), On least squares and linear combination of observations, Proc. Roy. Soc. Edin., Vol. 55, pp. 42-48.

2. Box, G. E. P. (1959), The effects of errors in the factor levels for some simple experimental designs. Paper presented to the 1959 Christmas meetings of the American Statistical Association.

3. Daniel, C. (1957), Fractional replication in industrial experimentation. Transactions of the Eleventh Annual Convention of the American Society for Quality Control.

4. Davies, O. L. (Editor, 1954), Design and Analysis of Industrial Experiments, Oliver and Boyd, London.

5. Dykstra, O. (1959), Partial duplication of fractional experiments, Technometrics, I, pp. 63-75.

6. Yates, F. (1937), Imperial Bureau of Soil Science Technical Communication, No. 35.