RANDOM ALLOCATION DESIGNS I: ON GENERAL CLASSES OF ESTIMATION METHODS[1]

By A. P. Dempster

Harvard University

1. Summary. Certain linear estimation procedures for randomized experimental designs are evaluated relative to the criteria of bias, variance and mean square error. For the designs considered, treatment combinations are randomly allocated to experimental units, the randomness being subject only to a wide symmetry condition. Statistical properties refer to the discrete probabilities induced by the randomization hypothesis. Section 2 defines the basic statistical model and discusses the question of conditional inference relative to this model. Certain vectorial notation and terminology is introduced in Section 3. Although the theory of the paper applies directly to k-factor designs with general k, the notation is set up in Section 3 for a three factor design, and the three factor notation is used throughout, except for Section 5 which discusses an even simpler example. Two general classes of linear unbiased estimators are defined in Section 4 and illustrated in Section 5. In Section 6 it is shown that estimators of the types defined in Section 4 have optimum properties in a wide class of linear estimators. Finally, the theory for the basic model is generalized in Section 7 to cover the case of observations with error.

Formal proofs of stated theorems are to be found at the ends of Sections 4.2, 4.3 and 6.

2. Introduction. Consider a completely crossed k-factor design with R levels of factor 1, C levels of factor 2, $\cdots$, L levels of factor k. In recent years F. E. Satterthwaite [6] has proposed designing experiments by drawing at "random" from such an array, usually with some restrictions on the random choice, plus some replication, and then performing the experiments indicated by the random choice of treatment combinations. This has been termed "random balance experimentation" or "random allocation experimentation." For example an experiment with n observations could be designed by choosing independently n sets of treatment combinations each according to the following simple rule: the level of factor 1 is chosen at random from the R possibilities, independently a choice of one of the C levels of factor 2 is made, and so on. The general technique appears to have been introduced primarily for application to very large arrays where only a very small fraction can be contemplated, and so is often thought of as a competitor of highly fractionated factorials which deliberately confound certain effects with others. In the case of random allocation designs, confound-

Received May 25, 1959; revised June 27, 1960.

[1] This research has been supported in part by the United States Navy through the Office of Naval Research under contract Nonr 1866(37). Reproduction in whole or in part is permitted for any purpose of the United States Government.

ing is random, or at least partly random. The proper place in practice of random balance designs relative to more conventional designs is a controversial matter (see [1], [3], [6], [10]). The results presented in this paper were motivated by the need for a theoretical framework within which random balance designs could be compared and evaluated relative to more conventional designs. The results pertain, however, to a general class of models which can be applied not only to the random balance experiments referred to above but also to a wide range of other designs including (i) fixed fractions where the labels of the different levels of each factor are assigned at random, (ii) arrays which are complete except for randomly selected "missing cells," and (iii) conventional "completely randomized designs" as discussed in [9]. Subsequently in this paper the term random allocation will be used rather than random balance, partly because the wider applicability than simply to Satterthwaite's random balance designs makes a more neutral term desirable, and partly because the term random allocation seems to be more descriptive.

The basic statistical model is as follows. Corresponding to each of the $N = R \times C \times \cdots \times L$ cells of the complete k-way design there exists a fixed quantity to be thought of as the result of an experiment performed with the corresponding factor level combinations. By a *design* we mean a subset of n of the N factor level combinations, and an experiment performed using a given design provides the values of the n fixed quantities corresponding to the n cells of the design. For the mathematics of this paper a *random allocation scheme* is a method of selecting a design as a random subset of n of the N possible factor level combinations where the only restriction on the probability of selection of a particular set of n is that all other sets of n obtained from a given set by permutation of factor levels shall have the same probability of selection. It will be convenient to use group-theoretic language in dealing with this definition of random allocation. Suppose $\mathfrak{G}$ denotes the group of all permutations of the levels within all factors, so that $\mathfrak{G}$ has order

$$P = (R!)(C!) \cdots (L!).$$

By applying all the elements of $\mathfrak{G}$ to a given design one obtains a set of designs which we call a symmetry class of designs under group $\mathfrak{G}$. In this way all $\binom{N}{n}$ possible designs are classified into mutually exclusive symmetry classes. Our definition of a random allocation scheme states that all the designs of any one symmetry class must be equiprobable, but no restriction is placed on the probabilities of selection of the different symmetry classes. The different weights allowable for different symmetry classes result in the wide range of possible types of designs alluded to above. Possibly the simplest example occurs when the n selected combinations are a simple random sample without replacement from the N possible combinations, and this we call *simple random allocation*. Two types of modifications of simple random allocation which may appear separately or in combination may be termed *random allocation with partial balance* and *random allocation with partial confounding*. An often recommended example of

the first type is defined by the restriction of the random choice so as to require for each factor that each of its levels appear an equal (or as near equal as possible) number of times. It is clearly possible to extend this technique to balancing with regard to combinations of factors rather than simply with regard to factors one at a time. For example, in an experiment of size $n = 32$ on an array with 20 factors each at 2 levels, one could divide the factors into 4 groups of 5 each and balance the experiment with regard to 4 complete 2^5 experiments, i.e., the experiment would be a complete 2^5 in 4 ways, the 4 ways being randomly superposed on one another. Partially confounded random allocation, as defined here, would arise if one were to choose a fixed fractional factorial from the whole array where certain effects are deliberately confounded, subsample at random in some sense from the fixed fraction, and finally randomly permute the labels of rows, columns, etc., by choosing at random a member of $\mathfrak{G}$. Clearly partial balance can be built into the second of the three stages of choice of a partially confounded random allocation design. Alternatively, if the second stage of choice is omitted, partially confounded random allocation includes any complete standard fractional factorial provided that the experimenter has seen fit to randomly permute the labels of factor levels independently for each factor. Thus any comparison between, say, simple random allocation and a prechosen-then-randomized fraction can be made entirely within our class of random allocation designs. A final example, to show the breadth of our definition, is the class of completely randomized designs where t treatment combinations are applied at random to t experimental units chosen at random from $r \geqq t$ experimental units. This is clearly a random allocation design where the t treatment combinations are regarded as the t levels of one factor and the r experimental units are regarded as the r levels of a second factor.

A generalization of the basic model will be considered in Section 7. The theory extends immediately to this generalization, but in the interest of clarity the main presentation will use the basic model. The generalization allows the fixed array to become random through the addition of a random error with zero mean and arbitrary variances and covariances. The generalization can be made to cover certain methods of design where some of the cells are replicated in the design.

Our aims are to provide, for the basic statistical model, methods of estimating linear combinations of the N fixed cell-values, and to search for optimum methods of estimation. The same discussion will apply to all of the types of random allocation schemes within our definition. The criteria for good estimators will be the usual criteria, "unbiasedness", "unbiasedness with minimum variance" and "minimum mean square error." These criteria have been placed in quotes to emphasize that they are not yet well-defined and indeed that they may be defined in several ways. Controversy over what statistical properties may be properly associated with randomized designs has a long history as may be seen in the opposing contentions of Fisher and "Student" [2], [8], and such issues do not appear to be definitely resolved even today. The question usually comes down to: how conditional should the inference be? One point of view is that, having

made our design random, it is only sensible to use this randomness, by averaging over the random choice of design, when defining statistical properties like means and variances. The opposing point of view states that the randomness in the choice of design does not depend on the unknown quantities to be estimated, i.e., that the chosen design is an ancillary statistic in the sense of Fisher [5], and so we must make inferences conditional on the design actually used. Curiously enough the two opposing principles, namely the principle of basing inferences on the random properties of randomized designs and the principle that one must make inferences conditional on ancillary statistics, are both associated with the name of Fisher. The author believes the latter principle to be desirable in theory but not always practicable. In the case of our basic model where n of N fixed constants are provided by the random allocation scheme, to condition on the design is to eliminate all randomness from the model. There are two methods of restoring randomness to the model. The first method, and the one adopted for our theory, is to relax the conditioning requirement. The second method is to assume a more structured model for the observations; for example we might assume a model I fixed effects analysis of variance model with fewer than n fixed effects to be estimated. The choice of method poses a dilemma, for the logically more satisfying second method may yield incorrect results because the more structured model makes incorrect assumptions.

In this paper we develop theory for the general unstructured model and take averages over the random choice of design. We can also, however, take one step towards conditioning on the design and condition on the observed symmetry class of designs, i.e., rather than average over all designs we can just average over the observed symmetry class. This conditioning is equivalent to the assumption that the random allocation scheme used has just the observed symmetry class with probability one. Since all of the designs of one symmetry class have similar confounding patterns the procedure of conditioning on the observed symmetry class has the intuitive appeal of averaging only over designs with confounding patterns similar to that observed. In any case we will be dealing with estimators which are unbiased in one of two senses: they may be (i) *U_1-unbiased*, i.e., conditionally unbiased given the symmetry class, or (ii) *U_2-unbiased*, i.e., unbiased under averaging over symmetry classes as well as within. Clearly any U_1-unbiased estimator is also U_2-unbiased.

The consequences of adopting the method of averaging over designs may at first appear startling. For example it becomes possible to find unbiased estimates of every linear combination of the N fixed quantities for every random allocation design, even with $n = 1$. Suppose the N fixed quantities are denoted by $v_1, v_2, \cdots, v_N$. Then a random allocation experiment with $n = 1$ amounts to observing one of the v_i chosen at random. For each i an unbiased estimator of v_i can be defined as

$$\hat{v}_i = Nv_i \quad \text{if } v_i \text{ is observed}$$
$$= 0 \quad \text{otherwise.}$$

Consequently an unbiased estimator of $\sum_{i=1}^{N} \alpha_i v_i$ is $\sum_{i=1}^{N} \alpha_i \hat{v}_i$. Of course, such an estimator based on $n = 1$ would have so large a variance that it would be useless, but still it retains theoretical validity.

3. Terminology and notation. Although the theory applies directly to designs with any number of factors, suppose, to simplify notation, that we use as prototype a design with three factors having R, C and L levels and so $N = RCL$ cells altogether. Suppose the cells have associated numbers

$$v_{ijk} \ (i = 1, 2, \cdots, R; j = 1, 2, \cdots, C; k = 1, 2, \cdots, L)$$

to be regarded as fixed quantities or parameters. The quantities v_{ijk}, together with all linear combinations of these quantities, may be taken as the values assumed by a particular linear functional f_t over an N-dimensional vector space E. The vector space E is defined abstractly in terms of N basis vectors

$$\mathbf{V}_{ijk} \ (i = 1, 2, \cdots, R; j = 1, 2, \cdots, C; k = 1, 2, \cdots, L),$$

which are in one to one correspondence with the cells of the basic array; and f_t, called the *total functional*, is defined by

$$f_t\Big(\sum_{i,j,k} \alpha_{ijk} \mathbf{V}_{ijk}\Big) = \sum_{i,j,k} \alpha_{ijk} v_{ijk} .$$

In particular $f_t(\mathbf{V}_{ijk}) = v_{ijk}$ for all i, j and k. The random allocation experiment provides observations for a random subset of n of the N cells. Suppose the corresponding n vectors $\mathbf{V}_{ijk}$ span subspace E_p of E. Note that the experiment provides the values of $f_t(\mathbf{V})$ for $\mathbf{V} \ \varepsilon \ E_p$ only.

An alternative method of introducing vectorial terminology would be to regard the v_{ijk} as defining an N-dimensional vector. Such a vector lies in the dual space (see [7]) of the vector space E introduced in the preceding paragraph. Since we wish to work with vectors in E we prefer the terminology which calls the set of v_{ijk} a linear functional rather than a vector.

A Euclidean metric may be inserted in E by regarding each $\mathbf{V}_{ijk}$ to be a unit vector orthogonal to all other $\mathbf{V}_{ijk}$, i.e., the metric is such that the vector

$$\sum_{i,j,k} \alpha_{ijk} \mathbf{V}_{ijk}$$

has squared length $\sum_{i,j,k} \alpha_{ijk}^2$. This metric will be referred to throughout as the *formal metric* and, unless otherwise stated, orthogonality relationships and lengths will be relative to the formal metric.

In accordance with standard analysis of variance ideas, the space E can be expressed as the direct sum of eight mutually orthogonal subspaces, E_M, E_R, E_C, E_L, E_{RC}, E_{RL}, E_{CL} and E_{RCL}, of dimensions respectively 1, $R - 1$, $C - 1$, $L - 1$, $(R - 1)(C - 1)$, $(R - 1)(L - 1)$, $(C - 1)(L - 1)$ and

$$(R - 1)(C - 1)(L - 1).$$

For example, the space E_{RC} consists of all vectors $\sum_{i,j,k} \alpha_{ijk} \mathbf{V}_{ijk}$ such that analysis of variance of the array α_{ijk} produces zero mean squares for all effects

except possibly the RC interactions. Similar definitions apply to the other subspaces.

A class of metrics may be introduced in E by stretching or shrinking the space along the subspaces $E_M, \cdots, E_{RCL}$. More precisely, suppose $\mathbf{V} \varepsilon E$ is expressed as

$$\mathbf{V} = \alpha_M \mathbf{V}_M + \alpha_R \mathbf{V}_R + \cdots + \alpha_{RCL} \mathbf{V}_{RCL}$$

where $\mathbf{V}_M \varepsilon E_M, \cdots, \mathbf{V}_{RCL} \varepsilon E_{RCL}$ and these are unit vectors according to the formal metric. Then, according to the λ-*metric* defined by $(\lambda_M, \lambda_R, \cdots, \lambda_{RCL})$, $\mathbf{V}$ has squared λ-length

$$\alpha_M^2 \lambda_M^2 + \cdots + \alpha_{RCL}^2 \lambda_{RCL}^2 .$$

Note that the formal metric is the particular λ-metric where

$$\lambda_M = \lambda_R = \cdots = \lambda_{RCL} = 1.$$

Metrical properties relative to a general λ-metric will be referred to as λ-*properties*, e.g., a vector has a λ-*length* or a pair of vectors have a λ-*angle*.

4. General classes of estimation methods.

4.1. *Motivation.* Our purpose is to find unbiased estimators of $f_t(\mathbf{V})$ for any $\mathbf{V}$ based on data giving the values of $f_t(\mathbf{V})$, where $\mathbf{V}$ belongs to the random subspace E_p. The values of $f_t(\mathbf{V})$ for $\mathbf{V}$ belonging to one of the subspaces

$$E_M, \cdots, E_{RCL}$$

have special interpretations and are of special interest. For example a typical $\mathbf{V} \varepsilon E_R$ is

$$\mathbf{V}_0 = \frac{1}{CL} \sum_{j,k} (\mathbf{V}_{i_1 jk} - \mathbf{V}_{i_2 jk})$$

and its associated f_t value, namely

$$v_0 = \frac{1}{CL} \sum_{j,k} (v_{i_1 jk} - v_{i_2 jk}),$$

is of special interest as the difference of two row main effects. A brief discussion of methods of estimating v_0 will help to motivate subsequent general methods.

The first unbiased estimator of v_0 which comes to mind is probably the difference of two means, the mean of those v_{ijk} which were observed with $i = i_1$, and the mean of those with $i = i_2$. Provided that each possible design yields at least one observation in every row, this estimator is evidently both defined and U_1-unbiased for any random allocation scheme. If, however, it is suspected that large column effects are present, and if the design permits, one would probably apply the foregoing method to the data corrected for column main effects. In both cases these estimators are values of f_t for vectors in E_p, in the first case for a vector in E_p which is (i) perpendicular to E_M, and (ii) perpendicular to

that part of E_R orthogonal to $\mathbf{V}_0$, and which makes the smallest angle with $\mathbf{V}_0$ subject to conditions (i) and (ii). In the second case a third condition must be added: (iii) perpendicular to E_C. Both of these methods and their natural extensions are discussed by Anscombe [1]. The first method avoids any influence of the grand mean on an estimate, whereas the second method avoids any influence of the grand mean or column effects. Extensions to the correction for other effects are evident. This example is intended to illustrate the following heuristic viewpoint: in estimating an effect v_0 one will use $f_t(\mathbf{V})$ for a vector $\mathbf{V}$ in E_p chosen as a compromise between being near $\mathbf{V}_0$ and not too near other directions with large associated effects. In the example, the device for keeping away from dangerous directions is to require the used direction to be orthogonal to dangerous directions.

A generalization of this device with much greater flexibility is to use a vector in E_p which makes minimum λ-angle with $\mathbf{V}_0$ for λ-metric $(\lambda_M, \cdots, \lambda_{RCL})$. The two methods described above may be shown to be limiting cases of this general method where

(a) $\lambda_M = \lambda_R \to \infty$ and $\lambda_C = \lambda_{RC} = \cdots = \lambda_{RCL} = 1$

and

(b) $\lambda_M = \lambda_R = \lambda_C \to \infty$ and $\lambda_{RC} = \cdots = \lambda_{RCL} = 1$ respectively. The general method was motivated by the belief that it would be better to stop between these extremes. In fact the author's heuristic feeling led him to conjecture that one should stretch (or shrink) the metric until the effects corresponding to λ-unit vectors in the directions $E_M, \cdots, E_{RCL}$ are in some sense equalized, or more specifically until

$$\lambda_M^2 = (MS)_M, \lambda_R^2 = (MS)_R, \cdots, \lambda_{RCL}^2 = (MS)_{RCL}$$

where the (MS) values are the mean squares appearing in an analysis of variance table for the complete array v_{ijk}. The sense in which such a procedure is optimum will be indicated in Section 6.

There is one direction in E_p λ-nearest to $\mathbf{V}_0$, but there are many vectors in this direction. The question of which to use in defining the estimator has been up to now mostly ignored in our discussion. We now proceed to more precise definitions of estimators.

4.2. *λ-minimum extensions.* The objective is to find an unbiased (U_1- or U_2-) estimator of $f_t(\mathbf{V})$ for all $\mathbf{V} \varepsilon E$. We shall consider only estimators which are themselves linear functionals over E, i.e., if $\hat{f}_t(\mathbf{V}_1)$ estimates $f_t(\mathbf{V}_1)$ and $\hat{f}_t(\mathbf{V}_2)$ estimates $f_t(\mathbf{V}_2)$ then the estimator of $f_t(\alpha\mathbf{V}_1 + \beta\mathbf{V}_2)$ is $\alpha\hat{f}_t(\mathbf{V}_1) + \beta\hat{f}_t(\mathbf{V}_2)$. The estimators which we shall consider are conveniently expressed in terms of a random linear functional denoted by $f_\lambda(\mathbf{V})$ and called the *λ-minimum extension of f_t from E_p to E.*

The linear functional f_λ is determined by the observations and the λ-metric $(\lambda_M, \cdots, \lambda_{RCL})$. It is defined by

$$\begin{aligned} f_\lambda(\mathbf{V}) &= f_t(\mathbf{V}) && \text{for } \mathbf{V} \varepsilon E_p \\ &= 0 && \text{for } \mathbf{V} \lambda\text{-orthogonal to } E_p. \end{aligned}$$

These two statements define f_λ completely, since any $\mathbf{V} \varepsilon E$ has a unique expression as $\mathbf{V}_1 + \mathbf{V}_2$ where $\mathbf{V}_1 \varepsilon E_p$ and $\mathbf{V}_2$ is λ-orthogonal to E_p and so

$$f_\lambda(\mathbf{V}) = f_\lambda(\mathbf{V}_1) + f_\lambda(\mathbf{V}_2) = f_t(\mathbf{V}_1).$$

It is clear that an alternative characterization of f_λ is to define $f_\lambda(\mathbf{V})$ for any $\mathbf{V} \varepsilon E$ to be $f_t(\mathbf{V}_1)$ where $\mathbf{V}_1$ is that vector in E_p at minimum λ-distance from $\mathbf{V}$. This characterization ties in with the description in Section 4.1 of estimators of v_0 as f_t values for vectors in E_p nearest to $\mathbf{V}_0$.

A third, and quite different, characterization of f_λ is as follows. *The values of f_λ for $\mathbf{V}_{ijk}$ corresponding to the unobserved cells are those numbers which, together with the known $f_\lambda (= f_t)$ values for $\mathbf{V}_{ijk}$ corresponding to the observed cells, produce that full array which minimizes the expression*

$$\frac{(SS)_M}{\lambda_M^2} + \frac{(SS)_R}{\lambda_R^2} + \cdots + \frac{(SS)_{RCL}}{\lambda_{RCL}^2}$$

where $(SS)_M, \cdots, (SS)_{RCL}$ are the sums of squares arising from the analysis of variance of the full array. To prove that this characterization agrees with the first given, consider the class F_p of linear functionals which agree with f_t for $\mathbf{V} \varepsilon E_p$ but are otherwise arbitrary. Pick any basis of E consisting of λ-unit and λ-orthogonal vectors $\mathbf{W}_1, \mathbf{W}_2, \cdots, \mathbf{W}_N$ and consider for any $f \varepsilon F_p$ the property

$$S(f) = \sum_{i=1}^{N} [f(\mathbf{W}_i)]^2,$$

which may easily be checked to be independent of the choice of basis. According to the first definition, f_λ is that member of F_p which is zero over the subspace $\tilde{E}_p$ of E where $\tilde{E}_p$ consists of those $\mathbf{V} \varepsilon E$ λ-orthogonal to E_p. In terms of a basis $\mathbf{W}_1, \cdots, \mathbf{W}_N$ such that $\mathbf{W}_1, \cdots, \mathbf{W}_n \varepsilon E_p$ and $\mathbf{W}_{n+1}, \cdots, \mathbf{W}_N \varepsilon \tilde{E}_p$ it is seen to be equivalent to say that f_λ is that member of F_p which minimizes $S(f)$. On the other hand, when a basis $\mathbf{W}_1, \cdots, \mathbf{W}_N$ is selected which lies entirely within the subspaces $E_M, E_R, \cdots, E_{RCL}$, it becomes evident that

$$S(f) = \frac{(SS)_M}{\lambda_M^2} + \frac{(SS)_R}{\lambda_R^2} + \cdots + \frac{(SS)_{RCL}}{\lambda_{RCL}^2},$$

as required.

4.3. *Unbiased estimators.* For a given random allocation scheme and a given λ-metric $(\lambda_M, \lambda_R, \cdots, \lambda_{RCL})$ we define two unbiased estimators of $f_t(\mathbf{V})$, called the class 1 estimator and the class 2 estimator. The class 1 estimator is U_1-unbiased and the class 2 estimator is U_2-unbiased. The definitions of these estimators rest on the following theorem which is proved at the end of this section.

THEOREM. *Suppose Δ is a generic symbol representing one of the subscript combinations $M, R, \cdots, RCL$. Suppose the symbol "ave$\{\cdots\}$" denotes average over the random choice of design of a random allocation scheme. Then there exist*

constants $\gamma_M , \gamma_R , \cdots , \gamma_{RCL}$, *depending in general on the random allocation scheme, such that*

$$\text{ave}\{f_\lambda(\mathbf{V}_\Delta)\} = \gamma_\Delta f_t(\mathbf{V}_\Delta)$$

for any $\mathbf{V}_\Delta \varepsilon E_\Delta$ *and for all* Δ. *Further, if* $\lambda_M , \cdots , \lambda_{RCL}$ *are all finite and non-zero, then* $\gamma_M , \cdots , \gamma_{RCL}$ *are also all finite and non-zero.*

The definition of the class 2 estimator follows immediately: any $\mathbf{V} \varepsilon E$ has a unique expression as

$$\mathbf{V} = \sum_\Delta \mathbf{V}_\Delta \quad \text{where} \quad \mathbf{V}_\Delta \varepsilon E_\Delta$$

and, from the theorem, the *class* 2 *estimator* defined as

$$f_\lambda(\mathbf{V}) = \sum_\Delta \gamma_\Delta^{-1} f_\lambda(\mathbf{V}_\Delta)$$

is U_2-unbiased for $f_t(\mathbf{V})$ for all $\mathbf{V} \varepsilon E$. The definition of the class 1 estimator follows from an application of the theorem to the random allocation scheme which restricts designs to a single symmetry class G under $\mathfrak{G}$. This results in a set of constants $\gamma_{\Delta G}$ for each symmetry class G. Then the *class* 1 *estimator* is defined as

$$\hat{f}_\lambda(\mathbf{V}) = \sum_\Delta \gamma_{\Delta G}^{-1} f_\lambda(\mathbf{V}_\Delta)$$

for any $\mathbf{V} = \Sigma_\Delta \mathbf{V}_\Delta \varepsilon E$, where G denotes the observed symmetry class. The class 1 estimator is clearly U_1-unbiased. In general the class 1 and class 2 estimators are different and the class 2 estimator is not U_1-unbiased. Notice, however, that if the random allocation scheme allows only one symmetry class with probability one, then the class 1 and class 2 estimators coincide as do the concepts of U_1- and U_2-unbiasedness.

To prove the theorem stated above we shall express ave $\{f_\lambda(\mathbf{V}_{ijk})\}$ successively in terms of the following three sets of eight quantities each:

$$(v_{(i)(j)(k)} , v_{i(j)(k)} , v_{(i)j(k)} , \cdots , v_{ijk})$$

$$(v\ldots , v_{i\cdot\cdot} , v_{\cdot j\cdot} , \cdots , v_{ijk})$$

$$(m\ldots , m_{i\cdot\cdot} , m_{\cdot j\cdot} , \cdots , m_{ijk}).$$

The first set represents a method of breaking $\Sigma_{r,s,t} v_{rst}$ into

$$v_{(i)(j)(k)} = \sum_{\substack{r \neq i \\ s \neq j \\ t \neq k}} v_{rst} ,$$

$$v_{i(j)(k)} = \sum_{\substack{s \neq j \\ t \neq k}} v_{ist} ,$$

etc. In the second set a dot means that a mean is taken over the corresponding index, e.g.,

$$v\ldots = \frac{1}{RCL}\sum_{r,s,t} v_{rst}\,,\quad v_{i}\ldots = \frac{1}{CL}\sum_{s,t} v_{ist}\,,$$

etc. The third set gives the standard representation of the observation v_{ijk} in terms of its mean effect, row effect, etc.,:

$$m\ldots = v\ldots\,,$$

$$m_{i}\ldots = v_{i}\ldots - v\ldots\,, \qquad \text{and so on to}$$

$$m_{ijk} = v_{ijk} - v_{ij\cdot} - v_{i\cdot k} - v_{\cdot jk} + v_{i}\ldots + v_{\cdot j\cdot} + v_{\cdot\cdot k} - v\ldots\,.$$

It is easily seen that any one of these three sets can be expressed as linear functions of any other of the three, where the coefficients do not depend on the particular i, j and k involved. Thus

$$v\ldots = (1/RCL)(v_{(i)(j)(k)} + v_{i(j)(k)} + \cdots + v_{ijk}),$$

$$v_{i}\ldots = (1/CL)(v_{i(j)(k)} + v_{ij(k)} + v_{i(j)k} + v_{ijk}),$$

etc.

Now ave $\{f_\lambda(\mathbf{V}_{ijk})\}$ is evidently expressible as a certain linear combination of the v_{rst}. The symmetry under $\mathfrak{G}$ of the defining property of all random allocation schemes implies the equality of those coefficients in this linear combination corresponding to v_{rst} values comprising a particular sum of the eight sums $(v_{(i)(j)(k)}, v_{i(j)(k)}, \cdots, v_{ijk})$. Hence there are just eight different coefficients. The symmetry further implies that these eight coefficients do not depend on i, j and k. Going over to m-quantities, it follows that there exist constants $\gamma_M, \gamma_R, \cdots, \gamma_{RCL}$ independent of i, j and k such that

$$\text{ave}\,\{f_\lambda(\mathbf{V}_{ijk})\} = \gamma_M m\ldots + \gamma_R m_{i}\ldots + \cdots + \gamma_{RCL} m_{ijk}\,.$$

By taking linear combinations of both sides it follows that

$$\text{ave}\,\{f_\lambda(\mathbf{V}_\Delta)\} = \gamma_\Delta f_t(\mathbf{V}_\Delta)$$

for any $\mathbf{V}_\Delta \,\varepsilon\, E_\Delta$.

Since $f_\lambda(\mathbf{V}_\Delta) = f_t(\mathbf{V})$ for that $\mathbf{V} \,\varepsilon\, E_p$ which is λ-nearest to $\mathbf{V}_\Delta$, the contribution to $\gamma_\Delta f_t(\mathbf{V}_\Delta)$ from any particular E_p must be a positive or zero multiple of $f_t(\mathbf{V}_\Delta)$. Thus $\gamma_\Delta = 0$ only if all E_p are λ-orthogonal to $\mathbf{V}_\Delta$, which is in turn possible only if all $\mathbf{V}_{ijk}$ are λ-orthogonal to $\mathbf{V}_\Delta$. Since $\mathbf{V}_\Delta$ is a linear combination of $\mathbf{V}_{ijk}$ this last cannot happen unless the λ-metric degenerates e.g., by some λ-values becoming zero or infinite. This completes the proof of the theorem.

In degenerate cases where $\gamma_\Delta = 0$ for one or several Δ our estimators are undefined. Such cases can be of practical interest. In a later paper [4] we shall discuss our estimators when several λ_Δ tend to infinity and relate our estimators to well-known least squares estimators. The non-existence of our estimators

can in these cases be related to the inability to estimate by least squares more parameters than there are observations.

5. Simple examples. No mention has yet been made of the difficulty of computation of general estimators of the kinds just defined. To compute f_λ for a given set of observations and a given arbitrary λ-metric in general requires that the n observations be orthogonalized with respect to the λ-metric. Further, no general formula has been found for the correction factors γ_Δ for unbiasedness, which are different for different random allocation schemes and, as the theory now stands, must be computed directly for each scheme. In this section we illustrate the computation of estimates in the case of simple random allocation designs with $n = 2$ and $n = 3$ from an underlying array with two factors at two levels each, i.e., $N = 2^2$. The complexity of the general method is such that even these very simple examples are not quite trivial, and so they serve to illustrate the definitions of Section 4. Our objective is to find expressions for the class 1 and class 2 estimators and for their variances.

For the underlying 2^2 array, the four fixed numbers $v_{ij} (i = 1, 2 \text{ and } j = 1, 2)$ define f_t for the four unit orthogonal basis vectors $\mathbf{V}_{ij}$ of E. The vectors

$$\mathbf{V}_M = \tfrac{1}{2}(\mathbf{V}_{11} + \mathbf{V}_{12} + \mathbf{V}_{21} + \mathbf{V}_{22})$$

$$\mathbf{V}_R = \tfrac{1}{2}(\mathbf{V}_{11} + \mathbf{V}_{12} - \mathbf{V}_{21} - \mathbf{V}_{22})$$

$$\mathbf{V}_C = \tfrac{1}{2}(\mathbf{V}_{11} - \mathbf{V}_{12} + \mathbf{V}_{21} - \mathbf{V}_{22})$$

$$\mathbf{V}_{RC} = \tfrac{1}{2}(\mathbf{V}_{11} - \mathbf{V}_{12} - \mathbf{V}_{21} + \mathbf{V}_{22})$$

define unit vectors in the directions of mean effect, row effect, column effect and row-column interaction effect. These directions define the one-spaces E_M, E_R, E_C and E_{RC} respectively. The corresponding f_t values v_M, v_R, v_C and v_{RC} can be similarly expressed in terms of the v_{ij}. We shall denote class 1 and class 2 estimators of v_Δ by $\hat{v}_\Delta$ and $\hat{\hat{v}}_\Delta$, respectively, for each Δ.

Take first the case where the observations are a random sample of three from the four possible cells. Here there are four possible samples of three, each with probability one quarter. Since all four possible samples belong to one symmetry class, class 1 and class 2 estimators are the same. Suppose λ-metric (λ_M, λ_R, λ_C, λ_{RC}) is adopted and we set out to find $f_\lambda(\mathbf{V}_\Delta)$ using the second characterization of f_λ given in Section 4.2. Thus we seek $\mathbf{V} \varepsilon E_p$ λ-nearest to $\mathbf{V}_\Delta$. This need only be done for one choice of Δ, say R, and one sample of three, say v_{11}, v_{12} and v_{21}, since it will follow by symmetry for other choices. Let us therefore find the vector $a_{11}\mathbf{V}_{11} + a_{12}\mathbf{V}_{12} + a_{21}(-\mathbf{V}_{21})$ subject to the restriction $a_{11} + a_{12} + a_{21} = 1$ which makes the smallest λ-angle with $\mathbf{V}_R$. Each of $\mathbf{V}_{11}$, $\mathbf{V}_{12}$ and $-\mathbf{V}_{21}$ has the same λ-inner product with $\mathbf{V}_R$ and hence this same inner product is shared with $a_{11}\mathbf{V}_{11} + a_{12}\mathbf{V}_{12} - a_{21}\mathbf{V}_{21}$ where $a_{11} + a_{12} + a_{21} = 1$. Thus to minimize the λ-angle of this vector with $\mathbf{V}_R$ we need only minimize its λ-length. Its (λ-length)^2 is found from summing the properly weighted (component)^2 along $\mathbf{V}_M$, $\mathbf{V}_R$, $\mathbf{V}_C$

and $\mathbf{V}_{RC}$, to be

$$\lambda_M^2(a_{11}+a_{12}-a_{21})^2+\lambda_R^2(a_{11}+a_{12}+a_{21})^2 + \lambda_C^2(a_{11}-a_{12}-a_{21})^2+\lambda_{RC}^2(a_{11}-a_{12}+a_{21})^2.$$

It is easily seen that this expression is minimized when

$$a_{11}=K\left(\frac{1}{\lambda_M^2}+\frac{1}{\lambda_{RC}^2}\right),\ a_{12}=K\left(\frac{1}{\lambda_M^2}+\frac{1}{\lambda_C^2}\right),\ a_{21}=K\left(\frac{1}{\lambda_C^2}+\frac{1}{\lambda_{RC}^2}\right),$$

where K is a normalizing constant. By symmetry we could now write down the coefficients for each of the other three possible samples, with the same normalizing constant K. In order to define the class 1 or class 2 estimator $\hat{v}_R$ it is only necessary to choose factor K to produce unbiasedness. By direct algebra this K is found to be

$$K=\left(\frac{1}{\lambda_M^2}+\frac{1}{\lambda_C^2}+\frac{1}{\lambda_{RC}^2}\right)^{-1}$$

so that

$$\hat{v}_R=\left(\frac{1}{\lambda_M^2}+\frac{1}{\lambda_C^2}+\frac{1}{\lambda_{RC}^2}\right)^{-1}\left[\left(\frac{1}{\lambda_M^2}+\frac{1}{\lambda_{RC}^2}\right)v_{11} + \left(\frac{1}{\lambda_M^2}+\frac{1}{\lambda_C^2}\right)v_{12}-\left(\frac{1}{\lambda_C^2}+\frac{1}{\lambda_{RC}^2}\right)v_{21}\right]$$

when v_{11}, v_{12}, and v_{21} are observed. Similar expressions could be written for the other three possible samples. Also by direct algebra we find

$$\operatorname{var}(\hat{v}_R)=\left(\frac{v_M^2}{\lambda_M^4}+\frac{v_C^2}{\lambda_C^4}+\frac{v_{RC}^2}{\lambda_{RC}^4}\right)\Big/\left[\left(\frac{1}{\lambda_M^2}+\frac{1}{\lambda_C^2}+\frac{1}{\lambda_{RC}^2}\right)^2\right].$$

The various symmetries permit deduction of the corresponding estimators $\hat{v}_M$, $\hat{v}_C$ and $\hat{v}_{RC}$. For example, formulas for the case of v_{11}, v_{12} and v_{21} observed are

$$\hat{v}_M=\left(\frac{1}{\lambda_R^2}+\frac{1}{\lambda_C^2}+\frac{1}{\lambda_{RC}^2}\right)^{-1}\left[\left(\frac{1}{\lambda_R^2}+\frac{1}{\lambda_C^2}\right)v_{11}+\left(\frac{1}{\lambda_R^2}+\frac{1}{\lambda_{RC}^2}\right)v_{12}+\left(\frac{1}{\lambda_C^2}+\frac{1}{\lambda_{RC}^2}\right)v_{21}\right],$$

$$\hat{v}_C=\left(\frac{1}{\lambda_M^2}+\frac{1}{\lambda_R^2}+\frac{1}{\lambda_{RC}^2}\right)^{-1}\left[\left(\frac{1}{\lambda_M^2}+\frac{1}{\lambda_{RC}^2}\right)v_{11}-\left(\frac{1}{\lambda_R^2}+\frac{1}{\lambda_{RC}^2}\right)v_{12}+\left(\frac{1}{\lambda_M^2}+\frac{1}{\lambda_R^2}\right)v_{21}\right]$$

and

$$\hat{v}_{RC}=\left(\frac{1}{\lambda_M^2}+\frac{1}{\lambda_R^2}+\frac{1}{\lambda_C^2}\right)^{-1}\left[\left(\frac{1}{\lambda_R^2}+\frac{1}{\lambda_C^2}\right)v_{11}-\left(\frac{1}{\lambda_M^2}+\frac{1}{\lambda_C^2}\right)v_{12}-\left(\frac{1}{\lambda_M^2}+\frac{1}{\lambda_R^2}\right)v_{21}\right].$$

Corresponding formulas for var $(\hat{v}_M)$, var $(\hat{v}_C)$ and var $(\hat{v}_{RC})$ could be immediately written down.

If var $(\hat{v}_R)$ is minimized over different choices of λ-metric $(\lambda_M, \cdots, \lambda_{RC})$ it is

easily seen that the resulting minimum is

$$\left(\frac{1}{v_M^2}+\frac{1}{v_C^2}+\frac{1}{v_{RC}^2}\right)^{-1}$$

occurring when

$$\frac{|v_M|}{\lambda_M}=\frac{|v_C|}{\lambda_C}=\frac{|v_{RC}|}{\lambda_{RC}}.$$

Thus, by symmetry, the choice of a λ-metric such that

$$\frac{|v_M|}{\lambda_M}=\frac{|v_R|}{\lambda_R}=\frac{|v_C|}{\lambda_C}=\frac{|v_{RC}|}{\lambda_{RC}}$$

minimizes var $(\hat{v}_\Delta)$ for all Δ. This result agrees with the general theory of Section 6 and the heuristic feeling of Section 4.1. It is to be noted that this optimum is not available as a practical method since the choice of λ-metric depends on unknown quantities.

To illustrate the difference between class 1 and class 2 estimators we discuss the case of samples of two from the 2^2 array. Here, in the case of simple random allocation, there are six equiprobable pairs of observations defining three symmetry classes of size two each, say G_1, G_2 and G_3:

$$G_1: (v_{11}, v_{12}) \quad \text{and} \quad (v_{21}, v_{22})$$

$$G_2: (v_{11}, v_{21}) \quad \text{and} \quad (v_{12}, v_{22})$$

$$G_3: (v_{11}, v_{22}) \quad \text{and} \quad (v_{12}, v_{21}).$$

If a λ-metric $(\lambda_M, \cdots, \lambda_{RC})$ is selected and we seek a $\mathbf{V} \varepsilon E_p$ making minimum λ-angle with $\mathbf{V}_R$ we find that in the six cases such vectors are:

$$\mathbf{V}_{11}+\mathbf{V}_{12}=\mathbf{V}_R+\mathbf{V}_M,\ -\mathbf{V}_{21}-\mathbf{V}_{22}=\mathbf{V}_R-\mathbf{V}_M$$

$$\mathbf{V}_{11}-\mathbf{V}_{21}=\mathbf{V}_R+\mathbf{V}_{RC},\ \mathbf{V}_{12}-\mathbf{V}_{22}=\mathbf{V}_R-\mathbf{V}_{RC}$$

$$\mathbf{V}_{11}-\mathbf{V}_{22}=\mathbf{V}_R+\mathbf{V}_C,\ \mathbf{V}_{12}-\mathbf{V}_{21}=\mathbf{V}_R-\mathbf{V}_C.$$

In any scheme of weighting these to produce an unbiased estimator of v_R it is evidently necessary to weight the members of each pair of a symmetry class equally, so the question is how to choose three coefficients α, β and γ where the resulting estimator is:

$$\alpha(v_{11}+v_{12})=\alpha(v_R+v_M), \qquad \alpha(-v_{21}-v_{22})=\alpha(v_R-v_M)$$

$$\beta(v_{11}-v_{21})=\beta(v_R+v_{RC}), \qquad \beta(v_{12}-v_{22})=\beta(v_R-v_{RC})$$

$$\gamma(v_{11}-v_{22})=\gamma(v_R+v_C), \qquad \gamma(v_{12}-v_{21})=\gamma(v_R-v_C)$$

which, since each sample has probability 1/6, yields a U_2-unbiased estimator whenever $\alpha+\beta+\gamma=3$. The variance of this estimator is given by

$$\tfrac{1}{3}[\alpha^2(v_R^2+v_M^2)+\beta^2(v_R^2+v_{RC}^2)+\gamma^2(v_R^2+v_C^2)]-v_R^2.$$

For the class 2 estimator it is easily seen that

$$\frac{\alpha}{\lambda_R^2 + \lambda_M^2} = \frac{\beta}{\lambda_R^2 + \lambda_{RC}^2} = \frac{\gamma}{\lambda_R^2 + \lambda_C^2}$$

whence, using $\alpha + \beta + \gamma = 3$, the class 2 estimator is defined. Also, if the above general expression for the variance is minimized over choices of α, β and γ it is seen that the minimum variance occurs for

$$\frac{\alpha}{v_R^2 + v_M^2} = \frac{\beta}{v_R^2 + v_{RC}^2} = \frac{\gamma}{v_R^2 + v_C^2}$$

indicating again that the metric with $| v_\Delta |/\lambda_\Delta$ constant produces optimum estimators.

For U_1-unbiasedness it is necessary that $\alpha = \beta = \gamma = 1$ and so this choice produces the class 1 estimator. Note that this estimator happens to be independent of the λ-metric. Its variance is $\frac{1}{3}(v_M^2 + v_C^2 + v_{RC}^2)$. Since the class 1 estimator is conditionally unbiased given the symmetry class it makes sense to quote the conditional variances given the symmetry class, namely v_M^2, v_{RC}^2 or v_C^2 given symmetry class G_1, G_2 or G_3 respectively.

The formulas for these estimators and their variances are not as simple as one might expect from the simplicity of the examples. To give some feeling for

TABLE 1

Variances for the numerical example $N = 4$, $v_{11} = 8.5$, $v_{12} = .5$, $v_{21} = 2.5$ and $v_{22} = -1.5$: class 2 estimators for $n = 2$ and $n = 3$ and for various λ-metrics.

Metric	n	var $(\hat{v}_M)$	var $(\hat{v}_R)$	var $(\hat{v}_C)$	var $(\hat{v}_{RC})$
$\lambda_M = \lambda_R = \lambda_C$	2	18.67	21.67	15.00	25.67
$= \lambda_{RC} = c$	3	6.22	7.22	5.00	8.56
$\lambda_M \to \infty, \lambda_R = \lambda_C$	2	18.67	38.00	33.00	41.00
$= \lambda_{RC} = c$	3	6.22	10.00	5.00	13.00
$\lambda_M = \lambda_R \to \infty$	2	23.12	23.00	84.00	116.00
$\lambda_C = \lambda_{RC} = c$	3	10.00	10.00	4.00	36.00
$\lambda_M = \lambda_C \to \infty$	2	15.92	44.00	15.00	56.00
$\lambda_R = \lambda_{RC} = c$	3	5.00	4.00	5.00	16.00
$\lambda_M = \lambda_R = \lambda_C \to \infty$	2	15.88	16.44	15.15	25.67
$\lambda_{RC} = c$	3	4.00	4.00	4.00	8.56
$\lambda_M/5 = \lambda_R/4 =$	2	14.86	16.04	13.49	23.40
$\lambda_C/6 = \lambda_{RC}/2 = c$	3	2.94	3.14	2.84	7.67

TABLE 2

Variances for the numerical example $N = 4$, $v_{11} = 8.5$, $v_{12} = .5$, $v_{21} = 2.5$ *and* $v_{22} = -1.5$: *class* 1 *estimators for* $n = 2$ *and for any* λ*-metric*

Conditioning	var $(\hat{v}_M)$	var $(\hat{v}_R)$	var $(\hat{v}_C)$	var $(\hat{v}_{RC})$
Conditional on G_1	16.00	25.00	4.00	36.00
Conditional on G_2	36.00	4.00	25.00	16.00
Conditional on G_3	4.00	36.00	16.00	25.00
Unconditional	18.76	21.67	15.00	25.67

actual numbers produced by these methods consider the numerical example where $v_M = 5$, $v_R = 4$, $v_C = 6$ and $v_{RC} = 2$, i.e., the basic 2^2 array where

$$v_{11} = 8.5, \qquad v_{12} = .5, \qquad v_{21} = 2.5$$

and $v_{22} = -1.5$. Variances pertaining to this example are presented in Tables 1 and 2.

Table 1 illustrates how variances corresponding to various λ-metrics which might be used in practice are related to variances corresponding to the unknown optimum λ-metric $\lambda_M/5 = \lambda_R/4 = \lambda_C/6 = \lambda_{RC}/2$. A comparison of Table 1 and Table 2 shows that, unconditionally, var $(\hat{v}_\Delta) >$ var $(\hat{\hat{v}}_\Delta)$ for the optimum λ-metric, but that this inequality does not always hold for different λ-metrics. The inequality illustrates the general fact that the minimum variance of estimators in the class of U_2-unbiased estimators must be less than or equal to the minimum variance of estimators in the more restricted class of U_1-unbiased estimators.

6. Optimum properties. In this section we define a wide class of linear unbiased estimators of $f_t(\mathbf{V})$ for any random allocation scheme and show that certain estimators of the types defined in Section 4 are optimum in the wide class. The criteria of optimality are minimum symmetrized variance among U_2- or U_1-unbiased estimators and minimum symmetrized mean square error among all estimators. These "symmetrized" criteria will be defined shortly.

Suppose the random allocation scheme permits exactly d distinct n-spaces E_p each with positive probability. Consider nd arbitrary real numbers to be used as d sets of n coefficients applicable to the n observations corresponding to each possible E_p. The resulting linear combinations define the values of a random variable whose random properties are induced by the random choice of E_p. As the nd coefficients assume all real values they define an nd-dimensional vector space O of random variables. Any $v^* \varepsilon O$ has for its average some linear combination of the N quantities v_{ijk} of the underlying array, i.e.,

$$\text{ave}\{v^*\} = f_t(\mathbf{V}) \quad \text{for some} \quad \mathbf{V} \varepsilon E,$$

where averaging is over the complete random choice of designs, and so v^* is a U_2-unbiased estimator of $f_t(\mathbf{V})$. For any given $\mathbf{V} \varepsilon E$ we consider as our general

class of linear U_2-unbiased estimators of $f_t(\mathbf{V})$ all those $v^* \varepsilon O$ satisfying

$$\text{ave}\,\{v^*\} = f_t(\mathbf{V}),$$

and we seek the optimum in this general class. Denote this subset of O by $O(\mathbf{V})$.

One might first decide to seek that $v^* \varepsilon O(\mathbf{V})$ with minimum variance. Unfortunately the class $O(\mathbf{V})$ is sufficiently large to include estimators which have variance zero but which are of an uninteresting and asymmetrical type. Note that in general the variance of an estimator is a quadratic function of the v_{ijk} and what we are doing is finding estimators which are unbiased for any v_{ijk} but whose variance is minimum for a particular set of v_{ijk} (i.e., the "true" values). For example, consider a simple random allocation scheme observing n of the N quantities v_{ijk}. Let us now define an unbiased estimator of, say, v_{111} which will have variance zero when the v_{ijk} are in fact equal to a set of numbers x_{ijk}. The coefficients in this estimator will depend of course on the x_{ijk}. Suppose for simplicity that all the x_{ijk} are non-zero, and define $z_{ijk} = v_{ijk}/x_{ijk}$. Define random variable $\bar{z}$ to be the mean of the n observed values of v_{ijk}. Define random variable

$$\begin{aligned} \tilde{z}_{111} &= \frac{N-1}{n-1} z_{111} - \frac{N-n}{n-1} \bar{z} && \text{if } v_{111} \text{ is observed} \\ &= \bar{z} && \text{otherwise.} \end{aligned}$$

Then it may be easily checked that $x_{111}\tilde{z}_{111}$ is U_2-unbiased for v_{111} and has zero variance when $v_{ijk} = x_{ijk}$ for all i, j and k. Similarly we can define zero variance U_2-unbiased estimators for any v_{ijk} and thence for $f_t(\mathbf{V})$ for any $\mathbf{V} \varepsilon E$.

In order to have a more interesting optimum v^* we define the criterion of symmetrized variance of an estimator v^*. Symmetry here refers to symmetry under the group $\mathfrak{G}$ of $P = (R!)(C!)(L!)$ permutations of the R rows, C columns and L layers of the basic array. For any $g \varepsilon \mathfrak{G}$, any $\mathbf{V} \varepsilon E$ and any n-space E_p contained in E we denote by $g(\mathbf{V})$ the vector in E found by operating with g on $\mathbf{V}$ and we denote by $g(E_p)$ the subspace of E found by operating with g on E_p. If $v^* \varepsilon O$ is a U_2-unbiased estimator of $f_t(\mathbf{V})$ and $v^* = f_t(\mathbf{V}^*)$ where $\mathbf{V}^* \varepsilon E_p$, then for any $g \varepsilon \mathfrak{G}$ we can define $g(v^*) \varepsilon O$, a U_2-unbiased estimator of $f_t(g(\mathbf{V}))$, to be $g(v^*) = f_t(g(\mathbf{V}^*))$ when $g(E_p)$ is the subspace corresponding to the observations. Then we define the *symmetrized variance* of v^* to be

$$\rho^2(v^*) = \frac{1}{P} \sum_g \text{var}\,\{g(v^*)\}.$$

Thus the group $\mathfrak{G}$ breaks the space O into mutually exclusive symmetry classes of estimators such that the estimators of one class share a common symmetrized variance, namely the mean variance over the class. The use of the criterion $\rho^2(v^*)$ seems reasonable when one's *a priori* beliefs about the array of v_{ijk} are symmetrical under $\mathfrak{G}$. For, if v^* were adopted as the estimator of $f_t(\mathbf{V})$, then it would be only reasonable to adopt $g(v^*)$ as the estimator of $f_t(g(\mathbf{V}))$ and to judge all such estimators together by their mean variance. For similar reasons

we may wish to use the criterion of *symmetrized mean square error* of v^* defined analogously as

$$\frac{1}{P}\sum_{g} \text{ave } \{[g(v^*) - f_t(g(\mathbf{V}))]^2\}.$$

The following optimality Theorems 6.1, 6.2 and 6.3 will be proved in order at the end of this section. Suppose we select any λ-metric such that

$$\frac{\lambda_M^2}{(MS)_M} = \frac{\lambda_R^2}{(MS)_R} = \cdots = \frac{\lambda_{RCL}^2}{(MS)_{RCL}}$$

where the $(MS)_\Delta$ are the mean squares resulting from the analysis of variance of the complete array of v_{ijk}, and suppose we refer to this λ-metric as the *optimum λ-metric.*

THEOREM 6.1. *For any random allocation scheme and any $\mathbf{V} \varepsilon E$, that $v^* \varepsilon O$ with minimum symmetrized mean square error for $f_t(\mathbf{V})$ is given by the λ-minimum extension $f_\lambda(\mathbf{V})$ corresponding to the optimum λ-metric. The symmetrized mean square error of $f_\lambda(\mathbf{V})$ is minimum both unconditionally and conditionally with conditioning on each symmetry class G of designs.*

THEOREM 6.2. *For any random allocation scheme and any $\mathbf{V} \varepsilon E$, that $v^* \varepsilon O$ which is U_2-unbiased for $f_t(\mathbf{V})$ with minimum symmetrized variance is given by the class 2 estimator $\hat{f}_\lambda(\mathbf{V})$ corresponding to the optimum λ-metric.*

THEOREM 6.3. *For any random allocation scheme and any $\mathbf{V} \varepsilon E$, that $v^* \varepsilon O$ which is U_1-unbiased for $f_t(\mathbf{V})$ with minimum symmetrized variance is given by the class 1 estimator $\hat{f}_\lambda(\mathbf{V})$ corresponding to the optimum λ-metric. The symmetrized variance of $\hat{f}_\lambda(\mathbf{V})$ is minimum both unconditionally and conditionally with conditioning on each symmetry class G of designs.*

Note that in these theorems it is the same λ-minimum extension or class 2 estimator or class 1 estimator (i.e., the λ-minimum extension or class 2 estimator or class 1 estimator corresponding to the same λ-metric) which is optimum for all $\mathbf{V} \varepsilon E$. Note also that in each case the optimum estimator depends on the "true" underlying v_{ijk} through its choice of optimum λ-metric.

In the case of underlying arrays with factors at two levels each a slightly different corollary of each theorem can be stated. Here the subspaces E_Δ of E are one-dimensional and the principal aims are to estimate the $v_\Delta = f_t(\mathbf{V}_\Delta)$ where $\mathbf{V}_\Delta \varepsilon E_\Delta$. If v^* is unbiased for v_Δ then $g(v^*)$ is unbiased for $\pm v_\Delta$ for all $g \varepsilon \mathfrak{G}$, and so it is natural to consider only those $v^* \varepsilon O$ which are identical with $\pm g(v^*)$ as possible estimators. For such estimators variance and symmetrized variance are the same, and mean square error and symmetrized mean square error are the same. Thus, for example, as a corollary to Theorem 6.2 we have that, *in the case of factors at two levels each* (i.e., $N = 2^k$), *among all estimators $v^* \varepsilon O$ which are U_2-unbiased for v_Δ and symmetric in the sense that $v^* = \pm g(v^*)$ for all $g \varepsilon \mathfrak{G}$, the class 2 estimator $\hat{f}_\lambda(\mathbf{V}_\Delta)$ corresponding to the optimum λ-metric has minimum variance.* Similar corollaries clearly hold for Theorems 6.1 and 6.3.

We now prove Theorem 6.1. Suppose $\mathbf{W}_1, \cdots, \mathbf{W}_N$ is a set of basis vectors

of E which are unit orthogonal in the sense of the formal metric and where $\mathbf{W}_1$ spans E_M, $\mathbf{W}_2$ up to $\mathbf{W}_R$ span E_R, $\mathbf{W}_{R+1}$ up to $\mathbf{W}_{R+C-1}$ span E_C, and so on. Suppose we set out to estimate $f_t(\mathbf{V})$ where

$$\mathbf{V} = \sum_{i=1}^{N} \beta_i \mathbf{W}_i$$

and for a particular E_p suppose the estimator is

$$v^* = f_t(\mathbf{V}^*) \quad \text{where} \quad \mathbf{V}^* = \sum_{i=1}^{N} \alpha_i \mathbf{W}_i$$

and where $\mathbf{V}^* \varepsilon E_p$. We should like to minimize the contribution to the symmetrized mean square error from the symmetry class G of E_p by choosing $\mathbf{V}^* \varepsilon E_p$ to minimize

$$\frac{1}{P} \sum_{g} \left[\sum_{i=1}^{N} (\alpha_i - \beta_i) g(w_i) \right]^2$$

where $g(w_i)$ denotes $f_t(g(\mathbf{W}_i))$. Suppose $(MS)_\Delta$ is the mean square associated with E_Δ in the analysis of variance of the complete array of v_{ijk}, and suppose $(MS)_i$ is defined to be $(MS)_\Delta$ for the Δ such that $\mathbf{W}_i \varepsilon E_\Delta$. Then the desired result will follow if we show the last expression to be equal to

$$\frac{1}{P} \sum_{i=1}^{N} (\alpha_i - \beta_i)^2 (MS)_i,$$

for clearly this expression is minimized by choosing $\mathbf{V}^*$ to be that vector in E_p λ-nearest to $\mathbf{V}$ in the sense of the optimum λ-metric, and this amounts to choosing $v^* = f_\lambda(\mathbf{V})$ for the optimum λ-metric.

In order to prove the desired equality we need only show that

$$\frac{1}{P} \sum_{g} g(w_i) g(w_j) = 0 \qquad \text{for } i \neq j$$

and

$$\frac{1}{P} \sum_{g} [g(w_i)]^2 = (MS)_i.$$

One way of regarding this problem is to suppose that the fixed array v_{ijk} is made into a random array by choosing at random with equal probabilities an element $g \varepsilon \mathfrak{G}$ and applying g to the array. Under this scheme we are looking for average squares and average cross-products for the set of degrees of freedom corresponding to $\mathbf{W}_1, \cdots, \mathbf{W}_N$. The first equality for $\mathbf{W}_i$ and $\mathbf{W}_j$ in different subspaces E_Δ is easily seen directly; for example if $\mathbf{W}_i \varepsilon E_R$ and $\mathbf{W}_j \varepsilon E_{RCL}$ then summation over those elements of $\mathfrak{G}$ which leave rows unchanged is summation for which $g(w_i)$ is constant and hence this summation over $g(w_i)g(w_j)$ is zero by the well-known property that triple interactions sum to zero when summed over any of their indices. The remaining equalities are best shown indirectly.

Suppose $\mathbf{W}_i$ and $\mathbf{W}_j$ both belong to E_Δ. The left hand sides of both of the above sums are symmetric quadratic expressions and so can be expressed as linear combinations of $(MS)_M, \cdots, (MS)_{RCL}$. But the left hand sides are clearly unaffected by changes in any of these except $(MS)_\Delta$ so that the right hand sides must in each case be a constant times $(MS)_\Delta$. By supposing the v_{ijk} to be N independent $N(0, 1)$ variables and by averaging both sides over this normal variation we deduce that the constants are as shown. This completes the proof of Theorem 6.1.

To deduce Theorem 6.2 from Theorem 6.1 we need some relations between symmetrized variance and symmetrized mean square error. Suppose $v^* \varepsilon O(\mathbf{V})$, i.e., $v^* \varepsilon O$ and is U_2-unbiased for $f_t(\mathbf{V})$. Suppose v^* has symmetrized variance $\rho^2(v^*)$. We may define the symmetrized squared mean of v^* to be

$$\mu^2(v^*) = \frac{1}{P}\sum_g [\text{ave } \{g(v^*)\}]^2 = \frac{1}{P}\sum_g [f_t(g(\mathbf{V}))]^2$$

and so the symmetrized mean square error of v^* is

$$\frac{1}{P}\sum_g \text{ave } \{[g(v^*) - f_t(g(\mathbf{V}))]^2\} = \rho^2(v^*) + \mu^2(v^*).$$

Among the statistics $k\,v^*$ for different k suppose v^{**} is the one with minimum symmetrized mean square deviation from $f_t(\mathbf{V})$. It is easily seen that

$$v^{**} = \{[\mu^2(v^*)]/[\mu^2(v^*) + \rho^2(v^*)]\}v^*$$

with symmetrized mean square deviation from $f_t(\mathbf{V})$ given by

$$[2\mu^2(v^*)\rho^2(v^*)]/[\mu^2(v^*) + \rho^2(v^*)].$$

It may also be easily checked that v^* has minimum symmetrized variance in $O(\mathbf{V})$ if and only if the corresponding v^{**} has minimum symmetrized mean square error among estimators in O which are unbiased except for a constant factor. Thus in a sense it is immaterial whether we find the optimum v^* or the corresponding optimum v^{**}. Theorem 6.1 tells us that $f_\lambda(\mathbf{V})$ for the optimum λ-metric provides the minimum symmetrized mean square error estimator in O, and Section 4.3 tells us that, for $\mathbf{V}_\Delta \varepsilon E_\Delta$, $f_\lambda(\mathbf{V}_\Delta)$ is unbiased except for a constant factor. Thus, for any $\mathbf{V}_\Delta \varepsilon E_\Delta$, $f_\lambda(\mathbf{V}_\Delta)$ is the optimum v^{**} described above and hence the corresponding v^* is the class 2 estimator $\hat{f}_\lambda(\mathbf{V}_\Delta)$. This proves Theorem 6.2 for vectors $\mathbf{V}$ of the special type belonging to an E_Δ for some Δ.

To complete the proof we need only show that the minimum symmetrized variance estimator $v^* \varepsilon O(\mathbf{V})$ for any $\mathbf{V} \varepsilon E$ can be written

$$v^* = \sum_\Delta v^*_\Delta$$

where the corresponding $\mathbf{V}$ can be written

$$\mathbf{V} = \sum_\Delta \mathbf{V}_\Delta$$

with $\mathbf{V}_\Delta \varepsilon E_\Delta$ and where v_Δ^* is the minimum symmetrized variance U_2-unbiased estimator of $f_t(\mathbf{V}_\Delta)$. This gives the desired result since it is known that

$$\hat{f}_\lambda(\mathbf{V}) = \sum_\Delta \hat{f}_\lambda(\mathbf{V}_\Delta).$$

As may be easily checked, a Euclidean metric may be defined for vector space O by setting the squared length of $v^* \varepsilon O$ equal to its symmetrized variance $\rho^2(v^*)$. Define O_1 to be the subspace of O consisting of all $v^* \varepsilon O$ whose average is identically zero, and define O_2 to be the subspace of O orthogonal to O_1 according to the ρ-metric. Any $v^* \varepsilon O(\mathbf{V})$ can be written as $v^* = v_1^* + v_2^*$ where

$$v_i^* \varepsilon O_i (i = 1, 2).$$

Clearly $v_2^* \varepsilon O(\mathbf{V})$ and, since

$$\rho^2(v^*) = \rho^2(v_1^* + v_2^*) = \rho^2(v_1^*) + \rho^2(v_2^*) \geqq \rho^2(v_2^*),$$

v_2^* has minimum symmetrized variance in $O(\mathbf{V})$. If

$$\mathbf{V} = \sum_\Delta \mathbf{V}_\Delta$$

and if v_Δ^* is any element of $O(\mathbf{V}_\Delta)$ then

$$v^* = \sum_\Delta v_\Delta^* \varepsilon O(\mathbf{V}).$$

Also, if $v^* = v_1^* + v_2^*$ and $v_\Delta^* = v_{1\Delta}^* + v_{2\Delta}^*$ where v_i^* and $v_{i\Delta}^*$ belong to O_i $(i = 1, 2)$, then,

$$v_2^* = \sum_\Delta v_{2\Delta}^*$$

which is the desired result, completing the proof of Theorem 6.2.

Theorem 6.3 follows immediately from the application of Theorem 6.2 to random allocation schemes with just one symmetry class of designs.

7. Generalization of the theory for the basic model. In this section we suppose the observations v_{ijk} to be random such that $v_{ijk} = \nu_{ijk} + \epsilon_{ijk}$ where ν_{ijk} are constant and ave $\{\epsilon_{ijk}\} = 0$. We suppose the ϵ_{ijk} for all i, j and k to have arbitrary variances and covariances. The ϵ_{ijk} are assumed independent of the random choice of design. The same estimators used for the basic model can be considered for the generalized model, but when we compute their means and variances we will average over the randomness of the ϵ_{ijk} in addition to the randomness of the choice of design. For example, an estimator is now defined to be U_1-unbiased if it is unbiased under averaging over both ϵ_{ijk} and the random choice of E_p conditional on each symmetry class G of designs. U_2-unbiasedness is similarly defined omitting the conditioning provision.

The total functional $f_t(\mathbf{V})$ is now redefined in terms of ν_{ijk} rather than v_{ijk}. Thus

$$f_t\Big(\sum_{i,j,k} \alpha_{ijk}\mathbf{V}_{ijk}\Big) = \sum_{i,j,k} \alpha_{ijk}\nu_{ijk}.$$

Clearly any estimator which was U_2-unbiased for $f_t(\mathbf{V})$ in the basic model remains U_2-unbiased for $f_t(\mathbf{V})$, with the new definitions, under the generalized model. The same statement holds for U_1-unbiasedness. The optimality theorems of Section 6 remain valid along with their proofs, provided only that $(MS)_\Delta$ is replaced by ave $\{(MS)_\Delta\}$ where averaging is over the randomness induced by the ϵ_{ijk}. Thus the optimum λ-metric becomes a λ-metric such that

$$\lambda_\Delta^2/\text{ave}\,\{(MS)_\Delta\}$$

is constant for all Δ. It has now been shown that the entire theory given for the basic model generalizes with no gaps to the generalized model.

The generalized model covers a standard model I analysis of variance model, this being the case where the ϵ_{ijk} have common variance and zero covariance. An application with a more general set of variances and covariances would be as follows. Suppose our notion of a random allocation design were broadened to allow random replication of certain cells. We may suppose in this case our estimators of $f_t(\mathbf{V})$ to be based on cell means. If the basic observations have a model I structure then the cell means, with differing numbers of observations per cell, do not. However, if our random allocation scheme required those cells with no replication, those cells with one replicate, etc., each to have probability schemes symmetrical under $\mathfrak{G}$, then the cell means can evidently be treated as observations under the generalized model, and the theory applies.

REFERENCES

[1] F. J. Anscombe, "Quick analysis methods for random balance screening experiments," *Technometrics*, Vol. 1 (1959), pp. 195–209.

[2] S. Barbacki and R. A. Fisher, "A test of the supposed precision of systematic arrangements," *Ann. Eugenics*, Vol. 7 (1936), pp. 189–193.

[3] T. A. Budne, "The application of random balance designs," *Technometrics*, Vol. 1 (1959), pp. 139–155.

[4] A. P. Dempster, "Random allocation designs, II: approximate theory for simple random allocation," submitted for publication in the *Ann. Math. Stat.*

[5] R. A. Fisher, "Uncertain inference," *Proc. Amer. Acad. Arts and Sciences*, Vol. 71, No. 4 (1936), pp. 245–258, (reprinted as paper no. 27 in R. A. Fisher, *Contributions to Mathematical Statistics*, John Wiley and Sons, New York, 1950).

[6] F. E. Satterthwaite, "Random balance experimentation," *Technometrics*, Vol. 1 (1959), pp. 111–137.

[7] P. R. Halmos, *Finite Dimensional Vector Spaces*, D. Van Nostrand, Princeton, 1958.

[8] "Student," "Comparison between balanced and random arrangements of field plots," *Biometrika*, Vol. 29 (1937), pp. 363–379.

[9] M. B. Wilk and Oscar Kempthorne, "Some aspects of the analysis of factorial experiments in a completely randomized design," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 950–985.

[10] W. J. Youden, Oscar Kempthorne, J. W. Tukey, G. E. P. Box and J. S. Hunter, "Discussion of the papers of Messrs. Satterthwaite and Budne," *Technometrics*, Vol. 1 (1959), pp. 157–193.