DESIGNS FOR ESTIMATING VARIANCE COMPONENTS

R. L. Anderson

Institute of Statistics
Mimeograph Series No. 310
November, 1961

DESIGNS FOR ESTIMATING VARIANCE COMPONENTS [1/]

R. L. Anderson
Institute of Statistics
North Carolina State College

1. Introduction

The science of statistics is essentially the application of the mathematical theory of probability to the study of variation in experimental and operational data. In a paper prepared for the 1958 meeting of the International Statistical Institute in Brussels, I discussed the "Uses of Variance Component Analysis in the Interpretation of Biological Experiments." Some of this discussion is repeated here.

The identification of the sources of variation and development of methods of estimating the separate variances and testing hypotheses concerning them is one of the statistician's important tasks. In most experiments, some variation is imposed by the experimenter, in the form of different treatments, varieties, or practices. Often inferences are desired concerning only the particular treatments used in the experiment; it has been the custom to label these as <u>fixed</u> sources of variability. In other cases, it is assumed that the variability among the sources used in the experiment is representative of the variability in a much larger population; these are usually labeled as <u>random</u> sources of variability. Admittedly the particular representatives used in the experiment sometimes are not selected in an accepted random manner; however, the experimenter can often justify the assumption that order of selection is of little importance. Most experiments have a combination of fixed and random sources of variability; the <u>mixed category</u>. Often an experimenter will want to examine a given variable from both the random and fixed point of views.

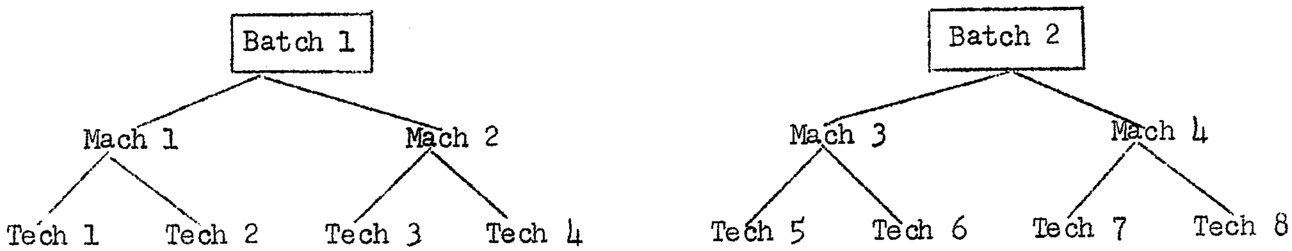
1/ Presented at the Seventh Conference on the Design of Experiments, October 18-20, 1961, U. S. Army Signal Research and Development Laboratory, Fort Monmouth, New Jersey.

In this paper I will discuss only situations in which all sources of variation are essentially random. Some examples are presented on the next page. There are two experimental and operational procedures to be considered: the nested or hierarchical type and the classification type. To illustrate the basic differences between these two types, suppose we have a large number (b) of batches of some material. These are to be processed by one of a large number of more or less identical machines. Suppose material from the first batch is subdivided into m_1 sub-batches and each sub-batch processed by a different machine; the second batch is processed by m_2 machines, all different from the first m_1; and so forth for all batches. Hence there are $m = \sum_{i=1}^{b} m_i$ machines used for processing. The processed material from each machine is tested, assayed or inspected by a corps of technicians, a different corps for each machine. The material is subdivided into samples, each sample handled by one technician. Let us assume there are t_{ij} technicians for the j^{th} machine processing the i^{th} batch, so that the total number of technicians is $t = \sum_{i,j} t_{ij}$. In the following example of two batches, we have assumed that each $m_i = 2$ and each $t_{ij} = 2$.

Now suppose that there are only, say, 4 processing machines and, say, 2 technicians at a given plant, but that these machines and technicians are "representative" of large populations of machines and technicians. Let us further assume that every batch is subdivided into 4 sub-batches, one for each machine. Then the product from each machine is divided into 2 parts, one part for each technician. This operation for two batches is also exemplified on the following page.

In general, the first operation will be a combination of the two types, since subsequent batches will be allocated to machines 1 and 2, 3 and 4, etc.

EXAMPLES OF RANDOM OPERATIONS



Nested or Hierarchical-Type Operation

* *

		Tech 1	Tech 2
Batch 1	Mach 1		
	Mach 2		
	Mach 3		
	Mach 4		
Batch 2	Mach 1		
	Mach 2		
	Mach 3		
	Mach 4		

Classification-Type Operation

* *

Streptomycin Experiment

Initial Incubation → Primary Inoculation → Secondary Inoculation → Fermentation → Assay

* *

Rubber Sampling

Producer's Estate → Day at Estate → Bale on given day → Sheet of Rubber from Bale → Sample from Sheet

A better example of the strictly nested-type operation might be one I used for the International Statistical Institute paper: A pilot study was considered to assess the various sources of variability in the production and assay for streptomycin before conducting an experiment on the efficacy of various molds. There were five stages in this process: an initial incubation in a test tube; a primary inoculation period in a petrie dish; a secondary inoculation period in another petrie dish; a fermentation period in a bath; and the final assay of the amount of streptomycin produced. Newton, et al. (1951) in a study of the variability of rubber considered the following sources of variation: producer's estates, days at a given estate, bales on a given day, sheets from a given bale, and samples from each sheet. Technicians then took measurements on the samples. In this paper I shall discuss only purely nested and classification-type operations but the same methods are applicable to the combination.

Returning to our batch problem, the final measurement made by the technician is designated as the process yield, Y. We are here interested in the variability of the Y's. Such variation is a result of variable batches, variable batch sampling and subsequent machine processing and variable product sampling and measurement. The variation in machine processing results from machine-to-machine differences and failure of a given machine to perform in the same manner on different batches, and similarly for the variation in the measuring process. The failure of a given machine to perform similarly on different batches is called a machine-batch interaction and similarly for the technician-batch and technician-machine interactions. We are assuming that only two-factor interactions need be considered; however, one might also be interested in a technician-machine-batch interaction. Interactions can be detected only with the classification-type operation.

It is assumed that the sources of variability act independently and additively; hence, one can write the mathematical model for the final yield as follows:

$$Y_j = \mu + \sum_{i=1}^{v} e_{ij} , \tag{1}$$

where there are $\underline{v}$ sources of variability, the $\{e_{ij}\}$ assumed independently distributed random variables with zero means and constant variances, σ_i^2; hence, μ is the product average. The variance of the final product is $\sigma^2 = \sum_{i=1}^{v} \sigma_i^2$. It is assumed that the experimenter or manufacturer is interested in reducing the magnitude of σ^2. In order to do this, a preliminary experiment is to be set up to estimate the individual variances, σ_i^2.

The estimates of the σ_i^2 will be designated as $\hat{\sigma}_i^2$. In studying the adequacy of these estimates, it will be assumed that the random variables (the e's) are normally distributed. Hence the observed yields are assumed normally distributed with the same mean μ and variance $\sigma^2 = \sum \sigma_i^2$. However, if there is more than one source of variation, the Y's (in general) will be correlated. Consider the nested process with 2 batches, 2 machines per batch and 2 technicians per machine. One might represent the measurements as follows:

$$Y_{111} = \mu + b_1 + m_1 + t_1, \; Y_{112} = \mu + b_1 + m_1 + t_2, \; Y_{123} = \mu + b_1 + m_2 + t_3,$$
$$\ldots, \; Y_{248} = \mu + b_2 + m_4 + t_8. \tag{2}$$

The usual short cut notation is to renumber the technicians within each machine, the machines within each batch, etc.; i.e.,

$$Y_{111} = \mu + b_1 + m_{11} + t_{111}, \; Y_{112} = \mu + b_1 + m_{11} + t_{112},$$
$$Y_{121} = \mu + b_1 + m_{12} + t_{121}, \; \ldots, \; Y_{222} = \mu + b_2 + m_{22} + t_{222} . \tag{3}$$

The latter notation is much simpler for complicated designs and will be used here. In any case, it is easy to see that the covariance between Y_{111} and Y_{112} is $\sigma_b^2 + \sigma_m^2$; between Y_{111} and Y_{121} is σ_b^2; and between Y_{111} and Y_{222} is 0. Hence it is possible to write the likelihood of the sample in the familiar multivariate normal form, with a constant mean (μ) and variance ($\sigma^2 = \sigma_b^2 + \sigma_m^2 + \sigma_t^2$) but with correlations of either $(\sigma_b^2 + \sigma_m^2)/\sigma^2$, σ_b^2/σ^2 or 0. The method of estimation would be the familiar maximum likelihood (ML).

The model for the classification-type process might be

$$Y_{ijk} = \mu + b_i + m_j + (bm)_{ij} + t_k + (bt)_{ik} + (mt)_{jk} + e_{ijk} , \qquad (4)$$

where e may be a three-factor interaction effect or technique error or a combination of the two. In this case we have variance components: $\sigma_b^2, \sigma_m^2, \sigma_{bm}^2, \sigma_t^2, \sigma_{bt}^2, \sigma_{mt}^2, \sigma_e^2$.

Again if the components are assumed to be normally distributed, one can set up the normal multivariate likelihood function, which will be considerably more complicated than the nested one.

The usual estimating procedure for variance components is based on the familiar, I hope, Analysis of Variance table. For balanced designs, the analysis of variance (ANOV) and maximum likelihood (ML) estimators are essentially the same; I will not take the time to discuss this difference here. For non-balanced designs the ML equations are complicated functions of the estimators; iterative methods are usually needed to solve them. There is no unique ANOV estimating procedure for non-balanced designs. The usual procedure is based on a pooling procedure first presented by Ganguli (1941) and summarized in Anderson and Bancroft (1952). This procedure produces unbiased estimators, but of uncertain efficiency. The total sum of squares can be subdivided into orthogonal sources but there remains the problem of how to best combine these sources. One probably would prefer unbiased estimators which are linear combinations of these mean squares in the ANOV table.

Hence one solution to the estimation problem would be to select that unbiased linear estimator which has the smallest variance. In all discussion which follows, a linear estimator of a variance component will refer to a linear function of the mean squares; obviously this is a quadratic function of the observations. Much theoretical work has been done on the construction of best quadratic estimators for balanced designs, but the results have been rather sterile for non-balanced designs.

2. Some Simple Examples.

To illustrate the complexity of the problem, let us consider a two-stage nested process with only 8 final measurements, coming from a = 2,3,..., or 7 classes, with the model

$$Y_{ij} = \mu + a_i + b_{ij}; \quad i = 1,2,\ldots,a. \tag{5}$$

The object of the investigation is to obtain a minimum variance unbiased linear estimator of σ_a^2. One possible design has a = 4 classes and 2 samples per class. The analysis of variance would be as follows ($\rho = \sigma_a^2/\sigma_b^2$) 2/.

Table 1. Analysis of Variance for 4 Classes (N=8)

Source of Variation (SV)	Degrees of Freedom (DF)	Mean Square (MS)	Expected Value of Mean Square (EMS)
Classes (A)	3	MSA	$\sigma_b^2 + 2\sigma_a^2 = \sigma_b^2(1+2\rho)$
Samples (B)	4	MSB	σ_b^2

The ANOV estimates of σ_a^2 and σ_b^2 are

$$\hat{\sigma}_a^2 = \frac{MSA-MSB}{2} \; ; \; \hat{\sigma}_b^2 = MSB. \tag{6}$$

2/ It is assumed that the reader is familiar with ANOV procedures. If not, reference is made to Anderson and Bancroft (1952, Chapter 22).

Under the normality assumption, MSA and MSB are distributed as multiples of independent χ^2-variates. In fact MSA $= \chi_1^2(\sigma_b^2 + 2\sigma_a^2)/3$ and MSB $= \chi_2^2\sigma_b^2/4$, where χ_1^2 and χ_2^2 are independent χ^2-variates with 3 and 4 respective degrees of freedom.
Hence

$$\mathrm{Var}\ (\hat{\sigma}_a^2) = \frac{2}{4}\left[\frac{(\sigma_b^2 + 2\sigma_a^2)^2}{3} + \frac{\sigma_b^4}{4}\right] = 2\sigma_b^4\left[7/48 + \rho/3 + \rho^2/3\right] . \tag{7}$$

In comparing designs and estimators, we will consider

$$V_a = \frac{\mathrm{Var}(\hat{\sigma}_a^2)}{2\sigma_b^4} = 7/48 + \rho/3 + \rho^2/3 . \tag{8}$$

$$\mathrm{Var}(\hat{\sigma}_b^2) = 2\sigma_b^4/4 \ ; \ V_b = 1/4 . \tag{9}$$

Another plan would have a = 2 classes and 4 samples per class, with the results given in Table 2 and equations (10). A comparison of these two designs shows that the one with 2 classes will be superior (in estimating σ_a^2) only when ρ is quite small. But we could consider 3, 5, 6 or 7 classes. Unfortunately all of these will involve non-balanced designs. As an example, consider the design with 7 classes, one sample per class for 6 classes and 2 samples from the seventh class; the ANOV is in Table 3. A_1 represents the variation among the 6 classes with one sample per class. A_2 represents the comparison of the sum of these first 6 observations and 3 times the sum of the two observations from the seventh class. In estimating σ_a^2, one must decide how to weight MSA_1 and MSA_2. If these are weighted directly as their DF, we have the results in equations (11).

Using estimators of this type, the values of V_a and V_b have been computed when a = 2,3,..., or 7 classes are used, for ρ = .05,.1,.2,.5,1,2,5 and 10(Table 4). As expected, one requires more classes to estimate σ_a^2 efficiently as ρ increases; at the same time, the variance of $\hat{\sigma}_b^2$ increases. Hence for large ρ, if one wishes good estimates of both σ_a^2 and σ_b^2, a compromise is necessary. This will be discussed in more detail later.

Table 2. Analysis of Variance for 2 Classes

SV	DF	MS	EMS
A	1	MSA	$\sigma_b^2 + 4\sigma_a^2 = \sigma_b^2\,(1 + 4\rho)$
B	6	MSB	σ_b^2

$$\hat{\sigma}_a^2 = \frac{\text{MSA} - \text{MSB}}{4}\ ;\quad \hat{\sigma}_b^2 = \text{MSB}\ ;\quad V_a = \frac{7}{96} + \frac{\rho}{2} + \rho^2\ ;\quad V_b = \frac{1}{6}\ . \tag{10}$$

* *

Table 3. Analysis of Variance for 7 Classes

SV	DF	MS	EMS
A_1	5	MSA_1	$\sigma_b^2 + \sigma_a^2 = \sigma_b^2(1 + \rho)$
A_2	1	MSA_2	$\sigma_b^2 + 1.75\ \sigma_a^2 = \sigma_b^2(1 + 1.75\rho)$
B	1	MSB	σ_b^2

$$\hat{\sigma}_a^2 = \frac{5\ MSA_1 + MSA_2 - 6\ MSB}{6.75}\ ;\ V_a = .922 + .2963\ \rho + .17695\ \rho^2\ ;\ V_b = 1/(8-a). \tag{11}$$

* *

Table 4. Values of V_a and V_b for Various Designs, N = 8 [a/]

No. Classes (a)	V_a for given ρ: .05	.10	.2	.5	1.0	2.0	5.0	10.0	V_b
2	$\underline{.100}$	$\underline{.133}$	.213	.573	1.57	5.07	27.6	105	.167
3	.121	.145	$\underline{.198}$	.419	.993	2.90	14.8	54.9	.200
4	.163	.182	.226	$\underline{.396}$	$\underline{.812}$	2.15	10.1	36.8	.250
5	.255	.274	.314	.467	.829	1.96	8.60	30.5	.333
6	.430	.447	.485	.623	.941	$\underline{1.90}$	7.43	25.4	.500
7	.937	.953	.988	1.11	1.40	2.22	$\underline{6.83}$	$\underline{21.6}$	1.000

a/ $\text{Var}(\hat{\sigma}_a^2) = 2V_a\sigma_b^4$; $\text{Var}(\hat{\sigma}_b^2) = 2V_b\sigma_b^4$. Underlined values are minima for a given $\rho = \sigma_a^2/\sigma_b^2$.

3. Selection of Best "Linear" Estimators.

As stated earlier, for non-balanced designs the choice of even the best linear unbiased estimator is not obvious. We might use an iterative weighted least squares procedure, using the mean squares as the dependent variable and the variance components as regression coefficients. Since the mean squares are multiples of χ^2-variates, they have unequal variances, both due to EMS and DF; i.e.,

$$\mathrm{Var}\ (\mathrm{MS}) = (\mathrm{EMS})^2/\mathrm{DF}.$$

For the above 7-class problem the least squares equations are given in equations (12), with limiting values for $\rho = 0$ and $\rho \to \infty$ in equations (13) and (15). In the latter two cases, explicit estimators of σ_a^2 and σ_b^2 are available. These and their variances are presented in equations (14) and (16). For the general case (12), a suggested procedure is to use a preliminary estimate of ρ, such as $\hat{\sigma}_a^2/\hat{\sigma}_b^2$ from the pooled ANOV results; then resolve for $\hat{\sigma}_a^2$ and $\hat{\sigma}_b^2$ using the least squares equations; continue the iteration until two successive sets of estimates agree. Unfortunately, it is almost impossible to obtain the variance of the final estimate of $\hat{\sigma}_a^2$; hence, a comparison of designs becomes very difficult. Very little research has been pursued along these lines; the ideas are merely tossed out here for consideration.

A comparison of the limiting estimators with those of Table 4 is presented in Table 5 for ρ = .05,.1,1 and 10. The first two estimators are about the same for all ρ but the third one is good only for very small ρ. It appears that one does not lose much information by use of MSB to estimate σ_b^2 and the pooled MSA_i to estimate σ_a^2, except possibly when ρ is very small. In subsequent discussions, it will be assumed that the estimators will be of the type used in Section 2.

Least Squares equations for 7-class problem.

$$\left[\frac{5}{(1+\rho)^2}+\frac{(1.75)^2}{(1+1.75\rho)^2}\right]\hat{\sigma}_a^2+\left[\frac{5}{(1+\rho)^2}+\frac{1.75}{(1+1.75\rho)^2}\right]\hat{\sigma}_b^2=\frac{5A_1}{(1+\rho)^2}+\frac{1.75A_2}{(1+1.75\rho)^2}$$

$$\left[\frac{5}{(1+\rho)^2}+\frac{1.75}{(1+1.75\rho)^2}\right]\hat{\sigma}_a^2+\left[\frac{5}{(1+\rho)^2}+\frac{1}{(1+1.75\rho)^2}+1\right]\hat{\sigma}_b^2=\frac{5A_1}{(1+\rho)^2}+\frac{A_2}{(1+1.75\rho)^2}+B\ . \qquad (12)$$

$\underline{\rho = 0}$

$$\left[5+(1.75)^2\right]\hat{\sigma}_a^2+6.75\hat{\sigma}_b^2=5A_1+1.75A_2$$

$$6.75\hat{\sigma}_a^2+7\hat{\sigma}_b^2=5A_1+A_2+B \qquad (13)$$

$$\hat{\sigma}_{ao}^2=\frac{10A_1+44A_2-54B}{87}\ ;\ V_{ao}=\frac{20(1+\rho)^2+1936(1+1.75\rho)^2+2916}{7569}$$

$$\hat{\sigma}_{bo}^2=\frac{35A_1-20A_2+43B}{58}\ ;\ V_{bo}=\frac{245(1+\rho)^2+400(1+1.75\rho)^2+1849}{3364} \qquad (14)$$

$\underline{\rho = \infty}$

$$\hat{\sigma}_{b,\infty}^2=B=\hat{\sigma}_b^2\ ;\ 6\hat{\sigma}_a^2+\left(5+\frac{1}{1.75}\right)\hat{\sigma}_b^2=5A_1+\frac{A_2}{1.75}\ ; \qquad (15)$$

$$\hat{\sigma}_{a,\infty}^2=\frac{35A_1+4A_2-39B}{42}\ ;\ V_{a,\infty}=\frac{245(1+\rho)^2+16(1+1.75\rho)^2+1521}{1764}\ . \qquad (16)$$

* *

Table 5. Comparison of Variances of Estimators (N=8)

	ρ			
V	.05	.5	1	10
V_a	.94	1.11	1.40	21.6
$V_{a,\infty}$	1.03	1.21	1.49	20.8
$V_{a,o}$	.69	1.29	2.32	88.2
V_b	1.00	1.00	1.00	1.00
$V_{b,o}$	.77	.80	.87	1.13

These methods could be extended to any design with any number of variance components, provided independent mean squares can be computed. In general let M_h represent the h^{th} mean square, with n_h D.F. and $E\, M_h = \sum_{i=1}^{v} k_{hi}\sigma_i^2$, where there are $\underline{v}$ variance components (some of the k_{hi} will be zero; none will be negative). The i^{th} normal equation will be

$$\sum_{j=1}^{v} \underset{h}{S}\left[w_h k_{hi} k_{hj}\right] \hat{\sigma}_j^2 = \underset{h}{S}\left[w_h k_{hi} M_h\right], \tag{17}$$

where $w_h = n_h/(EM_h)^2$. One can simplify the weights (w_h) by using ratios of variance components, as above.

4. Optimal Designs for a Two-Stage Nested Process.

P. P. Crump (1954) considered these two problems for the two-stage nested process:

(1) Given N, the total sample size, find the best design for use in estimating μ, σ_b^2, σ_a^2 or $\rho = \sigma_a^2/\sigma_b^2$, using the pooled ANOV considered in Section 2.

(2) For certain selected designs, compare some alternative estimators of μ and σ_a^2.

Only (1) will be discussed at this time.

The best design for estimating μ is a = N classes with 1 observation per class and for σ_b^2 is a = 1 class of N observations; the results for σ_a^2 and ρ are intermediate. Given $\underline{a}$ classes, Crump proved that for both σ_a^2 and ρ, the optimum allocation of N was to have $\underline{r}$ classes with p+1 observations per class and $\underline{a} - \underline{r}$ classes with $\underline{p}$ observations per class, where N = ap + r, r (an integer) $<$ a. However, the optimal $\underline{a}$ would be different for estimating σ_a^2 and ρ. It could not be proven rigorously but it was conjectured (and shown to be correct for all computed examples) that the following procedure should be followed to determine the optimal $\underline{a}$:

(1) If $\underline{a}$ did not have to be an integer, it has been shown by several authors that the optimal $\underline{a}$ (the one giving an unbiased minimum variance estimator based on the pooled ANOV) is

$$\text{i) for } \sigma_a^2\text{: } a = a_1 = \frac{N(N\rho+2)}{N(\rho+1)+1} \;;$$

$$\text{ii) for } \rho\text{ : } a = a_2 = 1 + \frac{(N-5)(N\rho+1)}{2N\rho + N - 3} \;. \qquad (18)$$

(2) Since a_1 and a_2 will not, in general, be integers, select the integers just larger and just smaller than a_1 (or a_2); compute the exact variance for each; then whichever variance is smaller, compute the variance for the next adjacent integral value of $\underline{a}$; continue this process until a smallest variance is obtained. In all examples computed, this was found for one of the two original integers. These calculations usually can be omitted since the variance profile is quite flat near the optimum; hence, I suggest using for $\underline{a}$ the integer closest to a_1 (or a_2).

(3) The true variances of $\hat{\sigma}_a^2$ and $\hat{\rho}$ for a given $\underline{a}$ ($N = ap + r$) are:

$$\mathrm{Var}(\hat{\sigma}_a^2) = 2\sigma_b^4(A\rho^2 + B\rho + C) \;,$$

$$\mathrm{Var}(\hat{\rho}) = 2(A'\rho^2 + B'\rho + C'). \qquad (19)$$

$$A = \frac{N^2S_2 - 2NS_3 + S_2^2}{(N^2 - S_2)^2}\;;\; B = \frac{2N}{N^2 - S_2}\;;\; C = \frac{N^2(N-1)(a-1)}{(N-a)(N^2 - S_2)^2}\;;$$

$$A' = \frac{(N-a-2)A+1}{N-a-4}\;;\; B' = \frac{(N-3)B}{N-a-4}\;;\; C' = \frac{(N-a)(N-3)}{(N-1)(N-a-4)}C;$$

$$S_2 = Np + r(p+1)\;;\; S_3 = Np^2 + r(p+1)(2p+1).$$

Crump presents values of A, B, C, A', B', and C' for selected values of $\underline{a}$ for $N = 10, 20, 30$, and 100, Crump compared the minimum variances for $\hat{\sigma}_a^2$ and $\hat{\rho}$ following the above procedure with the theoretical minimum disregarding the fact that a and the number of observations per class must be integers. Some of these efficiency ratios of variances (E_o) are given in Table 6 for $\hat{\sigma}_a^2$.

Table 6. Values of $E_o(\hat{\sigma}_a^2)$ [a]

N	ρ: 0	.50	1.00	2.00
10	.988	.981	.980	.940
20	.997	.987	.995	.943
100	1.000	.998	1.000	.947

[a] $E_o = \frac{\text{min. hypothetical variance}}{\text{min. realizable variance}}$

These results indicate that the procedure outlined above must be quite good.

Recognizing that the allocation which produces a good estimator of one parameter may be poor for estimating some other parameter, Crump computed the efficiency ratio (E_f) for all four parameters, for various allocation plans. These are presented in Table 7. This table presents (for each parameter) the ratio of the variance for the optimum allocation to estimate that parameter to the variance for the given allocation. For example with N = 10 and $\rho = 1$, the best allocation plans for μ, σ_b^2, σ_a^2 and ρ are, respectively, a = 10, 1,5 and 3. The minimum respective variances are $\sigma_b^2/5$; $2\sigma_b^4/9$; $2\sigma_b^4(.6125)$; $2(1.98775)$. The respective variances for a = 4 are $\sigma_b^2(.36)$; $2\sigma_b^4/6$; $2\sigma_b^4(.6950)$; $2(2.32285)$. The ratios of the minimum variances to those for a = 4 give the respective efficiency ratios, E_f: $5/9 = .56$; $2/3 = .67$; $.6125/.6950 = .88$; $1.98775/2.32285 = .86$.

Table 7. Efficiency Ratios (E_f) for $\hat{\mu}$, $\hat{\sigma}_b^2$, $\hat{\sigma}_a^2$, $\hat{\rho}$. 1/

N	ρ	a	$E_f(\hat{\mu})$	$E_f(\hat{\sigma}_b^2)$	$E_f(\hat{\sigma}_a^2)$	$E_f(\hat{\rho})$
10	0.5	2	.43	.89	.61	.88
		3	.56	.78	.89	1.00
		4	.65	.67	.99	.78
		5	.75	.56	1.00	.43
	1.0	2	.33	.89	.42	.79
		3	.45	.78	.71	1.00
		4	.56	.67	.88	.86
		5	.67	.56	1.00	.50
20	0.5	5	.50	.79	.90	1.00
		7	.61	.68	1.00	.95
		8	.65	.63	.99	.87
		10	.75	.53	.96	.69
	1.0	5	.40	.79	.70	.96
		7	.51	.68	.87	1.00
		8	.56	.63	.92	.96
		10	.67	.53	1.00	.82
		11	.70	.47	.98	.69
	2.0	5	.33	.79	.54	.90
		7	.44	.68	.72	1.00
		10	.60	.53	.95	.89
		11	.63	.47	.97	.82
		14	.71	.32	1.00	.36
100	0.5	10	.25	.91	.58	.71
		20	.43	.81	.88	.97
		25	.50	.76	.96	1.00
		33	.60	.68	1.00	.96
		35	.61	.66	.99	.94
		50	.75	.51	.91	.74
	1.0	25	.40	.76	.77	.96
		33	.49	.68	.89	1.00
		40	.56	.61	.94	.97
		50	.67	.51	1.00	.88
		60	.71	.40	.95	.72
	2.0	25	.33	.76	.61	.90
		38	.47	.63	.81	1.00
		50	.60	.51	.97	.97
		66	.69	.34	1.00	.72
		75	.82	.25	.98	.54

1/ $E_f = \dfrac{\text{Variance for optimum allocation at optimum a}}{\text{Variance for optimum allocation at given a}}$

An examination of Table 7 indicates that large gains can be achieved in the efficiency of estimates of σ_a^2 and ρ by suitable choice of the sampling plan, for fixed total number of observations, N. For N = 20, $\rho = 1.0$, one should use 7,8,9, or 10 classes depending on the relative importance of the two parameters, assuming μ and σ_b^2 were of secondary importance. Similarly if $\rho = 2.0$, one should use between 7 and 14 classes. And if $\rho = 0.5$, use 5, 6, or 7 classes. It is important to note that if one thought that $0.5 < \rho < 2.0$, he might logically choose 7,8 or 9 classes without departing too far from the optimum. For N = 100 and ρ somewhere between 0.5 and 2.0, he could choose around 40 classes without much risk of departing far from the optimum.

One can use the results in Table 7 to study the effects of using the incorrect ρ in making his allocations. Some of these comparisons plus others regarding the estimation of σ_a^2 have been made in Table 8. In general it appears that if the true ρ is 1 or less, one will not lose more than 10% in efficiency even if he uses 2ρ or $\rho/2$. The loss is considerably less when ρ is as large as 2.

5. Optimal Designs for a Two-Way Classification Model

5.1 Introduction. D. W. Gaylor (1960) considered methods of sampling, under certain restrictions, to minimize the variance of certain estimators of variance components for a two-way crossed classification

$$Y_{ijk} = \mu + r_i + c_j + (rc)_{ij} + s_{ijk}\,; \quad \begin{array}{l} i = 1,2,\ldots,r < N \\ j = 1,2,\ldots,c, < N, \end{array} \tag{20}$$

where the effects are assumed normally and independently distributed with zero means and respective variances, $\sigma_r^2, \sigma_c^2, \sigma_{rc}^2$ and σ_s^2 and n_{ij} observations are obtained in the (ij) cell. It is assumed that a total of $N = \sum\sum n_{ij}$ samples are allocated to the rc cells. The problem is to make the allocation so that the variances of the estimates of the variance components will be as small as

Table 8. Effect of Incorrect Estimate of ρ on the Efficiency of Estimates of σ_a^2

N	ρ	Optimal No. Classes (a)	Estimated ρ	No. Classes Used (a)	Efficiency
20	0.5	7	0.22	5	.903
			0.31	6	.964
			0.68	9	.985
			0.85	10	.960
			1.06	11	.897
	1.0	10	0.41	7	.869
			0.53	8	.918
			1.32	12	.957
			1.66	13	.921
			2.12	14	.869
	2.0	14	0.41	7	.725
			0.68	9	.868
			0.85	10	.947
			1.06	11	.969
			3.70	16	.951
100	0.5	33	0.14	14	.739
			0.26	22	.911
			0.59	38	.986
			0.97	50	.912
			1.46	60	.775
	1.0	50	0.31	25	.772
			0.47	33	.893
			0.64	40	.944
			1.46	60	.946
			1.90	66	.894
	2.0	66	0.64	40	.832
			0.97	50	.969
			2.95	75	.976
			3.94	80	.935

possible. This again poses the problem of deciding which parameters are most important. Gaylor studied optimal allocations for individual components and for certain combinations.

Given any quadratic estimator

$$Q_i = \underline{y}'M_i\,\underline{y} \tag{21}$$

where $\underline{y}$ is the $(N \times 1)$ vector of observations and $M_i = M_i'$ an $N \times N$ matrix of coefficients. In vector form, we can write

$$\underline{y}(N \times 1) = \underline{\mu}(N \times 1) + \underline{e}(N \times 1), \tag{22}$$

where $\underline{e}$ is the error or random effects vector. The variance-covariance matrix of the e's is

$$V = E(\underline{e}\,\underline{e}'). \tag{23}$$

The diagonal elements of V will be

$$\sigma_r^2 + \sigma_c^2 + \sigma_{rc}^2 + \sigma_s^2\ . \tag{24}$$

The off-diagonal elements will be

$$\sigma_r^2, \sigma_c^2, \sigma_r^2 + \sigma_c^2 + \sigma_{rc}^2 \text{ or } 0, \tag{25}$$

depending on whether the two corresponding observations are in the same row but different columns, the same column but different rows, the same row and column or in different rows and columns. It is easy to show that, for unbiased estimators,

$$E(Q_i) = tr(VM_i) = \sigma_i^2\ ;\ Var(Q_i) = 2tr(VM_i)^2\ , \tag{26}$$

where σ_i^2 is the variance component to be estimated and "tr" refers to the trace of the matrix. The lower bound to the variance is

$$L.B.Var(Q_i) = 2\sigma_i^4/(N-1). \tag{27}$$

For the following cases, this lower bound can be realized; for others, it usually is not possible.

Function estimated	r	c	n_{ij}
σ_s^2	1	1	N
$\sigma_s^2 + \sigma_{rc}^2 + \sigma_r^2$	N	1	1
$\sigma_s^2 + \sigma_{rc}^2 + \sigma_c^2$	1	N	1
$\sigma_s^2 + \sigma_{rc}^2 + \sigma_r^2 + \sigma_c^2$	N	N	1 or 0*

(28)

* each observation from a different row and column

The estimator in each case is $S(y - \bar{y})^2/(N-1)$. The last design is also the optimal design for estimating μ. The variance of $\bar{y}$ is then

$$(\sigma_s^2 + \sigma_{rc}^2 + \sigma_r^2 + \sigma_c^2)/N. \tag{29}$$

One special case is of importance. If $\sigma_{rc}^2 = 0$, the optimal design for estimating σ_r^2 or $\rho_r = \sigma_r^2/\sigma_s^2$ consists of one column; in this case the results by Crump apply. Similarly, if only σ_c^2 or ρ_c is of interest, use one row and apply Crump's results.

5.2 Estimation of σ_r^2 or ρ_r when $\sigma_{rc}^2 > 0$. It was not possible to develop a completely general class of designs to estimate σ_r^2 or $\rho_r = \sigma_r^2/\sigma^2$, where $\sigma^2 = \sigma_s^2 + \sigma_{rc}^2$, which could be proven optimal for all situations. It was first shown that if the design were limited to a class in which each $n_{ij} = 0$ or n(an integer) observations per cell, n should equal 1. Hence each cell should either be empty or contain one observation. The sums of squares used for estimating σ_r^2 and ρ_r were those given in an ANOV table based on the method of fitting constants (Table 9). The estimators of σ_r^2 and ρ_r and the variances of these estimators are given in equations (30) and (31). These results are based on the fact that $(r-1)R^*$ can be subdivided into r-1 independent quantities.

Table 9. Analysis of Variance for 2-Way Classification Data

SV	DF	MS	EMS
Columns	$c - 1$	C	$\sigma_s^2 + \sigma_{rc}^2 + c_1\sigma_r^2 + r_o\sigma_c^2$
Rows (adjusted for cols.)	$r - 1$	R^*	$\sigma_s^2 + \sigma_{rc}^2 + c_o\sigma_r^2$
Interaction (adjusted for rows and cols.)	$N-r-c+1$	I^*	$\sigma_s^2 + \sigma_{rc}^2$
Total	$N - 1$		

$$\hat{\sigma}_r^2 = (R^* - I^*)/c_o \;;\; \hat{\rho}_r = \left[\frac{N-r-c-1}{N-r-c+1} F' - 1\right] / c_o \,, \tag{30}$$

where $c_o = (N-c)/(r-1)$ and $F' = R^*/I^*$.

$$\frac{\text{Var}(\hat{\sigma}_r^2)}{2\sigma^4} = \frac{r-1}{m(N-c)} + \frac{2\rho_r}{N-c} + \frac{\rho_r^2}{r-1} + \frac{\rho_r^2}{(N-c)^2}\sum_{i=1}^{r-1}(d_i - \bar{d})^2 \;; \tag{31a}$$

$$\frac{\text{Var}(\hat{\rho}_r)}{2} = \frac{(r-1)(1+c_o\rho_r)^2(m+r-3) + (m-2)\rho_r^2\sum(d_i - \bar{d})^2}{c_o^2(r-1)^2(m-4)} \tag{31b}$$

where $m = N-r-c+1$ (>4 for $\hat{\rho}_r$), $\sigma^2 = \sigma_s^2 + \sigma_{rc}^2$, $\rho_r = \sigma_r^2/\sigma^2$ and

$$(r-1)R^* = \sigma^2\sum_{i=1}^{r-1}(1 + d_i\rho_r)\chi_i^2 \,.$$

*

If $N = rc$,

$$\text{Var}(\hat{\sigma}_r^2) = \frac{2\sigma^4}{r-1}\left[\frac{1}{c(c-1)} + \frac{2\rho_r}{c} + \rho_r^2\right]; \tag{32a}$$

$$\text{Var}(\hat{\rho}_r) = 2(1+c\rho_r)^2(N-c-2)/c^2(r-1)(N-r-c-3) \,. \tag{32b}$$

All statements refer to the above estimators. Admittedly other estimators might be more efficient for a given design.

If for a given N and r, N is an integral multiple of r, N = rc, then the best design with $n_{ij} = 0$ or n to estimate σ_r^2 or ρ_r consists of r rows and c columns with one observation per cell. Graybill and Wortham (1956) show that the resultant $\hat{\sigma}_r^2$ is a minimum variance unbiased quadratic estimator. In this case in Table 9, $c_1 = 0$, $c_o = c$ and $r_o = r$, and $d_i = \bar{d} = c$. The variances of the estimators are given in equations (32).

In general N/r will not be an integer. Suppose $N = r(k-1) + s$, $0 < s < r$. For a given r, a design of the type $n_{ij} = 0$ or 1 which minimizes $\mathrm{Var}(\hat{\sigma}_r^2)$ based on the above ANOV table consists of r rows by k-1 columns (using only N-s observations) or r rows by k-1 columns plus one column with s of the r rows. The variances are given in equations (33). This leads to the interesting result that a balanced design may lead to a smaller variance than an unbalanced design with more observations; this situation arises when ρ_r is large and indicates that presumably the estimators used are not the best for the unbalanced situation.

In order to find the optimal value of r, one should minimize (33a) or (33b). This is quite complicated because k and s are functions of r. Because of the near balance of the design, a good approximation to (33) is available by replacing each d_i in (31) by $\bar{d} = c_o = (N-c)/(r-1)$ and c by $c_o + 1/2$ (since $c_o \leq c \leq c_o + 1$). These approximate variances are given in equations (34). Setting the derivative of (34a) with respect to c_o equal to 0, we have

$$c_o \doteq \frac{\rho_r^2(N-1/2) + 2\rho_r - 1 + \sqrt{\rho_r^2(N-1/2)(N-3/2) + 2\rho_r(N-3/2) + 1}}{\rho_r^2(N-1/2) + 2\rho_r} \qquad (35)$$

This is approximated by (36a) with the approximate value of r in (37a). Corresponding results for $\hat{\rho}_r$ are presented in (36b) and (37b).

If $N = r(k-1) + s,\ 0 < s < r$

$$\frac{\mathrm{Var}(\hat{\sigma}_r^2)}{2\sigma^4} = \begin{cases} \left[\rho_r^2 + 2\rho_r/(k-1) + 1/(k-1)(k-2)\right]/(r-1), \text{ using } k-1 \text{ cols.} \\ \dfrac{\rho_r^2}{r-1} + \dfrac{\rho_r^2(s-1)(r-s)}{(N-k)^2(r-1)} + \dfrac{2\rho_r}{N-k} + \dfrac{r-1}{(N-k)(N-k-r+1)}, \text{ using } k \text{ cols.} \end{cases} \tag{33a}$$

$$\frac{\mathrm{Var}(\hat{\rho}_r)}{2} = \begin{cases} \left[1 + (k-1)\rho_r\right]^2(rk-r-k-1)/(k-1)^2(r-1)(rk-2r-k-2), \text{using } k-1 \text{ cols.} \\ \dfrac{(N-r-k-1)(s-1)(r-s)\rho_r^2 + \left[(r-1) + (N-k)\rho_r\right]^2(N-k-2)}{(N-k)^2(r-1)(N-r-k-3)}, \text{ using } k \text{ cols.} \end{cases} \tag{33b}$$

Using all N observations,

$$\frac{\mathrm{Var}(\hat{\sigma}_r^2)}{2\sigma^4} \doteq \frac{1 + 2c_o\rho_r + c_o^2\rho_r^2 - 2\rho_r - c_o\rho_r^2}{(c_o-1)(N-c_o-1/2)} . \tag{34a}$$

$$\frac{\mathrm{Var}(\hat{\rho}_r)}{2} \doteq \frac{(1 + c_o\rho_r)^2}{(c_o-1)(N-c_o-1/2)} . \tag{34b}$$

For $\hat{\sigma}_r^2$,

$$c_o \doteq \tilde{\tilde{c}}_o = \frac{\rho_r(N-1/2) + (N-1/2) + 1}{\rho_r(N-1/2) + 2} \longrightarrow 1 + 1/\rho_r \text{ for large } N, \tag{36a}$$

and

$$r \doteq \tilde{\tilde{r}} = (N - 1/2)/\tilde{\tilde{c}}_o . \tag{37a}$$

For $\hat{\rho}_r$,

$$c_o \doteq \tilde{c}_o = \frac{2\rho_r(N-1/2) + (N-1/2) + 1}{\rho_r(N-1/2) + \rho_r + 2} \longrightarrow 2 + 1/\rho_r \text{ for large } N, \tag{36b}$$

$$r \doteq \tilde{r} = (N - 1/2)/\tilde{c}_o . \tag{37b}$$

Again one can consider integers above and below $\tilde{\tilde{r}}$ or $\tilde{r}$ as the optimal r, testing by insertion in the true variance formula (33). As in Crump's case, the variance profile is so flat that the integer closest to *r* will probably be sufficient. Once *r* is determined, the number of columns will be found from the fact that $N = kr$ or $N = (k-1)r+s$, where $s < r$. In the latter case, the experimenter presumably should check whether to use a balanced design with $(k-1)r$ observations or an unbalanced one with *s* rows in the k^{th} column; since it requires a very large ρ_r to prefer the smaller balanced design, the unbalanced design usually will be better. In the latter case, the experimenter might wish to consider the use of enough more observations to have a $(k x r)$ balanced design; matters of this kind need further investigation, such as maximizing the information per observation rather than the total information.

Gaylor also investigated the loss of information due to an incorrect estimate of ρ_r, as given in Table 10. These results are of the same order of magnitude as found by Crump. As before, losses due to use of the incorrect ρ are less serious for large ρ. Gaylor checked some of these results against those found by use of exact variances; the agreement on the efficiency factor was very good; e.g., $E_{f2} = .907$ instead of $.909$ for $N = 30$, $\rho_r = 1.0$, $\rho_r' = 2.0$. It should be noted that Gaylor considered ratios of 4:1 for $\hat{\rho}_r$ and only 2:1 for $\hat{\sigma}_r^2$; hence, some of these efficiencies are lower than for $\hat{\sigma}_r^2$.

Table 11 presents the efficiency of various designs used to obtain information on σ_r^2 and ρ_r for $N = 30$. This table shows that extreme departures from the optimal design results in a considerable loss of efficiency; however, moderate departures usually result in small losses. Where ρ_r is small, a balanced design in the neighborhood of the optimal design has high efficiency for estimating both parameters. However, even for $\rho_r = 1$, the efficiency of each estimator is less than 90% when the optimal design for the other is used. On the other hand, one is able to find an intermediate design which is quite good for estimating both parameters when ρ_r is no larger than 2.

Table 10. Approximate Relative Efficiencies, E_f of the Restricted Optimal Designs for $\hat{\sigma}_r^2$ and $\hat{\rho}_r$ Based on Incorrect Values of ρ.

Results for $\hat{\sigma}_r^2$

N	Design Based on True ρ_r		Design Based on Incorrect ρ_r'					
	ρ_r	$\tilde{\tilde{c}}_o$	ρ'_{r-1}	$\tilde{\tilde{c}}'_{o1}$	E_{f1}	ρ'_{r-2}	$\tilde{\tilde{c}}'_{o2}$	E_{f2}
30	.25	4.04	.125	6.01	.928	.50	2.70	.905
	1.0	1.90	.50	2.70	.920	2.00	1.47	.909
	4.0	1.24	2.00	1.47	.943	8.00	1.12	.939
100	.25	4.67	.125	7.82	.907	.50	2.90	.900
	1.0	1.97	.50	2.90	.912	2.00	1.49	.909
	4.0	1.25	2.00	1.49	.942	8.00	1.12	.936

Results for $\hat{\rho}_r$

True ρ_r	ρ_r' Used	E_f	
		N = 30	N = 100
0.25	1.00	.849	.829
1.00	0.25	.847	.823
0.50	2.00	.900	.891
2.00	0.50	.907	.896
2.00	8.00	.970	.976
8.00	2.00	.993	.983

Table 11. Efficiency of Some Designs for Estimating σ_r^2 and ρ_r, N = 30 a/

ρ_r	r	c	s	$E_f(\hat{\sigma}_r^2)$	$E_f(\hat{\rho}_r)$
.25	2	15	0	.422	.446
	3	10	0	.687	.739
	5	6	0	.948	.975
	6	5	0	.999	1.000
	7	5	2	.984	.947
	7*	4*	0*	.941*	.889*
	8	4	6	1.000	.943
	9	4	3	.959	.871
	10	3	0	.965	.862
	15	2	0	.732	.557
.50	5	6	0	---	.889
	6	5	0	.833	.964
	7	5	2	.882	.968
	7*	4*	0*	.857*	.922*
	8	4	6	.938	1.000
	9	4	3	.954	.967
	10	3	0	1.000	.986
	11	3	8	.986	.927
	15	2	0	.933	.732
1.0	6	5	0	---	.852
	7*	4*	0*	---	.862*
	9	4	3	---	.956
	10	3	0	.877	1.000
	11	3	8	.899	.968
	14	3	2	.957	.833
	15	2	0	1.000	.844
	16	2	14	.984	---
	19	2	11	.907	---
2.0	6	5	0	---	.776
	7*	4*	0*	---	.814*
	9	4	3	---	.938
	10	3	0	.740	1.000
	11	3	8	---	.987
	12	3	6	.822	.969
	15	2	0	.974	.931
	18	2	12	.997	---
	19	2	11	1.000	.631
	20	2	10	.998	---
	24	2	6	.902	---
4.0	6	5	0	---	.735
	7*	4*	0*	---	.788*
	9	4	3	---	.927
	10	3	0	.586	1.000
	11	3	8	---	1.000
	12	3	6	---	.994
	15	2	0	.837	.991
	20	2	10	.943	.527
	23	2	7	.994	---
	24	2	6	1.000	---
	25	2	5	.993	---
	28	2	2	.658	---

$E_f(\hat{\rho}_r)$ only for ρ_r = 8 and 16

r	c	s	$\rho_r = 8$	$\rho_r = 16$
10	3	0	.974	.955
11	3	8	.981	.965
12	3	6	.982	.970
13	3	4	.977	.969
14	3	2	.963	.960
15	2	0	1.000	1.000
16	2	14	.945	.948
20	2	10	.634	.646

a/ r = No. rows ; c = No. columns. If s = o, N = rc; if s >o, N = r(c-1) + s;
* designs have only 28 observations.

The two sets of restricted optimal designs begin to diverge considerably as ρ_r increases beyond 2. One notes that the limiting value of $\underline{c}$ is $1 + 1/\rho_r$ for $\hat{\sigma}_r^2$ and $2 + 1/\rho_r$ for $\hat{\rho}_r$. For $\rho_r = 1$, the optimal designs have 2 and 3 columns, respectively. However, for $\rho > 1$, the optimal design for $\hat{\sigma}_r^2$ is non-balanced with 2 columns, whereas that for $\hat{\rho}_r$ approaches a balanced design with 2 columns.

The particular case of $c = 2$ has been investigated more thoroughly for $\hat{\sigma}_r^2$. Consider a two-column restricted optimal design with $N-r'$ rows in one column and $\underline{r}'$ of these same rows occupied in the second column; i.e., $s = 0$ if $r' = N/2$ or $2 \leq r' = s < N/2$. From (33a),

$$\frac{\operatorname{Var}(\hat{\sigma}_r^2)}{2\sigma^4} = \left[\frac{N + 2r' - 4}{N-2}\ \rho_r^2 + 2\rho_r + \frac{N - r' - 1}{r' - 1}\right] / (N-2). \tag{38}$$

If $r' = N/2$, (38) simplifies to $(\rho_r^2 + \rho_r + 0.5)/(r'-1)$. The value of $\underline{r}'$ which minimizes (38) is

$$r' = \begin{cases} 2 & \rho_r \geq (N-2)/\sqrt{2} \\ 1 + (N-2)/\rho_r\sqrt{2} & \sqrt{2} \leq \rho_r \leq (N-2)/\sqrt{2} \\ N/2 & \rho_r \leq \sqrt{2} \end{cases} \tag{39}$$

Hence when $\rho_r > \sqrt{2}$, a two-column design with one column shorter than the other is better than two columns of equal length.

These results suggest the following procedure for finding a nearly optimal design for $\hat{\sigma}_r^2$:

(i) If $\rho_r > \sqrt{2}$, use one column with $r = N - r'$ rows and a second column with r' of these rows, where r' is the integer (≥ 2) which is nearest $1 + (N-2)/\rho_r\sqrt{2}$.

(ii) When $\rho_r \leq \sqrt{2}$, use a balanced design with $\underline{c}$ as the integer above or below

$$\tilde{\tilde{c}}_0 = \left[\rho_r(N-1/2) + N + 1/2\right] / \left[\rho_r(N-1/2) + 2\right],$$

which minimizes

$$V(c) = \frac{1}{(N-c)(c-1)} + \frac{2\rho_r}{N-c} + \frac{c\rho_r^2}{N-c}, \tag{40}$$

and where $\underline{r}$ was set equal to N/c. Presumably little information is lost by simply using as $\underline{c}$ the integer closer to $\tilde{\tilde{c}}_o$.

Another approach to (ii) above is to determine intervals of ρ_r for which $c = 3$ is best, $c = 4$ is best, ... This is accomplished by determining $\rho_r = \rho(c)$ in the equation

$$V(c) = V(c-1). \tag{41}$$

When $\rho_r < \rho(c)$, $\underline{c}$ is preferred to c-1 columns. The solution to (41) is

$$\rho_1(c) = \frac{1}{\sqrt{(c-1)(c-2)}} - \frac{1}{N} - \frac{1}{N}\sqrt{\frac{c-1}{c-2}}\left[1 + 1/2N + O(1/N^2)\right] \tag{42}$$

For $N = 30$, the following are the first three separation points:

$$\rho_1(3) = .62, \quad \rho_1(4) = .33, \quad \rho_1(5) = .22.$$

Hence

$$c = 2 \text{ for } .62 < \rho_r < \sqrt{2};$$

$$c = 3 \text{ for } .33 < \rho_r < .62$$

$$c = 4 \text{ for } .22 < \rho_r < .33.$$

Note that for $\rho_r = .25$ in Table 11, $c = 4$ is slightly preferred to $c = 5$.

Table 11 indicates that for $\hat{\rho}_r$ there is a considerable range of ρ_r for which the best allocation plan will have 3 columns and N/3 rows. If N is not divisible by 3, some gain may be obtained by using an extra row with one column not filled. It appears that the best plan would be to either increase N or reduce N by enough to have it divisible by 3. As ρ_r increases, only 2 columns should be used with N/2 rows. Again add or subtract one to make N even.

The following general procedure is suggested for $\hat{\rho}_r$:

(i) Use a balanced design with c as the integer above or below (36b),

$$\tilde{c}_o = (2\rho_r N + N - \rho_r + 1/2)/(\rho_r N + \rho_r/2 + 2),$$

which minimizes (32b):

$$\frac{\mathrm{Var}(\hat{\rho}_r)}{2} = (1 + c\rho_r)^2 (N-c-2)/(N-c)(Nc-N-c^2-3c) \ . \tag{43}$$

(ii) Choose r so that $rc \doteq N$.

The same procedure of finding separation points on ρ_r could be used on (43) as was done above. The equation comparable to (41) is considerably more complicated; hence, all I am reporting here is the asymptotic result (for large N). In this case the solution is

$$\rho_2(c) = \frac{1 + \sqrt{(c-1)(c-2)}}{c^2 - 3c + 1} \ . \tag{44}$$

Some of the separation points are:

$\rho_2(3) = 1 + \sqrt{2} = 2.4$, $\rho_2(4) = (1 + \sqrt{6})/5 = .7$, $\rho_2(5) = (1 + 2\sqrt{3})/11 = .4$;

$\rho_2(6) = (1 + 2\sqrt{5})/19 = .28$.

Hence

$c = 3$ for $.7 < \rho_r < 2.4$

$c = 4$ for $.4 < \rho_r < .7$

$c = 5$ for $.28 < \rho_r < .4$

Note that in Table 11, $c = 5$ is slightly preferred to $c = 6$ for $\rho = .25$. The above separation points would be altered slightly if the adjustments of $O(1/N)$ were made.

5.3 Estimation of σ_c^2 or $\rho_c = \sigma_c^2/\sigma^2$. Use the results in Section 5.2 by reversing the roles of r and c. In Table 9, one computes mean squares, R and C^*, for Rows and Columns (adjusted for rows). The latter can be computed quickly by use of the identity:

$$(r-1)R^* + (c-1)C = (r-1)R + (c-1)C^* . \tag{45}$$

5.4 Simultaneous Estimation of σ_r^2 and σ_c^2. Gaylor obtained only some tentative results for this case but they are indicative of the problem. Since efficient estimation of σ_r^2 requires many rows and efficient estimation of σ_c^2 requires many columns, simultaneous estimation requires many rows and columns. Gaylor considered two special designs, called the L-design and the Balanced Disjoint Rectangles (BDR)-design, and compared them for the case $\sigma_r^2 = \sigma_c^2$.

The L-design is as follows:

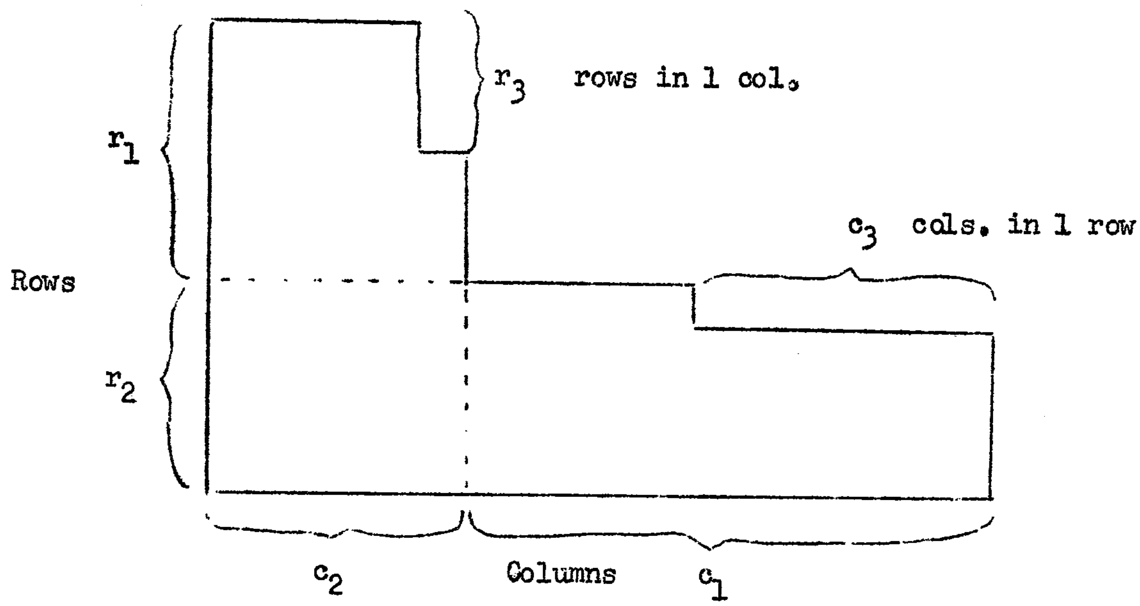


Figure 1. The L-Design

For $\sigma_r^2 = \sigma_c^2$, one should use $r_1 = c_1$, $r_2 = c_2$, $r_3 = c_3$. The total rows and columns used are $r = r_1 + r_2$ and $c = c_1 + c_2$. $N_1 = rc_2 - r_3$ observations contain one column with $r_3 (0 \leq r_3 < r_1)$ empty cells and $N_2 = cr_2 - c_3$ observations contain one row with c_3 $(0 \leq c_3 < c_1)$ empty cells. The N_1 observations are of the form of a restricted optimal design for $\hat{\sigma}_r^2$ and the N_2 observations for $\hat{\sigma}_c^2$. In general c_2 and r_2 will be small.

The BDR design consists of g distinct rectangles, each consisting of r rows by c colums, with one observation per cell. The rectangles are distinct in that each rectangle samples a different set of r rows and c columns. The ANOV for this design is given in Table 12.

If the component of variance for the rectangle effects is zero, $\sigma_g^2 = 0$, then a unique solution for the estimators of the components of variance is not obtained by equating the mean squares to their expected values. Then, there are five mean squares (equations) with which to estimate four components of variance. Since the best procedure in this case is unknown, the case will not be considered here, but this certainly warrants future investigation.

If G is neglected, the estimators and their variances are quite simple, as given in equations (46).

The two designs have been compared for the special case $\sigma_r^2 = \sigma_c^2$ and $\mathrm{Var}(\hat{\sigma}_r^2) = \mathrm{Var}(\hat{\sigma}_c^2)$. Some specific examples are given in Table 13 for $\rho = \sigma_r^2/\sigma^2 = \sigma_c^2/\sigma^2$, where $\sigma^2 = \sigma_s^2 + \sigma_{rc}^2$. The total number of observations, N, was allowed to vary slightly in order to obtain balance if desired. Thus comparisons were made on the basis of $N\mathrm{Var}(\sigma_r^2) = N\mathrm{Var}(\sigma_c^2)$ which is the reciprocal of the amount of information per observation. The estimator for σ_r^2 for the BDR-design is given in (46) and for the L-design in (30). The variances in the last column of Table 13, $N\dot{V}$, are based on the estimator, $\dot{\sigma}_r^2$, in which the last c_1 columns in Figure 1 for the L-design are ignored. The estimator, σ_r^2, could be improved slightly by pooling the error sum of squares from both legs of the L-design.

Several observations can be made from Table 13. When ρ is small, BDR-designs are better than the L-designs. When ρ is large, the L-design is better. For those designs where $\rho^2 \sum (d_i - \bar{d})^2$ is large, see (31a), the estimator, $\dot{\sigma}_r^2$,

Table 12. ANOV for a Balanced Disjoint Rectangles Design

SV	DF	MS	EMS
Rectangles	g-1	G	$\sigma_s^2 + \sigma_{rc}^2 + c\sigma_r^2 + r\sigma_c^2 + rc\sigma_g^2$
Rows	g(r-1)	R	$\sigma_s^2 + \sigma_{rc}^2 + c\sigma_r^2$
Columns	g(c-1)	C	$\sigma_s^2 + \sigma_{rc}^2 + r\sigma_c^2$
Interaction	g(r-1)(c-1)	I	$\sigma_s^2 + \sigma_{rc}^2$

$\hat{\sigma}_r^2 = (R-I)/c$; $\sigma_c^2 = (C-I)/r$;

$$\mathrm{Var}(\hat{\sigma}_r^2) = \frac{2r}{N(r-1)c}\left[(\sigma_s^2 + \sigma_{rc}^2 + c\sigma_r^2)^2 + \frac{(\sigma_s^2 + \sigma_{rc}^2)^2}{c-1}\right] ; \tag{46}$$

$$\mathrm{Var}(\hat{\sigma}_c^2) = \frac{2c}{N(c-1)r}\left[(\sigma_s^2 + \sigma_{rc}^2 + r\sigma_c^2)^2 + \frac{(\sigma_s^2 + \sigma_{rc}^2)^2}{r-1}\right] .$$

* * * * * * ** *

Table 13. Comparison of BDR and L Designs with $\mathrm{Var}(\hat{\sigma}_r^2) = \mathrm{Var}(\hat{\sigma}_c^2)$, Where $\sigma_r^2 = \sigma_c^2$.

	BDR-Design				L-Design					
ρ	g	r=c	N	NV a/	r=c	$r_2=c_2$	$r_3=c_3$	N	NV	$N\mathring{V}$
.5	4	3	36	3.38	10	2	0	36	4.24	5.00
.5	2	4	32	3.11	7	3	0	33	3.29	4.12
.5	1	6	36	3.24	6	6	0	36	3.24	3.24
1	9	2	36	10.00	10	2	0	36	10.89	10.00
1	4	3	36	8.25	7	3	0	33	9.46	10.08
1	2	4	32	8.44	6	6	0	36	9.84	9.84
2	9	2	36	26.0	14	2	8	36	26.8	25.6
2	4	3	36	24.8	10	2	0	36	34.8	26.0
2	2	4	32	27.1	7	3	0	33	31.6	30.2
4	9	2	36	82.0	16	2	12	36	72	79.8
4	4	3	36	84.8	10	2	0	36	125	82.0

a/ $V = \mathrm{Var}(\hat{\sigma}_r^2)/2\sigma^4$ and $\mathring{V} = \mathrm{Var}(\mathring{\sigma}_r^2)/2\sigma^4$.

which discards the last c_1 columns is often better than the estimator, $\hat{\sigma}_r^2$, which uses all of the observations. This is a result found previously where it was shown desirable to keep the numbers of observations per row nearly equal. Also, as was shown in Section 5.2, the L-design is desirable which has less than two complete columns in the N_1 observations and less than two complete rows in the N_2 observations when $\rho > \sqrt{2}$.

For an L-design with $r_2 = c_2 = 2$, $r = c$, and $r_3 = c_3 = 0$,

$$\mathrm{Var}\ (\tilde{\sigma}_r^2) = 2\sigma^4(2 + 4\rho_r + 4\rho_r^2)/N, \tag{47}$$

where $\rho_r = \sigma_r^2/\sigma^2$. For a BDR-design with $r = c = 2$ and $g = N/4$, from (46), $\mathrm{Var}(\hat{\sigma}_r^2)$ is the same as (47). This result is borne out in Table 13. Similarly, $\mathrm{Var}(\tilde{\sigma}_c^2) = \mathrm{Var}(\hat{\sigma}_c^2)$. No restriction that $\sigma_r^2 = \sigma_c^2$ has been imposed here.

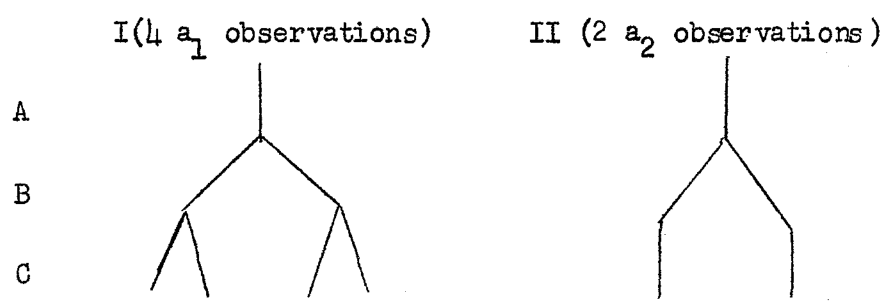
When $\rho \geq 2.87$, the best BDR-design, has $r = c = 2$ and $g = N/4$. Since $\rho \geq \sqrt{2}$, the L-design with $r_2 = c_2 = 2$ can be improved upon by letting $r_3 = c_3 > 0$. Hence, when $\rho \geq 2.87$, for the L-design $\mathrm{Var}(\tilde{\sigma}_r^2) \leq \mathrm{Var}(\hat{\sigma}_r^2)$ for a BDR-design. $\mathrm{Var}(\tilde{\sigma}_r^2)$ could be reduced still further by using the observations in the last c_1 columns to estimate σ^2 and pooling this with the estimate of σ^2 from the first c_2 columns. Also, when $r = c$ is small for a BDR-design, the number of rectangles, g, becomes large. If $\sigma_g^2 = 0$, considerable information may be lost here by not making use of the mean squares among rectangles.

A few tentative observations can be made. In general, as ρ increases, the L-design is favored over disjoint rectangles if $\sigma_g^2 > 0$. Also, the analysis which ignores the observations in the other leg of the L-design in order to achieve balance sometimes reduces the variance.

6. Additional Comments and Suggestions for Future Research

6.1 Extensions of Nested Sampling to More than Two Stages. One important research problem is to extend Crump's results to more than two stages of sampling. In Section 1, two five-stage problems were mentioned. In the Anderson-Bancroft book and in my International Statistical Institute paper, I proposed a so-called staggered design, which consists of a number of balanced designs but with incomplete sampling at some stages.

For three stages, two different groups (I and II) would be used with a_1 first-stage samples for I and a_2 for II, as follows:



ANALYSIS OF VARIANCE

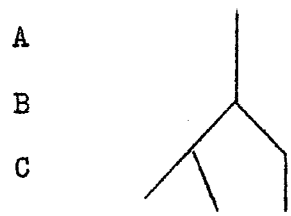
SV	DF	MS	EMS
Groups	1	MSA_3	$\sigma_c^2 + k_b\sigma_b^2 + k_a\sigma_a^2$
A_1	a_1-1	MSA_1	$\sigma_c^2 + 2\sigma_b^2 + 4\sigma_a^2$
A_2	a_2-1	MSA_2	$\sigma_c^2 + \sigma_b^2 + 2\sigma_a^2$
B_1	a_1	MSB_1	$\sigma_c^2 + 2\sigma_b^2$
B_2	a_2	MSB_2	$\sigma_c^2 + \sigma_b^2$
C	$2a_1$	MSC	σ_c^2

where $MSA_3 = (T_1/4a_1 - T_2/2a_2)^2 / (1/4a_1 + 1/2a_2)$, $k_b = \dfrac{2(a_1 + a_2)}{2a_1 + a_2}$,

$k_a = \dfrac{4(a_1 + a_2)}{2a_1 + a_2}$, $T_1 = SY_I$ and $T_2 = SY_{II}$.

One might consider using an additional third plan with a_3 first-stage samples, only one second-stage sample per first-stage and one third-stage sample per second-stage, giving a_3 observations.

R. R. Prairie is studying another plan, which would attempt to equalize the total number of observations at each stage. In this case there would be <u>a</u> first-stage samples, with each followed by the following second- and third-stages:



ANALYSIS OF VARIANCE

SV	DF	MS	EMS
A	a-1	MSA	$\sigma_c^2 + \frac{5}{3}\sigma_b^2 + 3\sigma_a^2$
B	a	MSB	$\sigma_c^2 + \frac{4}{3}\sigma_b^2$
C	a	MSC	σ_c^2

The variance of $\hat{\sigma}_a^2$ for this design is complicated by the fact that MSA and MSB are positively correlated. This is a feature of non-balanced designs which we were not anticipating. It indicates a fact which many of us often forget; namely, that the usually accepted concepts regarding the analysis of variance are based on a regression and not a variance components model. Naturally if one uses the ML approach, he does not start with such preconceived ideas. Also one never has any trouble when the design is balanced. Note that my staggered design is balanced within groups; the major difficulty there is how to use the between-groups mean square.

6.2 Other Suggestions

1) (a) Develop numerical methods, suitable for high-speed computers, to obtain, under normal theory, the ML estimates of variance components from unbalanced data.

   (b) Study the small sample properties of the ML estimators by empirical means, and, where possible, by analytic means.

   (c) It might be possible to find approximate formulas to estimate the biases and variances of these ML estimates. A comparison of asymptotic variances and empirically determined small sample variances would be useful.

   (d) Compare the small sample properties of the ML estimates with those of more convenient estimates, e.g., ANOV or iterated least squares.

2) Study the effect of non-normality on estimating procedures.

3) Introduce unequal cost factors into the problem.

4) Develop suitable criteria for use in simultaneous estimation of several parameters.

5) Develop a sequential estimation procedure in which the first stage (or stages) would be used to estimate ρ (or various ρ's) needed to design the remaining stages.

REFERENCES

Anderson, R. L. 1960. Uses of variance component analysis in the interpretation of biological experiments. Bulletin Int. Statist. Instit. 37:3:71-90.

Anderson, R. L. and T. A. Bancroft. 1952. Statistical Theory in Research. McGraw-Hill Book Company, Inc., New York.

Crump, P. P. 1954. Optimal designs to estimate the parameters of a variance component model. Unpublished Ph.D. Thesis, North Carolina State College, Raleigh, N.C.

Ganguli, M. 1941. A note on nested sampling. Sankhya 5:449-452.

Gaylor, D. W. 1960. The construction and evaluation of some designs for the estimation of parameters in random models. Unpublished Ph.D. Thesis, North Carolina State College, Raleigh. Institute of Statistics Mimeo Series No. 256.

Graybill, F. A. and A. W. Wortham. 1956. A note on uniformly best unbiased estimators for variance components. J. Amer. Stat. Assoc. 51:266-268.

Newton, R. G., M. W. Philpott, H. Fairfield-Smith and W. G. Wren. 1951. Variability of Malayan rubber. Industrial and Engineering Chemistry 43:329-334.

INSTITUTE OF STATISTICS

NORTH CAROLINA STATE COLLEGE

(Mimeo Series available for distribution at cost)

265. Eicker, Friedhelm. Consistency of parameter-estimates in a linear time-series model. October, 1960.
266. Eicker, Friedhelm. A necessary and sufficient condition for consistency of the LS estimates in linear regression. October, 1960.
267. Smith, W. L. On some general renewal theorems for nonidentically distributed variables. October, 1960.
268. Duncan, D. B. Bayes rules for a common multiple comparisons problem and related Student-t problems. November, 1960.
269. Bose, R. C. Theorems in the additive theory of numbers. November, 1960.
270. Cooper, Dale and D. D. Mason. Available soil moisture as a stochastic process. December, 1960.
271. Eicker, Friedhelm. Central limit theorem and consistency in linear regression. December, 1960.
272. Rigney, Jackson A. The cooperative organization in wildlife statistics. Presented at the 14th Annual Meeting, Southeastern Association of Game and Fish Commissioners, Biloxi, Mississippi, October 23-26, 1960. Published in Mimeo Series, January, 1961.
273. Schutzenberger, M. P. On the definition of a certain class of automata. January, 1961.
274. Roy, S. N. and J. N. Shrizastaza. Inference on treatment effects and design of experiments in relation to such inferences. January, 1961.
275. Ray-Chaudhuri, D. K. An algorithm for a minimum cover of an abstract complex. February, 1961.
276. Lehman, E. H., Jr. and R. L. Anderson. Estimation of the scale parameter in the Weibull distribution using samples censored by time and by number of failures. March, 1961.
277. Hotelling, Harold. The behavior of some standard statistical tests under non-standard conditions. February, 1961.
278. Foata, Dominique. On the construction of Bose-Chaudhuri matrices with help of Abelian group characters. February, 1961.
279. Eicker, Friedhelm. Central limit theorem for sums over sets of random variables. February, 1961.
280. Bland, R. P. A minimum average risk solution for the problem of choosing the largest mean. March, 1961.
281. Williams, J. S., S. N. Roy and C. C. Cockerham. An evaluation of the worth of some selected indices. May, 1961.
282. Roy, S. N. and R. Gnanadesikan. Equality of two dispersion matrices against alternatives of intermediate specificity. April, 1961.
283. Schutzenberger, M. P. On the recurrence of patterns. April, 1961.
284. Bose, R. C. and I. M. Chakravarti. A coding problem arising in the transmission of numerical data. April, 1961.
285. Patel, M. S. Investigations on factorial designs. May, 1961.
286. Bishir, J. W. Two problems in the theory of stochastic branching processes. May, 1961.
287. Konsler, T. R. A quantitative analysis of the growth and regrowth of a forage crop. May, 1961.
288. Zaki, R. M. and R. L. Anderson. Applications of linear programming techniques to some problems of production planning over time. May, 1961.
289. Schutzenberger, M. P. A remark on finite transducers. June, 1961.
290. Schutzenberger, M. P. On the equation $a^{2+n} = b^{2+m}c^{2+p}$ in a free group. June, 1961.
291. Schutzenberger, M. P. On a special class of recurrent events. June, 1961.
292. Bhattacharya, P. K. Some properties of the least square estimator in regression analysis when the 'independent' variables are stochastic. June, 1961.
293. Murthy, V. K. On the general renewal process. June, 1961.
294. Ray-Chaudhuri, D. K. Application of geometry of quadrics of constructing PBIB designs. June, 1961.
295. Bose, R. C. Ternary error correcting codes and fractionally replicated designs. May, 1961.
296. Koop, J. C. Contributions to the general theory of sampling finite populations without replacement and with unequal probabilities. September, 1961.
297. Foradori, G. T. Some non-response sampling theory for two stage designs. Ph.D. Thesis. November, 1961.
298. Mallios, W. S. Some aspects of linear regression systems. Ph.D. Thesis. November, 1961.
299. Taeuber, R. C. On sampling with replacement: an axiomatic approach. Ph.D. Thesis. November, 1961.
300. Gross, A. J. On the construction of burst error correcting codes. August, 1961.
301. Srivastava, J. N. Contribution to the construction and analysis of designs. August, 1961.
302. Hoeffding, Wassily. The strong laws of large numbers for u-statistics. August, 1961.
303. Roy, S. N. Some recent results in normal multivariate confidence bounds. August, 1961.
304. Roy, S. N. Some remarks on normal multivariate analysis of variance. August, 1961.
305. Smith, W. L. A necessary and sufficient condition for the convergence of the renewal density. August, 1961.
306. Smith, W. L. A note on characteristic functions which vanish identically in an interval. September, 1961.
307. Fukushima, Kozo. A comparison of sequential tests for the Poisson parameter. September, 1961.
308. Hall, W. J. Some sequential analogs of Stein's two-stage test. September, 1961.
309. Bhattacharya, P. K. Use of concomitant measurements in the design and analysis of experiments. November, 1961.