RANDOM ALLOCATION DESIGNS II: APPROXIMATE THEORY FOR SIMPLE RANDOM ALLOCATION[1]

By A. P. Dempster

Harvard University

1. Introduction.

1.1. *Aims.* In Section 2 of a previous paper [1] a viewpoint was described under which one can compare within one framework a wide class of randomized experimental designs, namely those designs called in [1] random allocation designs. To make such comparisons one needs to know how each design performs, which, from our point of view, means finding the variances of linear unbiased estimators under the randomization hypothesis. The aim of this paper is to take a beginning step towards finding some variances analytically.

The previous paper [1] defined some very general classes of techniques of linear unbiased estimation. The practical use of these general methods is inhibited by two difficulties. Firstly, the computations with data are laborious and unfamiliar, and, secondly, the calculation of the variances of the estimators presents formidable mathematical difficulties. In this paper we avoid the first difficulty by considering a smaller class of estimators, within the general class, consisting only of estimators which lead to familiar data computations. The mathematical difficulties remain however, so that the calculation of variances is attempted only by indirect approximate methods which apply only to *simple random allocation* designs as defined in [1]. The restriction to simple random allocation, although very stringent, does allow answers to some interesting questions, for this case is, in a sense, the most radical form of random balance design.

1.2. *The class of data analysis techniques.* We postulate data consisting of n quantities corresponding to a subset of n of the N cells of a complete crossed k-factor array, i.e.,

$$N = R \cdot C \cdots L,$$

where $R, C, \cdots, L$ are the numbers of levels of the first, second, $\cdots$, kth factors. In broad terms, the techniques under consideration have two stages, the first stage consisting of the least squares estimation of a selected set of effects, and the second stage consisting of further estimation and testing based on the residuals from the first stage.

When a mathematical statistician is confronted with an unbalanced fraction of a complete array, his first thought is to set up a linear model with various

Received December 2, 1959; revised September 26, 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.

selected main effect and interaction terms plus random errors, and then to estimate these effects by least squares. It is assumed that the reader knows how to do this (e.g., by setting up the so-called normal equations and solving them as in Wilks [4] p. 192). This is precisely the first stage of our analysis. We do not, however, commit ourself to the model which led to this first stage estimation, but proceed to look for further effects. In general, we denote by m the number of effects selected for fitting at the first stage, and we may choose any m such that $0 \leqq m \leqq n$. In practice one always chooses $m \geqq 1$, for one always fits at least the grand mean, and one generally allows $n - m$ to be moderately large in order to have some interesting variation left in the residuals.

The second stage of our analysis uses as input the n residuals after the first stage fitting. The simplest way to think of the second stage computations is to imagine a complete array of N cells whose n entries corresponding to the observed cells are the residuals from the first stage, and whose $N - n$ entries corresponding to the remaining cells are all zero. The second stage computations are simply the usual analysis of variance computations on this complete array of size N, i.e., the usual linear estimators and the usual mean squares. Of course, the linear estimators and mean squares corresponding to effects estimated at the first stage will be zero, and it is only the remaining effects, and not necessarily all of these, which are of interest at the second stage.

It is clear that the linear estimators calculated at the second stage are not unbiased, for the $N - n$ zeros in the array have the effect of reducing the absolute value of the average of such an estimator. Thus a correction factor is needed to make such a raw estimator into an unbiased estimator. The approximate theory of Section 2 suggests, at least for the case of simple random allocation, that an approximate correction factor for unbiasedness is

$$c = \frac{N - m}{n - m}. \tag{1.2.1}$$

It is also suggested, again for the case of simple random allocation, that ratios of mean squares coming from the second stage analysis of variance may be tested as F statistics with the same degrees of freedom as would be used if the full array of size N were observed. In effect this suggestion is saying that the F tests are sufficiently robust to be approximately valid when applied to non-normal data where the observations are zero with probability $1 - (n/N)$ and other quantities with probability n/N. This suggestion arises more precisely from the approximate theory of Section 2.

The foregoing is intended to be a verbal description of how to carry out data analysis for a general technique in the class of techniques under consideration. To be more concrete we take the example of a 3-factor design where $N = R \cdot C \cdot L$. The choice of a particular technique in the available class of techniques is made by deciding which effects to estimate by least squares. In our example let us decide to estimate by least squares the grand mean, the row main effects and the column main effects. Thus we observe a subset of n of the N quantities v_{ijk} for

$1 \leqq i \leqq R$, $1 \leqq j \leqq C$ and $1 \leqq k \leqq L$, and, as the first stage of analysis, we find $\hat{v}_{..}$, $\hat{v}_{i.}$ and $\hat{v}_{.j}$ which minimize the quantity

$$\sum (v_{ijk} - \hat{v}_{..} - \hat{v}_{i.} - \hat{v}_{.j})^2,$$

where summation is over the n observed cells. Because of linear constraints (e.g., we may require $\sum_1^R \hat{v}_{i.} = \sum_1^C \hat{v}_{.j} = 0$) we are here fitting $m = R + C - 1$ parameters. We shall assume for simplicity that all of the parameters are uniquely estimable. The first step in the second stage of analysis is to consider the array

$$Y_{ijk} = v_{ijk} - \hat{v}_{..} - \hat{v}_{i.} - \hat{v}_{.j} \qquad \text{if cell } (i, j, k) \text{ is observed}$$

$$= 0 \qquad \text{otherwise.}$$

From this array one computes in the usual way linear estimators $\hat{v}_{..k}$, $\hat{v}_{ij.}$, $\hat{v}_{i.k}$, $\hat{v}_{.jk}$ and $\hat{v}_{ijk}$, and also the corresponding mean squares $(MS)_L$, $(MS)_{RC}$, $(MS)_{RL}$, $(MS)_{CL}$ and $(MS)_{RCL}$. Here, for example,

$$\hat{v}_{..k} = \frac{1}{RC} \sum_{i=1}^{R} \sum_{j=1}^{C} Y_{ijk} - \frac{1}{RCL} \sum_{i=1}^{R} \sum_{j=1}^{C} \sum_{k=1}^{L} Y_{ijk}$$

$$= \frac{1}{RC} \sum_{i=1}^{R} \sum_{j=1}^{C} Y_{ijk},$$

and

$$(MS)_L = \frac{1}{L-1} \sum_{k=1}^{L} (\hat{v}_{..k})^2.$$

Supposing now that the n observations come from a simple random allocation scheme we multiply these second stage linear estimators by c from (1.2.1) to make them approximately unbiased, and we test the mean squares, for example by regarding $(MS)_L/(MS)_{RCL}$ as an F-statistic on $L - 1$ and $(R - 1)(C - 1)(L - 1)$ degrees of freedom.

1.3. *The theoretical approach.* The statistical model under consideration is as follows. Using 3-factor notation for simplicity, we suppose that, corresponding to each of the N cells (i, j, k) for $1 \leqq i \leqq R$, $1 \leqq j \leqq C$ and $1 \leqq k \leqq L$, there is a quantity v_{ijk} which is observed if an experiment is performed at levels (i, j, k). We also suppose that

$$v_{ijk} = \nu_{ijk} + e_{ijk} \tag{1.3.1}$$

where the ν_{ijk} are non-random quantities and the e_{ijk} are uncorrelated random variables with common mean zero and variance σ^2. We shall further assume that the e_{ijk} are normally distributed *only* when discussing F-ratios. In the case of simple random allocation we suppose that a simple random sample without replacement of n of the N quantities v_{ijk} is observed. Using these data and an arbitrary member of the class of techniques described above, we compute the

first and second stage linear estimators along with the second stage ratios of mean squares. Our basic objectives are to find the first and second moments of these linear estimators and the distributions of the ratios of mean squares. It should be emphasized that the randomness in these statistics arises from two sources, first, the discrete randomness induced by the randomization hypotheses, i.e. the random choice of the subset of n of the N cells, and second, the randomness induced by the e_{ijk}. These sources are assumed throughout to be statistically independent.

Unfortunately any attempt through mathematical analysis to meet the objectives posed runs into great difficulties. It just does not appear feasible to compute directly the distributions of the various statistics under the randomization hypothesis. For this reason the main mathematical results of this paper, presented in Section 2 and derived in Section 4, refer to a quite different statistical model underlying the randomness of the computed statistics. As statistical theory relating to a well-defined model these results stand on their own. On the other hand these results are intended as approximations to the corresponding results for the simple random allocation model. A completely satisfying discussion of the accuracy of these approximations is beyond the scope of this paper, and indeed seems feasible only by Monte Carlo methods. Thus any interpretations of the results, such as those given in Section 3, must be regarded as suggestions whose verification will require Monte Carlo methods.

The second or approximating model replaces the discrete randomization probability mechanism by an analogous continuous probability mechanism, and henceforth this model will be referred to as the *continuous analogue model.* The detailed definition of the continuous analogue model requires considerable care and is postponed to Section 4. Likewise the discussion of the motivation and partial justification of this model is deferred to Section 5.

2. Results.

2.1. *Formulas for means and variances.* This discussion will be carried out using the terminology and notation of Section 3 of [1] except that we shall now use general notation not restricted to a 3-factor case. Thus we suppose that the N factor level combinations are labeled 1 to N, and we define a Euclidean vector space E in terms of unit orthogonal basis vectors $\mathbf{V}_1, \mathbf{V}_2, \cdots, \mathbf{V}_N$ which are in correspondence with the N cells of the basic array. The cells have associated quantities

$$v_i = \nu_i + e_i \qquad \text{for} \quad 1 \leqq i \leqq N \tag{2.1.1}$$

where v_i is observed if the experiment corresponding to cell i is performed. Formula (2.1.1) is the same as (1.3.1) using different subscripts, so that the ν_i are fixed quantities and the e_i are uncorrelated random variables with means zero and variances all σ^2.

Our aim is to provide unbiased linear estimators for linear combinations of the ν_i, i.e., for quantities like $\sum_1^N c_i\nu_i$. These quantities may be regarded as the

values of a linear functional g_t over E defined by

$$g_t(\mathbf{V}) = \sum_{i=1}^{N} c_i \nu_i \tag{2.1.2}$$

where $\mathbf{V} = \sum_1^N c_i \mathbf{V}_i$ is any vector in E. Now the first stage of estimation, the least squares stage, estimates $g_t(\mathbf{V})$ where $\mathbf{V}$ belongs to a selected m-dimensional subspace of E which we shall denote by E_m. The second stage of estimation estimates $g_t(\mathbf{V})$ where $\mathbf{V}$ belongs to the $(N - m)$-dimensional subspace of E orthogonal to E_m, and this subspace we shall denote by $\tilde{E}_m$. It will be convenient to define $\mathbf{W}_1, \mathbf{W}_2, \cdots, \mathbf{W}_N$ to be an alternative unit orthogonal basis of E such that

$$\mathbf{W}_i \varepsilon E_m \qquad \text{for} \quad 1 \leqq i \leqq m,$$

and

$$\mathbf{W}_i \varepsilon \tilde{E}_m \qquad \text{for} \quad m + 1 \leqq i \leqq N,$$

and to denote $g_t(\mathbf{W}_i)$ by ω_i. Then the least squares stage of estimation provides an estimator $\hat{\omega}_a$ for any $\omega_a = \sum_1^m a_i \omega_i = g_t(\mathbf{W}_a)$, where $\mathbf{W}_a = \sum_1^m a_i \mathbf{W}_i$ is any vector in E_m. Similarly the second stage provides a raw (i.e., uncorrected for bias) estimator $\hat{\omega}_b$ for any $\omega_b = \sum_{m+1}^N b_i \omega_i = g_t(\mathbf{W}_b)$, where $\mathbf{W}_b = \sum_{m+1}^N b_i \mathbf{W}_i$ is any vector in $\tilde{E}_m$.

The following formulas are derived in Section 4.3 and are exact formulas relating to the continuous analogue model. Denote by A^2 and B^2 the quantities $\sum_1^m a_i^2$ and $\sum_{m+1}^N b_i^2$, and set

$$(M\Sigma)_{II} = \frac{1}{N - m} \sum_{i=m+1}^{N} \omega_i^2. \tag{2.1.3}$$

Then

$$\text{ave } \{\hat{\omega}_a\} = \omega_a, \tag{2.1.4}$$

$$\text{ave } \{(\hat{\omega}_a)^2\} = \omega_a^2 + \frac{N - n}{n - m - 1} A^2 (M\Sigma)_{II} + \frac{N - m - 1}{n - m - 1} A^2 \sigma^2, \tag{2.1.5}$$

$$\text{ave } \{\hat{\omega}_b\} = \frac{n - m}{N - m} \omega_b, \tag{2.1.6}$$

and

$$\begin{aligned} \text{ave } \{(\hat{\omega}_b)^2\} = \frac{n - m}{N - m - 1} \Bigg[& \left(\frac{n - m + 2}{N - m + 2} - \frac{1}{N - m} \right) \omega_b^2 \\ & + \left(1 - \frac{n - m + 2}{N - m + 2} \right) B^2 (M\Sigma)_{II} + \left(1 - \frac{1}{N - m} \right) B^2 \sigma^2 \Bigg]. \end{aligned} \tag{2.1.7}$$

Thus $\hat{\omega}_a$ is unbiased for ω_a with variance

$$\text{(2.1.8)} \qquad \text{var}\,(\hat{\omega}_a) = \frac{N-n}{n-m-1} A^2(M\Sigma)_{II} + \frac{N-m-1}{n-m-1} A^2\sigma^2,$$

and $(N-m)\hat{\omega}_b/(n-m)$ is unbiased for ω_b with variance

$$\text{(2.1.9)} \qquad \begin{aligned} \text{var}\left(\frac{N-m}{n-m}\hat{\omega}_b\right) &= \frac{N-m}{N-m+2} \\ &\cdot \frac{N-n}{n-m}\left[\frac{N-m-2}{(N-m)(N-m-1)}\omega_b^2 + \frac{N-m}{N-m-1}B^2(M\Sigma)_{II}\right] \\ &+ \frac{N-m}{n-m}B^2\sigma^2. \end{aligned}$$

The factor $(N-m)/(n-m)$ applied to $\hat{\omega}_b$ explains the factor c in formula (1.2.1). It is worth noting, without actually displaying the formulas, that the general theory of Section 4.3 also provides formulas for all covariances among such estimators.

For concreteness let us apply these formulas to the specific case discussed in Section 1.2 of a 3-factor design where grand mean, row main effects and column main effects are estimated at the first stage. In this case $m = R + C - 1$ and E_m is the subspace spanned by the three subspaces E_M, E_R and E_C defined in Section 3 of [1]. If we revert to the notation of (1.3.1) we may define $(M\Sigma)_M$, $(M\Sigma)_R$, $\cdots$, $(M\Sigma)_{RCL}$ to be the mean squares arising from the analysis of variance of the complete array v_{ijk}. Expressed in terms of these quantities, $(M\Sigma)_{II}$ of (2.1.3) is given by

$$\text{(2.1.10)} \qquad \begin{aligned} (N-m)(M\Sigma)_{II} = {} & (L-1)(M\Sigma)_L + (R-1)(C-1)(M\Sigma)_{RC} \\ & + (R-1)(L-1)(M\Sigma)_{RL} \\ & + (C-1)(L-1)(M\Sigma)_{CL} \\ & + (R-1)(C-1)(L-1)(M\Sigma)_{RCL}. \end{aligned}$$

Now a typical example of an ω_a to be estimated at the first stage is the difference of two row main effects, i.e.

$$\omega_a = \frac{1}{CL}\sum_{j=1}^{C}\sum_{k=1}^{L}(v_{i_1jk} - v_{i_2jk}),$$

for which $A^2 = 2/CL$, and hence the variance of the least squares estimator is given by (2.1.8) with $(M\Sigma)_{II}$ given by (2.1.10) and $A^2 = 2/CL$. Similarly a typical example of an ω_b to be estimated at the second stage is the difference of two layer main effects, i.e.,

$$\omega_b = \frac{1}{RC}\sum_{i=1}^{R}\sum_{j=1}^{C}(v_{ijk_1} - v_{ijk_2}),$$

for which $B^2 = 2/RC$ and formula (2.1.9) applies directly to give the variance of the unbiased estimator of ω_b.

It should be clear from this example how to write down variances for any estimator from any particular technique in the class considered, and for an array with any number of factors.

2.2. *Significance tests.* Consider the $\mathbf{W}_i$ and ω_i introduced above, in particular those entering at the second stage of analysis where $m + 1 \leqq i \leqq N$. Suppose that $N - m = M_1 + M_2 + M_3$, that we are willing to assume that $\omega_i = 0$ for $m + M_1 + M_2 + 1 \leqq i \leqq N$, and that we wish to test the null hypothesis that $\omega_i = 0$ for $m + 1 \leqq i \leqq m + M_1$ with ω_i arbitrary for $m + M_1 + 1 \leqq i \leqq m + M_1 + M_2$. Then, denoting by $\hat{\omega}_i$ the raw second stage estimator of ω_i, the natural test statistic for this null hypothesis is $(MS)_1/(MS)_3$, where

$$(MS)_1 = \frac{1}{M_1} \sum_{i=m+1}^{m+M_1} (\hat{\omega}_i)^2 \tag{2.2.1}$$

and

$$(MS)_3 = \frac{1}{M_3} \sum_{i=m+M_1+M_2+1}^{N} (\hat{\omega}_i)^2. \tag{2.2.2}$$

The main result of Section 4.4 states that, under the continuous analogue model and assuming normality of the e_i of (2.1.1), $(MS)_1/(MS)_3$ has exactly the F distribution on M_1 and M_3 degrees of freedom regardless of the values of ω_i for $m + M_1 + 1 \leqq i \leqq m + M_1 + M_2$.

Thus, for the standard 3-factor example, if we denote by $(MS)_L$, $(MS)_{RC}$, $\cdots$, $(MS)_{RCL}$ the mean squares arising from the second stage analysis of variance, and if we are willing to assume $(M\Sigma)_{RCL} = 0$, then we may test the null hypothesis that $(M\Sigma)_L = 0$ by regarding $(MS)_L/(MS)_{RCL}$ as an F statistic on $L - 1$ and $(R - 1)(C - 1)(L - 1)$ degrees of freedom. Similarly the second order interactions may be tested against the third order interaction.

Note that it is *not* true that the numerator and denominator of such a test statistic are distributed as multiples of independent χ^2 random variables, but only that the ratio has the stated F distribution. By more detailed arguments of the type given in Section 4.3 we could specify the distributions of the numerator and denominator, and we could specify the non-central distribution of the test statistic. Since the resulting distributions are not of familiar simple types this analysis has not been pursued. We can, however, say something simple which is related to the non-central distribution of the test statistic, for formula (2.1.7) allows us to write down formulas for the average values of the numerator and denominator mean squares. For example

$$\begin{aligned} \text{ave}\,\{(MS)_1\} = \frac{n - m}{N - m - 1}\Bigg[\left(\frac{n - m + 2}{N - m + 2} - \frac{1}{N - m}\right)(M\Sigma)_1 \\ + \left(1 - \frac{n - m + 2}{N - m + 2}\right)(M\Sigma)_{II} + \left(1 - \frac{1}{N - m}\sigma^2\right)\Bigg]. \end{aligned} \tag{2.2.3}$$

Obvious similar formulas hold for $(MS)_2$ and $(MS)_3$, or for $(MS)_L, \cdots, (MS)_{RCL}$ in the 3-factor example.

3. Interpretations. To show how the formulas of Section 2.1 may bear on a practical problem we now attempt a comparison between a simple random allocation scheme and a more orthodox fractional factorial. Suppose $n = 2^r$ observations are to be allowed on a factorial structure with $2^r - 1$ factors at 2 levels each, so that

$$N = 2^{2^r-1}.$$

It is well known (c.f., [2]) to be possible to determine a fixed fraction of size 2^r with the property that all $2^r - 1$ main effects are unconfounded and estimable using $2^r - 1$ simple orthogonal linear combinations of the data points. If such a fraction is chosen, and if, in advance of using it, the labels of the levels of each factor (i.e., 1 or 2) are assigned at random, then this design is a random allocation design within the definition given in [1]. On the other hand one could design a simple random allocation experiment by choosing a simple random sample without replacement of 2^r out of the 2^{2^r-1} factor level combinations. For values of r in the range $r = 4, 5, 6$ the question of the relative performance of these alternative designs is of some practical interest. From our point of view the way to compare these designs is to compare the variances of linear unbiased estimators where variances are found by averaging *both* over the randomness of the randomization hypothesis and over the randomness of the "error" superposed on the observations. (The pros and cons of this point of view were discussed in Section 2 of [1].)

In the case of the fixed fraction there is only one standard method of estimating main effects, and the computation of variances for these estimators is easy. For the simple random allocation fraction there are many different approaches to estimation, for example all of those described in Section 1.2, and the computation of variances is difficult. Thus we are obliged to use the approximate formulas of Section 2.1 and to treat the results as tentative and subject to checking by Monte Carlo methods. The plan is to make a first comparison in terms of (a) an initial model underlying the data and (b) an initial method of estimation for the unbalanced data. Further comparisons will be made modifying both (a) and (b). The initial model is a main effect model, i.e., the model of (2.1.1) with the restriction that v_i is made up only of main effect terms so that

$$v_i = \mu + \sum_{t=1}^{2^r-1} (\pm\Delta_t), \tag{3.1}$$

where Δ_t is the main effect of factor t and the sign of Δ_t is $+$ or $-$ according as v_i is an observation at the upper or lower level of factor t. The initial method of estimation is the simplest practical method in the class described in Section 1.2, namely the method where only the grand mean μ is estimated at the first stage and all other effects are estimated at the second stage.

Under the initial model the estimator of Δ_t from the balanced fraction data is

$\frac{1}{2}$ (mean of the 2^{r-1} observations at the upper level of factor t − mean of the 2^{r-1} observations at the lower level of factor t).

This clearly has variance

$$2^{-r}\sigma^2. \tag{3.2}$$

Notice that this estimator is unbiased with the given variance conditional on the particular design chosen under randomization, and so has the same properties when averaging is carried out over the random choice of design. The comparable formula for the simple random allocation design is given by (2.1.9) where

$$m = 1, \qquad N = 2^{2^{r-1}}, \qquad n = 2^r,$$
$$B^2 = \tfrac{1}{4}\cdot N\cdot(2/N)^2 = 2^{-2^r+1},$$

and

$$B^2(M\Sigma)_{II} = \frac{1}{N-m}\sum_{1}^{2^r-1}\Delta_t^2.$$

If this formula is simplified by ignoring distinctions between $n - m$ and n, $N - n$ and N, $N - m + 2$ and N, etc., then the variance from (2.1.9) is seen to be approximately

$$2^{-r}\left(\Delta_t^2 + \sum_{u=1}^{2^r-1}\Delta_u^2 + \sigma^2\right). \tag{3.3}$$

Now it is clear that the latter variance (3.3) is worse than the former (3.2), and even that it could be enough worse to destroy the value of the estimator. This is not the whole story, however, for the proponent of simple random allocation may argue that he can eliminate from the variance (3.3) as many of the offending Δ_u^2 terms as he wishes, simply by altering his method of analysis to fit the corresponding Δ_u main effects by least squares in the first stage of analysis. This is true, and corresponds to what would be done in practice, but two new disadvantages of simple random allocation appear at this point. Firstly, it will not be known which Δ_u are offending except from prior beliefs or from trial analyses of the data. Secondly, there is a price to be paid in additional variance for the least squares removal of the offending Δ_u terms. Note that both (2.1.8) and (2.1.9) contain factors of $(n - m)^{-1}$ or $(n - m - 1)^{-1}$ which were treated as n^{-1} in (3.3). However, if m comes to be an appreciable fraction h of n, then the variance in (3.3) should be altered not only by omitting the fitted Δ_u^2 terms but also by multiplying by factor $(1 - h)^{-1}$. The first disadvantage is less a criticism of the technique than it is a statement that trial analyses affect the properties of the technique in an unknown way. The second disadvantage is more illuminating, and could be important if a substantial proportion of the main

effects were large. The author believes that, *provided we can accept the main effect model*, the discussion of this paragraph gives a clear picture of how the simple random allocation fraction yields efficiency to the balanced fraction.

However, the proponent of simple random allocation may claim that interaction effects should not be ignored, and he may insist that we compare variances after putting in the terms corresponding to interactions. In the case of the balanced fraction each main effect Δ_t is confounded with

$$p = (2^{2^r-1}/2^r) - 1$$

interaction effects which may be denoted by $\Delta_{t,s}$ for $1 \leqq t \leqq 2^r - 1$ and $1 \leqq s \leqq p$' and we may generalize the model in (3.1) to

$$\nu_i = \mu + \sum_{t=1}^{2^r-1} \left(\pm\Delta_t + \sum_{s=1}^{p} \pm \Delta_{t,s} \right). \tag{3.4}$$

When the formula generalizing (3.2) is sought, it is found that, on account of the confounding, the estimators for the balanced fraction design are not even unbiased until averaging is carried out over the random choice of design, and also that the confounded effects enter into the variance in the obvious way resulting in variance

$$\sum_{s=1}^{p} \Delta_{t,s}^2 + 2^{-r}\sigma^2. \tag{3.5}$$

The formula generalizing (3.3) comes as before from (2.1.9), the only difference being the inclusion of all of the Δ^2 terms in $(M\Sigma)_{II}$. Thus (3.3) becomes

$$2^{-r}\left[\Delta_t^2 + \sum_{u=1}^{2^r-1} \left(\Delta_u^2 + \sum_{s=1}^{p} \Delta_{u,s}^2 \right) + \sigma^2 \right]. \tag{3.6}$$

Formula (3.6) differs from (3.5) in that it has approximately 2^r times as many Δ^2 terms as (3.5), these additional terms being compensated for by the factor 2^{-r}. Thus, if interaction effects are entering substantially into a few of the variances (3.5), then these effects will be greatly spread out when (3.6) applies, and it is no longer at all clear that the balanced fraction is superior to the simple random allocation fraction.

4. The theory of the continuous analogue.

4.1. *Geometrical considerations.* We now describe the general two stage estimation procedure in geometrical terms, following the notation of Section 3 of [1] as introduced in Section 2.1 of this paper. We suppose that a linear functional f_t over E is defined by

$$f_t\left(\sum_{i=1}^{N} c_i \mathbf{V}_i \right) = \sum_{i=1}^{N} c_i v_i, \tag{4.1.1}$$

where $\sum_1^N c_i\mathbf{V}_i$ is any vector in E. The data provide the values of $f_t(\mathbf{V})$ for any vector $\mathbf{V}$ in the n-dimensional subspace E_p of E spanned by the $\mathbf{V}_i$ corresponding

to the n observed cells. In fact, we may regard the information in the data as providing the functional $f_{t,p}$, where

$$\begin{aligned} f_{t,p}(\mathbf{V}) &= f_t(\mathbf{V}) \qquad \text{for} \quad \mathbf{V} \varepsilon E_p \\ &= 0 \qquad\quad\ \ \text{for} \quad \mathbf{V} \perp E_p . \end{aligned} \tag{4.1.2}$$

The first stage of estimation provides the least squares estimators $\hat{\omega}_a$ of $\omega_a = g_t(\mathbf{W}_a)$ for any $\mathbf{W}_a$ in E_m, and the second stage of estimation provides raw estimators $\hat{\omega}_b$ of $\omega_b = g_t(\mathbf{W}_b)$ for any $\mathbf{W}_b$ in $\tilde{E}_m$. Our immediate purpose is to characterize vectors $\mathbf{Z}_a$ and $\mathbf{Z}_b$, both in E_p, with the properties that

$$\hat{\omega}_a = f_t(\mathbf{Z}_a) \quad \text{and} \quad \hat{\omega}_b = f_t(\mathbf{Z}_b). \tag{4.1.3}$$

As a preliminary we define a one-to-one correspondence between linear functionals f and vectors $\mathbf{F}$. Given any linear functional f over E defined by

$$f(\mathbf{U}_i) = u_i$$

for $1 \leqq i \leqq N$, where the $\mathbf{U}_i$ are a unit orthogonal basis of E, we may define an associated vector $\mathbf{F}$ as the vector with components u_i relative to the basis $\mathbf{U}_i$. Conversely, from $\mathbf{F}$ one may recover f. It may be easily checked that the correspondence thus defined does not depend on the particular choice of basis $\mathbf{U}_i$. It is also clear that, if f_1 and f_2 are two functionals with corresponding vectors $\mathbf{F}_1$ and $\mathbf{F}_2$, then $\alpha_1 f_1 + \alpha_2 f_2$ has corresponding vector $\alpha_1 \mathbf{F}_1 + \alpha_2 \mathbf{F}_2$. Thus the set of all linear functionals over E and the set of all vectors in E form isomorphic vector spaces in an obvious manner and we may use interchangeable languages. For example, we may regard the information in the data as $f_{t,p}$ or its corresponding $\mathbf{F}_{t,p}$.

Now the well-known geometrical interpretation of least squares (c.f., Scheffé [3], p. 12) uses vector language and states that the process of least squares fitting is equivalent to splitting $\mathbf{F}_{t,p}$ into

$$\mathbf{F}_{t,p} = \mathbf{F}_I + \mathbf{F}_{II} \tag{4.1.4}$$

where $\mathbf{F}_I$ and $\mathbf{F}_{II}$ are both in E_p but $\mathbf{F}_{II}$ is in $E_p \cap \tilde{E}_m$ and $\mathbf{F}_I$ is perpendicular to $E_p \cap \tilde{E}_m$. $\mathbf{F}_I$ represents the fitted variation and $\mathbf{F}_{II}$ represents the residual variation. However the complete solution of the least squares problem requires the determination of a functional $f_{I'}$ such that

$$\begin{aligned} f_{I'}(\mathbf{W}_a) &= \hat{\omega}_a \qquad \text{for} \quad \mathbf{W}_a \varepsilon E_m \\ f_{I'}(\mathbf{V}) &= 0 \qquad\ \ \text{for} \quad \mathbf{V} \varepsilon \tilde{E}_m . \end{aligned} \tag{4.1.5}$$

The crucial property which this functional must possess is that it must reproduce the fitted variation on E_p, i.e., we must define $f_{I'}$, satisfying (4.1.5) and

$$f_{I'}(\mathbf{V}) = f_I(\mathbf{V}) \qquad \text{for} \quad \mathbf{V} \varepsilon E_p , \tag{4.1.6}$$

where f_I is the functional corresponding to the vector $\mathbf{F}_I$.

Given any $\mathbf{V}$ in E_p we may remove its component along $E_p \cap \tilde{E}_m$ and have left a vector $\mathbf{Z}_a$ in E_p but orthogonal to $E_p \cap \tilde{E}_m$. Then it is clear that

$$f_t(\mathbf{Z}_a) = f_{t,p}(\mathbf{Z}_a) = f_I(\mathbf{V}). \tag{4.1.7}$$

From $\mathbf{Z}_a$ we may further remove its remaining component along $\tilde{E}_m$ and have left a vector $\mathbf{W}_a$ in E_m. Conversely, given such a $\mathbf{W}_a$ we can determine uniquely its corresponding $\mathbf{Z}_a$ as follows: $\mathbf{Z}_a$ is that vector in $E_{p,a}$ which (i) differs from $\mathbf{W}_a$ by a vector in $\tilde{E}_m$ and which (ii) is minimum distance from $\mathbf{W}_a$ subject to (i). Here $E_{p,a}$ is the subspace of E_p formed by the intersection of E_p with the space spanned by $\mathbf{W}_a$ and $\tilde{E}_m$. This situation is pictured in Figure 1. If we assume that every ω_a is estimable, i.e., that no vector in E_m is orthogonal to E_p, then every $\mathbf{W}_a$ in E_m can be reached in this way by some $\mathbf{Z}_a$ in E_p and we have a one-to-one linear correspondence between every vector $\mathbf{W}_a$ in E_m and its corresponding $\mathbf{Z}_a$. If we now define

$$\begin{aligned} f_{I'}(\mathbf{W}_a) &= f_t(\mathbf{Z}_a) \qquad \text{for} \quad \mathbf{W}_a \,\varepsilon\, E_m \\ f_{I'}(\mathbf{V}) &= 0 \qquad\qquad \text{for} \quad \mathbf{V} \,\varepsilon\, \tilde{E}_m, \end{aligned} \tag{4.1.8}$$

then it follows from (4.1.7) and (4.1.8) that

$$f_{I'}(\mathbf{V}) = f_{I'}(\mathbf{W}_a) = f_t(\mathbf{Z}_a) = f_I(\mathbf{V})$$

for any $\mathbf{V}$ in E_p, as required by (4.1.5), and we conclude that

$$\hat{\omega}_a = f_t(\mathbf{Z}_a)$$

is the least squares estimator of $\omega_a = g_t(\mathbf{W}_a)$.

The situation with the second stage estimators is simpler, for the raw estimator of $\omega_b = g_t(\mathbf{W}_b)$ is simply $f_{II}(\mathbf{W}_b)$ where $\mathbf{F}_{II}$ is defined in (4.1.4). Since $f_{II}(\mathbf{V}) = 0$ for any $\mathbf{V}$ in E_m it is clear, as stated in Section 1.2, that the second stage analysis of variance simply gives zero for those effects ω_a estimated at the first stage. Also, if $\mathbf{Z}_b$ is defined to be the component of $\mathbf{W}_b$ in $E_p \cap \tilde{E}_m$, then

$$f_{II}(\mathbf{W}_b) = f_{II}(\mathbf{Z}_b) = f_{t,p}(\mathbf{Z}_b) = f_t(\mathbf{Z}_b),$$

so that

$$\hat{\omega}_b = f_t(\mathbf{Z}_b),$$

where $\hat{\omega}_b$ is the raw second stage estimator of $\omega_b = g_t(\mathbf{W}_b)$.

4.2. *The model for the continuous analogue.* The formulas (4.1.3) indicate how the estimators $\hat{\omega}_a$ and $\hat{\omega}_b$ may be expressed in terms of the underlying functional f_t and the subspace E_p where E_p is at the choice of the experimenter. The subspace E_p used in an actual experiment is necessarily one of the discrete set of $\binom{N}{n}$ subspaces E_p determined by which n of the N cells are used. No other E_p can be observed with n observations. However one can postulate a model under which f_t is observed on other n-dimensional subspaces with positive probability, and (4.1.3) provides reasonable definitions of $\hat{\omega}_a$ and $\hat{\omega}_b$ for such a model.

In the simple random allocation scheme the random subspace E_p is that spanned by a simple random sample of n of the N unit vectors $\mathbf{V}_1, \mathbf{V}_2, \cdots, \mathbf{V}_N$. In the *continuous analogue model* all directions, not just those of a unit orthogonal set, are regarded as "equally likely" in the following sense: E_p is taken to be that random n-dimensional subspace of E which is spanned by n independent *spherically distributed vectors*. A random vector in E is said to be spherically distributed if the distribution of its direction is invariant under any orthogonal transformation of E leaving the origin fixed. The simplest analytical realization of a spherically distributed vector is a vector whose components relative to a unit orthogonal coordinate system are independently $N(0, 1)$. The random subspace E_p defined in this way may also be called a *spherically distributed random subspace* of dimension n.

The author believes that, for reasonably large N and n, the distributions of $\hat{\omega}_a$ and $\hat{\omega}_b$ for the simple random allocation scheme may be reasonably well approximated by the corresponding distributions under the continuous analogue model. This issue was discussed briefly in Section 1.3 and is discussed further in Section 5. Sections 4.3 and 4.4 derive certain distribution properties of the continuous analogue model.

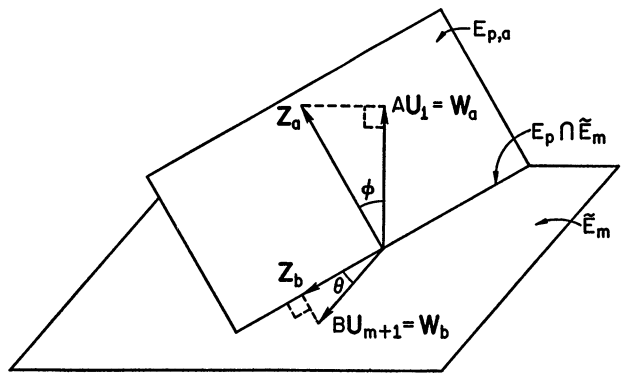

FIGURE 1.

Geometrical picture of $\mathbf{Z}_a$, $\mathbf{Z}_b$, $\mathbf{W}_a$, $\mathbf{W}_b$ *and various subspaces.*

4.3. *Derivation of formulas.* Formulas are now derived for the first and second moments of $\hat{\omega}_a$ and $\hat{\omega}_b$ under the continuous analogue model. For given $\omega_a = \sum_1^m a_i\omega_i = g_t(\mathbf{W}_a)$ and $\omega_b = \sum_{m+1}^N b_i\omega_i = g_t(\mathbf{W}_b)$ we introduce a new basis $\mathbf{U}_1, \cdots, \mathbf{U}_N$ of unit orthogonal vectors in E with the properties that

$$\mathbf{W}_a = A\mathbf{U}_1, \qquad \mathbf{U}_i \varepsilon E_m \qquad \text{for} \quad 1 \leqq i \leqq m,$$

$$\mathbf{W}_b = B\mathbf{U}_{m+1}, \quad \text{and} \quad \mathbf{U}_i \varepsilon \tilde{E}_m \qquad \text{for} \quad m+1 \leqq i \leqq N.$$

Here, as before, $A^2 = \sum_1^m a_i^2$ and $B^2 = \sum_{m+1}^N b_i^2$. Figure 1 gives a picture of the various relevant vectors and subspaces of E, except that $E_{p,a}$, $E_p \cap \tilde{E}_m$ and $\tilde{E}_m$ are pictured as having 2, 1 and 2 dimensions rather than $n - m + 1$, $n - m$ and $N - m$ dimensions respectively.

Suppose $\mathbf{Z}_a$ makes angle Φ with $\mathbf{U}_1$. The spherical distribution of E_p in E induces a spherical distribution of $E_{p,a}$ in the space spanned by $\mathbf{U}_1$ and $\tilde{E}_m$ so that Φ is distributed like the angle between a spherically random $(n - m + 1)$-plane and a fixed direction in $(N - m + 1)$-space. This is the same as the distribution of the angle between a fixed $(n - m + 1)$-plane and a spherically random direction, i.e.,

$$\cos^2\Phi \sim \beta_{\frac{1}{2}(n-m+1),\frac{1}{2}(N-n)} \qquad (0 \leqq \Phi \leqq \tfrac{1}{2}\pi).$$

Similarly if θ is the angle from $\mathbf{Z}_b$ to $\mathbf{U}_{m+1}$, then

$$\cos^2\theta \sim \beta_{\frac{1}{2}(n-m),\frac{1}{2}(N-n)} \qquad (0 \leqq \theta \leqq \tfrac{1}{2}\pi),$$

and this is valid conditional on any given Φ so that θ and Φ are independent.

Now we may write $\mathbf{Z}_a = A\mathbf{U}_1 + A \tan\Phi\mathbf{Z}_1$ and $\mathbf{Z}_b = B \cos^2\theta\mathbf{U}_{m+1} + B \cos\theta \sin\theta\, \mathbf{Z}_2$ where $\mathbf{Z}_1$ is a unit vector in $\tilde{E}_m$ orthogonal to $\mathbf{Z}_b$ and $\mathbf{Z}_2$ is a unit vector in $\tilde{E}_m$ orthogonal to $\mathbf{U}_{m+1}$. Since $\hat{\omega}_a + \hat{\omega}_b = f_t(\mathbf{Z}_a + \mathbf{Z}_b) = Af_t(\mathbf{U}_1) + A \tan\Phi\, f_t(\mathbf{Z}_1) + B \cos^2\theta f_t(\mathbf{U}_{m+1}) + B \cos\theta \sin\theta\, f_t(\mathbf{Z}_2)$, the only unspecified random elements in the expression for this statistic are $f_t(\mathbf{Z}_1)$ and $f_t(\mathbf{Z}_2)$. The marginal distribution of $\mathbf{Z}_1$ given θ and Φ but not $\mathbf{Z}_2$ is simply spherical in $\tilde{E}_m$, and the marginal distribution of $\mathbf{Z}_2$ given θ and Φ but not $\mathbf{Z}_1$ is spherical in the $(N - m + 1)$-dimensional subspace of $\tilde{E}_m$ orthogonal to $\mathbf{V}_{m+1}$. The joint distribution of $\mathbf{Z}_1$ and $\mathbf{Z}_2$ is much more difficult to specify but for purposes of second joint moments this is not necessary. For suppose $\mathbf{Z}_1$ makes angle ξ_i with $\mathbf{U}_i$ $(i = m + 1, \cdots, N)$ and $\mathbf{Z}_2$ makes angle η_i with $\mathbf{U}_i$ $(i = m + 2, \cdots, N)$. Since $\cos\xi_i$ or $\cos\eta_i$ are symmetrically distributed about 0 their averages are 0. Similarly, given ξ_i the unknown conditional distribution of $\cos\xi_i$ or $\cos\eta_k$ is still symmetrical about 0 and thence the following relations hold:

$$\begin{aligned}
&\text{ave}\,\{\cos\xi_i \cos\xi_j\} = 0 \quad \text{if} \quad i \neq j \\
&\text{ave}\,\{\cos\xi_i \cos\eta_j\} = 0 \\
&\text{ave}\,\{\cos\eta_i \cos\eta_j\} = 0 \quad \text{if} \quad i \neq j.
\end{aligned}$$

The marginal distributions of $\cos\xi_i$ and $\cos\eta_j$ are given by

$$\cos^2\xi_i \sim \beta_{\frac{1}{2},\frac{1}{2}(N-m-1)}$$

and

$$\cos^2\eta_j \sim \beta_{\frac{1}{2},\frac{1}{2}(N-m-2)}.$$

If we set $f_t(\mathbf{U}_i) = u_i$ for $i = 1, \cdots, N$ we have

$$f_t(\mathbf{Z}_1) = \sum_{i=m+1}^{N} u_i \cos\xi_i$$

and

$$f_t(\mathbf{Z}_2) = \sum_{i=m+2}^{N} u_i \cos\eta_i.$$

Thus we can write

$$\hat{\omega}_a + \hat{\omega}_b = Au_1 + A \tan \Phi \sum_{i=m+1}^{N} u_i \cos \xi_i + B \cos^2 \theta u_{m+1} + B \cos \theta \sin \theta \sum_{i=m+2}^{N} u_i \cos \eta_i .$$

Now we are in a position to average over the randomness induced by random E_p, regarding f_t as fixed. Since Φ and θ are independent of the ξ_i and η_j we need only replace the trigonometric functions by their averages. Thus

$$\begin{aligned} \text{ave } \{\hat{\omega}_a + \hat{\omega}_b\} &= Au_1 + Bu_{m+1} \text{ ave } \{\cos^2 \theta\} \\ &= Au_1 + \frac{n-m}{N-m} Bu_{m+1} \\ &= w_a + \frac{n-m}{N-m} w_b , \end{aligned}$$

where $w_a = g_t(\mathbf{W}_a)$ and $w_b = g_t(\mathbf{W}_b)$. Also

$$\begin{aligned} \text{ave } \{(\hat{\omega}_a + \hat{\omega}_b)^2\} &= A^2u_1^2 + 2Au_1Bu_{m+1} \text{ ave } \{\cos^2 \theta\} \\ &\quad + B^2u_{m+1}^2 \text{ ave } \{\cos^4 \theta\} + A^2 \text{ ave } \{\tan^2 \Phi\} \sum_{i=m+1}^{N} u_1^2 \text{ ave } \{\cos^2 \xi_i\} \\ &\quad + B^2 \text{ ave } \{\cos^2 \theta \sin^2 \theta\} \sum_{i=m+2}^{N} u_i^2 \text{ ave } \{\cos^2 \eta_i\} \\ &= A^2u_1^2 + 2 \frac{n-m}{N-m} Au_1Bu_{m+1} \\ &\quad + \frac{(n-m)(n-m+2)}{(N-m)(N-m+2)} B^2u_{m+1}^2 + A^2 \frac{N-n}{n-m-1} \sum_{i=m+1}^{N} u_i^2 \frac{1}{N-m} \\ &\quad + B^2 \frac{(n-m)(N-n)}{(N-m)(N-m+2)} \sum_{i=m+2}^{N} u_i^2 \frac{1}{N-m-1} \\ &= w_a^2 + 2 \frac{n-m}{N-m} w_aw_b + \frac{(n-m)(n-m+2)}{(N-m)(N-m+2)} w_b^2 \qquad (4.3.1) \\ &\quad + A^2 \frac{N-n}{n-m-1} (MS)_{II} \\ &\quad + \frac{(n-m)(N-n)}{(N-m-1)(N-m+2)} \left[B^2(MS)_{II} - \frac{1}{N-m} w_b^2 \right] \\ &= w_a^2 + 2 \frac{n-m}{N-m} w_aw_b + \frac{n-m}{N-m-1} \left(\frac{n-m+2}{N-m+2} - \frac{1}{N-m} \right) w_b^2 \\ &\quad + A^2 \frac{N-n}{n-m-1} (MS)_{II} \\ &\quad + B^2 \frac{n-m}{N-m-1} \left(1 - \frac{n-m+2}{N-m+2} \right) (MS)_{II}, \end{aligned}$$

where

$$(MS)_{II} = \frac{1}{N-m}\sum_{i=m+1}^{N} w_i^2 .$$

Finally we average over the randomness of f_t, i.e.,

$$\text{ave}\,\{w_a\} = \omega_a , \qquad \text{ave}\,\{w_b\} = \omega_b ,$$
$$\text{ave}\,\{w_a^2\} = \omega_a^2 + A^2\sigma^2, \qquad \text{ave}\,\{w_a w_b\} = \omega_a\omega_b ,$$
$$\text{ave}\,\{w_b^2\} = \omega_b^2 + B^2\sigma^2 \quad \text{and} \quad \text{ave}\,\{(MS)_{II}\} = (M\Sigma)_{II} + \sigma^2$$

where

$$(M\Sigma)_{II} = \frac{1}{N-m}\sum_{i=m+1}^{N} \omega_i^2 . \tag{4.3.2}$$

Thus

$$\text{ave}\,\{\hat{\omega}_a + \hat{\omega}_b\} = \omega_a + \frac{n-m}{N-m}\,\omega_b \tag{4.3.3}$$

and

$$\begin{aligned}\text{ave}\,\{(\hat{\omega}_a + \hat{\omega}_b)^2\} &= \omega_a^2 + 2\,\frac{n-m}{N-m}\,\omega_a\omega_b \\ &\quad + \frac{n-m}{N-m-1}\left(\frac{n-m+2}{N-m+2} - \frac{1}{N-m}\right)\omega_b^2 \\ &\quad + A^2\,\frac{N-n}{n-m-1}\,(M\Sigma)_1 + A^2\,\frac{N-m-1}{n-m-1}\,\sigma^2 \\ &\quad + B^2\,\frac{n-m}{N-m-1}\left(1 - \frac{n-m+2}{N-m+2}\right)(M\Sigma)_1 \\ &\quad + B^2\,\frac{n-m}{N-m-1}\left(1 - \frac{1}{N-m}\right)B^2\sigma^2 .\end{aligned} \tag{4.3.4}$$

Formulas (2.1.4) and (2.1.6) are simply special cases of (4.3.3) and formulas (2.1.5) and (2.1.7) are simply special cases of (4.3.4). Note also that, if $\omega_{a'}$ and $\omega_{b'}$ are alternative parameters estimated by $\hat{\omega}_{a'}$ and $\hat{\omega}_{b'}$ at the first and second stages, then from formulas (4.3.3) and (4.3.4) we can find var $(\hat{\omega}_a + \hat{\omega}_b)$, var $(\hat{\omega}_{a'} + \hat{\omega}_{b'})$ and var $([\hat{\omega}_a + \hat{\omega}_{a'}] + [\hat{\omega}_b + \hat{\omega}_{b'}])$ and hence deduce cov $(\hat{\omega}_a + \hat{\omega}_b ,\ \hat{\omega}_{a'} + \hat{\omega}_{b'})$.

4.4. *The distribution of the ratio of mean squares.* The purpose of this section is to prove the following theorem. *Suppose the random functional f_t is defined by* (4.1.1) *and* (2.1.1) *where the e_i are normally distributed. Suppose E_p is spherically random according to the continuous analogue model. Suppose $\hat{\omega}_i$ is the raw second stage estimator of ω_i for $m + 1 \leqq i \leqq N$ and $(MS)_1$ and $(MS)_3$ are as defined in* (2.2.1) *and* (2.2.2). *Suppose, according to the null hypothesis, that $\omega_i = 0$ for*

$m + 1 \leqq i \leqq m + M_1$ and $m + M_1 + M_2 + 1 \leqq i \leqq N$, but that otherwise the ω_i are arbitrary. Then $(MS)_1/(MS)_3$ has the F distribution on M_1 and M_3 degrees of freedom.

Denote by E_1 the M_1-dimensional subspace of $\tilde{E}_m$ spanned by $\mathbf{W}_i$ for $m + 1 \leqq i \leqq m + M_1$, by E_3 the M_3-dimensional subspace of $\tilde{E}_m$ spanned by $\mathbf{W}_i$ for $m + M_1 + M_2 + 1 \leqq i \leqq N$, and by $E_{1,3}$ the space spanned by E_1 and E_3 together. Denote by $\mathbf{F}_{1,3}$ the vector in $E_{1,3}$ whose components are $\hat{\omega}_i$ along $\mathbf{W}_i$ for $m + 1 \leqq i \leqq m + M_1$ and $m + M_1 + M_2 + 1 \leqq i \leqq N$, i.e., $\mathbf{F}_{1,3}$ is the component of $\mathbf{F}_{II}$ of (4.1.4) in $E_{1,3}$. Under the hypothesis of the theorem, $f_t(\mathbf{W}_i) = d_i$, where the d_i are independently $N(0, \sigma^2)$ for $m + 1 \leqq i \leqq m + M_1$ and $m + M_1 + M_2 + 1 \leqq i \leqq N$, so that the distribution of f_t is invariant under any orthogonal transformation of $E_{1,3}$. Similarly the distribution of E_p is invariant under any orthogonal transformation of $E_{1,3}$, and E_p and f_t are assumed independent. Since f_t and E_p determine $\mathbf{F}_{1,3}$ it follows that $\mathbf{F}_{1,3}$ is spherically distributed in $E_{1,3}$, and hence that

$$\frac{(MS)_1}{(MS)_3} = \frac{M_3\,(\text{component of } \mathbf{F}_{1,3} \text{ in } E_1)^2}{M_1\,(\text{component of } \mathbf{F}_{1,3} \text{ in } E_3)^2}$$

has an F distribution on M_1 and M_3 degrees of freedom.

5. Discussion of the continuous analogue model. In constructing the continuous analogue model an arbitrary choice was made, namely the choice that, under the continuous model, E_p should be the subspace spanned by n independent sample vectors from *some* multivariate normal distribution over E. This choice was made partly for mathematical convenience and partly because of the author's not infallible intuition in N dimensions. The particular choice of the *spherical* multivariate normal distribution was dictated by the requirement that, as with simple random allocation, the distribution of E_p should be invariant under all $N!$ permutations of the coordinate axes $\mathbf{V}_i$.

For moderately large N and n this particular continuous analogue has some intuitive appeal as an approximation to simple random allocation, for a discrete distribution with a large number $\binom{N}{n}$ of equi-probable n-spaces and with symmetry under any permutation of the $\mathbf{V}_i$ is approximated by a continuous distribution with analogous continuous uniformity and symmetry properties. Of course, intuition in N-dimensional space is uncertain. Certainly, statistics can be found whose discrete and approximating continuous distributions under the randomization hypothesis are quite different, especially in the sense that the discrete distribution is very discrete and so not fitted well by any continuous distribution. For example,

$$f_{t,p}(\mathbf{V}_1) = v_1 \quad \text{with probability} \quad n/N$$

$$= 0 \quad \text{with probability} \quad 1 - (n/N)$$

under the discrete hypothesis, but under the continuous hypothesis is fitted with a continuous distribution with some dependence on $v_2, \cdots, v_n$ as well as on v_1. However, the real issue for our purposes is whether the approximation is adequate for first stage estimators $\hat{\omega}_a$, second stage estimators $\hat{\omega}_b$, or second stage ratios of mean squares. This issue is, for the most part, beyond the scope of this paper.

One comparison between discrete and continuous case formulas is easy, and this we now carry out. We now compute the discrete case formulas analogous to (2.1.6) and (2.1.7) appropriate for the special method of data analysis where the first stage of analysis is empty, i.e., $m = 0$. In this case every ν_i and all linear combinations $\sum_1^N c_i\nu_i$ are particular cases of ω_b parameters, and $\hat{\omega}_b = \sum_1^N c_i\hat{\nu}_i$ where

$$\hat{\nu}_i = v_i \quad \text{if the } i\text{th cell is observed}$$
$$= 0 \quad \text{otherwise.}$$

Under simple random sampling

$$\text{ave } \{\hat{\nu}_i\} = \frac{n}{N} v_i, \qquad \text{ave } \{(\hat{\nu}_i)^2\} = \frac{n}{N} v_i^2,$$

and

$$\text{ave } \{\hat{\nu}_i \hat{\nu}_j\} = \frac{n(n-1)}{N(N-1)} v_i v_j,$$

Thus, under the randomization hypothesis,

$$\text{ave } \{\hat{\omega}_b\} = \frac{n}{N} \sum_{i=1}^N c_i v_i,$$

and

$$\text{ave } \{(\hat{\omega}_b)^2\} = \frac{n}{N} \sum_{i=1}^N c_i^2 v_i^2 + \frac{n(n-1)}{N(N-1)} \sum_{\substack{i=1 \\ i \neq j}}^N \sum_{j=1}^N c_i c_j v_i v_j.$$

$$= \frac{n(n-1)}{N(N-1)} \left(\sum_{i=1}^N c_i v_i \right)^2 + \frac{n(N-n)}{N(N-1)} \sum_{i=1}^N c_i^2 v_i^2.$$

Further averaging over the randomness of $v_i = \nu_i + e_i$, we get

$$\text{ave } \{\hat{\omega}_b\} = \frac{n}{N} \omega_b \tag{5.1}$$

and

$$\text{ave } \{(\hat{\omega}_b)^2\} = \frac{n}{N} \left[\left(\frac{n-1}{N-1} \right) \omega_b^2 + \left(1 - \frac{n-1}{N-1} \right) \sum_{i=1}^N c_i^2 \nu_i^2 + B^2 \sigma^2 \right], \tag{5.2}$$

where $B^2 = \sum_1^N c_i^2$. If $m = 0$ is substituted in (2.1.6) and (2.1.7) it is seen that (5.1) agrees with (2.1.6), thereby justifying the correction factor $c = N/n$ in this case, but that (5.2) differs from (2.1.7) in two ways. Firstly, the coefficients

depending on N and n differ, but their ratios approach unity as n and N increase. Secondly the expression $B^2(M\Sigma)_{II}$ in (2.1.7) is replaced by $\sum_1^N c_i^2\nu_i^2$ in (5.2). The second difference is due to the symmetrizing effect of the continuous analogue model. If symmetrized average squares or symmetrized variances are considered as in Section 6 of [1], then the difference disappears. In particular, when the basic array consists of k factors at two levels each, i.e., $N = 2^k$, and when main effects or interaction terms are estimated, then the c_i^2 are all identical and so $B^2(M\Sigma)_{II} = \sum_1^N c_i^2\nu_i^2$. In this case variance and symmetrized variance are the same.

Further analytical comparisons of the above type become very complex; even the trivial case of least squares estimation of the grand mean with $m = 1$ results in very messy algebraic detail. Spot-checking by Monte Carlo seems to be the only method available.

6. Details on the class of data analysis techniques. It was stated in Section 1.1 that the class of techniques described in Section 2.1 falls within the more general class described in [1]. Furthermore we have tacitly assumed throughout that estimators like $\hat{\omega}_a$ and $\hat{\omega}_b$ are unbiased except for constant scale factors. This is true for any type of random allocation design, with the justification following from the theory of [1] together with a proof that the methods of Section 1.2 are covered by the general methods of [1].

The required proof now follows. Suppose, as before, that $\mathbf{W}_a$ and $\mathbf{W}_b$ are general vectors in E_m and $\tilde{E}_m$ respectively, and consider the statistic $\hat{\omega}_a + \hat{\omega}_b$. We claim that this statistic is a special case of what was referred to in [1] as a λ-minimum extension, i.e.,

$$\hat{\omega}_a + \hat{\omega}_b = f_\lambda(\mathbf{W}_a + \mathbf{W}_b),$$

for a particular choice of λ-metric. In [1] unbiased estimators were defined from f_λ just as we have defined them from $\hat{\omega}_a$ and $\hat{\omega}_b$ in this paper.

Consider the second characterization of f_λ given in Section 4.2 of [1]. Consider also the characterization of $\hat{\omega}_a$ and $\hat{\omega}_b$ in (4.1.3) of this paper. These two characterizations coincide for the limiting choice of λ-metric where the λ-values corresponding to $\mathbf{W}_1, \cdots, \mathbf{W}_m$ all tend to ∞, and the λ-values corresponding to $\mathbf{W}_{m+1}, \cdots, \mathbf{W}_N$ are all equal to unity. For $\mathbf{Z}_a + \mathbf{Z}_b$ can be characterized as that vector in E_p which is at minimum distance from $\mathbf{W}_a + \mathbf{W}_b$ subject to the condition that the components corresponding to $\lambda = \infty$ are not allowed to count, i.e., $\mathbf{Z}_a + \mathbf{Z}_b$ is that vector in E_p nearest to $\mathbf{W}_a + \mathbf{W}_b$ subject to the condition that $\mathbf{Z}_a + \mathbf{Z}_b - \mathbf{W}_a - \mathbf{W}_b$ has zero component along E_m. This completes the proof.

REFERENCES

[1] A. P. Dempster, "Random allocation designs I: on general classes of estimation methods," *Ann. Math. Stat.*, Vol. 31 (1960), pp. 885–905.

[2] R. L. Plackett and J. P. Burman, "The design of optimum multifactor experiments," *Biometrika*, Vol. 33 (1946), pp. 305–325.

[3] Henry Scheffé, *The Analysis of Variance*, John Wiley and Sons, New York, 1959.

[4] S. S. Wilks, *Mathematical Statistics*, Princeton University Press, Princeton, 1943.