WEIGHING DESIGNS AND BALANCED INCOMPLETE BLOCKS

By K. S. Banerjee
Pusa, Bihar, India

1. Introduction. Following a paper by Hotelling [1] on the weighing problem, Kishen [4] and Mood [2] furnished generalized solutions. This note consists of some additional remarks on the weighing problem when the weighing is restricted to be made on one pan.

Hotelling remarked that when the problem was to determine a particular difference or any other linear function of the weights, a different design should be sought to minimize the variance. An account of efficient designs of this kind has also been furnished in this note. The notations used by Hotelling and Mood have been used here.

2. Chemical balance problem. It has been shown by Mood that when $N \equiv 0 \pmod 4$, an optimum design exists if a Hadamard matrix H_N exists, and is obtained by using any p columns of H_N. When $N \equiv i \pmod 4$, $(i = 1, 2, 3)$, very efficient designs are obtained either by adding to or deleting from the rows of H_{4K}, making the resultant number of rows equal to N.

It has further been shown by Mood in connection with this class of designs that arrangements[1] are available which are more efficient than the one obtained by repeating the row of ones. As a matter of fact, if any row other than the row of ones be repeated, this will lead to a design of the same efficiency as in the case of repeated addition of the row of ones; for, the determinant of $X'X$ will remain exactly identical. That this is so, will be clear from the following properties showing the connection of the matrix X with the determinant $|a_{ij}|$:

(i) Any two rows of the matrix X can be interchanged without changing the determinant $|a_{ij}|$.

(ii) Any two columns of the matrix X can be interchanged without changing the determinant $|a_{ij}|$.

(iii) The signs of all the elements in a column of the matrix X may be changed without changing the determinant $|a_{ij}|$.

3. Spring balance problem. Mood has exhaustively discussed the designs when $N > p$. Efficient designs under this class will, however, be available from the arrangements afforded by balanced incomplete block designs discussed in [3]. These designs will be represented by certain of the efficient submatrices of the P_k of Mood.

Usually v and b are used to denote respectively the number of varieties and the number of blocks in the above mentioned designs. Here v will take the place of

[1] This had been independently shown by me before the paper of A. M. Mood was brought to my notice by H. Hotelling.

p, the number of objects to be weighed and b that of N, the number of weighings that can be made. The matrix $X'X$ in this case will take the form

$$\begin{bmatrix} r & \lambda & \lambda & \cdots & \lambda \\ \lambda & r & \lambda & \cdots & \lambda \\ \lambda & \lambda & r & \cdots & \lambda \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \lambda & \lambda & \lambda & \cdots & r \end{bmatrix} \tag{1}$$

The variance of the estimated weight of each of the p objects for such a design can be easily seen to be

$$\frac{r + \lambda(p - 2)}{(r - \lambda)\{r + \lambda(p - 1)\}} \sigma^2 \qquad \text{for zero bias,} \tag{2}$$

where p is the number of objects to be weighed and r and λ have meanings similar to those in connection with balanced incomplete block designs; that is, r is the number of times each object is weighed, and λ is the number of times each pair of objects is weighed together.

Though the *minimum minimorum* of σ^2/N can never be attained by the objects to be weighed under such designs, σ^2/N may however be kept as the standard with which the efficiency of a given design may be calculated. The efficiency of the above design will therefore for zero bias be

$$\frac{(r - \lambda)\{r + \lambda(p - 1)\}}{N\{r + \lambda(p - 2)\}}. \tag{3}$$

The identities well known in the theory of balanced incomplete blocks,

$$bk = vr, \qquad \lambda(v - 1) = r(k - 1),$$

may, upon replacing b by N and v by p to accord with the notation of weighing designs, be written

$$r = Nk/p, \qquad \lambda = r(k - 1)/(p - 1).$$

Upon substituting these in (3) we obtain the efficiency factor in the form

$$\frac{k^2(p - k)}{p(pk - 2k + 1)}, \tag{4}$$

where k is the number of plots per block or the number of objects that can be weighed at a time.

If instead of adopting repetitions of P_K, only $\binom{p}{K}$ weighings be made in all, the efficiency factor calculated for such a combinatorial design would be

$$\frac{(r - \lambda)\{r + \lambda(v - 1)\}}{b\,\{r + \lambda(v - 2)\}}, \qquad \text{for zero bias.}$$

where

$$r = \binom{v-1}{K-1}, \qquad \lambda = \binom{v-2}{K-2}$$

and $b = \binom{v}{K}$. The above expression on simplification reduces to (4).

It will be noticed that the efficiency of such designs depends only upon the total number of objects to be weighed and the number of such objects that can be weighed at a time.

These designs have the advantage that all the weights are estimated with equal precision. If a slightly larger number of weighing than what is afforded by the number of blocks in a balanced incomplete block design has to be made, all the objects may be weighed together and this weighing be repeated as many times as required. This will be equivalent to the repeated addition of the row of ones. The repetition of the row of ones in particular is necessary to make the weights estimable with equal precision, which however, may be demanded at times as a matter of necessity in certain experiments. Otherwise, any other single row or different rows of the matrix X may be repeated, making the number of rows of the matrix X equal to the number of weighings proposed to be made in all.

From the practical point of view also, it will be advantageous to connect the designs for weighing with the already existing balanced incomplete block designs, which have been highly developed in recent years and are being extensively used in agro-biological investigations.

4. Spring balance design for small p. Under this class of designs, Mood has found the most efficient design for $p = 7$. It is given by

$$L_7 = \begin{bmatrix} 1&0&1&0&1&0&1 \\ 0&1&1&0&0&1&1 \\ 0&0&0&1&1&1&1 \\ 1&1&0&0&1&1&0 \\ 0&1&1&1&1&0&0 \\ 1&0&1&1&0&1&0 \\ 1&1&0&1&0&0&1 \end{bmatrix}.$$

This L_7 is easily recognized to be the design for $k = 4, b = 7, v = 7, r = 4, \lambda = 2$, given by an orthogonal series [3]. It is therefore seen that Hadamard matrices will lead to a new method of constructing balanced incomplete block designs of a certain class. For example H_{16} and H_{20} will lead respectively to the designs for $k = 8, b = 15, v = 15, r = 8, \lambda = 4$ (or for $k = 7, b = 15, v = 15, r = 7, \lambda = 3$) and for $k = 10, b = 19, v = 19, r = 10, \lambda = 5$ (or $k = 9, b = 19, v = 19, r = 9, \lambda = 4$). These designs also satisfy the condition of maximum efficiency, by virtue of the fact that $| L_N |$ will have the value

$$(N + 1)^{\frac{1}{2}(N+1)}/2^N,$$

as shown by Mood.

6. Determination of a linear function of the objects. An orthogonalized design which is cent percent efficient to determine individually the weight of p unknown objects is not necessarily the design of maximum efficiency for the estimation of a linear function of the objects. To illustrate this, let there be three objects, the weights O_1, O_2, O_3, of which have to be estimated on a balance corrected for zero bias and let us, for this purpose, concentrate on the design characterized by the matrix given below.

$$X = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{pmatrix}. \tag{5}$$

As has been indicated in the previous papers, the variance of each of the unknown objects comes out to be $\frac{1}{4}\sigma^2$, which is the *minimum minimorum* and as such the above design enjoys the cent percent efficiency, when the question of individual estimation is concerned. But in estimating a linear function of the objects, for instance the total weight, designs more efficient than this are available.

The variance of $l_1O_1 + l_2O_2 + l_3O_3$ is known to be

$$\sum_{i,j=1}^{3} l_i l_j C_{ij} \sigma^2 \tag{6}$$

where C_{ij} denotes the elements of the matrix reciprocal to the matrix $X'X$. As the above design furnishes the estimates of the unknown objects orthogonally, the variance of the estimated total weight of the three objects will be given by $\frac{3}{4}\sigma^2$. If, however, the design given by the matrix

$$x = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \tag{7}$$

be adopted, the variance of the estimate of the total weight may be easily seen to be $(3/7)\sigma^2$, by putting $l_1 = l_2 = l_3 = 1$. $(3/7)\sigma^2$ is evidently less than $\frac{3}{4}\sigma^2$. Therefore with four weighings, the design characterized by (7) is more efficient in estimating the total weight than that characterized by (5). A still more efficient design for getting the total weight is simply to weigh all the objects together four times.

6. Designs with arrangements afforded by balanced incomplete blocks. The necessity for an efficient design to estimate any linear function of the objects (or to be precise, say to estimate the total weight) will perhaps arise only when the objects cannot all be weighed at a time collectively on a single pan. Here also, an efficient design under the supposition that all the objects cannot be weighed together is afforded by the arrangements in balanced incomplete blocks. In such a design, the diagonal elements in the matrix reciprocal to $X'X$ will be all positive and equal to

$$\frac{r + \lambda(p - 2)}{(r - \lambda)\{r + \lambda(p - 1)\}}, \tag{8}$$

while the remaining elements in the reciprocal matrix will be negative and equal to

$$\frac{-\lambda}{(r - \lambda)\{r + \lambda(p - 1)\}}. \tag{9}$$

Using the generalized form of (6) and admitting of the possibility that any of the arbitrary constants l_i may be negative, the variance of the linear function $\sum_{i=1}^{p} l_i O_i$ may be easily seen to be

$$\left\{\frac{\Sigma l_i^2}{r - \lambda} - \frac{\lambda(\Sigma l_i)^2}{(r-\lambda)\{r+(p-1)\lambda\}}\right\}\sigma^2. \tag{10}$$

If, however, in the above expression, the coefficients l_i are equal to 1, (10) is the variance of the estimated total weight, and reduces to

$$\frac{p}{r + (p - 1)\lambda}\sigma^2. \tag{11}$$

When there are N weighings in all, the minimum variance that can be reached is σ^2/N and will be attained, it appears, only when all the objects are weighed together and the weighing is repeated N times. The efficiency of a given design may therefore be calculated with reference to σ^2/N. Remembering that the number of weighings takes the place of the number of blocks and p the place of v, the efficiency of the design will reduce to $(k/p)^2$, where k is the number of plots per block i.e. the number of objects that can be weighed at a time.

If, however, the combinatorial arrangement is adopted weighing all possible combinations of k objects and making $\binom{p}{k}$ weighings in all, the same efficiency as above will be obtained for such a design.

Given k, the above expression of efficiency will therefore be the deciding factor for choice between an arrangement of balanced incomplete block design and all possible combinations of k objects.

7. Design of maximum efficiency. Designs leading to the matrix $X'X$ of the type (1) have certain advantages inasmuch as the variances of the individual objects are equal, as are also the covariances between all possible pairs. The variance of the estimated total weight in such a design is given by (11). To minimize the variance thus obtained, the expression

$$r + (p - 1)\lambda \tag{12}$$

has to be the maximum for a given value of p. In an arrangement of the balanced incomplete block type or in an arrangement with all possible combinations of k objects being weighed at a time, (12) would reduce to rk and would therefore increase with the increasing value of rk. This shows that the estimation of the total weight will have increased precision if more of the objects are weighed at a time.

If all the objects could be weighed at a time and both the pans be used for the purpose, some of the elements in the matrix X will be -1 instead of 0. This would increase the value of r but would decrease the value of λ. To devise the best possible design therefore, account will have to be taken simultaneously of r and λ.

REFERENCES

[1] HAROLD HOTELLING, "Some improvements in weighing and other experimental techniques", *Annals of Math. Stat.*, Vol. 15 (1944), pp. 297–306.

[2] A. M. MOOD, "On Hotelling's weighing problem", *Annals of Math. Stat.*, Vol. 17 (1946), pp. 432–446.

[3] R. A. FISHER AND F. YATES, *Statistical Tables for Biological, Agricultural and Medical Research*, Oliver and Boyd, London, 1938, pp. 10–13.

[4] K. KISHEN, "On the design of experiments for weighing and making other types of measurements", *Annals of Math. Stat.*, Vol. 16 (1945), pp. 294–300.

[5] C. R. RAO, "On the most efficient designs in weighing", *Sankhyā*, Vol. 7 (1946), pp. 440