THE FACTORIAL APPROACH TO THE WEIGHING PROBLEM[1]

By O. Kempthorne

Iowa State College

1. Summary. The weighing problem is discussed from the point of view of factorial experimentation. The paper contains a brief description of the fractional replication of the 2^n factorial system. It is shown that optimum designs for the weighing problem may easily be obtained with this approach. This approach is valuable in indicating the structure of weighing problem designs, and the limited conditions under which such designs can give results of value.

2. Introduction. Considerable attention has been given recently to the problem of weighing a number of light objects on a scale [1, 2, 3, 6]. The problem was originally proposed by Yates in his paper on complex experiments [4] as an example of a factorial experiment in which interactions between the factors tested would not be expected to exist: that is, the weight of say two objects could be assumed to be the sum of the weights of the objects weighed separately, after taking account of any necessary zero corrections. Such a situation is comparatively rare in biological research when, for example, the effect on yield of a particular crop from the joint application of two nutrients is usually different from the sum of the effects of separate applications. In recent years attention has been given to the use of fractional replication in factorial experiments [7, 8, 9] and it is proposed in this paper to consider the weighing problem from this point of view.

3. The 2^n factorial system. A full description of the 2^n factorial system was given by Yates in his technical communication *The Design and Analysis of Factorial Experiments* [5]. Yates was particularly concerned with the analysis of such experiments and with the evolution of systems of confounding in order to reduce the number of plots in each block. The following brief account is given in order to facilitate the discussion of the weighing problem.

In a single replication of the 2^n system all combinations of n factors each at two levels are tested. With three factors, a, b, c, for example, the following eight combinations are tested: (1) a, b, ab, c, ac, bc, and abc, where (1) denotes the control, a the application of treatment a only, ab the application of treatments a and b, and so on. A set of seven independent comparisons between the eight test results is given formally by the expansion of the formula

$$\tfrac{1}{4}(a \pm 1)(b \pm 1)(c \pm 1),$$

[1] Journal Paper no. J 1548 of the Iowa Agricultural Experiment Station, Ames, Iowa. Project No. 890.

where at least one of the signs is negative. If, for instance the first sign only is taken to be negative, a formal expansion gives the expression

$$\tfrac{1}{4}\{abc - bc + ab - b + ac - c + a - 1\},$$

and this contrast of the observations gives the effect of the factor a averaged over the presence and absence of the factors b and c, which is denoted by effect A. Similarly taking the negative sign in the second bracket only, we get the average effect B,

$$B = \tfrac{1}{4}\{abc - ac + ab - a + bc - c + b - 1\}.$$

Taking negative signs in the first and second brackets we obtain the interaction AB

$$AB = \tfrac{1}{4}\{abc + c + ab + (1) - ac - bc - a - b\}, \quad \text{and so on.}$$

The definition of effects and interactions may be presented very simply in geometrical terminology, by representing the treatment combinations as points of an n-dimensional lattice, each axis of the lattice having two points at unit distance apart. The control treatment will have coordinates $(0, 0, 0, \cdots, \cdots, 0)$, the treatment consisting of a only will have co-ordinates $(1, 0, 0, \cdots, 0)$ and so on. The effect A is then the difference of the mean yield of the treatments corresponding to the points lying on the hyperplane

$$x_1 = 0,$$

and the mean yield of those represented by points lying on the hyperplane

$$x_1 = 1.$$

The interaction of two factors a and b, represented by the axes x_1 and x_2 respectively, will be obtained from the difference of the mean yields of those plots for which

$$x_1 + x_2 = 0, \qquad \text{or } x_1 + x_2 = 2,$$

and those for which

$$x_1 + x_2 = 1.$$

The extensions to the above for three-factor and higher order interactions are simple. The interaction of factors a, b, and c, which are represented by coordinate axes x_1, x_2, and x_3, is given by the difference between the mean of plots represented by points for which

$$x_1 + x_2 + x_3 = 0 \text{ or } 2,$$

and those represented by points for which

$$x_1 + x_2 + x_3 = 1 \text{ or } 3;$$

in other words, it is the difference of the mean yields of those plots for which

$$x_1 + x_2 + x_3 = 0 \pmod 2$$

and of those for which

$$x_1 + x_2 + x_3 = 1 \pmod 2.$$

Each effect or interaction is then defined as the mean difference of two sets of plots, each set being represented by points on parallel hyperplanes, and the planes of one set of parallel hyperplanes lying between the planes of the other set. It is necessary to specify only the direction cosines of the hyperplanes in order to specify the effect or interaction, and the usual terminology for effects and interactions follows, in that the interaction of factors a, b, c, for example may be represented by the symbol ABC.

In the same way as effects and interactions are defined in terms of the yields of the several treatment combinations, the expected yield from each treatment combination may be expressed in terms of the mean level of yield and the true effects and interactions. If the full set of combinations of the factorial scheme is tested, the best estimate of each true effect and interaction is the same function of the observed yields that the true effect or interaction is of the true yields. This fact is one of the advantages which follow from the use of the full factorial scheme.

We are not concerned here with factorial experiments in which the factors have more than two levels, but when the number of levels of each factor is the same prime number, effects and interactions may be represented by products of powers of the symbols for the factors. In the case of two factors (a, b) at three levels, for example, the main effects may be represented by A, B, and the interactions by AB and AB^2, each symbol referring to the two independent contrasts between three sets, each of three plots.

As an example of the use of the above representation, we may consider confounding, that is, the arrangement of the treatment combinations in blocks in order to reduce the experimental error. Suitable arrangements are such that contrasts between the blocks represent high order interactions which the experimenter is confident will be of negligible size.

If treatment combinations for which

$$\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n = 0 \pmod 2$$

and for which

$$\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n = 0 \pmod 2$$

are arranged in a particular block, then the coordinates of the treatment combinations in this block also lie on the hyperplane

$$(\alpha_1 + \beta_1)x_1 + (\alpha_2 + \beta_2)x_2 + \cdots + (\alpha_n + \beta_n)x_n = 0 \pmod 2,$$

where the coefficients $(\alpha_1 + \beta_1)$ must be reduced modulo two. If, therefore, the treatments are arranged in blocks so that two comparisons are block contrasts, then the generalised interaction of these contrasts is also a block contrast.

4. Fractional replication. The principle of fractional replication follows very simply. Suppose only those treatment combinations whose yields all occur either in the positive or the negative part of a particular contrast are represented in the experiment, that is only those combinations represented by the points of the lattice for which say

$$\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n = 0 \pmod 2.$$

Then the comparison between the yields of those plots represented by

$$\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n = 0 \pmod 2$$

and by

$$\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n = 1 \pmod 2$$

will be identical with the comparison between the yields of plots represented by

$$(\alpha_1 + \beta_1)x_1 + (\alpha_2 + \beta_2)x_2 + \cdots + (\alpha_n + \beta_n)x_n = 0 \pmod 2$$

and by

$$(\alpha_1 + \beta_1)x_1 + (\alpha_2 + \beta_2)x_2 + \cdots + (\alpha_n + \beta_n)x_n = 1 \pmod 2.$$

The former of these two comparisons may be represented by the symbol $x_1^{\beta_1}x_2^{\beta_2} \cdots x_n^{\beta_n}$, and the latter by $x_1^{\alpha_1+\beta_1}x_2^{\alpha_2+\beta_2} \cdots x_n^{\alpha_n+\beta_n}$, where $x_1, \cdots, x_n$ are no longer coordinates but symbols for the n factors, which satisfy the relations, $x_i^\alpha = 1$, if $\alpha = 0 \pmod 2$. The equivalence of the two comparisons may be obtained by the use of an identity relationship in the symbols for the factors

$$I = x_1^{\alpha_1}x_2^{\alpha_2} \cdots x_n^{\alpha_n}$$

where I is interpreted as unity, and only those combinations whose coordinates $(x_1, x_2, \cdots, x_n)$ satisfy one of the equations

$$\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n = 0, \qquad \text{or} = 1 \pmod 2,$$

are represented in the experiment. If this identity relationship is multiplied by the symbol $x_1^{\beta_1}x_2^{\beta_2} \cdots x_n^{\beta_n}$ by ordinary commutative algebra, reducing the powers modulo 2 where necessary, we obtain

$$x_1^{\beta_1}x_2^{\beta_2} \cdots x_n^{\beta_n} = x_1^{(\alpha_1+\beta_1)}x_2^{(\alpha_2+\beta_2)} \cdots x_n^{(\alpha_n+\beta_n)}.$$

It is more convenient to revert to the common use of capital letters A, B, C, etc. for effects corresponding to small letters a, b, c, etc. for the factors tested. An experiment in half-replicate is then represented formally by an equation of the type

$$I = A^\alpha B^\beta C^\gamma \cdots$$

In such an experiment on n factors only 2^{n-1} treatment combinations will be tested. Of the $2^n - 1$ independent comparisons in a fully replicated experiment, information on one comparison is lost completely since only those treatments which appear in the comparison with the same sign are represented: the remaining $2^n - 2$ independent comparisons of a fully replicated experiment are identical in pairs giving $2^{n-1} - 1$ independent comparisons. Each comparison is then said to have two aliases and measures the sum (or difference, depending on which half of the treatment combinations are used) of two effects, an effect and an interaction, or two interactions.

A quarter-replicated experiment can by the same process be represented by an identity relationship of the form

$$I = A^{\alpha_1}B^{\beta_1}C^{\gamma_1} \cdots = A^{\alpha_2}B^{\beta_2}C^{\gamma_2} \cdots = A^{(\alpha_1+\alpha_2)}B^{(\beta_1+\beta_2)}C^{(\gamma_1+\gamma_2)} \cdots$$

It is useful in the evolution of fractional designs to note that the elements in the identity relationship form an Abelian group.

Fractionally replicated experiments are formally identical with confounded experiments in that block differences may be regarded as additional factors in the confounded experiment. A 2^n experiment arranged in 2^p blocks, for example, may be regarded as a 1 in 2^p design of a 2^{n+p} experiment. Considerable care needs to be exercised in the use of fractionally replicated designs, but they have been found to be very useful in agricultural and biological research.

5. The weighing problem. The problem of weighing a number of objects may be regarded as the problem of the estimation of the effects of a number of factors which do not interact. To take a simple case, consider the estimation of the effects of factors a, b, and c for which one complete replicate would consist of the combinations

$$(1)\ a,\ b,\ ab,\ c,\ ac,\ bc, \text{ and } abc.$$

Suppose a half replicate design is used, based on the identity relationship

$$I = ABC.$$

The combinations tested would then consist of either the set $\{a, b, c, abc\}$ or the set $\{(1), ab, ac, bc\}$. If the former set were chosen, the comparison estimating the effect A could also be ascribed to the interaction BC, that estimating effect B also to the interaction AB, and that estimating effect C to the interaction AC, as can be observed by multiplying the identity relationship by A, B, and C in turn. If the experimenter is confident that the two-factor interactions are negligible, then any effect given by each comparison would be ascribed to the main effect.

6. Discussion of a particular case. We give the derivation of a design for weighing a particular number of objects, say ten. Let the objects be denoted by $a, b, c, d, e, f, g, h, k, l$. Then the total number of combinations which could be tested is 2^{10}, that is 1024, but as we are confident that interactions are negligible, it is necessary only to estimate main effects.

A fractionally replicated design must consist of a number 2^p of combinations and this will be a 1 in 2^{10-p} design. A suitable fractionally replicated design consisting of 16 combinations will exist if it is possible to evolve an identity relationship for a 1 in 64 design, such that each term in the relationship involves at least three letters. A possible identity relationship for such a design contains the numbers of the Abelian group obtained from all combinations of the elements 1, ABC, CDE, EFG, GHK, ADL, and AFH, with the rule that the square of each letter is to be equated to unity. Each possible comparison may then be due to any of the 64 effects or interactions which may be derived from this identity relationship. In other words, each comparison has 64 aliases: in the case of ten of the comparisons, only one of the aliases is a main effect, and for the remaining five comparisons the aliases are all interactions of at least two factors. The actual design may be written down by finding combinations of the letters which have the same number of letters in common with the unit element and the six three-factor interactions. These are themselves a group consisting of all combinations of unity and four combinations of letters. The sixteen combinations with an even number of letters in common with all the members of the identity group are found to be the following:

(1)	*abdef*,	*acefl*,	*bcdl*,
abfgkl,	*degkl*,	*bcegk*,	*acdfgk*,
fgh,	*abdegh*,	*aceghl*,	*bcdefghl*,
abhkl,	*defhkl*,	*bcefhk*,	*acdhk*

The estimation of effects from the results of the sixteen weighings is particularly easy; the weight of object a will be one-eighth of the difference between those weighings containing a and those not containing a. There are ten such contrasts which estimate the effects, and the remaining five contrasts may be used to obtain an estimate of the experimental error. If σ^2 is the variance of each weighing, the variance of the weight of a, that is, the effect A will be $(1/8 + 1/8)\sigma^2 = (1/4)\sigma^2$. The precision can be increased fourfold in the weighing problem with a chemical balance by interpreting the absence of each letter as the placing of the object in the left hand pan and the presence as the placing of the object in the right hand pan. Each effect will then measure twice the weight of the corresponding object and the estimated weight of each object will have a variance of $\sigma^2/16$, that is, the same precision as if each had been weighed by itself eight times in each pan, or sixteen times in all.

7. General case. The rules by which fractional designs may be constructed have been exemplified above and the procedure is simple, though laborious in the case of a large number of objects. It does not, therefore, seem worth while to enumerate particular designs for the weighing of particular numbers of objects. A general procedure in considering the design for a particular problem is as follows. Taking the case of a number n of objects, the experimenter should form a rough idea of the order of magnitude of the experimental error, σ say, and decide what accuracy he requires for his estimates of the weights, a standard error say of s. Then if he weighs 2^p combinations of the objects, the standard error of the estimate of each weight will be $2^{-p/2}\sigma$ in the case of the chemical balance. This serves to determine $2^{p/2}$ and therefore p, and it is then necessary to design a 2^n experiment of fraction 2^{n-p}. Alternatively, a design of higher fraction which can provide estimates may be replicated the corresponding number of times. In the case of the spring balance the corresponding standard error is $2^{-(p-1)/2}\sigma$ necessitating a design of higher fraction.

Designs of the type described above have some useful properties:
(1) the design automatically takes care of any bias in the balance,
(2) the effects or weights may be computed easily as indicated above,
(3) the effects or weights are uncorrelated,
(4) all the effects are measured with the same precision, and
(5) an estimate of the experimental error which is independent of the effects may be computed from the results.

In considering the use for a particular problem of a design like the one discussed, it is important to understand completely the structure of the design. Such designs may well have real value for the weighing problem, but it is not easy to visualize other problems for which they would not give results capable of various interpretations. The use of the above designs depends on an assumption that interactions between pairs of factors are negligible, and this is generally not the case, for example, in biological research work, in which the experimenter may well be confident that interactions between three or more factors are negligible. In the particular case described in detail, there are only fifteen independent comparisons between the sixteen results which will be obtained, and it follows from the identity relationship that the comparison giving the effect A, also measures the two factor interactions BC, DL, and FH. If therefore the factor a has no effect and there is an interaction between factors b and c or the other two pairs of factors, the experimenter will conclude that the factor a has an effect. It is clear that under these circumstances the experimental results are worthless.

8. Efficiency of designs. In [2], Mood states for optimum designs such that when N weighings are made, the variance of the estimates of the weights are of the order of $\frac{\sigma}{\sqrt{N}}$ in the case of the chemical balance and $\frac{2\sigma}{\sqrt{N}}$ in the case of the spring balance, where σ^2 is the variance of a single weighing. As indicated above, when N is a power of 2, the fractional factorial designs result in the same variances. These designs are similar to those proposed by Kishen [6].

Mood dealt with the case in which the number of weighings was restricted, for example to be equal to the number of objects, and defined a best design as that which gave the smallest confidence region in the p-dimensional space for the estimates of the p-weights.

To take an example for the weighings of 3 objects with a spring balance with no bias he suggests the following design:

$$X = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

where the rows of the array refer to weighing operations and the columns refer to objects. If the results of the weighings are y_1, y_2, and y_3 respectively, the estimates of the weights b_1, b_2, and b_3 are given by the equation

$$\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} 1/2 & 1/2 & -1/2 \\ 1/2 & -1/2 & 1/2 \\ -1/2 & 1/2 & 1/2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}.$$

If σ^2 is the variance of a single weighing, then the variance of each estimate will be $[(1/2)^2 + (1/2)^2 + (-1/2)^2]\sigma^2 = (3/4)\sigma^2$: or if N $(= 0(\text{mod } 3))$ weighings are made by replicating the above system $N/3$ times, the variance of each estimate will be $9\sigma^2/4N$. The covariance of any two estimates is $(-1/4)\sigma^2$ so that the square of the correlation between any two estimates is $-1/9$. The fractional factorial design will yield estimates which have a somewhat higher variance, namely $4\sigma^2/N$. This increase in precision obtained in Mood's design has been obtained at the expense of obtaining correlated estimates which in addition are subject to any bias which the measuring instrument may have. It is doubted whether the use of such designs for any practical problem can be justified.

It is of interest to note that the concept of fractional replication may be extended to give designs requiring a number of weighings other than a power or two. For the weighing of 3 objects for example, a factorial design of fraction 3/4 could be used: it could consist of a half-replicate based on the identity $I = ABC$, and a quarter replicate based on the identity

$$I = A = BC = ABC.$$

There is, however, little point in discussing such designs for the weighing problem, as their efficiency as measured by the total number of weighings needed to achieve a particular degree of accuracy is lower than for the designs described in this paper.

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. L. PLACKETT AND J. P. BURMAN, "The design of optimum multifactorial experiments", *Biometrika*, Vol. 33 (1946), pp. 305–325.

[4] F. YATES, "Complex experiments", *Roy. Stat. Soc. Jour. Suppl.*, Vol. 2 (1935), pp. 181–247.

[5] F. YATES, "The design and analysis of factorial experiments", *Technical Communication No.* 35, Imperial Bureau of Soil Science, Harpenden, Herts, England, 1937.

[6] K. KISHEN, "On the design of experiments for weighing", *Annals of Math. Stat.*, Vol. 14 (1945), pp. 294–301.

[7] D. J. FINNEY, "The fractional replication of factorial arrangements", *Annals of Eugenics*, Vol. 12 (1945), pp. 291–301.

[8] D. J. FINNEY, "Recent developments in the design of field experiments. III. Fractional replication", *Jour. of Agric. Sci.*, Vol. 36 (1946), pp. 184–191.

[9] O. KEMPTHORNE, "A simple approach to confounding and fractional replication in factorial experiments", *Biometrika*, Vol. 34 (1947), pp. 255–272.