

S. R. Searle<sup>2/</sup>

BU-308-M

Texas A and M University, College Station, Texas

and Cornell University, Ithaca, N.Y.

### Abstract

This paper is intended as an expository review of variance components. Part I discusses the models underlying variance component analysis, describing them by means of illustrative examples in biology rather than just through their mathematical definitions. Part II deals with balanced data (equal subclass numbers) and is mainly concerned with the analysis of variance method of estimating variance components from such data. Discussion of its properties includes consideration of unbiased estimators, negative estimates, minimum variance, and the consequences of normality assumptions. Definition of interactions in mixed models is also considered. Part III deals with methods of estimating variance components from unbalanced data (unequal subclass numbers):

→ The analysis of variance method, adjusting for bias with mixed models, the fitting constants method, analysis of means methods, symmetric sums, maximum likelihood and best quadratic unbiased estimation. Specific results for a few easy cases are quoted, as a point of departure for discussing development of more general (and often lengthy) results, to which references are given. Comment is also made on several other topics relating to variance component estimation, such as the many quadratic forms available and the designing of experiments for this purpose, the plotting of the likelihood function and the dispensing with mean unbiasedness as an estimation criterion.

<sup>1/</sup> Paper BU-197 in the Biometrics Unit, Department of Plant Breeding and Biometry, Cornell University, Ithaca, N.Y.

<sup>2/</sup> On leave from Cornell, 1968-9.

*also BU-308-M - Title - The concepts and estimation of variance components: an expository review*

*The concepts and estimation of variance components :  
an expository review.*

~~TOPICS IN VARIANCE COMPONENT ESTIMATION~~

S. R. Searle

Table of Contents

PART I: MODELS

1. Fixed effects models . . . . .	2
2. Random effects models. . . . .	5
3. Mixed models . . . . .	8
4. Fixed or random? . . . . .	10
4.1 Finite populations . . . . .	12

PART II: BALANCED DATA

5. The analysis of variance method of estimation . . . . .	15
6. Random models. . . . .	19
6.1 The 1-way classification . . . . .	20
6.2 The 2-way cross classification . . . . .	22
7. Mixed models . . . . .	24
7.1 The 2-way cross classification . . . . .	24
7.2 Defining elements of the model . . . . .	25
7.3 Estimating fixed effects . . . . .	27
8. Fixed effects models . . . . .	29
9. Distribution-free characteristics. . . . .	31a
9.1 Unbiasedness . . . . .	32
9.2 Sampling variances . . . . .	33
9.3 Minimum variance . . . . .	34
9.4 Negative estimates . . . . .	34
10. Normality assumptions. . . . .	39
10.1 Tests of hypotheses. . . . .	40
10.2 Confidence intervals . . . . .	41
10.3 Distributions of estimators. . . . .	45
10.4 Sampling variances of estimators . . . . .	48
10.5 Sufficiency. . . . .	51
10.6 Maximum likelihood . . . . .	52

PART III: UNBALANCED DATA

11. General quadratic forms. . . . .	55
11.1 Expected values. . . . .	56
11.2 Normality assumptions. . . . .	58
12. The analysis of variance method. (Henderson's Method 1) . . . . .	59
12.1 Analogous sums of squares. . . . .	60
(i) Empty cells. . . . .	61
(ii) Balanced data. . . . .	61
(iii) A negative sum of squares! . . . . .	61
(iv) Uncorrected sums of squares. . . . .	62

Ill. 197 also Ill. 308-M

12.2	Expected values. . . . .	63
12.3	General results. . . . .	65
12.4	Mixed models . . . . .	68
12.5	Sampling variances of estimators . . . . .	70
12.6	Synthesis. . . . .	78
13.	Adjusting for bias in mixed models (Henderson's Method 2) . . . . .	80
14.	The fitting constants method. . . . .	84
14.1	Reductions in sums of squares. . . . .	84
14.2	Expectations . . . . .	87
14.3	Estimation . . . . .	88
14.4	Calculation. . . . .	89
14.5	Too many equations . . . . .	92
14.6	General results. . . . .	95
14.7	Sampling variances of estimators . . . . .	97a
15.	Analysis of means methods . . . . .	97a
16.	Symmetric sums methods. . . . .	101
17.	Infinitely many quadratics. . . . .	102
18.	Maximum likelihood methods. . . . .	106
18.1	Estimating fixed effects . . . . .	107
18.2	Fixed effects and variance components. . . . .	110
18.3	Large sample variances . . . . .	112
19.	Best quadratic unbiased estimation. . . . .	114
20.	Designing experiments to estimate variance components . . . . .	116
21.	Other methods and ideas . . . . .	117
21.1	Bayes estimation . . . . .	117
21.2	Plotting the likelihood. . . . .	118
21.3	Sequential estimation. . . . .	119
21.4	Other models . . . . .	119
21.5	Harmonic components. . . . .	120
21.6	Dispensing with unbiasedness . . . . .	120

# TOPICS IN VARIANCE COMPONENT ESTIMATION<sup>1/</sup>

S. R. Searle<sup>2/</sup>

Texas A&M University, College Station, Texas  
and Cornell University, Ithaca, N.Y.

## Introduction

Variance component models are described in many places; among them Eisenhart [1947], Crump [1946 and 1951], Plackett [1960], and Sheffé [1959] come particularly to mind. Although these descriptions vary in their mathematical content, with consequent variation in their appeal to biologists, they are mostly concerned with the mathematics of the models rather than with interpretive illustration thereof. Since biology is a discipline wherein variance components models have widespread application it is apposite to begin a survey of topics in variance component estimation by describing the underlying models in terms of examples that may have some appeal to biologists. Such is the purpose of Part I of this paper. It attempts to describe variance component models by means of illustrative examples presented alongside familiar analysis of variance situations. Emphasis is placed not upon mathematical details but on the meaning and use of different models. General analysis of variance procedures and their associated statistics are assumed known, and attention is directed to the meaning of variance components.

---

<sup>1/</sup> Paper BU-197 in the Biometrics Unit and number 586 in the Department of Plant Breeding and Biometry, Cornell University, Ithaca, N.Y.

<sup>2/</sup> On leave from Cornell, 1968-9.



Parts II and III of the article survey some of the procedures used for estimating variance components from balanced and unbalanced data, respectively, with discussion of some of the problems associated with these procedures and the progress that has been made towards solving them. A few explicit results are given in detail, mostly of easy cases. They provide a framework for discussing more general results, many of which are quite lengthy and to which reference only is given.

Crump [1951] so well summarized the status of variance components at that time that only the progress made since then is mentioned here in any detail. The reader is referred to Crump [1951] for a more complete discussion of the state of the art prior to the fifties.

## PART I: MODELS

### 1. Fixed effects models

One experiment traditionally carried out in agricultural research concerns testing the efficacy of nitrogen (N), potash (P) and potassium (K) on crop yield, tomatoes, say. Suppose an experiment of this kind involved  $2^4$  plants, with each of 6 plants receiving one of the fertilizers and 6 getting no fertilizer at all, these being considered as control. The 4 different kinds of fertilizer (including none) will be referred to as treatments N, P, K and C (for control). The skeleton analysis of variance for this experiment is shown in Table 1.

Table 1

Skeleton analysis of variance for experiment of  
4 fertilizer treatments each used on 6 plants

Source of variation	Degrees of freedom
Treatments	3
Error	20
Total	23

The equation of a model for this analysis is

$$y_{ij} = \mu + t_i + e_{ij} \quad (1)$$

where  $y_{ij}$  is the  $j$ 'th observation ( $j = 1, 2, \dots, 6$ ) on the  $i$ 'th treatment ( $i = N, P, K$  or  $C$ ) with  $\mu$  being a general mean yield,  $t_i$  being the effect of fertilizer  $i$  on yield, and  $e_{ij}$  being an error term representing the difference between an observation  $y_{ij}$  and  $\mu + t_i$ .

Experiments of this nature are often used for providing evidence of whether or not fertilizer increases yield. In carrying out such an experiment, the 4 treatments (3 fertilizers and control) are 4 very specific treatments of interest; and in using them we have no thought for any other fertilizers, for interest lies solely in studying N, P and K in relation to each other and to no fertilizer. This is the concept of fixed effects. Our concentration is fixed upon just the treatments in our experiment, upon these and no others. In this context the treatment effects  $t_i$  in the model (1) are called fixed effects, and the model is correspondingly called the fixed effects model. It is often referred to as Model I, so named by Eisenhart [1947].

In any inference-making situation the manner of obtaining data affects the inferences that can be drawn from them. We therefore consider a sampling process pertinent to the fixed model (1) in which the  $t$ 's relate to the 4 specific treatments C, N, P and K. The data available are envisaged as being one of the many possible sets of data involving these same treatments that could be derived in repetitions of the experiment, repetitions in which the  $e$ 's on each occasion would be a random sample from a population of error terms that has zero mean and variance  $\sigma_e^2$ . It is the randomness associated with obtaining the  $e$ 's from a population that leads to our being able to make inferences about the differences between the  $t_i$ 's. We can also make inferences about the average difference among all 4 treatments, namely the average squared deviation from the mean,  $\frac{1}{3} \sum_{i=1}^4 (t_i - \frac{1}{4} \sum_{i=1}^4 t_i)^2$ , an expression that can sometimes be conveniently thought of as the variance among the 4  $t$ 's; so, too, can inferences be made about  $\sigma_e^2$ .

An expanded form of the above experiment would be when more than one variety of tomatoe was tested in combination with the 4 fertilizers. Suppose, for example, the 4 treatments were used on 6 plants of each of 3 varieties of tomatoe, early ripening, mid-season and late ripening. A suitable model for  $y_{ijk}$ , the yield of the  $k$ 'th plant of the  $j$ 'th variety receiving the  $i$ 'th treatment would be

$$y_{ijk} = \mu + t_i + v_j + (tv)_{ij} + e_{ijk} \quad (2)$$

where  $\mu$  is a general mean,  $t_i$  is the effect on yield due to the  $i$ 'th treatment,  $v_j$  is the effect of the  $j$ 'th variety,  $(tv)_{ij}$  is the effect of the interaction between the  $i$ 'th treatment and the  $j$ 'th variety, and  $e_{ijk}$  is the usual error term. Just as treatment effects are fixed effects, so too are variety effects, the  $v_j$ , for in this experiment interest centers solely on the three varieties being used. Thus both the  $t_i$  and the  $v_j$  and their interactions are considered as fixed effects.

## 2. Random effects models

Young et al [1965] report an experiment concerning the maternal ability of mice, using the litter weight of ten-day-old litters as a measure of maternal ability. Suppose the experiment had been based on four dams (all of the one breed) that had each had six litters. With the 24 litter weights we could carry out an analysis of variance having the skeleton shown in Table 2.

Table 2

Skeleton analysis of variance for experiment  
of 4 dams (mice) each having 6 litters

Source of variation	Degrees of freedom
Dams	3
Residual	20
Total	23

The equation of the model here is  $y_{ij} = \mu + d_i + e_{ij}$  where  $y_{ij}$  is the weight of the  $j$ 'th litter of the  $i$ 'th dam,  $d_i$  is the effect on litter weight due to the litter being from the  $i$ 'th dam, and  $e_{ij}$  is the deviation of  $y_{ij}$  from  $\mu + d_i$ .

Consider the  $d_i$ 's and the dams they represent. The data relate to maternal ability, a variable that is subject to biological variation from animal to animal. In this context, prime interest in the experiment is unlikely to center on only the 4 female mice available in the laboratory. They are only a sample from a population of (female) mice.

In comparing the mouse and fertilizer experiments, each mouse corresponds to a fertilizer treatment; but whereas each fertilizer is something of interest, with no thought for it being a sample from a population of fertilizers, each dam of the mouse data is merely a sample (of one) from a population of mice. Nothing important has conditioned our choosing any one mouse over another, and there is no interest in differences between specific pairs of mice as there was in differences between pairs of fertilizers in the fertilizer experiment. In contrast, our interest in the mouse experiment lies in the extent to which maternal ability varies throughout the whole population of mice from which our 4 are deemed to be a random sample. It is to studying this variation that our model is directed.

The sampling process involved in obtaining the mice data is envisaged as such that any one of many possible sets of data could be derived from repetitions of the data-gathering process. But now, in conceiving of repetitions, we do not confine ourselves to always having the same 4 mice — we imagine getting a random sample of 4 on each occasion from the population of mice. And furthermore, for whatever 4 mice we get on any occasion we envisage getting a random sample of  $e$ 's from a population of errors, along the same lines as in the fixed model. Thus our concept of the error terms in the two cases is essentially the same. But whereas in the fixed model we conceived of always having the same treatments, now, in the case of the mice data, we think of taking a random sample of mice on each occasion. Thus the  $d_i$ 's of our data are a random sample from a population of  $d_i$ 's. Hence, so far as the data are concerned, the  $d_i$ 's therein are random variables, or, as they are usually called in this context, random effects. And the model is correspondingly called the random effects model or the random model. Eisenhart [1947], in his detailed discussion of fixed and random models called the random model Model II, a name that continues to receive widespread use.

In both models the error terms have the same characteristics - that they are a random sample from a population that has zero mean and variance  $\sigma_e^2$ . But, whereas in the fixed effects model the  $t$ 's represented effects of specific treatments, in the random model the  $d$ 's are a random sample from a population, assumed as having a variance  $\sigma_d^2$ . It is also assumed that the  $d$ 's have zero mean. Otherwise we could define  $\mu^* = \mu + E(d_i)$  and  $\delta_i = d_i - E(d_i)$  and write the model as

$$y_{ij} = \mu + d_i + e_{ij} = \mu + E(d_i) + d_i - E(d_i) + e_{ij} = \mu^* + \delta_i + e_{ij} \quad (3)$$

where  $E(\delta_i) = 0$  and the form of  $\mu^* + \delta_i + e_{ij}$  is indistinguishable from  $\mu + d_i + e_{ij}$ . Thus the random model involving the  $d$ 's is described as

$$y_{ij} = \mu + d_i + e_{ij} \quad (4)$$

where  $\mu$  is an overall mean, <sup>and</sup>  $d_i$  is the effect due to the  $i$ 'th dam, this being a random effect from a population of  $d$ 's that has zero mean and variance  $\sigma_d^2$ ; and the  $e_{ij}$ 's are random error terms from a population having zero mean and variance  $\sigma_e^2$ . In this context inferences are sought about  $\sigma_d^2$  and  $\sigma_e^2$ . These and  $\mu$  are the parameters of the random model. Since, from (4),  $\sigma_y^2 = \sigma_d^2 + \sigma_e^2$ , the name "components of variance" is given to  $\sigma_d^2$  and  $\sigma_e^2$ ; each is a variance in its own right, and is a component of  $\sigma_y^2$ , the variance of an observation.

Fixed effects models can involve several factors and their interactions. So also can random effects models. An example arises in dairy cow breeding. Through artificial insemination, bulls can sire offspring in many herds simultaneously. When the females among these progeny calve and start to produce milk, analyses of their milk yields can be made. A suitable model for  $y_{ijk}$ , the yield of milk by the  $k$ 'th daughter of the  $i$ 'th bull in the  $j$ 'th herd is

$$y_{ijk} = \mu + b_i + h_j + (bh)_{ij} + e_{ijk}. \quad (5)$$

In this model  $b_i$  is the effect on yield of a cow's having been sired by the  $i$ 'th bull;  $h_j$  is the effect of her being in the  $j$ 'th herd,  $(bh)_{ij}$  is the interaction effect and  $e_{ijk}$  is the customary error term. In this case all effects are considered random: the bulls on whom we have data are taken to be a random sample of bulls from some population; the herds involved in the data are assumed to be a random sample from a population of herds, and the interaction effects are assumed to be random, too, and of course the error terms are also taken as random. All these effects are assumed uncorrelated, with variances  $\sigma_b^2$ ,  $\sigma_h^2$ ,  $\sigma_{bh}^2$  and  $\sigma_e^2$ . These are the components of variance of the model (5), from which  $\sigma_y^2 = \sigma_b^2 + \sigma_h^2 + \sigma_{bh}^2 + \sigma_e^2$ . They and  $\mu$  are the parameters about which we wish to make inferences.

### 3. Mixed models

Consider the tomatoe and fertilizer experiment again. Instead of having three varieties of tomatoe suppose we have 20 replicate crosses of two varieties, the early ripening and late ripening varieties. The equation of a model for data from such an experiment would be

$$y_{ijk} = \mu + t_i + r_j + (tr)_{ij} + e_{ijk} \quad (6)$$

where  $t_i$  is the effect on yield of the  $i$ 'th treatment,  $r_j$  is the effect of the  $j$ 'th replicate cross,  $(tr)_{ij}$  is the interaction effect and  $e_{ijk}$  is the error term. As before, the  $t_i$  are fixed effects. However, the  $r_j$ , representing the effects of the replicate crosses between the two varieties, are random effects because they represent just those crosses that happen to have been made for this experiment,

the 20 crosses available from the infinite population of crosses that could be made; and we assume these 20 to be a random sample from that infinite population. Hence we have a model containing fixed effects  $t_i$  and random effects  $r_j$ . This is called a mixed model: a model containing a mixture of both fixed effects and random effects.

Model (6) includes effects  $(tr)_{ij}$  for interactions between treatments and replicates. Since the latter are being taken as a random it is reasonable that these interactions also be treated as random. Thus the model has  $t_i$  as fixed effects and the  $r_j$  and  $(tr)_{ij}$  as random effects having zero means and variances  $\sigma_r^2$  and  $\sigma_{tr}^2$  respectively. The parameters of the model, which inferences will be made about, are thus  $\mu$  and the  $t_i$  and the variance components  $\sigma_r^2$ ,  $\sigma_{tr}^2$  and  $\sigma_e^2$ .

A second example of a mixed model is a mouse experiment involving three special diets, the object being to study the effect of these diets on litter weight. In this case a suitable model for  $x_{ijk}$ , the  $k$ 'th litter weight on the  $i$ 'th dam when receiving diet  $j$  would be

$$y_{ijk} = \mu + d_i + f_j + (df)_{ij} + e_{ijk} \quad (7)$$

where  $d_i$  is as before,  $f_j$  is the effect of the  $j$ 'th diet (feed),  $(df)_{ij}$  is an interaction and  $e_{ijk}$  is an error term. The  $d_i$  are random effects, but the  $f_j$  representing the three diets are fixed effects; and the  $(df)_{ij}$ -interactions between the fixed and random effects are random. This is a mixed model, with fixed effects  $f_j$  and variance components  $\sigma_d^2$ ,  $\sigma_{df}^2$  and  $\sigma_e^2$ .



#### 4. Fixed or random?

Apart from the symbols used, equation (2) for the treatments-varieties experiment is indistinguishable from (6), that for the treatments-replicates data. But the models involved are different, because of the interpretation attributed in (6) to the replicate crosses. Whereas varieties are considered fixed effects in (2) the replicate crosses in (6) are considered as random. Similarly, equations (5) for the bulls-herds data and (7) for the dams-diets data have the same formal appearance but their interpretations are not the same: in (5) both bulls and herds are random but in (7) dams are random and diets fixed.

In the above examples the classification of effects as either fixed or random appears to be quite straightforward. This is not always so. For example, consider the herd effects in equation (5). Data of the nature envisaged there usually involve numerous herds that are considered a random sample from some population of herds. Were there to be just a few herds, five or six say, wherein the sole interest concerning herds lay in just those 5 or 6, then the herd effects in the model would more appropriately be taken as fixed and not random. Thus the context of a model is the deciding factor in determining whether factors are fixed or random.

Many situations arise where effects can be judged fixed in one context and random in another. For example, consider the mouse experiment again. Suppose three different laboratory technicians managed an experiment and that technician effects were to be included in the model. If one object of the experiment was to assess differences between the three technicians, with a view to deciding who was the best, then the three technicians involved would be the only technicians of interest. They would in no way be considered a random sample from a population

of technicians and so technician effects would be considered as fixed. On the other hand, a laboratory experiment has to be cared for by someone, and we might reasonably agree that so far as the experiment itself is concerned there is little interest in who the technician is. Any available technicians might well be considered as just a random sample of technicians and in this case all we want to do is to assess the contribution to the general variability of the mouse data caused by technicians. In this case the technician effects of the model would be considered random and a variance component attributable to technicians would be the parameter of interest so far as technicians were concerned.

Year effects in studies of agricultural production are a good example of something that might be considered either fixed, or random. Years themselves are unlikely to be random, for they will probably be a group of consecutive years over which data have been gathered. But the effects of years on yield may reasonably be considered random - unless one is interested in comparing specific years for some purpose, in which case the year effects will be fixed.

In deciding whether a set of effects is to be assumed fixed or random the important question is that of inference: are inferences going to be drawn about just the treatments that occur in the data? "Yes" - then the treatments are to be considered as fixed effects. "No" - then presumably inferences will be made about some population of treatments, from which those in the data are presumed to be a random sample; and so the treatment effects are considered random. Thus when inferences will be confined to the effects in the data the effects are considered fixed; and when inferences will be made about a population of effects from which those in the data are considered a random sample, then the effects are taken as random.

As soon as the effects in a model have been categorized as either fixed or random, the model is then in one of three classes: fixed, random, or mixed. It is a fixed model when all elements except the error terms are fixed effects. It is a random model when all elements except  $\mu$  are random effects. And it is a mixed model when, other than  $\mu$  or the error terms, some of the effects are fixed and some are random. In point of fact, of course, all models having both  $\mu$  and error terms are mixed models because  $\mu$  is a fixed effect and the errors are random. However, the customary distinction between fixed, random and mixed models is usually maintained.

Note that the assumption of randomness in a random effects model does not carry with it the assumption of normality. Frequently this assumption is made, as we shall see, but it is a separate assumption made subsequent to that of assuming effects are random. Most estimation procedures for variance components do not require normality, but when distributional properties of estimators are to be investigated normality is usually assumed.

#### 4.1 Finite populations

Random effects occurring in data are assumed to be a random sample from a population of effects. Usually the population is considered infinite, as for example, is the population of all possible crosses between two varieties of wheat: they could be crossed an infinite number of times. However, the definition of random effects does not necessitate that their population be infinite. It may be finite. Furthermore, although finite it may be very large, indeed so large as to be considered infinite for most practical purposes, e.g., all Holstein cows

in the U.S.A. on May 1, 1970. Thus for random effects models the conceptual populations can be of three kinds, depending on their size: infinite, or finite, or finite but so large as to be deemed infinite. Those that are infinite or finite but large enough to be taken as infinite are the ones most frequently encountered, especially in biology.

Models having random effects with finite populations differ from those with infinite populations largely through consequences of the definition of variance in a finite population. For a population of effects  $\alpha_i$  for  $i = 1, 2, \dots, N$

having zero mean,  $\sum_{i=1}^N \alpha_i = 0$ ; and the population variance is then defined by

$$\sigma_{\alpha}^2 = \frac{\sum_{i=1}^N \alpha_i^2}{(N - 1)}. \quad \text{An easily derived consequence is } \sum_{i \neq i'} \alpha_i \alpha_{i'} = -(N - 1) \sigma_{\alpha}^2.$$

If  $\alpha_r$  is a random effect from this population its variance is  $E(\alpha_r^2) = N^{-1} \sum_{i=1}^N \alpha_i^2 =$

$(1 - N^{-1}) \sigma_{\alpha}^2$ ; and the covariance between it and another random effect,  $\alpha_s$ , is

$E(\alpha_r, \alpha_s) = -\sigma_{\alpha}^2 / N$ . Thus in models having random effects with finite populations

the effects are not uncorrelated as they are with infinite populations. The

consequences can be seen, for example, in Bennett and Franklin [1954, cf p.404].

They are discussed in greater detail by Wilk [1955] and by Wilk and Kempthorne

[1955, 1956, 1957] and also by Cornfield and Tukey [1956]. Some of these papers

are very much concerned with randomization theory and they all deal with balanced data only. For unbalanced data, Gaylor and Hartwell [1969] show the effects of having finite populations in the case of one particular nested classification, and Searle and Fawcett [1970] give general rules for converting infinite population situations into finite populations. These rules are applicable to nested or crossed classifications or to a mixture of both, in either random or mixed models.

A special case of a finite population is when the effects occurring in one's data constitute the whole population of effects. This is often used (e.g. Wilk and Kempthorne, loc cit) as a convenient representation of fixed effects.

## PART II: BALANCED DATA

The dichotomy of balanced data and unbalanced data is now introduced. Balanced data are those in which every one of the sub-most sub-classes of the model has the same number of observations; i.e. equal numbers of observations in all the subclasses. Opposingly, unbalanced data are those wherein the numbers of observations in the subclasses of the model are not all the same; i.e. unequal numbers of observations in the subclasses, including cases where

there are no observations in some subclasses. Thus "unbalanced data" refers not only to situations where all subclasses have some data, namely filled subclasses, but also to cases where some subclasses are empty, with no data in them. Since the estimation of variance components from unbalanced data is more complicated than from balanced data, we deal with the two cases separately, discussing balanced data first.

#### 5. The analysis of variance method of estimation

Analysis of variance is traditionally employed in situations involving fixed effects models, such as the simple experiment of testing 4 fertilizers described above, the analysis for which is shown in Table 3.

Table 3

Analysis of variance for experiment of  
4 fertilizer treatments each used on 6 plants  
(See Table 1.)

Source of variation	Degrees of freedom	Sum of Squares	Mean Square	
			Observed Value	Expected Value
Between treatments	3	$SSB = 6 \sum_{i=1}^4 (\bar{y}_{i.} - \bar{y}_{..})^2$	$MSB = SSB/3$	$\frac{6}{3} \sum_{i=1}^4 (t_i - \frac{1}{4} \sum t_i)^2 + \sigma_e^2$
Error	20	$SSE = \sum_{i=1}^4 \sum_{j=1}^6 (y_{ij} - \bar{y}_{i.})^2$	$MSE = SSE/20$	
Total	23	$SST = \sum_{i=1}^4 \sum_{j=1}^6 (y_{ij} - \bar{y}_{..})^2$		

The right-hand column of this table, frequently not shown in analyses of variance of fixed effects models, contains expected values of the mean squares of the analysis. Its first term,  $E(MSB)$ , involves the fixed effects for treatments, the  $t_i$ , that are part of the model (1) on which the analysis is based. Without this column of expected values the prime purpose of the table is to summarize the calculations needed for obtaining the F-statistic  $MSB/MSE$  for testing the hypothesis that all  $t_i$ 's are equal. However, appending the column to the table highlights both the nature of the F-test and the familiar procedure of estimating  $\sigma_e^2$  by the error mean square:

$$E(MSE) = \sigma_e^2; \text{ and so } \hat{\sigma}_e^2 = MSE \quad (8)$$

is an unbiased estimator of  $\sigma_e^2$ .

An extension of the estimation procedure used in (8) leads to estimation of variance components in random (and mixed) models generally. To demonstrate this suppose we analyze the mouse data (4 dams with 6 litters each) in the manner of Table 3, as shown in Table 4.

Table 4

Analysis of variance for experiment  
of 4 dams each having 6 litters  
(See Table 2)

Source of variation	Degrees of freedom	Sum of Squares	Mean Square	
			Observed Value	Expected Value
Between dams	3	$SSB = \sum_{i=1}^4 6(\bar{y}_{i.} - \bar{y}_{..})^2$	$MSB = SSB/3$	$6\sigma_d^2 + \sigma_e^2$
Error	20	$SSE = \sum_{i=1}^4 \sum_{j=1}^6 (y_{ij} - \bar{y}_{i.})^2$	$MSE = SSE/20$	$\sigma_e^2$
Total	23	$SST = \sum_{i=1}^4 \sum_{j=1}^6 (y_{ij} - \bar{y}_{..})^2$		



The equation of the model for these data is (4):  $y_{ij} = \mu + d_i + e_{ij}$ ; and that for the fertilizer experiment is (1):  $y_{ij} = \mu + t_i + e_{ij}$ . In both cases  $\mu$  is a mean and the e's are random error terms with zero means and variances  $\sigma_e^2$  and zero covariances. However, in contrast to the  $t_i$  of the fertilizer experiment being fixed effects, the  $d_i$  of the mice data are random effects with zero means, variance  $\sigma_d^2$  and zero covariances both with themselves and with the e's. The effect of this in E(MSB) of Table 4 is to have a term in  $\sigma_d^2$  instead of the term in the  $t_i$ 's as in Table 3. This is the only feature that distinguishes Table 4 from Table 3; and it arises solely because of the model. Table 4 relates to a random model and Table 3 to a fixed model.

The tables can, of course, be made to look even more alike. For, defining

$$s_t^2 = \frac{1}{3} \sum_{i=1}^4 (t_i - \bar{t})^2$$

in Table 3 gives  $E(\text{MSB}) = 6s_t^2 + \sigma_e^2$ , similar to  $E(\text{MSB}) = 6\sigma_d^2 + \sigma_e^2$  value of Table 4.

Defining and using  $s_t^2$  in this fashion

emphasizes that the quadratic in the  $t_i$ 's in E(MSB) in the fixed effects model is tantamount to a sample variance of the treatment effects  $t_1, t_2, t_3$  and  $t_4$ . While this kind of relationship is true for balanced data it does not hold for unbalanced data. More importantly, the one-to-one correspondence between  $s_t^2$  and  $\sigma_d^2$  so illustrated does not always occur, even with balanced data, in more complex experimental designs than the one considered here. It can therefore be misleading to use  $s_t^2$  in the belief that this apparent correspondence is universal. It is not.

From Table 3 the estimator of  $\sigma_e^2$  is  $\hat{\sigma}_e^2 = \text{MSE}$ , as in equation (8). The same procedure applied to Table 4 gives estimators of  $\sigma_e^2$  and  $\sigma_d^2$ :

$$\text{MSE} = \hat{\sigma}_e^2 \quad \text{and} \quad \text{MSB} = 6\hat{\sigma}_d^2 + \hat{\sigma}_e^2, \quad (9)$$

Hence  $\hat{\sigma}_e^2 = \text{MSE}$  and  $\hat{\sigma}_d^2 = (\text{MSB} - \text{MSE})/6$

The above example is the simplest illustration of estimating variance components from balanced data by the analysis of variance method, so called because of its reliance on the analysis of variance. It involves the following steps: (i) In the analysis of variance appropriate to the model equate observed mean squares to their expected values; because these expected values are linear functions of the unknown variance components, the resulting equations will be a set of simultaneous linear equations in the variance components. (ii) Solve the equations established in (i): the solutions are the estimators of the variance components.

## 6. Random Models

Estimating variance components by the analysis of variance method is readily achieved in any situation of crossed and/or nested classifications---any situation, that is, which is one of balanced data and the random effects model. However many lines there are in the analysis of variance table there will be the same number of components to estimate. The equations obtained by equating observed mean squares to the expected values will always be consistent, and so solutions (the estimators) can always be found. Details of applying this method are given in many places; e.g. Anderson and Bancroft (1952, Chapter 22), Scheffé (1959, Chapters 7 and 8) and Graybill (1961, Chapters 16 and 17), to name but three. The main step in the method is deriving the expected value of the mean square in each line of the analysis of variance table. The above references (and many others) contain the details of these expectations for a variety of experimental

designs. In addition, when balanced data fall into none of the well-documented designs, general rules of thumb are available for deriving the necessary expected values. These rules, expressed in various ways, are to be found in Lum (1954), Schultz (1955), Henderson (1959) and Millman and Glass (1967). The Henderson paper includes rules for determining what the lines in the analysis of variance shall be, and what their degrees of freedom and sums of squares are; and Millman and Glass have a rule for writing down the computing formula for each sum of squares. These rules, it should be noted, apply whether the model be fixed, mixed or random.

Although details of many specific cases are readily available in the literature we give two examples here, both to indicate how the method operates and to provide a basis for further discussion.

### 6.1 The 1-way classification

In the general 1-way classification (of which Table 4 is an example) we suppose there are  $n$  observations in each of  $a$  classes, that  $y_{ij}$  is the  $j$ 'th observation in the  $i$ 'th class, and that the equation of the model is

$$y_{ij} = \mu + \alpha_i + e_{ij},$$

where  $i = 1, 2, \dots, a$  and  $j = 1, 2, \dots, n$ . As usual,  $\mu$  is a general mean, the  $\alpha_i$  are random variables with zero means, zero covariances and variance  $\sigma_\alpha^2$ ; and the random error terms have zero means, zero covariances and variance  $\sigma_e^2$ . Covariances between the  $\alpha_i$  and  $e_{ij}$  are all assumed zero.

The analysis of variance table uses the following terms

$$T_o = \sum_{i=1}^a \sum_{j=1}^n y_{ij}^2, \quad T_a = n \sum_{i=1}^a \bar{y}_{i.}^2 = \sum_{i=1}^a y_{i.}^2 / n \quad \text{and} \quad T_{\mu} = a n \bar{y}_{..}^2 = y_{..}^2 / a n. \quad (10)$$

This notation is introduced because it refers to the basic calculations required, it simplifies writing of the analysis of variance table and it extends directly to unbalanced data. Each T-term is a total uncorrected sum of squares with subscript indicating the factor it refers to: o for the observations as they stand, a for the  $\alpha$ -factor and  $\mu$  for 'correction for the mean'. The analysis of variance is then as shown in Table 5.

Table 5

Analysis of variance for 1-way classification,  
random model, n observations in each of a classes

Source of variation	Degrees of freedom	Sum of squares	Mean square	Expected value of mean square
Between classes	a - 1	SSB = $T_a - T_{\mu}$	MSB = SSB/(a - 1)	$n\sigma_{\alpha}^2 + \sigma_e^2$
Error	a(n - 1)	SSE = $T_o - T_a$	MSE = SSE/a(n - 1)	$\sigma_e^2$
Total	an - 1	SST = $T_o - T_{\mu}$		

From the last two columns the estimators of the variance components are, similar to those of (9),

$$\hat{\sigma}_e^2 = \text{MSE} \quad \text{and} \quad \hat{\sigma}_{\alpha}^2 = (\text{MSB} - \text{MSE})/n. \quad (11)$$

## 6.2 The 2-way crossed classification

The general 2-way classification is typified by a treatments-by-blocks situation, where the equation of the model is

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + e_{ijk} \quad (12)$$

with T treatments,  $i = 1, 2, \dots, T$ , and B blocks,  $j = 1, 2, \dots, B$ , and n observations in each treatment-by-block cell. The general mean is  $\mu$ , and the  $\tau_i$ ,  $\beta_j$  and  $\tau\beta_{ij}$  are treatment, block and interaction effects respectively. The  $e_{ijk}$  are error terms in the usual way, with  $E(e_{ijk}) = 0$  and  $\text{var}(e_{ijk}) = \sigma_e^2$  for all i, j and k, and all covariances zero. The familiar analysis of variance table is shown in Table 6.

Table 6

Analysis of variance for 2-way classification data of  
T treatments and B blocks, with n observations in  
each treatment-block subclass

Source of variation	Degrees of freedom	Sum of Squares	Mean Square
Treatments	T-1	$S_T = \sum_{i=1}^T Bn(\bar{y}_{i..} - \bar{y}_{...})^2$	$MST = S_T / (T-1)$
Blocks	B-1	$S_B = \sum_{j=1}^B Tn(\bar{y}_{.j.} - \bar{y}_{...})^2$	$MSB = S_B / (B-1)$
Interaction	(T-1)(B-1)	$S_{TB} = \sum_{i=1}^T \sum_{j=1}^B n(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2$	$MSTB = S_{TB} / (T-1)(B-1)$
Error	TB(n-1)	$S_E = \sum_{i=1}^T \sum_{j=1}^B \sum_{k=1}^n (\bar{y}_{ijk} - \bar{y}_{ij.})^2$	$MSE = S_E / TB(n-1)$
Total	nTB-1	$\sum_i \sum_j \sum_k y_{ijk}^2 - nTB\bar{y}_{...}^2$	

Customary use of the word "treatments" implies that they are fixed effects. However, the word is used generically here, simply to denote one of the classifications of the model (12): at one time we think of treatments as being fixed effects and at another as random effects. "Blocks" is used in the same manner for the other classification.

For the random model we think of both treatments and blocks being random. An example is (5), the bulls-by-herds illustration, in which bulls play the part of treatments and herds are the blocks. With all effects except  $\mu$  assumed random we have

$$E(\tau_i) = E(\beta_j) = E(\tau\beta_{ij}) = 0, \text{ var}(\tau_i) = \sigma_\tau^2, \text{ var}(\beta_j) = \sigma_\beta^2 \text{ and } \text{var}(\tau\beta_{ij}) = \sigma_{\tau\beta}^2, \quad (13)$$

for all  $i$  and  $j$ ; and all covariances between these terms and between them and the  $e$ 's are zero. Under these conditions expected values of the mean squares of Table 6 are those shown in Table 7.

Table 7

Expected values of mean squares of Table 6  
Random model

$E(MST)$	$= Bn\sigma_\tau^2$	$+ n\sigma_{\tau\beta}^2$	$+ \sigma_e^2$
$E(MSB)$	$Tn\sigma_\beta^2$	$+ n\sigma_{\tau\beta}^2$	$+ \sigma_e^2$
$E(MSTB)$		$n\sigma_{\tau\beta}^2$	$+ \sigma_e^2$
$E(MSE)$			$\sigma_e^2$

These familiar results can be found in many places, e.g., Anderson and Bancroft [1952, p.316], Scheffé [1959, p. 244] and Steel and Torrie [1960, p.144]. The

analysis of variance method of estimating the variance components of the model then involves equating mean squares to their expected values, and from Table 7 this gives the following estimators:

$$\begin{aligned}\hat{\sigma}_e^2 &= \text{MSE} , & \hat{\sigma}_\beta^2 &= (\text{MSB} - \text{MSTB})/\text{Tn} , \\ \hat{\sigma}_{\tau\beta}^2 &= (\text{MSTB} - \text{MSE})/n, \text{ and } \sigma_\tau^2 &= (\text{MST} - \text{MSTB})/\text{Bn} .\end{aligned}\tag{14}$$

## 7. Mixed Models

The analysis of variance method of estimating variance components can also be used for mixed models, with balanced data. Among the expected mean squares there will be as many having no terms in the fixed effects as there are variance components in the model. This is illustrated by a variation of the example just considered.

### 7.1 The 2-way crossed classification

In the treatments-by-blocks example just discussed suppose that the treatment effects are considered as fixed effects. The equation of the model is still (12), and the assumptions in (13) still hold with the exception of  $E(\tau_i) = 0$  and  $\text{var}(\tau_i) = \sigma_\tau^2$  which are no longer used because the  $\tau_i$ 's are now fixed effects. The analysis of variance in Table 6 is also still appropriate with expected mean squares as shown in Table 8.

Table 8

Expected values of mean squares of Table 6  
Mixed Model  
 with treatment effects,  $\tau_i$ , fixed

---

$E(MST)$	$= \frac{Bn}{T-1} \Sigma(\tau_i - \bar{\tau})^2$	$+ n\sigma_{\tau\beta}^2 + \sigma_e^2$
$E(MSB)$	$=$	$Tn\sigma_{\beta}^2 + n\sigma_{\tau\beta}^2 + \sigma_e^2$ *
$E(MSTB)$	$=$	$n\sigma_{\tau\beta}^2 + \sigma_e^2$
$E(MSE)$	$=$	$\sigma_e^2$

---

\* The term  $n\sigma_{\tau\beta}^2$  is discussed in the text.

Apart from the quadratic in the  $\tau_i$ 's in  $E(MST)$  the entries in Table 8 are the same as those for the random model, in Table 7. There is, however, the matter of the term  $n\sigma_{\tau\beta}^2$  in  $E(MSB)$ . Depending on the definition of the random effects,  $n\sigma_{\tau\beta}^2$  does or does not occur in  $E(MSB)$ . This is discussed in Section 7.2. The important thing is that, whether  $n\sigma_{\tau\beta}^2$  occurs in  $E(MSB)$  or not, the components of variance of the model can be estimated from MSB, MSTB and MSE, the three mean squares whose expected values contain no fixed effects. Thus, so far as these mean squares are concerned, the analysis of variance method of estimating variance components applies to this mixed model just as readily as to the random model.

## 7.2 Defining elements of the model

Variations in defining elements of (12) for the mixed model lead to variations in the form of the expected mean squares shown in Table 8. First consider

$$E(MST) = \frac{Bn}{T-1} \Sigma(\tau_i - \bar{\tau})^2 + n\sigma_{\tau\beta}^2 + \sigma_e^2, \quad (15)$$



a result more frequently found in the form

$$E(MST) = \frac{Bn}{T-1} \sum_{i=1}^T \tau_i^2 + n\sigma_{\tau\beta}^2 + \sigma_e^2 . \quad (16)$$

The change from (15) to (16) arises when the  $\tau_i$  of (12) are defined so that  $\sum_{i=1}^T \tau_i = 0$ . This definition has not been used in deriving Table 8, nor is it necessary to use it. Nevertheless, if the  $\tau_i$  are defined as deviations of treatment effects from their mean,  $\sum_{i=1}^T \tau_i = 0$  applies and  $E(MST)$  takes the form of (16), with the  $\tau_i$  so defined.

We now turn to the term  $n\sigma_{\tau\beta}^2$  in  $E(MSB)$  of Table 8. Perusal of the literature reveals the presence of this term in some places and its absence in others. For example, Mood [1950, p.344], Steel and Torrie [1960, p.144] and Kirk [1968, p.137] are texts that include it, whereas Anderson and Bancroft [1952, p.339], Scheffé [1959, p.269], Graybill [1961, p.398], Fryer [1966, p.334] and Snedecor and Cochran [1967, p.367] exclude it. Mood and Graybill [1963] do not discuss the topic.

Exclusion of  $n\sigma_{\tau\beta}^2$  from  $E(MSB)$  of Table 8 can arise from the way in which block effects are defined. If that for block  $j$  is defined as the effect of block  $j$  plus the average interaction effect for that block averaged over treatments then, as detailed in Henderson and Searle [1970], this leads to  $n\sigma_{\tau\beta}^2$  being absent from  $E(MSB)$ . But in that case  $\sigma_{\beta}^2$  represents the variance not of the  $\beta_j$  but of block effects defined as

$$\beta'_j = \beta_j + \frac{1}{T} \sum_{i=1}^T (\tau\beta)_{ij} . \quad (17)$$

Thus when  $n\sigma_{\tau\beta}^2$  is omitted from  $E(MSB)$  of Table 8, the term  $Tn\sigma_{\beta}^2$  becomes  $Tn\sigma_{\beta'}^2$ , with

$$\sigma_{\beta'}^2 = \sigma_{\beta}^2 + \sigma_{\tau\beta}^2/T .$$

The preceding discussion relates to balanced data. Mixed models for unbalanced data do not usually define block effects to include average interactions. [To do so would involve, in expressions like (18), just those interactions corresponding to the  $T_j$  treatments occurring with block  $j$ .] Consequently, if in the expression for  $E(\text{MSB})$  with unbalanced data, given by Searle and Henderson [1961], we put all the subclass numbers equal - i.e., reduce the unbalanced data case to one of balanced data - then the  $\sigma^2_{\tau\beta}$ -term in  $E(\text{MSB})$  does not vanish identically but has the value  $n\sigma^2_{\tau\beta}$ . Using a model for balanced data that omits  $n\sigma^2_{\tau\beta}$  from  $E(\text{MSB})$  in Table 8 therefore introduces a discontinuity as between unbalanced and balanced data, as discussed by Hartley and Searle [1969]. For additional reasons of conformity as between the random and mixed models with balanced data, i.e. for conformity between Tables 7 and 8, Yates [1967] also prefers the model that puts  $n\sigma^2_{\tau\beta}$  into  $E(\text{MSB})$ . For the general sake of consistency it therefore seems, to this writer, more appropriate to use as the mixed model that which retains  $n\sigma^2_{\tau\beta}$  in  $E(\text{MSB})$  of Table 8.

### 7.3 Estimating fixed effects

Discussion of the mixed model for balanced data would be incomplete without briefly indicating how to estimate the fixed effects of the model. Without embarking on a lengthy discussion of linear estimation generally, we write the model as

$$\underline{y} = \underline{X}\underline{\beta} + \underline{\eta} \quad (18)$$

where  $\underline{y}$  is the vector of observations,  $\underline{\beta}$  is the vector of fixed effects,  $\underline{X}$  is the incidence matrix of the fixed effects in the data and  $\underline{\eta}$  is the vector of the sums of random effects, including error terms, that occur in the elements of  $\underline{y}$ . For

estimating  $\underline{\beta}$  or functions of its elements, two cases must be distinguished: /  $\underline{V}$ , when the variance-covariance matrix of  $\underline{y}$  (i.e., of  $\underline{\eta}$ ) is known, and when it is not.

When  $\underline{V}$  is known, generalized least squares applied to (18) yields normal equations

$$\underline{X}'\underline{V}^{-1}\underline{X}\underline{\beta}^{\circ} = \underline{X}'\underline{V}^{-1}\underline{y} . \quad (19)$$

Solutions for  $\underline{\beta}^{\circ}$  provide the means of obtaining best linear unbiased estimators of estimable functions in the usual way. [The symbol  $\underline{\beta}^{\circ}$  is used to emphasize that when  $\underline{X}'\underline{V}^{-1}\underline{X}$  is singular, solutions to (19) are just solutions to the normal equations and not necessarily estimators of the elements of  $\underline{\beta}$ .]

The difficulty with (19) is that  $\underline{V}$  is usually not known because its elements are linear combinations of the variance components corresponding to the random effects of the model. Therefore, in order to utilize (19), these components in  $\underline{V}$  must be replaced by estimates, which are obtainable as indicated in Section 7.1. If the resulting value of  $\underline{V}$  is called  $\hat{\underline{V}}$  then the equations

$$\underline{X}'(\hat{\underline{V}})^{-1}\underline{X}\underline{\beta}^{\circ\circ} = \underline{X}'(\hat{\underline{V}})^{-1}\underline{y} \quad (20)$$

can be solved to yield solutions  $\underline{\beta}^{\circ\circ}$  which, although not solutions to (19), would be were  $\hat{\underline{V}}$  the true value of  $\underline{V}$ . However, since  $\hat{\underline{V}}$  is only an estimate of  $\underline{V}$ , properties of  $\underline{\beta}^{\circ\circ}$  in terms of providing estimators of estimable functions of  $\underline{\beta}$  are unknown.

Sometimes it is unnecessary to use (20). This occurs when  $\underline{V}$  is such that its elements factor out of (19), which may then provide estimators of estimable functions of  $\underline{\beta}$  identical to those of the fixed effects model. An example of this is the 2-way nested (hierarchical) classification having a model

$y_{ijk} = \mu + \beta_i + \gamma_{ij} + e_{ijk}$ . If the  $\beta$ -effects are fixed and the  $\gamma$ -effects are random, then  $\underline{\beta}$  of (18) involves  $(\mu + \beta_i)$ -terms, and  $\underline{\eta}$  involves  $(\gamma_{ij} + e_{ijk})$ -terms. Simplification of (19) results in  $\sigma_\gamma^2$  and  $\sigma_e^2$  factoring out, leading to the best linear unbiased estimator of  $\mu + \beta_i$  as  $\widehat{\mu + \beta_i} = \bar{y}_{i..}$ , the same as in the fixed effects model. This is for balanced data. For unbalanced data the estimator is

$$\widehat{\mu + \beta_i} = \frac{\sum_{j=1}^{N_i} \bar{y}_{ij.}}{\sum_{j=1}^{N_i} \frac{1}{\sigma_\gamma^2 + \sigma_e^2/n_{ij}}} \bigg/ \frac{\sum_{j=1}^{N_i} 1}{\sum_{j=1}^{N_i} \frac{1}{\sigma_\gamma^2 + \sigma_e^2/n_{ij}}},$$

where there are  $N_i$   $\gamma$ -classes in the  $i$ -th  $\beta$ -class, and  $n_{ij}$  observations in the  $j$ -th  $\gamma$ -class of the  $i$ -th  $\beta$ -class.

## 8. Fixed effects models

The only variance in a fixed effects model is the error variance  $\sigma_e^2$ , whose estimator is usually taken as  $\hat{\sigma}_e^2 = \text{MSE}$ . This poses no problem. Nevertheless, the fixed model case of the treatments-by-blocks example merits brief discussion for comparison with the random and mixed models cases just discussed.

The equation of the fixed effects model is (12), but none of the assumptions in (13) are made, because the  $\tau_i$ ,  $\beta_i$  and  $(\tau\beta)_{ij}$  are all fixed effects. Table 6 is still the analysis of variance, but it has the expected mean squares shown in Table 9.

Table 9

Expected values of mean squares of Table 6  
Fixed effects model

---


$$E(MST) = \frac{Bn}{T-1} \sum_{i=1}^T [\tau_i + (\overline{\tau\beta})_{i.} - \bar{\tau} - (\overline{\tau\beta})_{..}]^2 + \sigma_e^2$$

$$E(MSB) = \frac{Tn}{B-1} \sum_{j=1}^B [\beta_j + (\overline{\tau\beta})_{.j} - \bar{\beta} - (\overline{\tau\beta})_{..}]^2 + \sigma_e^2$$

$$E(MSTB) = \frac{n}{(T-1)(B-1)} \sum_{i=1}^T \sum_{j=1}^B [(\tau\beta)_{ij} - (\overline{\tau\beta})_{i.} - (\overline{\tau\beta})_{.j} + (\overline{\tau\beta})_{..}]^2 + \sigma_e^2$$

$$E(MSE) = \sigma_e^2$$


---

Table 9 has been derived without defining the effects of the model in such a way as to imply what are so often called the "usual restrictions", namely

$$\sum_{i=1}^T \tau_i = 0, \quad \sum_{j=1}^B \beta_j = 0, \quad \sum_{i=1}^T \tau\beta_{ij} = 0 \text{ for all } j, \text{ and } \sum_{j=1}^B \tau\beta_{ij} = 0 \text{ for all } i. \quad (21)$$

When the effects are defined so that (21) applies then the expected mean squares of Table 9 simplify to more familiar forms. That for MST, for example, becomes

$$E(MST) = \frac{Bn}{T-1} \sum_{i=1}^T \tau_i^2 + \sigma_e^2. \quad (22)$$

The expectation in (22) differs from  $E(MST)$  of Table 9 because (21) is assumed to hold in deriving (22) but not in deriving Table 9. Nevertheless, the two expressions for  $E(MST)$  have the same meaning insofar as interpreting the F-

statistic MST/MSE is concerned. This is so because it is possible to define the effects of the model in such a way that (21) and (22) are satisfied and so that the  $\tau_i$  in (22) has the same meaning as does  $\tau_i + (\overline{\tau\beta})_{i.} - [\overline{\tau}_{.} + (\overline{\tau\beta})_{..}]$  in  $E(MST)$  of Table 9. Hence when interpreting the F-statistic MST/MSE it is immaterial whether one uses the expected mean squares of Table 9 or those akin to (22) based on (21). In both cases the interpretation is the same: MST/MSE is testing the significance of treatment effects in the presence of interactions or, equivalently, of treatment effects plus their average interaction effects. All this, of course, has been well known for a long time, but bears repetition in terms of relating the fixed effects model to the random and mixed models.

Defining elements of the model so that equations (21) are true has the consequences just described only with balanced data. With unbalanced data (21) has no such simplifying effects. This is because the sums of squares appropriate to unbalanced data do not have expected values that involve means of effects in such a simple manner as they are involved with balanced data. Sometimes weighted restrictions are suggested but when some of the cells are empty, as is often the case with unbalanced data, this has no simplifying effect on the expectation of

mean squares either.

Differences in the expected values of the mean squares of Table 6 can be noted for the three different models. With the random model (Table 7) every expectation is a linear function of variance components; for the mixed model (Table 8) some expectations involve only variance components whereas others involve quadratic functions of the fixed effects as well; and for the fixed effects model (Table 9) every expectation is the error variance plus a quadratic of fixed effects.

Surveying the many subtleties of analysis of variance models is not the purpose of this paper, as is obvious from the preceding discussion of models and expected values. (One author is reported as having at one stage 512 alternative sets of assumptions about the 2-way classification!) The reader who wishes to pursue these topics in detail has many opportunities to so do: Sheffé [1959], for example, provides a good entrée to the relevant literature, which includes numerous papers in the late 1950's, especially those by M. B. Wilk and O. Kempthorne, J. Cornfield and J. W. Tukey, and H. Scheffé himself. The references listed in these papers will lead the interested reader to many others. He is to be cautioned, however, that this literature deals largely with balanced data, giving little attention to unbalanced data.

#### 9. Distribution-free characteristics

Up to this point no mention has been made of distributional properties for the random elements of our models, other than that they have zero means and covariances, and finite variances. Although the analysis of variance

table for data from a fixed effects model is basically just a convenient summary of the arithmetic involved in calculating F-statistics, the use of those F-statistics in tests of hypotheses is founded upon normality assumptions. However, expected values of mean squares in analysis of variance tables do not use these normality assumptions, and so variance components estimators obtained by the analysis of variance method of estimation do not, of themselves, depend upon normality assumptions. The expected values apply to any distributions that have zero means and covariances, and finite variances, and subject to this mild requirement the analysis of variance method of estimation can therefore be used regardless of distributional properties. However, resulting estimators have limited properties: they are unbiased, they are minimum variance among the class of quadratic unbiased estimators (see Harville [1969a]) but, they can also yield negative estimates and, even under normality assumptions, their distributions are unknown.

### 9.1 Unbiasedness

Estimators of variance components derived by the analysis of variance method from balanced data are always unbiased. This is so whether the model be random or mixed. It is also a property of estimators obtained by the same method from unbalanced data, but for random models only; for mixed models such estimators are biased. We return to this point later, noting here that not even this simplest of properties, unbiasedness, is universally true for analysis of variance estimators of variance components.

Proof of unbiasedness of estimators from balanced data is simple. Suppose  $\underline{m}$  is the vector of mean squares and  $\underline{\sigma}^2$  is the vector of components to be



estimated. Then, if  $E(\underline{m}) = \underline{P}\sigma^2$  for some non-singular matrix  $\underline{P}$ , we take  $\underline{m} = \underline{P}\hat{\sigma}^2$  as the equations for estimating  $\sigma^2$ . Hence

$$\hat{\sigma}^2 = \underline{P}^{-1}\underline{m} \quad (23)$$

and

$$E(\hat{\sigma}^2) = \underline{P}^{-1}E(\underline{m}) = \underline{P}^{-1}\underline{P}\sigma^2 = \sigma^2;$$

i.e.  $\hat{\sigma}^2$  is unbiased. In a random model  $\underline{m}$  is the vector of all mean squares in the analysis of variance, whereas in a mixed model  $\underline{m}$  is just those mean squares whose expected values involve no fixed effects.

## 9.2 Sampling Variances

Sampling variances of variance components estimators involve relatively fearsome formulae in all except the simplest of situations, namely balanced data and models that include normality assumptions. We therefore quote no results here but just indicate where they can be found. For balanced data, without the use of normality but including the case of sampling from finite populations for the random effects, Tukey ([1956]) gives results for the 1- and 2-way classifications, the Latin Square and balanced incomplete block designs. He also indicates procedures for k-way classifications generally; and Tukey ([1957b]) gives third moments of estimators of a 1-way classification. Hooke ([1956]) deals with similar problems and, as Scheffe ([1959, p. 346]) comments, "obtained some results for the more realistic treatment of interactions (more realistic than independence in the non-normal case), but they look discouragingly complicated". Quartiles have also been looked at, in a manner of speaking, for situations that might be

called partially normal - normally distributed error terms and non-normal random effects. In this situation Roy and Cobb [1960] discuss replacing the random effects by a "substitute variate" having  $k$  equally probable discrete values, which they then show how to make inferences about. Their work is presaged by the thought that whereas a variance however suitable it may be as a measure of dispersion for the normal distribution "may be inappropriate for other distributions".

### 9.3 Minimum Variance

Variance component estimators obtained by the analysis of variance method are minimum variance quadratic unbiased, as shown by Graybill and Hultquist [1961]. This means that among all estimators of  $\sigma^2$  which are both quadratic functions of the observations and unbiased, those derived by the analysis of variance method have the smallest variance. Furthermore, this is true under fairly wide conditions, and not just under the normality assumptions. When normality is assumed the estimators are not just minimum variance quadratic unbiased but are minimum variance unbiased, - i.e., from among all unbiased estimators they have minimum variance. Graybill [1954] demonstrated this property for nested classifications and Graybill and Wortham [1956] showed it for crossed classifications under normality, as does Furukawa [1959]. The Graybill and Hultquist [1961] paper relaxes the normality conditions. Harville [1969a and b] also considers the minimum variance property, in terms of quadmissability, for unbalanced data in the 1-way classification.

### 9.4 Negative estimates

Variance components are, by definition, positive. Despite this, estimates obtained by the analysis of variance method can be negative. This can occur in even the simplest of cases; for example, suppose data consist of the following four observations in two classes.

Table 10

Hypothetical data in 2 classes

<u>Class 1</u>	<u>Class 2</u>	
20	18	
<u>2</u>	<u>10</u>	
<u>y<sub>1.</sub> = 22</u>	<u>y<sub>2.</sub> = 28</u>	<u>y<sub>..</sub> = 50</u>

Then, as in (10)

$$T_a = \frac{1}{2}(22^2 + 28^2) = 634,$$

$$T_u = 50^2/4 = 625,$$

and  $T_o = 20^2 + 2^2 + 18^2 + 10^2 = 828.$

Table 5 for these data is then as shown in Table 11.

Table 11

Analysis of variance of hypothetical data

Source	d.f	Sum of squares	Mean Square	Expected Mean Square
Between classes	1	$634 - 625 = 9$	9	$2\sigma_{\alpha}^2 + \sigma_e^2$
Error	2	$828 - 634 = 194$	97	$\sigma_e^2$
Total	3	$828 - 625 = 203$		

Hence, as in (11),

$$\hat{\sigma}_e^2 = 97 \quad \text{and} \quad \hat{\sigma}_\alpha^2 = \frac{1}{2}(9 - 97) = -44.$$

Although this might appear to be a pathological example, it illustrates how negative estimates can arise from the analysis of variance method. There is nothing intrinsic in the method to prevent it. This is so, not only in a simple case such as this, but also in many-factored situations, both with balanced data and with unbalanced data.

To estimate a variance as negative is clearly embarrassing. The question therefore arises as to what can be done when this occurs. Several possibilities exist, few of them satisfactory.

(i) Report the estimate obtained, and learn to live with it, like dandelions in the lawn. It can be taken as evidence that the true value of the component is zero. This is unsatisfying when one wants to utilize an estimate of a sum of variance components, one of which has a negative estimate. The estimated sum including that component is then less than the sum without it. In the above example  $\hat{\sigma}_y^2 = \hat{\sigma}_\alpha^2 + \hat{\sigma}_e^2 = -44 + 97 = 53 < \hat{\sigma}_e^2$ .

(ii) Accepting a negative estimate as evidence that the true value of the corresponding component is zero suggests changing the estimate which is negative to be zero. This seems logical enough, but as part of the estimation procedure such truncation disturbs the properties of the estimates. For example, they are no longer unbiased.

(iii) Taking a negative estimate as indication of a zero component could

also lead to simply ignoring that component in the model, although retaining the factor so far as the lines in the analysis of variance table are concerned. This would mean ignoring the component estimated as negative and re-estimating the others. Thompson [1961 and 1962] gives rules for doing this, known as "pooling minimal mean squares with predecessors", and discusses applications of these rules in Thompson and Moore [1963].

(iv) Interpreting a negative estimate as indication of a wrong model is another possible course of action. Models require care in their specification, as Anderson [1967] points out, and a negative variance component estimate might be the signal to reconsider one's model, in terms of applicability to data and their origin. One possibility is to consider models that have finite rather than infinite populations (see Section 4.1). This is briefly discussed by Searle and Fawcett [1970] for the 1-way classification of Table 5. If  $\hat{\sigma}_{\alpha}^2$  is negative, imputing a finite population to the error terms leads to a positive estimator of  $\sigma_{\alpha}^2$ . However, this may not be a very reasonable assumption in many situations.

In contrast to reconsidering one's model as a result of a negative variance estimate, Nelder [1954] suggests that at least for split plot and randomized blocks designs, the randomization model indicates that there are situations when a negative variance component could arise. This apparent inconsistency can be an outcome of the correlation between plots in the same block being less than the correlation between plots of different blocks.

(v) Estimation procedures other than the analysis of variance method are also available alternatives whenever the latter yields negative estimates.

Bayes estimators, which are non-negative, have been suggested by Hill [1965, 1967] and by Tiao and Tan [1965, 1966] and Tiao and Box [1967]. An excellent commentary on these and on Bayes estimation in general in the variance component context is given by Harville [1969b].

Maximum likelihood estimators (e.g., Herbach [1959] and Thompson [1962]) are also non-negative; these, like many of the Bayes estimators, are usually based on assuming normality of the data, an assumption that is not required when using analysis of variance estimators.

Federer [1968] has also suggested a non-negative estimator which, for  $\sigma_{\alpha}^2$  of Table 5, is  $\hat{\sigma}_{\alpha}^2 + n^{-1}\hat{\sigma}_e^2 \exp(-F)$  for  $F = \text{MSB}/\text{MSE}$ . Although non-negative, it is biased, with a bias that tends to zero as  $F$  increases but which is also largest when  $\hat{\sigma}_{\alpha}^2$  is negative, namely when  $F < 1$ . Furthermore, as Harville [1969b] points out, the estimator is not necessarily admissible.

(v) If, in one's ire at obtaining a negative estimate, all else fails then perhaps the only course left is that of the statistician's last hope: collect more data and repeat the analysis, either on the new data or on the new and old pooled together. If the estimate from the pooled data is negative, that would be additional evidence that the corresponding component has a true value of zero.

#### 10 Normality Assumptions

We now introduce normality and follow its consequences in terms of distributional properties of variance components estimators.

The assumptions usually made are that the  $e$ 's and each set of random effects in the model are normally distributed with zero means and variance-covariance structure set out earlier (see Sections 6.1 and 6.2). Then, for the random model with balanced data, it can be shown that the sums of squares in the analysis of variance are distributed independently of each other; and each sum of squares divided by the expected value of its mean square is a central  $\chi^2$ -distribution with degrees of freedom equal to those shown in the analysis of variance table. Thus if  $df$ ,  $SS$  and  $MS$  are the terms in a line of the table

$$MS = SS/df \quad (24)$$

and, independently of one another,

$$\frac{SS}{E(MS)} \sim \chi_{df}^2; \quad \text{i.e.} \quad \frac{df}{E(MS)} \cdot MS \sim \chi_{df}^2. \quad (25)$$

This is for the random model; for the mixed model, (25) will apply for all sums of squares whose expected values do not involve fixed effects; those that do involve fixed effects will be non-central  $\chi^2$ 's.

### 10.1 Tests of hypotheses

Expected values of mean squares suggest which mean squares are the appropriate denominators for testing hypotheses that certain variance components are zero. Thus in Table 6, MSTB/MSE is the appropriate F-statistic for testing the hypothesis  $H: \sigma_{\tau\beta}^2 = 0$ ; and MSB/MSTB is the F-statistic for testing  $H: \sigma_{\beta}^2 = 0$ . In the random model all ratios of mean squares have central F-distributions, because all sums of squares follow (25). In the mixed model the same is true of ratios of those mean squares whose expected values contain no fixed effects.

In some situations the table of expected values will not suggest in any "obvious" fashion the appropriate denominator for testing a hypothesis. For example, suppose that the expected value of a mean square is  $k_1 \sigma_1^2 + k_2 \sigma_2^2 + \dots + k_t \sigma_t^2$  but that no other mean square in the analysis has expectation  $k_2 \sigma_2^2 + \dots + k_t \sigma_t^2$ . There would then be no "obvious" denominator for testing the hypothesis  $H: \sigma_1^2 = 0$ . For situations such as this the Satterthwaite (1949) procedure for calculating a ratio that has approximately a central F-distribution is as follows. Suppose  $M_i$  is one of the mean squares, with  $f_i$  degrees of freedom. And suppose

$$E(M_r + \dots + M_s) = \lambda_1 \sigma_1^2 + \lambda_2 \sigma_2^2 + \dots + \lambda_t \sigma_t^2$$

with

$$E(M_m + \dots + M_n) = \lambda_2 \sigma_2^2 + \dots + \lambda_t \sigma_t^2$$

for some set of constants  $\lambda_i$ ,  $i = 1, \dots, t$ , and where no mean square occurs in both of these expectations. Then for testing the hypothesis  $H: \sigma_1^2 = 0$



$$F = \frac{M_r + \dots + M_s}{M_m + \dots + M_n}, \text{ is approximately } F_{p,q}$$

$$\text{where } p = \frac{(M_r + \dots + M_s)^2}{(M_r^2/f_r + \dots + M_s^2/f_s)} \text{ and } q = \frac{(M_m + \dots + M_n)^2}{(M_m^2/f_m + \dots + M_n^2/f_n)}.$$

The basis of this test is that the numerator and denominator of  $F$  are independent, and each is distributed approximately as a multiple of a central  $\chi^2$ . A more general form of the test is also available, in using linear combinations of the  $M$ 's instead of simple sums of them. In that case  $c_i M_i$  is used in place of  $M_i$ , where the  $c_i$  are positive constants. In all cases  $p$  and  $q$  are not necessarily integer-valued and interpolation is often necessary when entering the  $F$ -table for  $F_{p,q}$ . Properties of this Satterthwaite procedure have been considered by Cochran [1951] and, more recently, by Gaylor and Hopper [1969], who give particular attention to the case when mean squares are involved negatively.

## 10.2 Confidence intervals

The procedures just described also provide a method for obtaining approximate confidence intervals for linear functions of variances, more particularly for linear functions of expected mean squares. The method is as follows, adapted from Graybill [1961]. Define  $\chi_{n,L}^2$  and  $\chi_{n,U}^2$  as the lower and upper limits of a fraction  $1 - \alpha$  of the  $\chi_n^2$  distribution such that

$$P(\chi_{n,L}^2 \leq \chi_n^2 \leq \chi_{n,U}^2) = 1 - \alpha. \quad (26)$$

Then, for any set of constants  $\lambda_i$ , the approximate confidence interval on  $\Sigma \lambda_i E(M_i)$  is, provided  $\Sigma \lambda_i M_i > 0$ ,

$$P\left(\frac{n \Sigma \lambda_i M_i}{\chi_{n,U}^2} \leq \Sigma \lambda_i E(M_i) \leq \frac{n \Sigma \lambda_i M_i}{\chi_{n,L}^2}\right) = 1 - \alpha$$

where

$$n = (\sum \lambda_i M_i)^2 / \sum (\lambda_i^2 M_i^2 / f_i),$$

analogous to  $p$  and  $q$  previously. Since  $n$  will seldom be an integer,  $\chi_{n,U}^2$  and  $\chi_{n,L}^2$  are obtained from tables using either interpolation or the nearest (or next largest) integer to  $n$ . Graybill (1961) suggests that when  $n < 30$  a correction provided by Welch (1956) be made to the values of  $\chi_{n,U}^2$  and  $\chi_{n,L}^2$  taken from the tables.

Another approximation is given by Bulmer (1957) for the situation where

$$E(M_1) = \theta + \sigma^2 \quad \text{and} \quad E(M_2) = \sigma^2.$$

His confidence interval for  $\theta$  is given by

$$P \left[ \frac{M_2(F - F_1)(F + F_1 - F_2)}{F(F_2)} \leq \theta \right] \doteq \alpha$$

where  $F = M_2/M_1$  and  $F_1, F_2$  are the lower  $\alpha$ -points of the  $F_{f_1, f_2}$  and  $F_{f, \infty}$  distributions respectively. This is also the approximation suggested by Scheffe (1959, equation 7.2.31). Other approximations, based on Taylor expansions, have also been suggested; Bulmer (1957), Scheffe (1959) and Plackett (1960) give references. Approximate fiducial intervals have been considered by Bross (1950).

The above results all relate to intervals with approximate confidence coefficients. One <sup>always</sup> case where the coefficients are exact is for the error variance, because  $\hat{\sigma}_e^2 = \text{MSE}$  has, apart from a constant, a  $\chi^2$ -distribution

on  $f$  degrees of freedom, say, and so a confidence interval on  $\sigma_e^2$  is, from (25)

$$P\left[\frac{SSE}{\chi_{f,U}^2} \leq \sigma_e^2 \leq \frac{SSE}{\chi_{f,L}^2}\right] = 1 - \alpha ,$$

where  $\chi_{f,L}^2$  and  $\chi_{f,U}^2$  are defined as in (26).

Other cases of exact confidence coefficients arise in the 1-way classification. One of particular interest to biologists concerns the intra-class correlation,  $\sigma_\alpha^2/(\sigma_\alpha^2 + \sigma_e^2)$ , a parameter often used in heritability studies by geneticists and animal breeders. In this case, as shown in Table 5,  $E(MSB) = n\sigma_\alpha^2 + \sigma_e^2$  and  $E(MSE) = \sigma_e^2$ , and from (25), with  $F = MSB/MSE$ , we have

$$\frac{\sigma_e^2 F}{\sigma_\alpha^2 + n\sigma_e^2} \text{ distributed as } F_{a-1, a(n-1)},$$

the F-distribution on  $a - 1$  and  $a(n - 1)$  degrees of freedom. Now define  $F_L$  and  $F_U$  as the lower and upper limits which enclose  $1 - \alpha$  of the  $F_{a-1, a(n-1)}$ -distrib-

bution; i.e.

$$P_r \left( F_L \leq \frac{\sigma_e^2 F}{\sigma_\alpha^2 + n\sigma_e^2} \leq F_U \right) = 1 - \alpha.$$

Then from this, as Graybill [1961, p. 379] indicates, an exact confidence interval for the ratio  $\sigma_e^2/(\sigma_\alpha^2 + \sigma_e^2)$  can be derived as

$$P \left[ \frac{n}{n + F/F_L - 1} \leq \frac{\sigma_e^2}{\sigma_\alpha^2 + \sigma_e^2} \leq \frac{n}{n + F/F_U - 1} \right] = 1 - \alpha, \quad (27)$$

Similarly a confidence interval for  $\sigma_\alpha^2/(\sigma_\alpha^2 + \sigma_e^2)$ , the intra-class correlation, is

$$P \left[ \frac{F/F_U - 1}{n + F/F_U - 1} \leq \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_e^2} \leq \frac{F/F_L - 1}{n + F/F_L - 1} \right] = 1 - \alpha \quad (28)$$

and one for  $\sigma_\alpha^2/\sigma_e^2$  is

$$P \left[ \frac{F/F_U - 1}{n} \leq \frac{\sigma_\alpha^2}{\sigma_e^2} \leq \frac{F/F_L - 1}{n} \right] = 1 - \alpha \quad (29)$$

as given in Scheffe [1959, p. 229].

Confidence intervals for  $\sigma_\alpha^2$  in this case have also been considered. If  $\sigma_e^2$  is known, they can be derived both from (29), and from

$$P \left( \frac{SSB}{\chi_{a-1,U}^2} \leq \sigma_e^2 + n\sigma_\alpha^2 \leq \frac{SSB}{\chi_{a-1,L}^2} \right) = 1 - \alpha \quad (30)$$

which originates from SSB being a multiple of the  $\chi_{a-1}^2$ -distribution as in (25). Both require knowing  $\sigma_e^2$  in order to have a confidence interval on  $\sigma_\alpha^2$ . To

overcome this difficulty, Williams [1962] has combined (29) and (30) to give an interval on  $\sigma_\alpha^2$  as

$$P \left[ \frac{SSB(1 - F_U/F)}{n\chi_{a-1,U}^2} \leq \sigma_\alpha^2 \leq \frac{SSB(1 - F_L/F)}{n\chi_{a-1,L}^2} \right] \geq 1 - 2\alpha. \quad (31)$$

The extent to which the left-hand side of (31) exceeds  $1 - 2\alpha$  is discussed by Williams; it depends upon incomplete gamma functions, and Williams comments that the width of the interval "is often only slightly different from the  $1 - \alpha$  interval" that can be derived from (30) when  $\sigma_e^2$  is known.

It is clear from the form of (28), (29), and (31) that the limits of the intervals therein can have negative values. In discussing the difficulty of interpreting negative bounds Williams [1962], in reference to (31), suggests that when the lower limit is negative it should be put equal to zero; and when both limits are zero,  $\sigma_\alpha^2$  should be taken as zero. Scheffé [1959], in considering this problem relative to (29) suggests leaving negative limits untouched even though several ways of amending them are available. As he rightly says, leaving an interval estimate like that from -5 to 2 is "stronger evidence that the true value of a non-negative parameter is zero than that from" 0 to 2.

Simultaneous confidence intervals for all the variance components ratios  $\sigma_1^2/\sigma_e^2$  available in a random model (balanced data) have recently been suggested by Broemeling [1969a] with an additional example in Broemeling [1969b]. They rely on a property of the F-statistics given by Kimball [1951].

### 10.3 Distributions of estimators

The analysis of variance method of estimation yields estimators of variance components that are linear functions of mean squares. And under normality

assumptions those mean squares are distributed independently as multiples of central  $\chi^2$ -variables, in accord with (24) and (25). Hence the variance component estimators are linear functions of these  $\chi^2$ -variables, some of them with negative coefficients. No closed form exists for the distribution of such functions; furthermore, the coefficients are themselves functions of the population variance components. For example, in  $\hat{\sigma}_\alpha^2 = (\text{MSB} - \text{MSE})/n$  of the 1-way classification (Table 5), using (25) and  $\chi_k^2$  to mean a variable having the  $\chi^2$ -distribution with k degrees of freedom, we have

$$\hat{\sigma}_\alpha^2 = \frac{\text{MSB} - \text{MSE}}{n} = \frac{n\sigma_a^2 + \sigma_e^2}{n(a-1)} \chi_{a-1}^2 - \frac{\sigma_e^2}{an(n-1)} \chi_{a(n-1)}^2. \quad (32)$$

The exact form of the distribution in (32) cannot be derived because of the negative coefficient of the second term. Furthermore, since both coefficients involve the unknown variance components, they are unknown and this also contributes to the distribution of  $\hat{\sigma}_\alpha^2$  being unknown. For assumed values of the components, the methods of Robinson [1965] or of Wang [1967] can be employed to obtain the distributions.

The one exception to the general intractability of the distribution of variance component estimators is, of course, that of the error variance:

$$\hat{\sigma}_e^2 = \text{MSE} \sim \frac{\sigma_e^2}{f_{\text{SSE}}} \chi_{f_{\text{SSE}}}^2,$$

where SSE has degrees of freedom  $f_{\text{SSE}}$ .

Empirical evidence of the nature of the distribution of variance components estimators has been reported by Leone et. al. [1966], for a single 4-way nested classification with 5 levels of the main classification and two each of the others. Computer simulation of 1000 experiments of sampling from populations of

known variances revealed considerable variability in the estimators obtained. Three different populations were used, normal, rectangular and exponential, each with eight different sets of the four variance components involved. It is suggested that Perason Type III curves might be suitable for the distributions of the estimators from the normal and rectangular populations. Discussion is also given of the frequency of negative estimates: empirical evidence for all three populations and analytic results for the normal indicate that in certain circumstances negative estimates can be expected to occur 25% of the time.

Under normality assumptions, Leone et al (1966) give a useful procedure for deriving the probability of a negative estimate in nested classifications. Let us suppose that in the analysis of variance

$$E(M_1) = k_1 \sigma_1^2 + k_2 \sigma_2^2 + \dots + k_t \sigma_t^2$$

and  $E(M_2) = k_2 \sigma_2^2 + \dots + k_t \sigma_t^2$ .

Then

$$\hat{\sigma}_1^2 = (M_1 - M_2)/k_1 < 0 \text{ if } M_1/M_2 < 1.$$

Hence

$$P(\hat{\sigma}_1^2 < 0) = P(M_1/M_2 < 1)$$

$$= P\left[ \frac{M_1}{E(M_1)} \cdot \frac{E(M_2)}{M_2} < \frac{E(M_2)}{E(M_1)} \right]$$

$$= P\left[ F_{f_1, f_2} < \frac{k_2 \sigma_2^2 + \dots + k_t \sigma_t^2}{k_1 \sigma_1^2 + k_2 \sigma_2^2 + \dots + k_t \sigma_t^2} \right]. \quad (33)$$

Example In the 1-way classification (33) is

$$P(\sigma_{\alpha}^2 < 0) = P\left[F_{a-1, a(n-1)} < \frac{\sigma_e^2}{\sigma_e^2 + n\sigma_{\alpha}^2}\right] = P\left[F_{a-1, a(n-1)} < 1/(1 + n\rho)\right]$$

where  $\rho = \sigma_{\alpha}^2 / \sigma_e^2$ .

Calculation of (33) demands knowing the variance components that are being estimated. However, for a series of arbitrary values of the components, (33) can be calculated and used to provide some general indication of the probability of obtaining a negative estimate.

It is clear that a result analogous to (33) could also be developed based on the approximate F-statistic discussed in Section 10.1.

#### 10.4 Sampling variances of estimators

Although distributions of the estimators are, to any practical extent, unobtainable, variances of the estimators can be derived. They do, of course, involve the unknown components.

In general we write a variance component estimator as

$$\hat{\sigma}^2 = \sum_i k_i M_i,$$

a linear function of mean squares, its variance being

$$\text{var}(\hat{\sigma}^2) = \sum_i k_i^2 \text{var}(M_i)$$

because the mean squares are independent;



and because, as in (25),  $M_i$  has a distribution that is a multiple of a  $\chi^2$  (the variance of which is twice its degrees of freedom),

$$\text{var}(M_i) = 2f[E(M_i)]^2/f_i^2 = 2[E(M_i)]^2/f_i \quad (34)$$

so that the sampling variance of  $\hat{\sigma}^2$  is

$$\text{var}(\hat{\sigma}^2) = 2\sum_i k_i^2 [E(M_i)]^2 / f_i \quad (35)$$

In this expression the terms  $[E(M_i)]^2$  are just squares of expected mean squares, and as such are obtainable from analysis of variance tables like Table 5. They are linear functions of the variance components.

Example In the 1-way classification

$$\hat{\sigma}_e^2 = \text{MSE}/a(n-1) \text{ and } \hat{\sigma}_\alpha^2 = \text{MSB}/n - \text{MSE}/n$$

and so

$$\text{var}(\hat{\sigma}_e^2) = \frac{2[E(\text{MSE})]^2}{a(n-1)} \quad \text{and} \quad \text{var}(\hat{\sigma}_\alpha^2) = \frac{2}{n^2} \left\{ \frac{[E(\text{MSB})]^2}{a-1} + \frac{[E(\text{MSE})]^2}{a(n-1)} \right\} \quad (36)$$

Hence, using Table 5,

$$\text{var}(\hat{\sigma}_e^2) = \frac{2\hat{\sigma}_e^4}{a(n-1)} \quad \text{and} \quad \text{var}(\hat{\sigma}_\alpha^2) = \frac{2}{n^2} \left[ \frac{(n\sigma_\alpha^2 + \sigma_e^2)^2}{a-1} + \frac{\sigma_e^4}{a(n-1)} \right] .$$

The procedure of (35) applies to any variance component estimator obtained from balanced data by the analysis of variance method. It also extends to other moments. Each estimator is a linear combination of multiples of independent  $\chi^2$ -variables and the  $r$ 'th moment of any one of these combinations is a linear function of the  $r$ 'th moments of the  $\chi^2$ -variables. The unknown components will, of course, be involved in these linear functions, just as they are in (35) and (36).

The mean squares in an analysis of variance of balanced data are, under normality assumptions, independent. But estimators of variance components are not. In the above example

$$\text{cov}(\hat{\sigma}_e^2, \hat{\sigma}_\alpha^2) = -\text{var}(\text{MSE})/n = -2\sigma_e^4/(\text{an}(n-1)) .$$

Obtaining optimum-property estimators of these sampling variances and covariances of variance component estimators is complicated by the fact that the sampling variances and covariances are quadratic functions of the very variance components being estimated. Unbiasedness is about the only property that can be achieved. This is done by replacing  $[E(M_i)]^2/f_i$  in the expressions for the variances and covariances by  $M_i^2/(f_i + 2)$ . Thus from (35)

$$\hat{\text{var}}(\hat{\sigma}^2) = 2 \sum_i \frac{k_i M_i^2}{f_i + 2}$$

is an unbiased estimator of  $\text{var}(\hat{\sigma}^2)$ . The reason is as follows. By definition, and from (34),

$$\text{var}(M_i) = E(M_i^2) - [E(M_i)]^2 = 2[E(M_i)]^2/f_i ,$$

so that

$$E(M_i^2) = (1 + 2/f_i)[E(M_i)]^2 .$$

Hence

$$\frac{M_i^2}{f_i + 2} \text{ is an unbiased estimator of } \frac{[E(M_i)]^2}{f_i} .$$

Example An unbiased estimator of  $\text{var}(\hat{\sigma}_\alpha^2)$  in the 1-way classification is, from (36),

$$\hat{\text{var}}(\hat{\sigma}_\alpha^2) = \frac{2}{n^2} \left[ \frac{(n\hat{\sigma}_\alpha^2 + \hat{\sigma}_e^2)}{a + 1} + \frac{\hat{\sigma}_e^4}{a(n - 1) + 2} \right] .$$

### 10.5 Sufficiency

On the basis of the normality assumptions, the analysis of variance estimators of variance components obtained from balanced data are based on minimal sufficient statistics. For example, in the 1-way classification it can be shown that the likelihood of the sample of observation reduces to

$$\frac{\exp - \frac{1}{2} \left[ \frac{\text{SSE}}{\sigma_e^2} + \frac{\text{SSA}}{\sigma_e^2 + n\sigma_\alpha^2} + \frac{an(\bar{y} - \mu)^2}{\sigma_e^2 + n\sigma_\alpha^2} \right]}{(2\pi)^{\frac{1}{2}an} (\sigma_e^2)^{\frac{1}{2}a(n-1)} (\sigma_e^2 + n\sigma_\alpha^2)^{\frac{1}{2}a}} .$$

It is clear from this that SSE, SSB and  $\bar{y}$  are a set of sufficient statistics, and it is from SSE and SSB that  $\hat{\sigma}_e^2$  and  $\hat{\sigma}_a^2$  are derived. Graybill and Hultquist [1961] establish wide conditions under which this kind of result is true generally for analysis of variance estimators of variance components. They also consider the property of completeness. Weeks and Graybill [1961, 1962] also consider minimal sufficient statistics for balanced incomplete block designs, as do Kapadia and Weeks [1963] for the interaction case thereof.

#### 10.6 Maximum likelihood

Initial establishment of the analysis of variance method of estimating variance components relied upon nothing more than its intuitive appeal of equating observed mean squares to their expected values. The unbiased estimators so derived have since been shown to have certain optimal conditions, especially under normality assumptions. We might well ask, however, about the aptness of using the method of maximum likelihood as a procedure for estimating variance components. In the case of estimating parameters of a fixed effects model it leads in many cases to the same estimators (under normality) as do the methods of least squares and best linear unbiased estimation. One might hope, therefore, that with variance components estimation it would lead to the analysis of variance estimators. But such is not the case, because these estimators can take negative values. Thus they cannot be maximum likelihood estimators since these would be derived by maximizing the likelihood over the parameter space, which is non-negative so far as variance components are concerned. Maximum likelihood estimators have therefore to be non-negative.

Equating to zero the partial differentials of the likelihood functions yields, for balanced data, the same equations as can be used to obtain analysis of variance estimators. However, to obtain maximum likelihood estimators, these equations have to be considered in the light of requiring that such estimators have always to be non-negative.

Herbach [1959] has considered this problem in some detail as has also Thompson [1962], who uses a restricted maximum likelihood procedure confined to just that portion of the set of sufficient statistics which is location invariant. From this he develops his procedure, mentioned in Section 9.4, of pooling minimal mean squares with predecessors when the analysis of variance method yields negative estimates. Indicative of results from these papers are those for the 1-way classification. On defining

$$\theta = \frac{a - 1}{a + \lambda} \text{MSB} - \text{MSE}$$

the estimators are, for  $\theta \geq 0$ ,

$$\tilde{\sigma}_{\alpha}^2 = \theta/n \text{ and } \tilde{\sigma}_e^2 = \text{MSE}$$

and for  $\theta < 0$ ,

$$\tilde{\sigma}_{\alpha}^2 = 0 \text{ and } \tilde{\sigma}_e^2 = \text{SST}/(an + \lambda) .$$

With  $\lambda = 0$  these results are Herbach's; with  $\lambda = -1$  they are Thompson's. When  $a$  is relatively large, estimates derived from these estimators differ little (save for using zero when  $\theta < 0$ ) from the estimate derived from the analysis of variance method, namely  $\theta/n$  with  $\lambda = -1$ .

Commenting on these results Robson [1965] suggests that an estimator for  $\sigma_{\alpha}^2$  having smaller mean square error is  $\theta/n$  with  $\lambda = +1$ , an estimator which is also considered by Zachs [1967]. Another estimator with uniformly still smaller mean square error is given by Klotz and Milton [1967] and further considered by Klotz, Milton and Zachs [1969]. These authors also compare the mean square error of these estimators numerically, for a variety of values of  $a$ ,  $n$  and  $\sigma_{\alpha}^2/\sigma_e^2$ .

## TOPICS IN VARIANCE COMPONENT ESTIMATION (continued)

S. R. Searle

### PART III: UNBALANCED DATA

Estimating variance components from unbalanced data is not as straightforward as from balanced data. This is so for two reasons. First, several methods of estimation are available (most of which reduce to the analysis of variance method for balanced data), but no one of them has yet been clearly established as superior to the others. Second, all the methods involve relatively cumbersome algebra; discussion of unbalanced data can therefore easily deteriorate into a welter of symbols, a situation we do our best (perhaps not successfully) to minimize here.

It is probably safe to describe the Henderson [1953] paper as the foundation paper dealing with variance component estimation from unbalanced data. The methods there described have, accordingly, often been referred to as Henderson's Methods 1, 2, and 3. As described in Searle [1968], Method 1 is simply an analogue of the analysis of variance method used with balanced data; Method 2 is designed to correct a deficiency of Method 1 that arises with mixed models; and Method 3 is based on the method of fitting constants so often used in fixed effects models. Prior to the publication of these methods Winsor and Clark [1940] had utilized the analysis of variance method in studying variation in the catch of plankton nets, Eisenhart [1947] had clearly specified distinctions between fixed, random and mixed models and Crump [1946, 1947 and 1951] had established sampling variances of the variance component estimators in the 1-way classification. Henderson [1953]

however, greatly extended the estimation procedures, especially in describing three different methods and in indicating their use in multi-way classifications. Since then, a number of developments have been made. Variances of estimated components have been considered by Tukey [1957a], Searle [1956, 1958 and 1961a], Mahamunulu [1963], Blischke [1966 and 1968], Harville [1969c], Rohde and Tallis [1969] and Low [1969]; defects in Henderson's Method 2 have been demonstrated by Searle [1968], and difficulties with the mixed model have been discussed by Searle and Henderson [1961], Cunningham and Henderson [1968] and Thompson [1969]; and other methods of estimation have been developed: maximum likelihood by Hartley and Rao [1967] and large sample variances therefrom by Searle [1970], symmetric sums by Koch [1967a and 1968] and best quadratic unbiased estimation by Townsend [1968]. Not all of these developments have been applied to all of even the most straightforward applications and some of them are more specialized than others. What is even more important, <sup>only a</sup> few comparative studies have been made on the relative merits of the different methods; Bush and Anderson [1963], Anderson and Crump [1967] and Kussmaul and Anderson [1967] are three such.

#### 11. General quadratic forms

All currently available methods for estimating variance components from unbalanced data use, in one way or another, quadratic forms of the observations. Before describing the methods we therefore outline properties of quadratic forms of observations coming from a general linear model. This is taken as

$$\underline{y} = \underline{X}\underline{\beta} + \underline{e}$$

where  $\underline{y}$  is a vector of observations,  $\underline{X}$  is a matrix of known values,  $\underline{\beta}$  is a vector of parameters (including both fixed and random effects) and  $\underline{e}$  is a vector of the customary error terms. The vector of means and the variance-

covariance matrix are taken respectively as

$$E(\underline{y}) = \underline{\mu} \quad \text{and} \quad \underline{V} = \text{var}(\underline{y}) = E(\underline{y} - \underline{\mu})(\underline{y} - \underline{\mu})' .$$

### 11.1 Expected values

The expected value under the above-mentioned model of the quadratic form  $\underline{y}'\underline{Q}\underline{y}$  is

$$E(\underline{y}'\underline{Q}\underline{y}) = \text{tr}(\underline{Q}\underline{V}) + \underline{\mu}'\underline{Q}\underline{\mu} \quad (37)$$

where 'tr' represents the trace operation on a matrix, that of summing its diagonal elements. This result is the basis of most methods of estimating variance components from unbalanced data. The general methodology is to obtain expected values of quadratic forms from (37) and to equate them to their observed values; i.e., to equate  $E(\underline{y}'\underline{Q}\underline{y})$  to the observed  $\underline{y}'\underline{Q}\underline{y}$ . This is exactly what is done with mean squares (which are quadratic forms of the observations) in the analysis of variance method for balanced data. But, whereas with balanced data there is "obviously" only one set of quadratic forms to use (the analysis of variance mean squares), and they lead to estimators that have some optimal properties, there are many sets of quadratics that can be used for unbalanced data. However, most of such sets lead to estimators that have few optimal properties and no particular set of quadratics has yet been established as more optimal than any other set.

Result (37) applies no matter what form of the model  $\underline{y} = \underline{X}\underline{\beta} + \underline{e}$  is used:  $\underline{\beta}$  always includes all the effects in the model, be they fixed or random or a mixture of both. In most situations we assume that  $E(\underline{e}) = 0$ , so that  $\text{var}(\underline{e})$  is  $E(\underline{e}\underline{e}') = \sigma_e^2 \underline{I}$ . In addition, when  $\underline{\beta}$  is a vector of fixed effects,  $E(\underline{\beta}\underline{\beta}') = \underline{\beta}E(\underline{e}') = 0$ ; and when  $\underline{\beta}$  includes elements that are random effects they are assumed to have zero mean and zero covariance with the elements in  $\underline{e}$ ; thus at all times we take



$$E(\underline{\beta}\underline{e}') = E(\underline{e}\underline{\beta}') = \underline{0}.$$

In a fixed effects model  $\underline{\beta}$  is a vector of fixed effects,  $E(y) = \underline{X}\underline{\beta}$  and  $\underline{V} = \text{var}(y) = \text{var}(\underline{e}) = \sigma_e^2 \underline{I}_N$ , where there are  $N$  observations, i.e.,  $y$  is  $N \times 1$ . Then (37) becomes

$$E(\underline{y}'\underline{Q}\underline{y}) = \underline{\beta}'\underline{X}'\underline{Q}\underline{X}\underline{\beta} + 2\sigma_e^2 \text{tr}(\underline{Q}).$$

In a mixed model  $\underline{\beta}'$  can be partitioned as

$$\underline{\beta}' = (\underline{\beta}'_1 \quad \underline{\beta}'_A \quad \underline{\beta}'_B \quad \dots \quad \underline{\beta}'_K)$$

where  $\underline{\beta}'_1$  contains all the fixed effects of the model (including the mean) and where the other  $\underline{\beta}'$ 's each represent the set of random effects for the factors A, B, C, ..., K, these random effects having zero means and zero covariances with the effects of any other set. [Although only single subscripts are used, interaction effects and/or nested-factor effects are not excluded by this notation. They are considered merely as factors, each identified by a single subscript rather than the letters of the appropriate main effects; for example, AB-interaction effects might be in the vector labeled  $\underline{\beta}'_F$ .] Then, with  $\underline{X}$  partitioned conformably with  $\underline{\beta}$ ,

$$\underline{y} = \underline{X}_1 \underline{\beta}_1 + \sum_{\theta=A}^K \underline{X}_\theta \underline{\beta}_\theta + e \quad (38)$$

$$E(\underline{y}) = \underline{X}_1 \underline{\beta}_1 \quad \text{and} \quad \underline{V} = \text{var}(\underline{y}) = \sum_{\theta=A}^K \underline{X}_\theta \text{var}(\underline{\beta}_\theta) \underline{X}_\theta' + \sigma_e^2 \underline{I}_N.$$

On making the usual assumption about the model that the random effects in each factor are uncorrelated and have the same variance, i.e., that

$$\text{var}(\underline{\beta}_\theta) = \sigma_{\theta}^2 \underline{I}_{N_\theta} \quad \text{for } \theta = A, B, \dots, K,$$

(37) then becomes

$$E(\underline{y}'\underline{Q}\underline{y}) = \underline{\beta}_1' \underline{X}_1' \underline{Q} \underline{X}_1 \underline{\beta}_1 + \sum_{\theta=A}^K \sigma_{\theta}^2 \text{tr}(\underline{Q} \underline{X}_{\theta} \underline{X}_{\theta}') + \sigma_e^2 \text{tr}(\underline{Q}) . \quad (39)$$

In a random model, all effects are random except  $\mu$ , so that  $\underline{\beta}_1 = \mu$  and  $\underline{X}_1 = \underline{1}$  and so from (39)

$$E(\underline{y}'\underline{Q}\underline{y}) = \mu^2 \underline{1}' \underline{Q} \underline{1} + \sum_{\theta=A}^K \sigma_{\theta}^2 \text{tr}(\underline{Q} \underline{X}_{\theta} \underline{X}_{\theta}') + \sigma_e^2 \text{tr}(\underline{Q}) , \quad (40)$$

where the coefficient of  $\mu^2$ , namely  $\underline{1}' \underline{Q} \underline{1}$ , is the sum of all elements of  $\underline{Q}$ . ( $\underline{1}$  is a vector of 1's.)

## 11.2 Normality assumptions

Distributional properties of quadratic forms  $\underline{y}'\underline{Q}\underline{y}$  are mostly known only when  $\underline{y}$  is normally distributed  $N(\underline{\mu}, \underline{V})$ . The well-known theorem (e.g. Graybill, [1961]) then applies, that  $\underline{y}'\underline{Q}\underline{y}$  has a non-central  $\chi^2$ -distribution if and only if  $\underline{Q}\underline{V}$  is an idempotent matrix. This theorem, however, has little use in the estimation of variance components from unbalanced data because few of the quadratics involved are such that  $\underline{Q}\underline{V}$  is idempotent. Despite this, the variance of  $\underline{y}'\underline{Q}\underline{y}$  is obtainable as

$$\text{var}(\underline{y}'\underline{Q}\underline{y}) = 2\text{tr}(\underline{Q}\underline{V})^2 + 4\underline{\mu}'\underline{Q}\underline{V}\underline{Q}\underline{\mu} \quad (41)$$

and the covariance between two quadratics  $\underline{y}'\underline{Q}\underline{y}$  and  $\underline{y}'\underline{R}\underline{y}$  is

$$\text{cov}(\underline{y}'\underline{Q}\underline{y}, \underline{y}'\underline{R}\underline{y}) = 2\text{tr}(\underline{Q}\underline{V}\underline{R}\underline{V}) + 4\underline{\mu}'\underline{Q}\underline{V}\underline{R}\underline{\mu} . \quad (42)$$

These results hold provided  $\underline{Q}$  and  $\underline{R}$  are symmetric. (41) is a special case of a more general result that the  $r$ 'th cumulant of  $\underline{y}'\underline{Q}\underline{y}$  is

$$K_r(\underline{y}'\underline{Q}\underline{y}) = 2^{r-1} (r-1)! [\text{tr}(\underline{Q}\underline{V})^r + r\underline{\mu}'(\underline{Q}\underline{V})^{r-1}\underline{V}^{-1}\underline{\mu}] .$$

No doubt (42) could be generalized in similar fashion.

12. The analysis of variance method (Henderson's Method 1)

The analysis of variance method of estimating variance components is essentially the only method used with balanced data. It is also the most frequently used method with unbalanced data, although its application is not as straightforward and its deficiencies are more pronounced. Nevertheless, it is likely to continue as an oft-used method and so considerable attention is devoted to it here. With balanced data the method consists of equating expected mean squares to their expected values. Essentially the same procedure is used with unbalanced data, as is now shown in terms of an example, the 2-way crossed classification with interaction.

The model for  $a$  levels of an A-factor crossed with  $b$  levels of a B-factor is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} \quad (43)$$

where  $y_{ijk}$  is the  $k$ 'th observation (for  $k = 1, 2, \dots, n_{ij}$ ) in the  $i$ 'th level of the A-factor and the  $j$ 'th level of the B-factor, where  $i = 1, 2, \dots, a$  and  $j = 1, 2, \dots, b$ . Thus  $n_{ij}$  is the number of observations in the  $(ij)$ 'th cell -- the  $i$ 'th level of A and the  $j$ 'th level of B. Since not all of the cells may contain observations we let  $s$  represent the number that do; i.e.,  $s$  is the number of  $n_{ij}$ 's for which  $n_{ij} > 0$ . Thus  $ab-s$  is the number of cells containing no data ( $n_{ij} = 0$ ). In (43),  $\mu$  is a general mean,  $\alpha_i$  is the effect due to the  $i$ 'th level of the A-factor,  $\beta_j$  is the effect due to the  $j$ 'th level of the B-factor,  $\gamma_{ij}$  is the interaction effect and  $e_{ijk}$  is the customary error term. In terms of (38) the model (43) is

$$y = \mu \underline{1} + X_{A-A} \beta_A + X_{B-B} \beta_B + X_{C-C} \beta_C + e$$

with  $\beta_A$ ,  $\beta_B$  and  $\beta_C$  being the vectors of effects corresponding respectively to the A and B-factors and to the AB-interactions. In random models these have variance matrices  $\sigma_{A-a}^2 I$ ,  $\sigma_{B-b}^2 I$  and  $\sigma_{AB-s}^2 I$  and covariances between different effects are all zero.

### 12.1 Analogous sums of squares

The analysis of variance of balanced data from the above model contains a line for the A-factor, which has a sum of squares (see Table 6)

$$\sum_{i=1}^a bn(\bar{y}_{i..} - \bar{y}_{...})^2 = \sum_{i=1}^a \frac{y_{i..}^2}{bn} - \frac{y_{...}^2}{abn}.$$

The analogous term for unbalanced data is

$$S_A = \sum_{i=1}^a \frac{y_{i..}^2}{n_{i.}} - \frac{y_{...}^2}{n_{..}}. \quad (44)$$

This is one of the terms used for estimating variance components by the analysis of variance method from unbalanced data: the computed value of  $S_A$  is equated to  $E(S_A)$  which can be derived from (40). It will be a linear function of the variances involved. The other terms used, all of them analogies of the balanced data analysis of variance sums of squares, are

$$S_B = \sum_{j=1}^b \frac{y_{.j.}^2}{n_{.j}} - \frac{y_{...}^2}{n_{..}}, \quad (45)$$

$$S_{AB} = \sum_{i=1}^a \sum_{j=1}^b \frac{y_{ij.}^2}{n_{ij}} - \sum_{i=1}^a \frac{y_{i..}^2}{n_{i.}} - \sum_{j=1}^b \frac{y_{.j.}^2}{n_{.j}} + \frac{y_{...}^2}{n_{..}}, \quad (46)$$

and

$$SSE = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{n_{ij}} y_{ijk}^2 - \sum_{i=1}^a \sum_{j=1}^b \frac{y_{ij}^2}{n_{ij}}. \quad (47)$$

Estimators of  $\sigma_A^2$ ,  $\sigma_B^2$ ,  $\sigma_{AB}^2$  and  $\sigma_e^2$  are derived by equating each of (44)-(47) to their expected values, which will be linear in the four variances. Before looking at the form of these expected values, which are derived from (40), certain characteristics of the S's are worth noting.

(i) Empty cells. Since  $n_{ij}$  is the number of observations in a cell it can, as mentioned, be zero. The summations in  $S_{AB}$  and SSE that involve  $n_{ij}$  in the denominator are therefore defined only for the  $(i, j)$  combinations for which  $n_{ij}$  is non-zero; i.e., the summing is over only those cells having observations in them. The possibility of zero denominators is thus removed. This is standard practice in unbalanced data analysis.

(ii) Balanced data. It is clear that when the data are balanced, i.e.  $n_{ij} = n$  for all  $i$  and  $j$ , then  $S_A$  reduces to  $\sum n(\bar{y}_{i..} - \bar{y}_{...})^2$ . In similar fashion  $S_B$ ,  $SS_{AB}$  and SSE of (45)-(47) also reduce to the familiar analysis of variance sums of squares for balanced data.

(iii) A negative 'sum of squares'. Suppose  $S_{AB}$  is computed for the data of Table 11.

Table 11

5 observations in a row-by-column analysis

Row	Column		Row
	1	2	Total
1	6	4	10
2	6, 42	12	60
Column Total	54	16	70

From (46)

$$S_{AB} = 6^2 + 4^2 + 48^2/2 + 12^2 - 10^2/2 - 60^2/3 - 54^2/3 - 16^2/2 + 70^2/5 = -22,$$

i.e.  $S_{AB}$  is negative. Thus  $S_{AB}$  is not a sum of squares; it is a quadratic form, but not positive semi-definite. Although this is not so with all S's of the example this illustration shows that, in general, the analogies of balanced data analyses of variance sums of squares may not always be sums of squares in unbalanced data. They might therefore be referred to as analogous sums of squares.

(iv) Uncorrected sums of squares. Because, as has just been demonstrated, the S's of unbalanced data estimation are not all positive semi-definite, it is convenient to deal with them in terms of uncorrected sums of squares, denoted by T's, as introduced for balanced data in Section 6.1. Thus with

$$T_A = \sum_{i=1}^a \frac{y_{i..}^2}{n_{i.}}, \quad T_B = \sum_{j=1}^b \frac{y_{.j.}^2}{n_{.j}}, \quad T_{AB} = \sum_{i=1}^a \sum_{j=1}^b \frac{y_{ij.}^2}{n_{ij}},$$

$$T_O = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{n_{ij}} y_{ijk}^2, \quad \text{and} \quad T_\mu = \frac{y_{...}^2}{n_{..}},$$
(48)

the S-terms of (44)-(47) are

$$S_A = T_A - T_\mu \quad \text{and} \quad S_B = T_B - T_\mu,$$

$$S_{AB} = T_{AB} - T_A - T_B + T_\mu \quad \text{and} \quad SSE = T_O - T_{AB}.$$
(49)

Although expected values of S's can be derived using expressions such as (44)-(47) directly in (40), they are obtained more easily by using (48) in (40), and then (49). This is so because of the tractability of the matrices involved in writing T's in the form  $\underline{y}'\underline{Q}\underline{y}$ . For example, that for  $T_A$  is the direct (Kronecker) sum of matrices  $(1/n_{i.})J_{n_{i.}}$ , where  $J_{n_{i.}}$  is a square matrix of order  $n_{i.}$  with every element unity. Indeed, as will be shown in (53), these matrices enable one to write down the expected value of any T quite generally from (40). Similarly, (41) and (42) can be used for deriving the variances and covariances of the T's, and these, by means of (49), lead to variances of the S's which in turn lead to variances of the variance component estimators. Thus the T's are a convenient means of deriving the estimators and their properties.

## 12.2 Expected values

Indication of the nature of expected values of S's can be gained from  $E(S_A)$  which, for the customary random model described above, is

$$\begin{aligned}
 E(S_A) = & \left( n_{..} - \frac{\sum_{i=1}^a n_{i.}^2}{n_{..}} \right) \sigma_A^2 + \left( \sum_{i=1}^a \frac{\sum_{j=1}^b n_{ij}^2}{n_{i.}} - \frac{\sum_{j=1}^b n_{..}^2}{n_{..}} \right) \sigma_B^2 \\
 & + \left( \sum_{i=1}^a \frac{\sum_{j=1}^b n_{ij}^2}{n_{i.}} - \frac{\sum_{i=1}^a \sum_{j=1}^b n_{ij}^2}{n_{..}} \right) \sigma_{AB}^2 + (a-1)\sigma_e^2.
 \end{aligned}
 \tag{50}$$

The most noticeable characteristic of this expression is the relative complexity of the coefficients of the  $\sigma^2$ 's - especially compared to their counterparts in balanced data shown in Table 7, to which they simplify (after division by  $a-1$  to reduce  $S_A$  to a mean square) when  $n_{ij} = n$  for all  $i$  and  $j$ .

A second point about (50) is its non-zero coefficient for every  $\sigma^2$  in the model. This contrasts with Table 7 where, for example,  $E(MST)$  has no term in  $\sigma_\beta^2$ : with unbalanced data from crossed classifications the expected value of every  $S$  contains every variance component of the model. [With nested classifications this is modified slightly in that an  $E(S)$ -term involves only the variance component of the factor it pertains to and those of the factors nested within it.]

Numeric methods of obtaining coefficients of variance components in expected values of  $S$ 's are given by Hartley [1967] and extended by Rao [1968]. Gaylor et al [1970] also show numeric procedures. Algebraic forms of these coefficients for specific models are to be found in a variety of places. A selection of references (with no attempt at completeness) is listed for convenience in Table 12.

Table 12

A selection of references to expected values of  $S$ -terms ("sums of squares") in random models with unbalanced data.

Classification	References
1-way	Graybill [1961, p. 351]
2-way, nested	Graybill [1961, p. 354]
2-way, crossed, no interaction	Graybill [1961, p. 359]
2-way, crossed, interaction	Searle [1958]
3-way, nested	Anderson and Bancroft [1952, p. 327]
3-way, crossed	Blischke [1968]
k-way, nested	Gates and Shiue [1962]
Finite populations	
3-way, nested	Gaylor and Hartwell [1969]
2, 3 & 4-way nested and/or crossed	Searle and Fawcett [1970]



Of the references listed, that by Gates and Shiue [1962] is noteworthy because it gives procedures for nested classifications of any order at all, and indicates relationships between coefficients of the  $\sigma^2$ 's in the  $E(S)$ -terms. Gower [1962] suggests somewhat the same thing, for a specific example, both papers making reference to Ganguli [1941]. The Gaylor and Hartwell [1969] paper is also important for it appears to be the first mention of analyzing unbalanced data from finite populations; Bennett and Franklin [1954] had considered them for balanced data, as had Cornfield and Tukey [1956] and Wilk and Kempthorne [1956]. General rules for changing to finite populations from infinite populations are given by Searle and Fawcett [1970].

### 12.3 General results

In terms of the general model (38) the T-term  $T_A$  for a factor A in a random model can be expressed in terms of  $y_{\cdot}(A_i)$  and  $n(A_i)$  defined respectively as the total and the number of the observations in the i'th level of the A-factor. Then

$$T_A = \sum_{i=1}^{N_A} \frac{[y_{\cdot}(A_i)]^2}{n(A_i)} \quad (51)$$

where  $N_A$  is the number of levels of the A-factor. On appropriately ordering the elements in  $y$  this is also

$$T_A = y' Q_A y \quad \text{with} \quad Q_A = \sum_{i=1}^{N_A} \frac{1}{n(A_i)} J_{n(A_i)} \quad (52)$$

$Q_A$  being the direct sum (denoted by  $\Sigma^+$ ) of  $N_A$  matrices  $[1/n(A_i)] J_{n(A_i)}$ . Hence from (40)

$$E(T_A) = N_{\mu}^2 + \sum_{\theta=A}^K \left( \sum_{i=1}^{N_A} \frac{\sum_{j=1}^{N_{\theta}} [n(A_i, \theta_j)]^2}{n(A_i)} \sigma_{\theta}^2 \right) + N_A \sigma_e^2 \quad (53)$$

where  $n(A_i, \theta_j)$  is the number of observations in the  $i$ 'th level of <sup>the</sup>  $A$ -factor and the  $j$ 'th level of the  $\theta$ -factor. With appropriate interpretation of  $n(A_i, \theta_j)$ ,  $n(A_i)$  and  $N_A$ , <sup>(53)</sup> result <sup>A</sup> applies quite generally to any  $T$  in any random model. Thus for  $T_o$ , the total sum of squares, it reduces to

$$E(T_o) = N\mu^2 + N \sum_{\theta=A}^K \sigma_{\theta}^2 + N\sigma_e^2$$

and for  $T_{\mu}$ , the correction for the mean, it becomes

$$E(T_{\mu}) = N\mu^2 + \sum_{\theta=A}^K \left\{ \sum_{j=1}^{N_{\theta}} [n(\theta_j)]^2 \right\} \sigma_{\theta}^2 / N + \sigma_e^2.$$

Clearly the term  $N\mu^2$  occurs in the expectation of every  $T$ . But because  $S$ 's involve only differences between  $T$ 's, expectations of  $S$ 's do not contain  $N\mu^2$ , and their coefficients of  $\sigma_e^2$  are correspondingly their "degrees of freedom".

When the number of observations in any set of data is  $N$  and the number of sub-most cells having data in them is  $s$ , the within-cell sum of squares SSE has expectation  $E(SSE) = (N - s)\sigma_e^2$ . Then, on defining

$\underline{s}$  = vector of  $S$ 's, but not SSE,

$\underline{\sigma}^2$  = vector of  $\sigma^2$ 's, but not  $\sigma_e^2$ ,

and  $\underline{f}$  = vector of "degrees of freedom", the coefficients of  $\sigma_e^2$  in  $E(\underline{s})$ ,

the expected values of all the  $S$ -terms involved in any random effects model can be written as

$$E \begin{bmatrix} \underline{s} \\ SSE \end{bmatrix} = \begin{bmatrix} \underline{P} & \underline{f} \\ \underline{0} & N-s \end{bmatrix} \begin{bmatrix} \underline{\sigma}^2 \\ \sigma_e^2 \end{bmatrix}. \quad (54)$$

Hence the analysis of variance estimators of  $\sigma_e^2$  and  $\sigma^2$  are

$$\hat{\sigma}_e^2 = \text{SSE}/(N - s) \quad (55)$$

and

$$\hat{\sigma}^2 = \underline{P}^{-1}(\underline{s} - \hat{\sigma}_e^2 \underline{f}). \quad (56)$$

In these equations the elements of  $\underline{P}$  are the coefficients of the  $\sigma^2$ 's (excluding  $\sigma_e^2$ ) in equations like (50). Clearly, these elements are of such a nature that there is no convenient explicit form for  $\underline{P}^{-1}$  and so the estimators are best left in the form of (55) and (56). For any particular situation one first derives  $E(\underline{s})$  and so obtains  $\underline{P}$  from  $E(\underline{s}) = \underline{P}\sigma^2 + \underline{f}\sigma_e^2$  of (54). Then (55) and (56) are used to obtain the estimators. These expressions are also useful in deriving variances of the estimators, as indicated in Section 12.5.

In random models all variance component estimators obtained by the analysis of variance method are unbiased. This is obviously so for  $\hat{\sigma}_e^2$  because  $E(\text{SSE}) = (N - s)\sigma_e^2$  in (55); and it is equally as clear for  $\hat{\sigma}^2$  because, from (56) and (54),

$$E(\hat{\sigma}^2) = \underline{P}^{-1}[E(\underline{s}) - \sigma_e^2 \underline{f}] = \underline{P}^{-1}[\underline{P}\sigma^2 + \sigma_e^2 \underline{f} - \sigma_e^2 \underline{f}] = \sigma^2.$$

This property of unbiasedness applies to all estimators from random models -- but not to those from mixed models, as is discussed in the next section.

An interesting variation on a random model is one considered by Harville (1967b)<sup>(1968)</sup>, wherein the analysis of variance estimators are biased. This is a model where dependence is assumed between the random effects of the model and the number of observations on them. Such a model is very realistic, for example, in dairy herd production, where the larger herds tend to be those with highest

production. If this dependence is taken into account the usual variance estimators are biased. Harville (1967b) discusses this situation for the 1-way classification and indicates the magnitude of the bias for a variety of situations. In doing so he assumes the bivariate distribution of the effects and their associated  $n$ 's to be of a normal-Poisson form. Similar discussion of the 2-way classification is given by Harville [1968].

#### 12.4 Mixed models

The expectation  $E(S_A)$  given in (50) is based on (43) being a random model. But suppose (43) were a mixed model with the  $\alpha_i$ 's being fixed and not random effects. The expectation would then be

$$E(S_A) = \left[ \sum_{i=1}^a n_{i.} \alpha_i^2 - \frac{(\sum_{i=1}^a n_{i.} \alpha_i)^2}{n_{..}} \right] + \left( \sum_{i=1}^a \frac{\sum_{j=1}^b n_{ij}^2}{n_{i.}} - \frac{\sum_{j=1}^b n_{.j}^2}{n_{..}} \right) \sigma_{\beta}^2 + \left( \sum_{i=1}^a \frac{\sum_{j=1}^b n_{ij}^2}{n_{i.}} - \frac{\sum_{i=1}^a \sum_{j=1}^b n_{ij}^2}{n_{..}} \right) \sigma_{\alpha\beta}^2 + (a-1) \sigma_e^2. \quad (57)$$

This is identical to (50) except for the first term which is now a function of the fixed effects, stemming from the first term of (39). Furthermore, this occurs in expected values of all  $S$ -terms (save that of  $SSE$ ).

More importantly, the function of the fixed effects is not the same from one  $E(S)$  term to the next. For example, with the  $\alpha_i$ 's fixed  $E(S_B)$ , of (49)

contains the term  $\sum_{j=1}^b \frac{(\sum_{i=1}^a n_{ij} \alpha_i)^2}{n_{.j}} - \frac{(\sum_{i=1}^a n_{i.} \alpha_i)^2}{n_{..}}$ , which differs from the first

term in  $E(S_A)$  of (57). Thus  $E(S_A - S_B)$  does not get rid of the fixed effects even though it does eliminate terms in  $\mu$ . This is true generally: in mixed

models, expected values of the S's contain functions of the fixed effects that cannot be eliminated by considering linear combinations of the S's. This means that the equations  $E(\underline{s}) = \underline{p}\sigma^2 + \sigma_e^2 \underline{f}$  of the random model take the form  $E(\underline{s}) = \underline{p}\sigma^2 + \sigma_e^2 \underline{f} + \underline{q}$  in the mixed model, where  $\underline{q}$  is a vector of quadratic functions of the fixed effects in the model. Hence  $\sigma^2$  cannot be estimated by this method; i.e. the analysis of variance method as applied to unbalanced data (Henderson's Method 1) cannot be used for mixed models. It yields biased estimators.

Two obvious ways of overcoming the above difficulty involve deviants from the true mixed model and must therefore be considered as unsatisfactory. The first is to ignore the fixed effects and eliminate them from the model: what remains is a random model for which variance components can be estimated. The second possibility is to assume the fixed effects are in fact random, and so treat the model as if it were random; in the resulting estimation process components for the fixed effect factors will be estimated and can be ignored. With each of these possibilities we deal with random models to which the estimation process is eminently suited, but the estimators will, in both cases, be biased because their expectations under the true mixed model will not equal the variance components of that model -- they will include functions (quadratics) of the fixed effects. Despite this, if the models which these approximations invoke are in any way acceptable alternatives to the mixed model then the approximations may be of some use. Furthermore, they utilize the relatively easy arithmetic of the method, which is sometimes advantageous in face of the greater complexity of analyses of mixed models that do yield unbiased estimators -- such as those discussed in Sections 13 and 14.

## 12.5 Sampling variances of estimators

Distributions of estimators of variance components from unbalanced data are unknown, except inasmuch as the estimators can sometimes be expressed as linear combinations of independent non-central  $\chi^2$ -variables, as discussed by Harville [1969d]. Some progress has been made in deriving sampling variances, however, although they are considerably more complicated than with balanced data.

For  $y$  having an arbitrary distribution, the only results available appear to be those of Hammersley [1949] and Tukey [1957a] for the 1-way classification. Hammersley considers infinite populations and Tukey, using polykays, deals with both infinite and finite populations. Tukey also considers the effect of weighting the group means in different ways in order to reduce the variance of the between-group variance estimator. From a limited numerical study he concludes that when the between component exceeds 50% (100%) of the within component in moderately (substantially) unbalanced data, then equal weighting of the group means rather than the customary weighting by group size seems to give a better estimate of the between component. Under normality assumptions (that the random effects in the model are normally distributed), Robertson [1962] suggests weighting inversely proportional to the square of group size might be more appropriate.

Since most derivations of sampling variances have been made under normality assumptions, further discussion is confined to this situation. Foremost is the obvious result that with normality the residual sum of squares always has, apart from a constant, a central  $\chi^2$ -distribution. Hence

$$\hat{\sigma}_e^2 = \text{SSE}/(N - s) \sim \chi_{N-s}^2 \left( \frac{\sigma_e^2}{N - s} \right) \text{ and so } v(\hat{\sigma}_e^2) = \frac{2\sigma_e^4}{N - s}. \quad (58)$$

just as with balanced data. This is the one exception to unknown sampling distributions.

For the 1-way classification Crump [1951] and Searle [1956] give the variance of the between-group component estimator. Unfortunately both results contain typographical errors. Crump's corrected formula (in his notation) is

$$v(\hat{\sigma}_A^2) = \frac{2\sigma_e^4}{n_o^2} \left\{ \frac{1}{(a-1)^2} \left[ \left( \frac{1}{n} \sum \frac{n_h^2}{w_h^2} \right)^2 + \sum \frac{n_h^2}{w_h^2} - \frac{2}{n} \sum \frac{n_h^3}{w_h^2} \right] + \frac{1}{n - a} \right\} \quad (59)$$

where  $n_h$  is the number of observations in the  $h$ 'th group ( $h = 1, 2, \dots, a$ ), with  $w_h = n_h \sigma_e^2 / (\sigma_e^2 + n_h \sigma_A^2)$  and  $n_o = (n - \sum n_h^2 / n) / (a - 1)$ . The published version of (59) omits the  $1/n$  term from inside the squared bracket of the first term. And Searle's corrected formula (in his notation) is

$$v(\hat{\sigma}_A^2) = \frac{1}{f^2} \left[ \frac{2\sigma_e^4(N-1)}{(c-1)(N-c)} + \frac{4\sigma_e^2\sigma_A^2(N^2 - S_2)}{N(c-1)^2} + \frac{2\sigma_A^4(N^2S_2 + S_2^2 - 2NS_3)}{N^2(c-1)^2} \right] \quad (60)$$

where  $N$  is the number of observations,  $c$  is the number of groups,  $S_2 = \sum n_h^2$ ,  $S_3 = \sum n_h^3$  and  $f = (N - S_2/N) / (c - 1)$ . The error in the published form of (60) is that of having 2 instead of 4 in the second term. The more familiar form of (60) is

$$v(\hat{\sigma}_A^2) = \frac{2\sigma_e^4 N^2 (N-1)(c-1)}{(N^2 - S_2)^2 (N-c)} + \frac{4\sigma_e^2 \sigma_A^2 N}{N^2 - S_2} + \frac{2\sigma_A^4 (N^2 S_2 + S_2^2 - 2NS_3)}{(N^2 - S_2)^2}.$$

It was consideration of the 1-way classification in Searle [1956] that led to applying results like (41) and (42) to the variance component problem. These results yield variances and covariances of the  $T$ 's which, if

we write  $\underline{t}$  for the vector of T's, constitute the elements of the matrix  $\text{var}(\underline{t})$ . If the relationship between  $\underline{t}$  and  $\underline{s}$  [e.g. (49)] is represented as  $\underline{s} = \underline{H}\underline{t}$ , the estimators in (56) become

$$\hat{\sigma}^2 = \underline{P}^{-1}(\underline{H}\underline{t} - \hat{\sigma}_e^2 \underline{f}) \quad \text{with } \underline{s} = \text{vector of S's}$$

$$\underline{t} = \text{vector of T's and } \underline{s} = \underline{H}\underline{t} .$$

Because  $\hat{\sigma}_e^2$  and each element of  $\underline{t}$  have zero covariance, the covariance matrix of  $\hat{\sigma}^2$  is

$$\text{var}(\hat{\sigma}^2) = \underline{P}^{-1}[\underline{H} \text{var}(\underline{t})\underline{H}' + v(\hat{\sigma}_e^2)\underline{f}\underline{f}']\underline{P}^{-1} , \quad (61)$$

where  $v(\hat{\sigma}_e^2)$  is given by (58); and the vector of covariances between  $\hat{\sigma}_e^2$  and  $\hat{\sigma}^2$  is

$$\text{cov}(\hat{\sigma}_e^2, \hat{\sigma}^2) = -\underline{P}^{-1}\underline{f}v(\hat{\sigma}_e^2) . \quad (62)$$

In these results  $\underline{H}$  and  $\underline{P}^{-1}$  are not necessarily symmetric — indeed they seldom will be.

The difficult part of (61) is  $\text{var}(\underline{t})$ .  $\underline{P}$  comes from the estimation procedure  $E(\underline{s}) = \underline{P}\sigma^2 + \underline{f}\sigma_e^2$ , and  $\underline{H}$  represents the relationship between the S's and T's; but  $\text{var}(\underline{t})$  is the matrix of variances and covariances of the T's that have to be developed from (41) and (42). Writing the T's in the form of (52),



it is easily shown for random effects models that the terms in  $\underline{\mu}$  in (41) and (42) are, for all T's,  $4\mu^2 \underline{1}' \underline{V} \underline{1}$ . Hence, because T's are involved only as differences of each other in S's, the last terms of (41) and (42) can be neglected in deriving variances and covariances of the T's to be used in  $\underline{H} \text{var}(\underline{t}) \underline{H}'$  of (61).

Expressions for variances and covariances of T's, which, from (41) and (42) are quadratics in the unknown  $\sigma^2$ 's, have been developed for several specific cases. The 2-way crossed classification with interaction is dealt with by Searle [1958], and the 2-way nested classification by Searle [1961]; and the 3-way nested classification is treated by Mahamunulu [1963]. Blischke [1966] deals with a specific 3-way classification and Blischke [1968] develops a general procedure applicable to any r-way crossed classification. An appendix to this latter paper contains specific expressions for  $\text{var}(\underline{t})$  of a 3-way crossed classification with all interactions, involving a matrix of order  $36 \times 36$ , of the coefficients of the squares and products of the  $\sigma^2$ 's in the variances and covariances of the T's. In an r-way crossed classification containing all interactions this matrix will be of order  $2^{r-1}(2^r + 1)$ , values of which are 3, 10, 36, 136 and 528 for  $r = 1, 2, 3, 4$  and 5 respectively.

The quadratics of the unknown variance components that are the elements of  $\text{var}(\underline{t})$  are "messy" functions of the  $n_{ij}$ 's of the data; for example, from Searle [1958], the variance of  $T_A$  of (48) is

$$\text{var}(T_A) = 2 \left\{ \sum_i \left[ \sum_j n_{ij} (n_{i.} \sigma_A^2 + n_{ij} \sigma_B^2 + n_{ij} \sigma_{AB}^2 + \sigma_e^2) / n_{i.} \right]^2 + \sum_i \sum_{i' \neq i} \frac{(\sum_j n_{ij} n_{i'j})^2}{n_{i.} n_{i'.}} \sigma_B^2 \right\}.$$

Although this expression can be expanded into a more computable form there appears to be no way of simplifying it analytically. Consequently explicit expressions for the variances and covariances of the estimated  $\sigma^2$ 's cannot be derived from (61) and (62); they must remain in that form. Implications of this are two-fold: the effects of unbalancedness (of different values of the  $n_{ij}$ 's) on the behaviour of variances of variance component estimators cannot be studied, and (ii) these variances, through being functions of the unknown variance components, can be evaluated in any particular case only after allocating to those components a set of numerical values. This can be done either by using estimated  $\sigma^2$ 's derived from the data (hoping, presumably, that none of them are negative), or by using any set or sets of arbitrary values that seem reasonable for the data at hand. Neither of these possibilities gives estimates of  $\text{var}(\hat{\sigma}^2)$  that have any known, optimal properties.

The elements of (61) and (62) are linear functions of the squares and products of the  $\sigma^2$ 's. In this situation Mahamudulu [1963] has pointed out a method for deriving unbiased estimators

of the elements of  $\text{var}(\hat{\sigma}_{\underline{e}}^2)$ . It consists of replacing in (61) and (62) every squared variance  $\sigma_A^4$  (including  $\sigma_e^2$ ) by  $\hat{\sigma}_A^4 - \hat{v}(\hat{\sigma}_A^2)$  and every product of variances,  $\sigma_A^2 \sigma_B^2$ , (also including error) by  $\hat{\sigma}_A^2 \hat{\sigma}_B^2 - \hat{\text{cov}}(\hat{\sigma}_A^2, \hat{\sigma}_B^2)$ . Calling the resulting expressions  $\hat{\text{var}}(\hat{\sigma}_{\underline{e}}^2)$  and  $\hat{\text{cov}}(\hat{\sigma}_{\underline{e}}^2, \hat{\sigma}_{\underline{e}}^2)$ , they are then equations in  $\hat{v}(\hat{\sigma}_A^2)$  and  $\hat{\text{cov}}(\hat{\sigma}_A^2, \hat{\sigma}_B^2)$ , the solutions for which are unbiased estimators of the variances and covariances of the estimated  $\sigma^2$ 's. For obtaining unbiased estimators explicitly this is clearly an intractable procedure in anything but the simplest case. Nevertheless, Ahrens (1965) has formulated it more succinctly. Suppose *that*  $v(\hat{\sigma}_e^2)$  of (58), the different elements of  $\text{var}(\hat{\sigma}_{\underline{e}}^2)$  in (61) and the elements of  $\text{cov}(\hat{\sigma}_{\underline{e}}^2, \hat{\sigma}_{\underline{e}}^2)$  in (62) are arrayed as a vector  $\underline{v}$  -- of order  $2^{r-1}(2^r + 1)$  in an  $r$ -way cross-classification. Then if, in similar manner, the squares and products of the  $\sigma^2$ 's are also arrayed in a vector  $\underline{y}$ , the equations (58), (61) and (62) can be written as

$$\underline{v} = \underline{A}\underline{y} \quad (63)$$

where  $\underline{A}$  is square of order  $2^{r-1}(2^r + 1)$ . [This is not the matrix referred to earlier in discussing  $\text{var}(\underline{t})$ ; there it was a case of writing the different elements of  $\text{var}(\underline{t})$  as vector  $\underline{B}\underline{y}$ ;  $\underline{B}$  differs from  $\underline{A}$ , though having the same order.] With this formulation an unbiased estimator of  $\underline{v}$  is

$$\hat{\underline{v}} = (\underline{I} + \underline{A})^{-1} \underline{A} \hat{\underline{y}} \quad (64)$$

where  $\hat{\underline{y}}$  is the vector of the squares and powers of the  $\hat{\sigma}^2$ 's, the estimated  $\sigma^2$ 's.

This procedure is similar to that of Section 10.4. It results from the fact that by its definition  $\hat{y}$  has expectation  $E(\hat{y}) = \bar{y} + y$  and so, by means of (63), the expected value of  $\hat{y}$  in (64) is  $y$ .

Example. For the 1-way classification, (63) is

$$\begin{bmatrix} v(\hat{\sigma}_e^2) \\ \text{cov}(\hat{\sigma}_A^2, \hat{\sigma}_e^2) \\ v(\hat{\sigma}_A^2) \end{bmatrix} = \begin{bmatrix} k_1 & 0 & 0 \\ k_2 & 0 & 0 \\ k_3 & k_4 & k_5 \end{bmatrix} \begin{bmatrix} \sigma_e^4 \\ \sigma_A^2 \sigma_e^2 \\ \sigma_A^4 \end{bmatrix}$$

where  $k_1 = 2/(N - c)$ , as in (58),  $k_2 = -2/[f(N - c)]$  and  $k_3, k_4$  and  $k_5$  are the coefficients of  $\sigma_e^4$ ,  $\sigma_A^2 \sigma_e^2$  and  $\sigma_A^4$  in (60). Then, by (64), unbiased estimators of the sampling variances and covariance are

$$\begin{bmatrix} \hat{v}(\hat{\sigma}_e^2) \\ \hat{\text{cov}}(\hat{\sigma}_A^2, \hat{\sigma}_e^2) \\ \hat{v}(\hat{\sigma}_A^2) \end{bmatrix} = \begin{bmatrix} 1+k_1 & 0 & 0 \\ k_2 & 1 & 0 \\ k_3 & k_4 & 1+k_5 \end{bmatrix}^{-1} \begin{bmatrix} k_1 & 0 & 0 \\ k_2 & 0 & 0 \\ k_3 & k_4 & k_5 \end{bmatrix} \begin{bmatrix} \hat{\sigma}_e^4 \\ \hat{\sigma}_e^2 \hat{\sigma}_A^2 \\ \hat{\sigma}_A^4 \end{bmatrix}$$

$$= \frac{1}{(1+k_1)(1+k_5)} \begin{bmatrix} k_1(1+k_5) & 0 & 0 \\ k_2(1+k_5) & 0 & 0 \\ k_3 - k_2k_4 & k_4(1+k_1) & k_5(1+k_1) \end{bmatrix} \begin{bmatrix} \hat{\sigma}_e^4 \\ \hat{\sigma}_e^2 \hat{\sigma}_A^2 \\ \hat{\sigma}_A^4 \end{bmatrix} \quad (65)$$

It is clear that (64) can be used only after (58), (61) and (62) have been written in the form of (63). Nested classifications, of which the above example is the simplest case, provide a relatively easy opportunity for this, for then the matrix  $P$  in (61) is triangular and, at least in simple cases, its inverse is easily obtained. Hence (61) and (62) can be readily expressed in the form of (63), from which (64) can be used. This is evident in the example.

Variances of variance component estimators appear, from the literature, to have been used very little on actual data. This is probably so because of their algebraic complexity. In one series of applications (e.g., Searle [1961b] and Evans [1968]) the variance of the interaction component estimator in a 2-way classification was very large. However, in the data used there was only a small fraction of the possible subclasses (5 - 15%) that contained observations, with those subclasses having an average of only 2 observations each. The paucity of such data for estimating interaction variances seems clearly apparent. Nevertheless,

the occurrence of such data highlights the need for improved methods of estimating variance components.

The nature of the distribution of variance component estimators has been studied through computer simulation by Leone et al [1968] for two special cases of unbalanced data. They studied nested designs that could be described as embodying planned unbalancedness, in contrast to the unbalancedness which is often fait accompli in survey data. The designs used are those of the "staggered" and "inverted" variety, as described by Bainbridge [1963]. The extent of the computer simulation was identical to that used by the same workers, Leone et al [1966], in studying a single balanced nested classification as discussed in Section 10.3. The main conclusion is that, for the specific designs studied, Pearson Type III curves might be suitable for describing the distributions. However, the limited degree of unbalancedness implicit in their designs precludes extrapolating conclusions to unbalanced data generally.

#### 12.6 Synthesis

The calculation, for any particular set of data, of coefficients of  $\sigma^2$ 's in expectations of S's and T's without first requiring the algebraic form of these coefficients can be achieved by a method developed by Hartley [1967], which he calls the method of 'synthesis'. The method also applies to calculating coefficients of squares and products of  $\sigma^2$ 's in variances and covariances

of quadratics. Furthermore, it applies not just to S's and T's but to any homogeneous quadratic form in the observations, and it requires no distributional properties of the model.

The method is easily described in terms of the general model (38) where, for random models, the only fixed effect is  $\mu$ . Then, writing  $T_A$  as in (52) and, from  $\underline{X}_\theta$  of (38), defining  $\underline{x}(\theta, j)$  as the  $j$ 'th column of  $\underline{X}_\theta$ , the method of synthesis obtains the coefficient of  $\sigma_\theta^2$  in  $E(T_A)$  as

$$k(\sigma_\theta^2, T_A) = \sum_{j=1}^{N_\theta} T_A[\underline{x}(\theta, j)] . \quad (66)$$

Thus the method uses each column in  $\underline{X}_\theta$  as a column of data (all 0's and 1's) to calculate  $T_A$ , and sums the results over all columns of  $\underline{X}_\theta$ . The sum, as in (66), is the coefficient of  $\sigma_\theta^2$  in  $E(T_A)$ .

This numerical procedure has no recourse to explicit algebraic forms of the coefficients. Since it applies to any quadratic form in the place of  $T_A$  it can be used directly on the S's, and so paraphrasing Hartley we can say: apply the analysis of variance method in turn to each of the  $N_\theta$  columns of  $\underline{X}_\theta$  used as data, and for a particular quadratic  $Q(\underline{y})$  form the sum of the  $Q(\underline{y})$  over the  $N_\theta$  analyses of variance, to obtain  $k[\sigma_\theta^2, Q(\underline{y})]$ , the coefficient of  $\sigma_\theta^2$  in  $E(Q(\underline{y}))$ .

Carrying out  $\sum_{\theta=A}^K N_{\theta}$  analyses of variance and summing them appropriately thus gives the coefficients of the  $\sigma^2$ 's in the expected values of all the S's. Since many of the 'observations' in these analyses will be zero, any computer procedure designed for this task should take account of this many-zeroed feature of the 'data'. Furthermore, the non-zero elements of  $\underline{x}(\theta, j)$  are all 1's, and this leads very readily to showing the equivalence of (65) to the coefficient of  $\sigma_{\theta}^2$  in (53).

The method also extends to finding coefficients of products of  $\sigma^2$ 's in expected values of products of quadratic forms. This involves sums of the form

$$\sum_{j=1}^{N_{\theta}} \sum_{j'=1}^{N_{\varphi}} T_A [\underline{x}(\theta, j) + \underline{x}(\varphi, j')] T_B [\underline{x}(\theta, j) + \underline{x}(\varphi, j')]$$

where a column of  $\underline{X}_{\theta}$  is added to one of  $\underline{X}_{\varphi}$ , in all possible combinations, and these sums are used as 'data' vectors in  $T_A$  and  $T_B$ . When the data are extensive and the  $\underline{X}$ 's have many columns (perhaps hundreds, as often occurs with random models) these formulae, although computationally well suited to obtaining coefficients numerically in specific situations, will nevertheless entail quite sizeable computer time and facilities. Extensions of the method to general design matrices and mixed models are considered by Rao [1968].

### 13 Adjusting for bias in mixed models (Henderson Method 2)

Mixed models involve dual estimation problems — estimating both fixed effects and variance components. For the moment attention is directed to estimating just the variance components. In some situations this is exactly what might be done in practice; with genetic data, for example, time or year



effects might be considered fixed and of little interest compared to the genetic variance components. On the other hand, time trends may be of very real interest in some data, in which case their estimation together with that of the variance components would be considered simultaneously. This dual estimation problem is considered subsequently.

As indicated in Section 12.4 the analysis of variance method for mixed models leads, with unbalanced data, to biased estimators of variance components. The method known as Method 2 in Henderson [1953] is designed to correct this deficiency. It uses the data to first estimate fixed effects of the model and then, using these estimators to adjust the data, variance components are estimated from the adjusted data, using the analysis of variance method. The whole procedure is designed so that the resulting variance component estimators are not biased by the presence of the fixed effects in the model, as they are with analysis of variance estimators derived from the basic data. So far as the criterion of unbiasedness is concerned, this is certainly achieved by Method 2. But the general method of analyzing data adjusted according to some estimator of the fixed effects is open to criticism on other grounds: It cannot be uniquely defined, and a simplified form of it, of which Henderson's Method 2 is a special case, cannot be used whenever the model includes interactions between the fixed effects and the random effects.

The general approach of Method 2 can be considered in terms of the model

$$\underline{y} = \mu \underline{1} + \underline{X}_f \underline{\beta}_f + \underline{X}_r \underline{\beta}_r + \underline{e}$$

where all fixed effects other than  $\mu$  are represented by  $\underline{\beta}_f$ , and all random

effects by  $\beta_r$ . As usual  $E(\beta_r) = 0$  and so  $E(\beta_r \beta_r') = V(\beta_r)$ , the variance-covariance matrix of the random effects. The general effect of correcting the data vector  $y$  according to an estimator of the fixed effects  $\beta_f$  is to suppose that such an estimator is  $\tilde{\beta}_f = Ly$  for some matrix  $L$  so that the vector of corrected data is  $z = y - X_f \tilde{\beta}_f$ . It can then be shown (Searle, [1968]) that the model for  $z$  contains no terms in  $\beta_f$  provided  $L$  is a generalized inverse of  $X_f$ . Under this condition the analysis of variance method applied to  $y - X_f \tilde{\beta}_f$ , will yield unbiased estimators of the variance components. However, the fact that  $L$  has only to be <sup>a</sup> generalized inverse of  $X_f$  indicates the arbitrariness of the method. This lack of specificity means the method is not uniquely defined and hence is impractical.

The model for  $z = y - X_f \tilde{\beta}_f$  just described contains no term in  $\beta_f$ . An additional restriction <sup>would be</sup> for the model to have the same term  $X_r \beta_r$  as does the model for  $y$ , as well as a mean term  $\mu_1 1$  where  $\mu_1$  is not necessarily equal to  $\mu$ . For this to occur  $L$  need not be a generalized inverse of  $X_f$  but it must satisfy three conditions:

$$X_f L X_f = 0; \text{ all row sums of } X_f L \text{ equal; all rows of } (X_f - X_f L X_f) \text{ equal.} \quad (67)$$

Although the non-unique condition on  $L$ , that  $X_f L X_f = X_f$ , has been replaced by these 3 conditions they too do not determine  $L$  uniquely. Furthermore, it has been shown <sup>(cf. [17])</sup> that a consequence of them is that the model for  $y$  cannot contain interactions between fixed and random effects. Whether such interactions are considered fixed or random their existence leads to meaningless conclusions if this simplified form of the general method is to be feasible. This is a severe limitation on the method.

Henderson's [1953] Method 2 is a particular way of carrying out the simpler form of the generalized method. It estimates  $\underline{\beta}_f$  as  $\underline{\beta}_f = \underline{L}\underline{y}$  with an  $\underline{L}$  that satisfies (67) derived from assuming, temporarily and just for the purpose of estimating  $\underline{\beta}_f$ , that  $\mu = 0$  and that the random effects are fixed. Then  $\tilde{\underline{\beta}}_f$  is obtained from the equations

$$\underline{X}'\underline{X} \begin{bmatrix} \tilde{\underline{\beta}}_f \\ \tilde{\underline{\beta}}_r \end{bmatrix} \equiv \begin{bmatrix} \underline{X}'_f\underline{X}_f & \underline{X}'_f\underline{X}_r \\ \underline{X}'_r\underline{X}_f & \underline{X}'_r\underline{X}_r \end{bmatrix} \begin{bmatrix} \tilde{\underline{\beta}}_f \\ \tilde{\underline{\beta}}_r \end{bmatrix} = \begin{bmatrix} \underline{X}'_f\underline{y} \\ \underline{X}'_r\underline{y} \end{bmatrix}. \quad (68)$$

It is the manner in which these are solved that leads to the solution being  $\tilde{\underline{\beta}}_f = \underline{L}\underline{y}$  with  $\underline{L}$  satisfying (67). The essential part of the solution is picking a generalized inverse of  $\underline{X}'\underline{X}$  by reducing it to full rank in such a way that in striking out rows and columns from  $\underline{X}'\underline{X}$  as many as possible are rows and columns through  $\underline{X}'_f\underline{X}_f$ . Details of this process, and the reasons for its satisfying (67) are given in Searle [1968]. Having obtained  $\underline{L}$  the variance components are estimated by using the analysis of variance method on  $\underline{y} - \underline{X}_f\tilde{\underline{\beta}}_f$ . Although  $\underline{y} - \underline{X}_f\tilde{\underline{\beta}}_f - \underline{X}_r\tilde{\underline{\beta}}_r$  is invariant to whatever solution of (68) is used for  $\tilde{\underline{\beta}}_f$  and  $\tilde{\underline{\beta}}_r$ , this invariance does not apply to  $\underline{y} - \underline{X}_f\tilde{\underline{\beta}}_f$ . The lack of unique specification of the method is thus readily apparent. Even though re-writing the model in some full rank form leads (with different definitions of  $\underline{X}_f$  and  $\underline{X}_r$ ) to a solitary solution for (68), doing so does not supersede the underlying non full rank model in terms of which the variance components are defined. Therefore discussion of Method 2 in terms of a full rank model, as is done by Oktaba [1968] for example, whilst appearing to avoid the lack of uniqueness of the method, does not really

do so. Because the method has this deficiency, and also because it cannot be used for models that contain interactions between fixed and random effects [a deficiency of any method for which (67) is satisfied], its use is discouraged by Searle [1968].

A variation on Henderson's Method 2 is to estimate the fixed effects by temporarily assuming the random effects do not exist. Then  $\tilde{\beta}_f$  is derived from  $\underline{X}'_f \underline{X}_f \tilde{\beta}_f = \underline{X}'_f \underline{y}$ , which removes the lack of uniqueness in  $\underline{z} = \underline{y} - \underline{X}_f \tilde{\beta}_f$  because  $\underline{X}_f (\underline{X}'_f \underline{X}_f)^{-} \underline{X}'_f$  is invariant to the choice of  $(\underline{X}'_f \underline{X}_f)^{-}$ . However, conditions (67) are no longer satisfied and the computational advantages which they imply are not gained.

#### 14 The fitting constants method (Henderson's Method 3)

The third method described by Henderson [1953] is based on the method of fitting constants traditionally used in fixed effects models. It uses reductions in sums of squares due to fitting different sub-groups of factors in the model, using them in exactly the same manner as the S's are used in the analysis of variance method, namely estimating the variance components by equating each computed reduction to its expected value. We illustrate the method in terms of an example, and then summarize the general case.

##### 14.1 Reductions in sums of squares

In the earlier example of the 2-way classification random model with interaction, the four quadratics equated to their expected values in the analysis of variance method are  $S_A$ ,  $S_B$ ,  $S_{AB}$  and SSE shown as functions of T's in equations (49). The fitting constants method also uses four quadratics, derived from fitting the model

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$$

shown in (43), and three sub-models

$$y_{ijk} = \mu + e_{ijk}, \quad y_{ijk} = \mu + \alpha_i + e_{ijk}, \quad \text{and} \quad y_{ijk} = \mu + \alpha_i + \beta_j + e_{ijk}.$$

In fitting these models the total sum of squares is  $T_0 = \sum \sum \sum y_{ijk}^2$  just as in (48). With each model there is a reduction in sum of squares,  $\underline{y}'\underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{y}$ , due to fitting the model. These reductions can be denoted by

$$R(\mu, A, B, AB), \quad R(\mu), \quad R(\mu, A) \quad \text{and} \quad R(\mu, A, B)$$

respectively, where the letters in parentheses indicate the factors fitted in the respective models. [In this notation AB represents, as usual, A-by-B interaction.] By way of example, the last of the above models fits  $\mu$ , A- and B-factors, and so the reduction is symbolized as  $R(\mu, A, B)$ . Writing the model as  $\underline{y} = \underline{X}\underline{b} + \underline{e}$ , where  $\underline{b}$  is the vector containing  $\mu$ , the  $\alpha$ 's and  $\beta$ 's, we have  $R(\mu, A, B) = \underline{y}'\underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{y}$  for that  $\underline{X}$ . The fitting constants method of estimating variance components uses  $T_0$  and these  $R(\ )$ -reductions by equating certain differences among them to their expected values, so setting up equations in the  $\sigma^2$ 's whose solutions are the estimators. These differences and their expectations are indicated in Table 13.

Table 13

Fitting constants . method of estimating variance components for a 2-way classification, random model.

Term	Definition	Expectation
$R(A,B,AB \mu)$	$= R(\mu,A,B,AB) - R(\mu)$	Linear in $\sigma_A^2$ , $\sigma_B^2$ , $\sigma_{AB}^2$ and $\sigma_e^2$
$R(B,AB \mu,A)$	$= R(\mu,A,B,AB) - R(\mu,A)$	Linear in $\sigma_B^2$ , $\sigma_{AB}^2$ and $\sigma_e^2$
$R(AB \mu,A,B)$	$= R(\mu,A,B,AB) - R(\mu,A,B)$	Linear in $\sigma_{AB}^2$ and $\sigma_e^2$
SSE	$= T_o - R(\mu,A,B,AB)$	$(N - s)\sigma_e^2$

Although for this example most of these reductions simplify in terms of the T's defined earlier, their presentation in Table 13 is in accord with the general application of the fitting constants method and illustrates properties of the method that make it important. The last entry in Table 13 is SSE, equal to the total sum of squares minus the reduction due to fitting the full model. This is always used and it provides, as usual, the estimator of  $\sigma_e^2$ . The second to last entry,  $R(AB|\mu,A,B)$ , is read as "the reduction due to AB after taking account of  $\mu$ , A and B"; or, more correctly "the reduction due to fitting  $\mu$ , A, B and AB, over and above that due to  $\mu$ , A and B", namely  $R(\mu,A,B,AB) - R(\mu,A,B)$  as shown in the table. These are the kinds of terms used generally in the fitting constants method. Others for the example are shown in Table 13, and the general form is discussed below.

The notation used here,  $R(AB|\mu,A,B)$  for example, is convenient and complete.

It conveniently indicates what each term is, and it completely specifies each term with no opportunity for confusion such as exists with other notations. For example, does  $R_{AB}$  mean  $R(AB)$ ,  $R(\mu, AB)$ ,  $R(\mu, AB) - R(\mu)$ , or  $R(AB|\mu, A, B)$ ? The latter notation carries no element of doubt.

Sub-models implied in any use of the  $R(.|.)$  notation must, in regard to reductions in sums of squares, be different from each other and from the full model. For example  $R(\mu, A, AB)$  cannot be used because it implies a model  $y_{ijk} = \mu + \alpha_i + \gamma_{ij} + e_{ijk}$  which, so far as reduction in sum of squares is concerned, is indistinguishable from the full model; i.e.  $R(\mu, A, AB) \equiv R(\mu, A, B, AB)$ . Similarly  $R(\mu, AB) \equiv R(\mu, A, B, AB)$ . Hence terms like  $R(B|\mu, A, AB)$  and  $R(A, B|\mu, AB)$  are identically zero and never part of the estimation process.

## 14.2 Expectations

Details of the expectations in Table 13 are not shown because they involve cumbersome expressions for the coefficients of the  $\sigma^2$ 's. However, the general form of the expectations merits emphasis.

All expectations are obtained under the full model. Thus although  $R(AB|\mu, A, B)$  involves  $R(\mu, A, B)$  which comes from the model  $y_{ijk} = \mu + \alpha_i + \beta_j + e_{ijk}$ , the expectation of  $R(AB|\mu, A, B)$  is obtained under the full model  $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$ . This is true of all expectations in Table 13 and applies generally to the fitting constants method: expectations are always taken under the full model.

Every  $R(.|.)$  term in Table 13 is the reduction for the full model minus that for some sub-model; i.e., is  $R(\mu, A, B, AB)$  minus the  $R(.)$  for a sub-model. As a result, each expectation contains  $\sigma_e^2$  and just the  $\sigma^2$ 's for those factors which precede the bar in  $R(.|.)$ ; e.g., the expectation of  $R(AB|\mu, A, B)$  involves only  $\sigma_{AB}^2$  in addition to  $\sigma_e^2$ , and similarly  $R(B, AB|\mu, A)$  involves only  $\sigma_B^2$ ,  $\sigma_{AB}^2$  and  $\sigma_e^2$ . This

is true only for  $R(.|.)$  terms which are of the form just described, namely  $R(.)$  for the full model minus the  $R(.)$  for a sub-model. Suppose the full model can be described in terms of two vectors of factors  $\beta_1$  and  $\beta_2$ . Then  $R(\beta_1, \beta_2)$  is the reduction for the full model, and the rule for expectations illustrated in Table 13 is that the expected value of  $R(\beta_2|\beta_1) = R(\beta_1, \beta_2) - R(\beta_1)$  always involves just  $\sigma_e^2$  and the  $\sigma^2$ 's of the factors in  $\beta_2$ . This general result is proven and discussed in Section 14.6. It is the basis of the results in Table 13.

The rule for the expected value of  $R(\beta_2|\beta_1)$  applies only when the full model, under which the expected value is taken, is represented by  $\beta_1, \beta_2$ . Thus it does not apply to  $R(A|\mu, B)$  when the full model consists of  $\mu$  and A-, B- and AB-factors. However, by expressing  $R(A|\mu, B)$  as the difference between two reductions to which the rule does apply, its expectation is readily derived. Thus

$$R(A|\mu, B) = R(\mu, A, B) - R(\mu, B) = R(A, AB|\mu, B) - R(AB|\mu, A, B)$$

and expectations of the second pair of reductions can be derived from the rule. Details of the expectations are derived from (70).

### 14.3 Estimation

In all cases  $\sigma_e^2$  is estimated as  $\hat{\sigma}_e^2 = \text{SSE}/(N - s)$  just as in the analysis of variance method. The other  $\sigma^2$ 's are estimated by equating the calculated  $R(.|.)$  terms to their expected values. And these equations are easily solved; they lead to obtaining the  $\hat{\sigma}^2$ 's successively from each equation. For example, working upwards in Table 13

$$E[R(AB|\mu, A, B)] = k_1 \sigma_{AB}^2 + k_2 \sigma_e^2 \quad \text{leads to} \quad \hat{\sigma}_{AB}^2 = [R(AB|\mu, A, B) - k_2 \hat{\sigma}_e^2]/k_1.$$

where the  $k$ 's are functions of the  $n_{ij}$ 's given by (70). Similarly

$$\hat{\sigma}_b^2 = [R(B, AB|\mu, A) - k_4 \hat{\sigma}_{AB}^2 - k_5 \hat{\sigma}_e^2]/k_3.$$



#### 14.4 Calculation

The preceding discussion illustrates the general use of the  $R(.|..)$  notation, applicable to any random model (and, as is shown in Section 14.6, to any mixed model with but one minor, though important, amendment). Once the expectations have been derived by the rule alluded to they <sup>are</sup> often more conveniently used in the form of a set of linearly independent functions of them. These will, of <sup>course,</sup> yield the same estimators. Thus for the 2-way classification of Table 13, linear combinations of the terms therein that are more familiar than those terms themselves are the expressions shown in Table 14.

Table 14

Terms that can be used for estimating variance components by the fitting constants method, for a 2-way classification, random model.

Linear combination of terms in Table 13		Equivalent forms	
$R(\mu, A) - R(\mu)$	$\equiv$	$T_A - T_\mu$	
$R(\mu, A, B) - R(\mu, A)$	$\equiv$	$R(\mu, A, B) - T_A$	
$R(\mu, A, B, AB) - R(\mu, A, B)$	$\equiv$	$T_{AB} - R(\mu, A, B)$	
$SSE = T_O - R(\mu, A, B, AB)$	$\equiv$	$T_O - T_{AB}$	
Total	$T_O - R(\mu)$	$\equiv$	$T_O - T_\mu$

Introduction of the T-symbols defined in (48) arises from their equivalence to certain reductions in sums of squares. For example,  $R(\mu) \equiv T_\mu$ ,  $R(\mu, A) = T_A$  and

$$R(\mu, A, B, AB) \equiv T_{AB}.$$

The only term of Table 14 that has to be computed additional to the T's of the analysis of variance method in (48) is  $R(\mu, A, B)$ . Calculation of this and of  $E[T_{AB} - R(\mu, A, B)]$  is given for the mixed model in Searle and Henderson [1961]. The results given there for these two terms also apply to the random model - a point not made in that paper. Because the formulae have since been corrected and have, in any case, yet to be succinctly displayed in readily computable format, they are repeated here in Table 15.

Table 15

Computing formulae for the terms needed in the fitting constants method of estimating variance components additional to those needed in the analysis of variance method;  
for the 2-way classification, mixed or random models.

To calculate $R(\mu, A, B)$ compute:	To calculate $E[T_{AB} - R(\mu, A, B)]$ compute:
<p>For <math>j = 1, \dots, b</math></p> $c_{jj} = n_{.j} - \sum_{i=1}^a \frac{n_{ij}^2}{n_{i.}}$ $c_{jj'} = - \sum_{i=1}^a \frac{n_{ij} n_{ij'}}{n_{i.}}, \quad j \neq j'$ <p>(Check: <math>\sum_{j'=1}^b c_{jj'} = 0</math>)</p> $r_j = y_{.j} - \sum_{i=1}^a n_{ij} \bar{y}_{i..}$ <p>(Check: <math>\sum_{j=1}^b r_j = 0</math>)</p> <p>For <math>j \neq j' = 1, 2, \dots, (b-1)</math></p> $\underline{c} = \{c_{jj'}\} \text{ and } \underline{c}^{-1} = \{c^{jj'}\}$ $\underline{r} = \{r_j\}$ <p>Then <math>t_B = \underline{r}' \underline{c}^{-1} \underline{r}</math></p> <p>and <math>R(\mu, A, B) = T_A + t_B</math>.</p>	<p>For <math>i = 1, \dots, a</math></p> $\lambda_i = \sum_{j=1}^b n_{ij}^2 / n_{i.}$ <p>For <math>i = 1, \dots, a</math> and <math>j, j' = 1, \dots, b</math></p> $f_{i,jj} = (n_{ij}^2 / n_{i.}) (\lambda_i + n_{i.} - 2n_{ij})$ $f_{i,jj'} = (n_{ij} n_{ij'} / n_{i.}) (\lambda_i - n_{ij} - n_{ij'})$ <p>for <math>j \neq j'</math>.</p> <p>(Check: <math>\sum_{j=1}^b f_{i,jj'} = 0</math>)</p> <p>For <math>i = 1, \dots, a</math> and <math>j, j' = 1, \dots, (b-1)</math></p> $\underline{F}_i = \{f_{i,jj'}\}$ <p>Then <math>k = \sum_{i=1}^a \lambda_i + \text{tr}(\underline{c}^{-1} \sum_{i=1}^a \underline{F}_i)</math> and</p> $E[T_{AB} - R(\mu, A, B)]$ $= (N - k) \sigma_{AB}^2 + (s - a - b + 1) \sigma_e^2.$

Calculations<sup>for</sup> multi-way classification data do not simplify as much as do those of the 2-way case. Each reduction  $R(\cdot)$  can be calculated as  $\underline{y}'\underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{y}$  for the appropriate value of  $\underline{X}$ . A general difficulty with this is that it can involve very large matrices, large because their order can equal the number of random effects in the data. These effects may be very numerous, several hundreds or thousands maybe, in which case the computing requirements are extensive. One fact of slight assistance is that  $\underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'$  is invariant to whatever generalized inverse of  $\underline{X}'\underline{X}$  is used for  $(\underline{X}'\underline{X})^{-1}$ , and so the easiest one to compute can always be used. The formulae of Table 15 for the 2-way classification are useful in this respect because they provide a means of avoiding a large matrix. When one factor has many more levels represented in the data than the other,  $a \gg b$  say, the procedures in Table 15 involve matrices whose order is only  $b - 1$ . This represents a considerable saving when, for example,  $a = 688$  and  $b = 4$  as it did in the studies of Searle and Henderson [1960].

#### 14.5 Too many equations

Table 13 contains no term  $R(\mu, B) = T_B$  that corresponds to fitting the model  $y_{ijk} = \mu + \beta_j + e_{ijk}$ . Yet, by the specifications set out in Sections 14.1 and 14.2, there is no reason why  $R(A, AB | \mu, B) = R(\mu, A, B, AB) - R(\mu, B) = T_{AB} - T_B$  could not be used: its expectation involves  $\sigma_A^2$ ,  $\sigma_{AB}^2$  and  $\sigma_e^2$ . With this available, one can set up an alternative to Table 14, as shown in Table 16.

Table 16

An alternative to Table 14

$R(\mu, B) - R(\mu)$	$= T_B - T_\mu$
$R(\mu, A, B) - R(\mu, B)$	$= R(\mu, A, B) - T_B$
$R(\mu, A, B, AB) - R(\mu, A, B)$	$= T_{AB} - R(\mu, A, B)$
$T_O - R(\mu, A, B, AB)$	$= T_O - T_{AB}$
Total $T_O - R(\mu)$	$= T_O - T_\mu$

With a fixed effects model the first two lines of Table 14 are, of course, the two ways of looking at sums of squares for "A ignoring B" and "B adjusted for A"; and those of Table 16 are for "B ignoring A" and "A adjusted for B". The choice of which table to use then depends on the nature of one's data. But with a random effects model, there appears to be no criteria for choosing between Tables 14 and 16. With either of them the variance components can be estimated by equating the lines therein to their expected values. And the resulting estimators for  $\sigma_A^2$  and  $\sigma_B^2$  will differ, depending on which table is used. (The estimators of  $\sigma_e^2$  and  $\sigma_{AB}^2$  will be the same, because in both tables the last two lines are the same, identical to the last two in Table 13 which, we have seen, determine  $\hat{\sigma}_{AB}^2$  and  $\hat{\sigma}_e^2$ .)

This is an unsolved difficulty with the fitting constants method: it can yield more equations than there are components to be estimated, and it provides no guidance as to which equations are to be used. Tables 14 and 16 have the

property of "adding up" to the total corrected sum of squares, in the manner of an analysis of variance table. This, on the face of it, may seem to recommend either of them as a reasonable set of quadratics to use, if only on the grounds that both of these tables reduce to the familiar analysis of variance (Table 6) for balanced data. Other possibilities stem from using parts of both tables. The analysis of variance method, for example, uses the first lines of each table, their last line, and  $T_{AB} = T_A + T_B + T_{\mu}$ . These are the terms in equation (49) and they too, "add up" and also reduce to Table 6 for balanced data. In contrast to these possibilities, Harville [1967a] suggests using the second lines of the two tables and their last two lines; these do not "add up" for unbalanced data, but they do for balanced data when they then also simplify to Table 6. Low [1964] follows the same course of action in the case of the 2-way classification without interaction, in which he suggests using  $R(\mu, A|B)$ ,  $R(\mu, B|A)$  as well as the customary SSE.

In general, the problem remains unsolved, as to which of the several quadratics that the fitting constants method provides should be used. It is a problem that is encountered not only with unbalanced data generally, but also with special cases thereof, such as the balanced incomplete block designs considered in this context by Low [1969] and Harville [1969c]. One possible deliverance from the dilemma, suggested by Robson [1957], is to apply 'least squares'. If the calculated reductions are arrayed in a vector  $\underline{r}$ , and  $E(\underline{r}) = \underline{A}\sigma^2$ , then  $\underline{r} = \underline{A}\hat{\sigma}^2$  are the equations we would like to have solved for the estimated  $\sigma^2$ 's. When we have more reductions in  $\underline{r}$  than there are variance components to be estimated we will usually find that the equations  $\underline{r} = \underline{A}\hat{\sigma}^2$  are not consistent.<sup>1/</sup> However, so long

---

<sup>1/</sup> Thanks go to D. A. Harville for bringing this to my attention.

as attention is confined to reductions that are linearly independent,  $\underline{A}$  will have full row column rank and the 'least squares' estimate of  $\underline{\sigma}^2$  to be derived from these equations would be  $\hat{\underline{\sigma}}^2 = (\underline{A}'\underline{A})^{-1}\underline{A}'\underline{r}$ .

#### 14.6 General results

General properties of the fitting constants method as detailed in Searle [1968] are now outlined. The general model  $\underline{y} = \underline{X}\underline{\beta} + \underline{e}$  is taken

$$\underline{y} = \underline{X}_1\underline{\beta}_1 + \underline{X}_2\underline{\beta}_2 + \underline{e} , \quad (69)$$

where the partitioning simply divides  $\underline{\beta}$  into two groups of effects,  $\underline{\beta}_1$  and  $\underline{\beta}_2$ , with no thought for whether the groups represent fixed or random effects. The only condition on the partitioning is the obvious one that no factor is to have some effects in  $\underline{\beta}_1$  and others in  $\underline{\beta}_2$ ; every factor is wholly in  $\underline{\beta}_1$  or  $\underline{\beta}_2$ . It has then been shown (loc cit) that, where  $R(\underline{\beta}_1)$  is the reduction in sum of squares due to fitting the sub-model  $\underline{y} = \underline{X}_1\underline{\beta}_1 + \underline{e}$ , the expected value of  $R(\underline{\beta}_2|\underline{\beta}_1)$  - reduction due to  $\underline{\beta}_2$  after accounting for  $\underline{\beta}_1$  - is

$$\begin{aligned} ER(\underline{\beta}_2|\underline{\beta}_1) &= E[R(\underline{\beta}_1, \underline{\beta}_2) - R(\underline{\beta}_1)] \\ &= \text{tr}\left\{\underline{X}_2' [\underline{I} - \underline{X}_1(\underline{X}_1'\underline{X}_1)^{-}\underline{X}_1']\underline{X}_2 E(\underline{\beta}_2\underline{\beta}_2')\right\} + \sigma_e^2[r(\underline{X}) - r(\underline{X}_1)] . \end{aligned} \quad (70)$$

In this result  $r(\underline{X})$  and  $r(\underline{X}_1)$  are the ranks of  $\underline{X}$  and  $\underline{X}_1$  and  $(\underline{X}_1'\underline{X}_1)^{-}$  is a generalized inverse of  $\underline{X}_1'\underline{X}_1$ .

The importance of (70) is that the only  $\underline{\beta}$ -term involved is  $\underline{\beta}_2$ ; i.e., the expectation of  $R(\underline{\beta}_2|\underline{\beta}_1)$  is a function simply of  $E(\underline{\beta}_2\underline{\beta}_2')$  and  $\sigma_e^2$ . It involves neither  $E(\underline{\beta}_1\underline{\beta}_1')$  nor  $E(\underline{\beta}_1\underline{\beta}_2')$ . This is important because (70) has been derived without any assumptions on the form of  $E(\underline{\beta}\underline{\beta}')$ , i.e.,  $\underline{\beta}$  can contain fixed or random

effects. Consequently, if the  $\beta$ -vector of one's model can be partitioned into two parts  $\beta_1$  and  $\beta_2$  where  $\beta_2$  contains just random effects, then  $ER(\beta_2|\beta_1)$  as given in (70) contains only  $\sigma_e^2$  and the variance components relating to those random effects. This is the reason for being able to write the expectations in Table 13 in the manner shown. Note however, as indicated in the discussion following Table (13), that (70) applies only to a difference between reductions that consists of the reduction for fitting the full model minus a reduction for fitting a sub-model. Then  $\sigma_e^2$  and the variance components that are not in the sub-model are involved in the expected value of  $R(.|.)$ .

The result in (70) is particularly pertinent in mixed models because, when  $\beta_1$  represents all the fixed effects,  $ER(\beta_2|\beta_1)$  contains no terms due to these fixed effects; the expectation is solely a function of  $\sigma_e^2$  and the variances of the random effects in  $\beta_2$ . This is the value of the fitting constants method of estimating variance components in the mixed model: it yields estimators of the variance components unaffected by the fixed effects. (70) is also the reason why the results of Searle and Henderson (1961) concerning  $T_{AB} - R(\mu, A, B) = R(AB|\mu, A, B)$  apply equally as well to a random as to a mixed model. By (70), the last three lines of Table 13 have the expectations shown there whether the A-effects are fixed or random. Hence in the mixed model, with the A-effects fixed, the variance components can be estimated from the last three lines of Table 13, equivalent to the last three lines of Table 14, the difficult portions of which are computed according to Table 15.

In general, the fact that (70) involves only  $E(\beta_2\beta_2')$  implies that  $E[R(\beta_2|\beta_1)]$  has no terms arising from any covariance between the elements of  $\beta_1$  and  $\beta_2$ . Hence, even if the model is such that terms in  $\beta_1$  are correlated with terms in  $\beta_2$  the



expectation in (70) does not involve this correlation - it depends solely on the second moments of the elements in  $\beta_2$  (and on  $\sigma_e^2$ ).

A general advantage of the fitting constants method would seem to be that in using it we are dealing with terms that are sums of squares - positive semi-definite quadratic forms in the observations that are known to have certain optimum properties in fixed effects models. One might hope that these optimum properties would carry over to some extent to random effects and mixed models, although to what extent they do so is unknown. One particular advantage of the method lies in its appropriateness for the mixed model, for which it yields variance component estimators that are unaffected and uncomplicated by the fixed effects. It is therefore the preferred method for mixed models. As already mentioned, its disadvantage is that it involves matrices that can be very large in models having large numbers of effects in them. This can be a difficulty not only in calculating reductions in sums of squares but also in deriving their expectations from (70). One instance of a simplification is the case discussed by Cunningham [1969], of having just one random effects factor in a mixed model. He shows, in effect, that for this case (70) reduces to

$$ER(\beta_2 | \beta_1) = [\text{tr}(\underline{K}) + \text{sm}(\underline{K})] \sigma_{\beta_2}^2 + (q - 1) \sigma_e^2$$

for appropriate choice of  $\underline{K}$ , where  $\text{sm}(\underline{K})$  is  $\underline{1}' \underline{K} \underline{1}$ , the sum of the elements of  $\underline{K}$ , and  $q$  is the number of levels of the random effects.

A variation of Henderson's Method 2 mentioned at the end of Section 13 is to use the analysis of variance method on the adjusted observations  $\underline{z} = \underline{y} - \underline{X}_f \tilde{\beta}_f$  where  $\tilde{\beta}_f = (\underline{X}_f' \underline{X}_f)^{-1} \underline{X}_f' \underline{y}$ . Another variation is to use the fitting constants method

on this  $\underline{z}$ . As indicated by Zelen [1968], this is equivalent to the fitting constants method used directly on  $\underline{y}$ , (see, e.g., Searle [1969]).

#### 14.7 Sampling variances of estimators

Variance component estimators derived by the fitting constants method of estimation are linear combinations of reductions in sums of squares due to fitting different models to the data. Each reduction can be expressed in the form  $\underline{y}'\underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{y}$  for an appropriate  $\underline{X}$ -matrix and hence its sampling variance can be derived from application of (41), and the sampling covariance between any two reductions can be derived from (42). In this way the sampling variance of linear combinations of these reductions can be obtained. The details are somewhat lengthy, involving considerable matrix manipulation. Low [1964], for the 2-way classification without interaction, random model, has derived sampling variances of variance component estimators obtained from  $R(A|\mu, B)$  and  $R(B|\mu, A)$  and SSE. Harville [1969c], for the same model but with interaction, provides the means for obtaining the variance of  $R(\mu, A, B)$  of Tables 14 and 16 and its covariances with other terms in those Tables. He then applies these results to the case of balanced incomplete block designs. For mixed models, Rohde and Tallis [1969] give matrix expressions for cases involving one or two random effects factors and indicate how the results can be extended to include in the model interactions among the random effects.

#### 15. Analysis of means methods

Data sometimes have every cell containing at least one observation. In such circumstances two methods of estimating variance components are available that

cannot be used when any of the cells are empty. Both methods use sums of squares of the cell means, equating them to their expected values and estimating the error variance from the within-cell mean square in the usual way. One method uses weighted sums of squares of cell means and the other uses unweighted sums. The two methods are illus-

trated below in terms of the 2-way classification example <sup>already</sup> discussed.

The unweighted means method uses the cell means (all  $n_{ij} > 0$ ) as if they were single observations. Thus, with the model for the data being

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk},$$

the variances  $\sigma_A^2$ ,  $\sigma_B^2$  and  $\sigma_{AB}^2$  are estimated from sums of squares of

$$x_{ij} = \bar{y}_{ij} = \sum_{k=1}^{n_{ij}} y_{ijk} / n_{ij}$$

as if they were single observations; and  $\sigma_e^2$  is estimated from the within-cell mean square of the  $y_{ijk}$ 's. The four terms used, and their expectations, are shown in Table 17.

Table 17

Terms used in the unweighted means method of estimating variance components from 2-way classification data, random or mixed models, with all  $n_{ij} > 0$ .

Term <sup>+</sup>	Expectation <sup>*</sup>
$SSA_u = b \sum_{i=1}^a (\bar{x}_{i.} - \bar{x}_{..})^2$	$E\left(\frac{SSA_u}{a-1}\right) = b\sigma_A^2 + \sigma_{AB}^2 + n_h \sigma_e^2$
$SSB_u = a \sum_{j=1}^b (\bar{x}_{.j} - \bar{x}_{..})^2$	$E\left(\frac{SSB_u}{b-1}\right) = a\sigma_B^2 + \sigma_{AB}^2 + n_h \sigma_e^2$
$SSAB_u = \sum_{i=1}^a \sum_{j=1}^b (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..})^2$	$E\left(\frac{SSAB_u}{(a-1)(b-1)}\right) = \sigma_{AB}^2 + n_h \sigma_e^2$
$SSE = \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{ij})^2$	$E\left(\frac{SSE}{N - ab}\right) = \sigma_e^2$

+ Means:  $x_{ij} = \bar{y}_{ij}$ ,  $\bar{x}_{i.} = \sum_{j=1}^b x_{ij} / b$ ,  $\bar{x}_{.j} = \sum_{i=1}^a x_{ij} / a$  and  $\bar{x}_{..} = \sum_{i=1}^a \sum_{j=1}^b x_{ij} / ab$ .

\*  $n_h = \frac{1}{ab} \sum_{i=1}^a \sum_{j=1}^b \frac{1}{n_{ij}}$ .

Estimators are derived from Table 17 by the usual process of equating observed values of its terms to their expectations. Known as the method of unweighted means it has been used, for example, by Mostafa (1967) in discussing the efficiency of designs wherein all the  $n_{ij}$ 's are either 1 or 2. In general application the method has the great advantage of being easy to compute; but it does demand that every cell have at least one observation. Other points of interest evident from Table 17 are as follows. (i) Its terms do not "add up" in the analysis of variance sense; the first three terms  $SSA_u$ ,  $SSB_u$  and  $SSAB_u$  add to  $\sum \sum x_{ij}^2 - \bar{x}_{..}^2 / ab$ , but all four do not add to  $\sum \sum \sum y_{ijk}^2 - \bar{y}_{...}^2 / n_{...}$ . (ii) The terms do not provide any F-tests of hypotheses because, under normality, none of the first three terms have  $\chi^2$ -distributions. (iii) Three of the four terms can be used in mixed models: for example, if the A-effects are fixed the last three terms provide estimators of  $\sigma_B^2$ ,  $\sigma_{AB}^2$  and  $\sigma_e^2$ . This is so because  $SSB_u$ , derived from  $\bar{x}_{.j} - \bar{x}_{..}$ , contains no  $\alpha_i$  terms, since both  $\bar{x}_{.j}$  and  $\bar{x}_{..}$  contain  $\bar{\alpha}_{.} = \sum_{j=1}^b \alpha_i / b$  arising from the fact that every cell in the data contains observations.

The weighted means analysis uses weighted sums of squares of the  $\bar{x}_{i.}$  and  $\bar{x}_{.j}$  in place of  $SSA_u$  and  $SSB_u$  in the above table. These terms and their expected values are shown in Table 18.

Table 18

Terms used in the weighted means method of estimating variance components from 2-way classification data, random or mixed models, with all  $n_{ij} > 0$ .

These terms are used in combination with  $SSAB_u$  and SSE of Table 17.

Term*	Expectation
$SSA_w = \sum_{i=1}^a w_i (\bar{x}_{i.} - \bar{x}_{(1)})^2$	$E\left(\frac{SSA_w}{a-1}\right) = \frac{1}{(a-1)b} \left( \sum w_i - \sum w_i^2 / \sum w_i \right) (b\sigma_A^2 + \sigma_{AB}^2) + \sigma_e^2$
$SSB_w = \sum_{j=1}^b v_j (\bar{x}_{.j} - \bar{x}_{(2)})^2$	$E\left(\frac{SSB_w}{b-1}\right) = \frac{1}{(b-1)a} \left( \sum v_j - \sum v_j^2 / \sum v_j \right) (a\sigma_B^2 + \sigma_{AB}^2) + \sigma_e^2$

$$* : w_i = \left( \frac{1}{b} \sum_{j=1}^b \frac{1}{n_{ij}} \right)^{-1}, \text{ and } \bar{x}_{(1)} = \sum w_i \bar{x}_{i.} / \sum w_i.$$

$$v_j = \left( \frac{1}{a} \sum_{i=1}^a \frac{1}{n_{ij}} \right)^{-1}, \text{ and } \bar{x}_{(2)} = \sum v_j \bar{x}_{.j} / \sum v_j.$$

Highlights of this table are as follows. (i) In no way do its terms, together with  $SSAB_u$  and SSE of Table 17, "add up" in the analysis of variance sense. (ii) Under conditions of normality

$$\frac{SSA_w}{SSE} \cdot \frac{N - ab}{a - 1} \quad \text{and} \quad \frac{SSB_w}{SSE} \cdot \frac{N - ab}{b - 1}$$

provide, in the fixed model, exact F-tests concerning the  $\alpha$ - and  $\beta$ -effects.

This was the context in which these weighted sums of squares were first

suggested by Yates (1934), who also mentions the unweighted means analysis as of Tables 17 and 18

providing approximate F-tests. Their use in variance components estimation is

more recent. (iii) In mixed models, with the A-effects fixed, the terms  $SSB_w$ ,  $SSAB_u$  and SSE can be used to estimate  $\sigma_B^2$ ,  $\sigma_{AB}^2$  and  $\sigma_e^2$ .

Extension of these methods of multi-way classifications is clear, particularly for the unweighted means method which is easy to compute. For balanced data, both methods reduce to the analysis of variance method.

#### 16. Symmetric sums methods

Estimation of variance components by utilizing symmetric sums of products of observations has been suggested by Koch [1967a]. The method is based on expected values of products of observations being functions of the components, and therefore sums of these products (and hence means of them) provide unbiased estimators of the components. Such estimators are relatively easy to compute. They are also unbiased and consistent, and are identical to the analysis of variance estimators in the case of balanced data. However, they have a singular disadvantage: their variances are functions of the general mean  $\mu$ . This deficiency is overcome by Koch [1968] who suggests that symmetric sums of squares of differences should be used instead of symmetric sums of products

The method is easily illustrated in terms of the familiar 1-way classification, random model, where  $y_{ij} = \mu + \alpha_i + e_{ij}$ , with  $E(\alpha_i) = E(e_{ij}) = 0$ ,  $E(\alpha_i^2) = \sigma_A^2$  and  $E(e_{ij}^2) = \sigma_e^2$  for all  $i$  and  $j$ , and all covariances zero. Expected values of squared differences of observations are:

$$E(y_{ij} - y_{i'j'})^2 = 2\sigma_e^2 \quad \text{when } i = i' \text{ and } j \neq j';$$

$$= 2(\sigma_e^2 + \sigma_A^2) \quad \text{when } i \neq i'.$$

These provide estimators

$$2\hat{\sigma}_e^2 = \frac{\sum_{i=1}^a \sum_{j=1}^{n_i} \sum_{j' \neq j}^{n_i} (y_{ij} - y_{ij'})^2}{\sum_{i=1}^a n_i (n_i - 1)}$$

and

$$2(\hat{\sigma}_e^2 + \hat{\sigma}_A^2) = \frac{\sum_{i=1}^a \sum_{i' \neq i}^a \sum_{j=1}^{n_i} \sum_{j'=1}^{n_{i'}} (y_{ij} - y_{i'j'})^2}{\sum_{i=1}^a \sum_{i' \neq i}^a n_i n_{i'}},$$

which are unbiased, have variances not dependent on  $\mu$ , and for balanced data ( $n_i = n$  for all  $i$ ) they are identical to the analysis of variance estimators.

In giving details of applying this method to specific cases, Koch [1968] discusses certain nearly-balanced situations, and Koch [1967b] presents a method for estimating  $\mu$  unbiasedly, based on an unbiased estimator of  $\mu^2$ .

## 17. Infinitely many quadratics

Indication has already been given of the many quadratic forms that can be used for estimating variance components from unbalanced data. Indeed, infinitely many quadratics are available for equating observed values of expected values and solving the resulting equations to get variance components estimators.



Although this is a widely used procedure it is one which imposes no conditions on the quadratics to be used, which sets up no criteria for deriving them, and which provides no rationale for determining what sets of quadratics are in any sense optimal. These are grave shortcomings, albeit ones that are not easily overcome, because of the algebraic complexity of unbalanced data. The only universal outcome of the procedure is that it provides unbiased estimators.

The sets of quadratics involved in each of the methods discussed above have been selected solely because they seemed "reasonable" in one way or another. The methods could therefore be compared on what constitutes "reasonableness" in each case, although this would make no comparison of the properties of the specific estimators that result from the different methods. The simplest comparison of this sort would be of sampling variances, except that this becomes buried in algebraic complexity. Harville [1969a], in considering such comparisons, has defined a quadratic estimator as being inadmissible if some other quadratic estimator exists having the same expectation and a smaller or equal sampling variance for all points in the parameter space, with the sampling variance being strictly smaller for at least one such point. He then derives conditions, for the 1-way classification, for an estimator to be quadmissible and shows that the analysis of variance estimator of  $\sigma_e^2$  is quadmissible.

Comparison of sampling variances of different estimators is difficult not only because the variances are in any way tractable only if normality is assumed but also, just as with balanced data, because the variances themselves involve the true variance components, as is evident in (60). Attempting to study the behaviour of such variances in terms of their being functions of the total number of observations, the number of levels of each factor, the number of observations in each cell, and as functions of the variance components - to do this is no small task, let alone to compare them with equally as complex functions that are variances of other estimators. Analytic comparison of different estimators by way of their

sampling variances therefore presents great difficulties. One instance of success is that by Harville [1969c] who develops explicit expressions for the differences between variances of analysis of variance estimators and fitting constants estimators for balanced incomplete block designs. Since these are functions of the variance components they can only be compared for specified values of those components.

It is interesting to note, I believe, how serious is the consequence of something that otherwise seems relatively insignificant - namely, of data being unbalanced rather than balanced. The change from one to the other might, on the surface, appear relatively small and yet it brings with it enormous changes in the tractability of methods and the properties of estimators. Not that everything is settled with balanced data, but at least the situation there is considerably simpler than that of unbalanced data.

Analytic comparison of estimators appearing to be so fruitless enforces recourse to numerical comparison. Unfortunately, results are difficult and costly to attain in this connection. A special form of the 1-way classification has been considered by Kussmaul and Anderson [1967] in which the compositing of samples in a 2-way nested classification is envisaged. As a result, the  $j$ -th observation in the  $i$ -th class is deemed to be an average of the  $n_{ij}$  observations "that the sample would have provided separately" had it not been composited prior to measurement. In this situation three methods of estimating the between-class variance component are compared numerically for a variety of  $n_{ij}$ -values and for various values of the ratio of the variance components. Comparison of the analysis of variance method of estimation with the unweighted means method has also been made for the between-class component of a 1-way classification, by Crump and Anderson [1967]. They found that the unweighted means estimator appears, for very unbalanced data, to be poorer (have larger variance) than the analysis of variance estimator for small

values of  $\rho = \sigma_A^2 / \sigma_e^2$ , but that it is superior (has smaller variance) for large  $\rho$ . Studies of the 2-way classification, with interaction, made by Bush and Anderson [1963] compare the analysis of variance method, the fitting constants method and the weighted means method, making comparisons by way of the variances of the different estimators. This was done for several sets of  $n_{ij}$ -values representing what might be called not wholly unbalanced data but designed unbalancedness. For example, in a case of 6 rows and 6 columns, three of the designs used were those shown in Table 19.

Table 19

Values of  $n_{ij}$  in three of the  $6 \times 6$  designs used  
by Anderson and Bush [1963]

<u>S22</u>	<u>C18</u>	<u>I24</u>
2 1 0 0 0 0	1 1 1 0 0 0	1 1 0 0 0 0
1 2 1 0 0 0	1 1 1 0 0 0	1 1 0 0 0 0
0 1 2 1 0 0	0 1 1 1 0 0	2 1 0 0 0 0
0 0 1 2 1 0	0 0 1 1 1 0	1 2 0 0 0 0
0 0 0 1 2 1	0 0 0 1 1 1	1 1 2 1 1 1
0 0 0 0 1 2	0 0 0 1 1 1	1 1 2 1 1 1

Designs of this nature were used to compare both the estimation procedures and the designs themselves, using a variety of values of the true components. The results indicate, at least for the designs used, that when the error variance is considerably larger than other components in the model then the analysis of variance method yields estimators with the smallest variances; otherwise the fitting constants method does.

Comparisons of the nature just discussed involve no mean effort. Yet, in terms of unbalanced data generally, the examples used are, by necessity, finite in extent. This highlights one of the great difficulties of numerical comparisons: designing sets of  $n_{ij}$ -values in such a way as to provide not only valid comparisons, but valuable ones, valuable in the sense of being informative about unbalanced data generally. For example, in the 1-way classification there are infinitely many sets of  $n_i$ -values that could be used in (60) for studying the behaviour of  $v(\hat{\sigma}_A^2)$  - and each such set could be used in combination with varying values of  $\sigma_A^2$  and  $\sigma_e^2$ . The difficulty is to plan a series of these values that "covers the field" in such a way as to provide a basis for drawing general conclusions. This difficulty is magnified, of course, when one comes to consider 2-way and higher-order classifications. Comparative studies of properties of estimators, either analytic or numeric, are therefore no easy task.

Aside from numerical comparisons, the grounds on which "reasonableness" was judged appropriate in establishing different estimation methods can be summarized as follows. The analysis of variance method commends itself because it is the obvious analogue of the analysis of variance of balanced data, and it is easy to use; but some of its terms are not sums of squares, and it gives unbiased estimators in

mixed models. The generalized form of Henderson's Method 2 makes up for this deficiency, but is not uniquely defined and his specific definition of it cannot be used when there are interactions between fixed and random effects. The fitting constants method uses sums of squares that have non-central  $\chi^2$ -distributions in the fixed effects model, and it gives unbiased estimators in mixed models; but it can involve more quadratics than there are components to be estimated; and it can also involve extensive computing. The symmetric sums method is easy to compute; it utilizes all possible products of the observations and their means. Little more than this can, at this time, be said by way of comparing the methods. Most of them give unbiased estimators and all of them reduce to the analysis of variance method for balanced data, - but they can all produce negative estimates. The general problem of comparison still awaits thorough investigation.

#### 18. Maximum likelihood methods

Maximum likelihood is best considered from the point of view that all models are mixed models, as is certainly so for models containing a general mean  $\mu$  (a fixed effect) and error terms  $e$  that are random. Although the fitting constants method of estimating variance components gives unbiased estimators with mixed models it gives no guidance on estimating the fixed effects of the model. Were the variance components known there would, of course, be no problem in estimating the fixed effects. They would come from the normal equations  $\underline{X}'\underline{V}^{-1}\underline{X}\hat{\underline{\beta}} = \underline{X}'\underline{V}^{-1}\underline{y}$  of the generalized least squares procedure, where  $\underline{V}$  is the variance-covariance matrix of  $\underline{y}$ , its elements being functions of the (assumed known) variance

components. However, when these components are unknown, as is usually the case, there exists the dual problem of estimating both the fixed effects and the variance components.

Two possible courses of action exist: (i) to do the estimation in two stages, by first estimating the variance components with the fitting constants method and then using the resulting estimates in place of the true components in  $\underline{V}$ , in the generalized least squares equations; or (ii) to estimate the fixed effects and the variance components simultaneously, with one unified procedure such as maximum likelihood. Although in both cases recourse has usually to be made to iterative procedures, for which the computing requirements can be extensive, some progress has been made analytically, the results of which are now summarized.

### 18.1 Estimating fixed effects

We write the model as

$$\underline{y} = \underline{X}\underline{\beta} + \underline{Z}\underline{\mu} + \underline{e} \quad (71)$$

where  $\underline{\beta}$  is the vector of fixed effects,  $\underline{\mu}$  is the vector of random effects,  $\underline{X}$  and  $\underline{Z}$  are the corresponding design matrices and  $\underline{e}$  is the vector of random error terms. Means of, and covariances between the random effects and the error terms are assumed to be zero, with variance-covariance matrices  $\text{var}(\underline{\mu}) = \underline{D}$  and  $\text{var}(\underline{e}) = \underline{R}$ . Then

$$\underline{V} = \text{var}(\underline{y}) = \underline{Z}\underline{D}\underline{Z}' + \underline{R} , \quad (72)$$

with which the generalized least squares normal equations are

$$\underline{X}'\underline{V}^{-1}\underline{X}\hat{\underline{\beta}} = \underline{X}'\underline{V}^{-1}\underline{y} . \quad (73)$$

By nature,  $\underline{V}$  is positive semi-definite and usually non-singular, and so we assume  $\underline{V}^{-1}$  exists. These equations can also be derived by the method of maximum likeli-

hood, on assuming normality and  $\underline{V}$  known, in which case  $\hat{\underline{\beta}}$  would be referred to as a maximum likelihood solution.

Calculation of (73) involves  $\underline{V}^{-1}$ , a matrix of order  $N$ , the number of observations, which may be very large, perhaps many thousands. In the fixed effects case  $\underline{V}$  usually has the form  $\sigma_e^2 \underline{I}$  or, with a little more generality, it may be diagonal: in either case, inversion is simple. But in general,  $\underline{V} = \underline{Z}\underline{D}\underline{Z}' + \underline{R}$  of (72) is not diagonal, even if  $\underline{D}$  and  $\underline{R}$  are, and so  $\underline{V}^{-1}$  is not necessarily easy to compute. However, as indicated in Henderson et al [1959], a set of equations alternative to (73) and not involving  $\underline{V}^{-1}$ , is

$$\begin{bmatrix} \underline{X}'\underline{R}^{-1}\underline{X} & \underline{X}'\underline{R}^{-1}\underline{Z} \\ \underline{Z}'\underline{R}^{-1}\underline{X} & \underline{Z}'\underline{R}^{-1}\underline{Z} + \underline{D}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\underline{\beta}} \\ \hat{\underline{u}} \end{bmatrix} = \begin{bmatrix} \underline{X}'\underline{R}^{-1}\underline{y} \\ \underline{Z}'\underline{R}^{-1}\underline{y} \end{bmatrix} \quad (74)$$

On eliminating  $\hat{\underline{u}}$  from these equations and utilizing  $\underline{V}$  from (72) the resulting equation simplifies to (73). The value of (74) however, is that it does not require the inverse of  $\underline{V}$ , but needs only the inverse of  $\underline{D}$  and  $\underline{R}$ , both of which are often diagonal. In that case, even though (74) involves more equations than (73) they are fewer than  $N$ , the order of  $\underline{V}$ , whose inverse is needed in (73). Hence (74) is often easier to solve than (73).

The format of equations (74) is easily described. They are simply the normal equations for the fixed model (i.e., assuming  $\underline{u}$  is a vector of fixed effects) with the inverse of the variance-covariance matrix of the random effects added to the sub-matrix that is the coefficient of  $\hat{\underline{u}}$  in the " $\underline{u}$ -equations" - i.e.,  $\underline{D}^{-1}$  added to  $\underline{Z}'\underline{R}^{-1}\underline{Z}$ , as in (74). This is patently simple, particularly in certain special cases. For example, taking  $\underline{R}$ , the variance-covariance matrix of the error terms, in its frequently-assumed form  $\underline{R} = \sigma_e^2 \underline{I}_N$  simplifies (74) to

$$\begin{bmatrix} \underline{X}'\underline{X} & \underline{X}'\underline{Z} \\ \underline{Z}'\underline{X} & \underline{X}'\underline{X} + \sigma_e^2 \underline{D}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\underline{\beta}} \\ \hat{\underline{u}} \end{bmatrix} = \begin{bmatrix} \underline{X}'\underline{y} \\ \underline{Z}'\underline{y} \end{bmatrix} .$$

Furthermore,  $\underline{D}$  is often diagonal of the form

$$\underline{D} = \text{diag} \left\{ \sigma_{\theta}^2 \underline{I}_{N_{\theta}} \right\} \text{ for } \theta = A, B, \dots, K,$$

in which case  $\sigma_e^2 \underline{D}^{-1}$  requires just adding  $\sigma_e^2 / \sigma_{\theta}^2$  to appropriate diagonal elements of  $\underline{Z}'\underline{Z}$ . In particular, if there is only one random factor the equations become

$$\begin{bmatrix} \underline{X}'\underline{X} & \underline{X}'\underline{Z} \\ \underline{Z}'\underline{X} & \underline{Z}'\underline{Z} + (\sigma_e^2 / \sigma_{\theta}^2) \underline{I} \end{bmatrix} \begin{bmatrix} \hat{\underline{\beta}} \\ \hat{\underline{u}} \end{bmatrix} = \begin{bmatrix} \underline{X}'\underline{y} \\ \underline{Z}'\underline{y} \end{bmatrix} . \quad (75)$$

This formulation of the maximum likelihood estimator is, of course, only applicable when the variance components are known, although in most models only their values relative to that of the error variance are required as for example, in (75). However, in combination with methods for estimating variance components independent of the fixed effects, such as the fitting constants method, (74) and its simplified forms provide a means of setting up iterative procedures for estimating both the fixed effects and the variance components of a mixed model. An example of these procedures is considered by Cunningham and Henderson [1968] with a correction given by Thompson [1969].

The origin of equations (74) is the joint density function of  $\underline{y}$  and  $\underline{u}$ . On the basis of normality this density is



$$f(\underline{y}, \underline{u}) = g(\underline{y}|\underline{u})h(\underline{u})$$

$$= C \exp \left[ -\frac{1}{2}(\underline{y} - \underline{X}\underline{\beta} - \underline{Z}\underline{u})' \underline{R}^{-1}(\underline{y} - \underline{X}\underline{\beta} - \underline{Z}\underline{u}) \right] \exp \left[ -\frac{1}{2}\underline{u}' \underline{D}^{-1}\underline{u} \right]$$

where  $C$  is a constant. Maximization of this with respect to  $\underline{\beta}$  and  $\underline{u}$  leads at once to (74).

In many situations not only is the solution of (74) for  $\underline{\beta}$  of interest, but so also is that for  $\hat{\underline{u}}$ , even though  $\underline{u}$  is a vector of random variables. This is so because  $\hat{\underline{u}}$  is the estimator of the conditional mean of  $\underline{u}$  given  $\underline{y}$ , i.e.,  $\hat{\underline{u}} = E(\underline{u}|\underline{y})$ , and is, as mentioned in Henderson et al [1959], the "estimated genetic merit" used by animal breeders. In their case  $\underline{u}$  is a vector of genetic merit values of a series of animals from whom  $\underline{y}$  is the vector of production records, and the problem is to use  $\underline{y}$  to get estimated values of  $\underline{u}$  in order to decide which animals are best in some sense. In this context  $\hat{\underline{u}}$  has been used extensively.

## 18.2 Fixed effects and variance components

Maximum likelihood equations for estimating variance components from unbalanced data cannot be solved explicitly even when ignoring the requirement that the variance component estimators be positive. Writing down the equations for just the simplest case, the 1-way classification, will soon convince the reader<sup>of</sup> their intractability. They involve, for example, terms like  $\sum_i n_i / (\tilde{\sigma}_e^2 + n_i \tilde{\sigma}_A^2)$  in equations that have to be solved for  $\tilde{\sigma}_e^2$  and  $\tilde{\sigma}_A^2$ . Explicit maximum likelihood estimators must therefore be despaired of. However, a general set of equations is given by Hartley and Rao [1967]. They write the model as

$$\underline{y} = \underline{X}\underline{\beta} + \sum_{\theta=A}^K \underline{Z}_{\theta} \underline{u}_{\theta} + \underline{e}$$

where the sum of the terms  $\sum_{\theta} \underline{Z}_{\theta} \underline{u}_{\theta}$  is  $\underline{Z} \underline{u}$  of (71), with  $\underline{u}_{\theta}$  being the vector of effects for the levels of the random factor  $\theta$ . Then, on defining

$$\gamma_{\theta} = \sigma_{\theta}^2 / \sigma_e^2 \text{ for } \theta = A, B, \dots, K \text{ and } \underline{H} = \underline{I}_N + \sum_{\theta=A}^K \gamma_{\theta} \underline{Z}_{\theta} \underline{Z}_{\theta}',$$

and assuming normality, the equations for the maximum likelihood estimators turn out to be

$$\underline{X}' \underline{H}^{-1} \underline{X} \underline{\beta} = \underline{X}' \underline{H}^{-1} \underline{y}, \quad (76)$$

$$\sigma_e^2 = (\underline{y} - \underline{X} \underline{\beta})' \underline{H}^{-1} (\underline{y} - \underline{X} \underline{\beta}) / N \quad (77)$$

and

$$\text{tr}(\underline{H}^{-1} \underline{Z}_{\theta} \underline{Z}_{\theta}') = (\underline{y} - \underline{X} \underline{\beta})' \underline{H}^{-1} \underline{Z}_{\theta} \underline{Z}_{\theta}' \underline{H}^{-1} (\underline{y} - \underline{X} \underline{\beta}) / \sigma_e^2 \text{ for } \theta = A, B, \dots, K. \quad (78)$$

These are the equations that have to be solved for the elements of  $\underline{\beta}$ , the error variance estimator  $\sigma_e^2$  and the variance components inherent in  $\underline{H}$ . Hartley and Rao [1967] indicate how this can be achieved, either by a method of steepest ascent, or by obtaining an alternative form for (78), which are the difficult equations to handle. (76) and (77) are, of course, recognizable as the familiar

maximum likelihood equations for the fixed effects and the error variance; and they are easily solved if values of the  $\gamma_{\theta}$ 's are available for  $\underline{H}$ . Thus is iteration established via equations (76), (77) and (78). It is further shown in the Hartley-Rao paper that if  $(\underline{\beta}', \sigma_e^2, \gamma')$  provides the global maximum of the likelihood then it is weakly consistent, as it also is if the starting value of the steepest ascent method is a weakly consistent estimator. Asymptotic efficiency is also established.

### 18.3 Large sample variances

Although the maximum likelihood estimators of variance components cannot be obtained explicitly, a general expression for their variances can be derived. To show this we revert to the model  $\underline{y} = \underline{X}\underline{\beta} + \underline{\epsilon}$  where  $\underline{\beta}$  is the vector of fixed effects and, as in (71),  $\underline{\epsilon} = \underline{Z}\underline{u} + \underline{e}$  with  $E(\underline{\epsilon}) = \underline{0}$  and  $\text{var}(\underline{\epsilon}) = \underline{V}$ . Then, on the basis of normality - i.e., of assuming  $\underline{\epsilon} \sim N(\underline{0}, \underline{V})$ , the variance-covariance matrix of the large sample maximum likelihood estimators of the  $p$  elements of  $\underline{\beta}$  and the  $q$  variance components is minus the inverse of the expected value of the Hessian matrix of the logarithm of the likelihood with respect to these  $p + q$

parameters, see Wald [1943]. Simplification of this matrix by Searle [1970] shows that the variance-covariance matrix of the large sample maximum likelihood (M.L.) estimators of the variance components is

$$\text{var}(\tilde{\sigma}^2) = 2 \left\{ \text{tr}(\underline{V}^{-1} \underline{V}_{-i}^{-1} \underline{V}_{-j}) \right\}^{-1} \quad \text{for } i, j = 1, 2, \dots, q, \quad (79)$$

where  $\underline{V}_{-i}$  and  $\underline{V}_{-j}$  are the partial differentials of  $\underline{V}$  with respect to  $\sigma_i^2$  and  $\sigma_j^2$ , for  $i$  and  $j = 1, 2, \dots, q$ , there being  $q$  variance components (including  $\sigma_e^2$ ) in the model. It is interesting to see that (79) is free of the fixed effects of the model and is solely a function of the variance-covariance matrix  $\underline{V}$  of the observations  $\underline{y}$ . Since  $\underline{V}$  in turn depends only on the occurrence of the random effects in the model we see that no matter what the fixed effects are, nor how they occur, they in no way affect  $\text{var}(\tilde{\sigma}^2)$ .

Two results concomitant with (79) are, firstly, that

$$\text{cov}(\tilde{\beta}, \tilde{\sigma}^2) = 0 ;$$

i.e., covariances between large sample M.L. estimators of fixed effects and variance components are zero. Since under conditions of normality the mean of a sample and its sum of squares are independent, this result is not surprising. A second result is that the variance-covariance matrix of the large sample M.L. estimators of the fixed effects is

$$\text{var}(\tilde{\beta}) = (\underline{X}' \underline{V}^{-1} \underline{X})^{-1}.$$

This corresponds to the M.L. solutions  $\hat{\beta}$  in (73), and is therefore no surprise either. Nevertheless, it is interesting to observe from this that even with unbalanced data from any mixed model, the variance-covariance matrix of the M.L. estimators of the fixed effects is exactly as it would be were the variance components known and did not have to be estimated simultaneously with the fixed effects.

One difficulty in applying (79) to specific cases is that it requires the inverse matrix  $\underline{V}^{-1}$ , which is not always readily amenable to an analytic form, as for example in the 2-way crossed classification random model either with or without interaction. Having unbalanced data is what makes the inverse so intractable. A second difficulty with (79) is that even if the <sup>terms</sup>  $\text{tr}(\underline{V}^{-1} \underline{V}_i \underline{V}^{-1} \underline{V}_j)$

can be obtained they are usually of such nature that the matrix inversion required in (79) precludes deriving analytic forms of the variances of the individual components. One application of (79) is to the 2-way nested classification by Searle [1970], where explicit expressions for the elements of (79) before inversion have been obtained. The first of the difficulties just described has been overcome in this case but the second has not.

#### 19. Best quadratic unbiased estimation

The analogue for variance components of best linear unbiased estimation of fixed effects is best quadratic unbiased estimation. This presents no problem with balanced data because the analysis of variance method estimators are unbiased and have minimum variance properties as discussed in Section 9. But with unbalanced data, estimators that are uniformly best do not yet exist. However, some initial progress has been made on this problem by Townsend [1968], who obtains best (but not uniformly best) quadratic unbiased estimators for the 1-way classification with  $\mu = 0$ . This is achieved by letting the desired estimators of  $\sigma_e^2$  and  $\sigma_A^2$  be  $\underline{y}' \underline{P} \underline{y}$  and  $\underline{y}' \underline{Q} \underline{y}$  respectively and, for  $\sigma_e^2 = \text{tr}(\underline{P} \underline{V})$  and  $\sigma_A^2 = \text{tr}(\underline{Q} \underline{V})$ , deriving  $\underline{P}$  and  $\underline{Q}$  such that  $2\text{tr}(\underline{P} \underline{V})^2$  and  $2\text{tr}(\underline{Q} \underline{V})^2$  are minimized,  $\underline{V}$  being the variance-covariance matrix of the observation vector  $\underline{y}$ . After lengthy algebraic manipulation, with

$$\rho = \sigma_A^2 / \sigma_e^2,$$

$$r = \sum_{i=1}^a \frac{1}{(1 + n_i \rho)^2} + N - a,$$

$$s = \sum_{i=1}^a \frac{n_i^2}{(1 + n_i \rho)^2} \quad \text{and} \quad t = \sum_{i=1}^a \frac{n_i}{(1 + n_i \rho)^2},$$

the estimators turn out to be

$$\hat{\sigma}_e^2 = \frac{1}{rs - t^2} \left[ \sum_{i=1}^a \frac{rn_i - t}{(1 + n_i \rho)^2} \frac{y_{i.}^2}{n_i} + s \left( \sum_{i=1}^a \sum_{j=1}^{n_i} y_{ij}^2 - \sum_{i=1}^a n_i \bar{y}_{i.}^2 \right) \right]$$

and

$$\hat{\sigma}_A^2 = \frac{1}{rs - t^2} \left[ \sum_{i=1}^a \frac{rn_i - t}{(1 + n_i \rho)^2} \frac{y_{i.}^2}{n_i} - t \left( \sum_{i=1}^a \sum_{j=1}^{n_i} y_{ij}^2 - \sum_{i=1}^a n_i \bar{y}_{i.}^2 \right) \right].$$

These estimators are functions of  $\rho = \sigma_A^2 / \sigma_e^2$  and not of the components individually. Furthermore, their variances are identical to those of the large sample maximum likelihood estimators; and their limits as  $\rho \rightarrow 0$  are the estimators given by Koch [1967a]. And the limit as  $\rho \rightarrow \infty$  of  $\hat{\sigma}_e^2$  is the analysis of variance estimator of  $\sigma_e^2$ .

Analytic comparison of these estimators with those obtained from the analysis of variance method is intractable, and so Townsend [1968] made numerical comparisons over a range of values of  $\rho$  both for the actual BQUE's (assuming  $\rho$  known) and for an approximate BQUE using a prior estimate or guess,  $\rho_0$ , for  $\rho$  in the estimation procedure. Indications are that, in certain situations, considerable reduction in the variance of estimates of  $\sigma_A^2$  can be achieved if the approximate BQUE is used rather than the ANOVA estimator.

Furthermore, this advantage can be had even when rather inaccurate prior estimates (guesses) of  $\rho$  are used as  $\rho_0$ . The reduction in variance appears to be greatest when the data are severely unbalanced and  $\rho$  is either small or large, and it appears smallest for values of  $\rho$  that are moderately small. For some sets of data there is actually no reduction in variance, when the ANOVA estimator is a BQUE for some specific  $\rho$ . A full discussion of these estimators and approximations to them for the model with  $\mu \neq 0$  are to be found in Townsend [1968].

#### 20. Designing experiments to estimate variance components.

Herzberg and Cox [1969], in an extensive review and bibliography of the design of experiments comment that "relatively little work has been done" on designing experiments for the purpose of estimating variance components. This is borne out by their bibliography of approximately 800 publications since 1957 which they have classified in nine overlapping groups according to subject matter: less than 3% are classed as dealing with designs for estimating variance components, more than half of which are concerned with genetics, mostly the diallel cross. Despite this, a need for designing experiments with this object in mind is highlighted by Anderson [1960a], who points out that in using the analysis of variance method of estimation in balanced nested designs the degrees of freedom for the innermost classifications are increasingly larger than those for the outermost classifications. Designs that rectify this imbalance are, for example, the "staggered" designs of Bainbridge [1963], which incorporate what may be called planned unbalancedness so as to make the degrees of freedom more nearly uniform. Examples and discussion of these designs is also to be found in Anderson [1960b].

Optimal designs for the 1-way classification random model are considered by Anderson and Crump [1967]. For a given number of observations  $N$  they suggest how to determine an optimum number of classes  $c$ , and then show that  $\text{var}(\hat{\sigma}_A^2)$  of (60) is minimized if  $p + 1$  observations are made in each of  $r$  classes and  $p$  in the remaining  $c - r$ , such that  $N = cp + r$  for  $r$  positive and less than  $c$ . Kussmaul and Anderson [1967] in discussing their compositing of samples, outlined above in Section 17, consider optimal designs for a given cost structure. In this context they write: "good designs for such general estimation purposes are extremely difficult to specify analytically", a comment that is also pertinent to the Anderson and Crump [1967] paper. Both papers show results of numerical studies of a number of different designs.

A limited form of planned unbalancedness is also considered by Mostafa [1967] for the 2-way classification with interaction. This involves having 2 observations in some cells and one in each of the others.

## 21. Other methods and ideas

### 21.1 Bayes estimation

The main papers dealing with Bayes procedures for variance component estimation are those by Hill [1965, 1967], Tiao and Tan [1965, 1966] and Box and Tiao [1967]. The latter deals with the 2-way nested classification whereas the other four are concerned with the 1-way case. They are comprehensively reviewed by Harville [1969b], to which the reader is directed for an exhaustive summary.

One problem of variance component estimation to which Bayes procedures pay particular attention is that of negative estimates. In this regard, Hill [1965] in summarizing his results writes "Bayesianly speaking, a large negative



unbiased estimate of the between variance component indicates an uninformative experiment in which the effective likelihood for that variance component is extremely flat, instead of strong evidence that the variance component is nearly zero." Hill indicates (op cit, p. 817) that when  $\hat{\sigma}_A^2 \ll 0$ , it "becomes crucial to carefully assess the prior", although in discussing a hypothetical numerical example he warns (p. 821) that in "any real problem the prior can always be assessed, although perhaps with difficulty." In this connection Tiao and Tan [1965 and 1966] use 'non-informative' priors, a procedure which Kempthorne [1968] finds unsatisfying. Evidently, just as with other applications of Bayesian estimation, so with variance components, both protagonists and adversaries can readily be found. Unfortunately, their arguments so often generate "more heat than light", as Harville [1969b] so aptly puts it.

#### 21.2 Plotting the likelihood.

Although maximum likelihood estimators (for unbalanced data) cannot be derived, use of the likelihood function can be made by plotting its value (for a given set of  $n_{ij}$ 's) over a range of values of the variance components. Kempthorne [1968] indicates that investigations of this nature are currently in progress and that when the usual estimate is negative "the likelihood function seems quite informative ... and points towards zero as the value of the variance component." However, he goes on farther to say that there are "no particularly compelling ways of interpreting the likelihood function, except in simple cases for which we can pick out aspects such as the point of maximum and can construct statistical tests using them." But such processes "seem totally non-viable with the cases that ... animal breeders usually meet," meaning, presumably, cases of badly unbalanced data.

### 21.3 Sequential estimation.

Sequential procedures for testing hypotheses about variance components are also to be found in the literature, although not to any great extent. Birnbaum [1958] gives a general sequential sampling method for comparing variances (by means of their ratio) of two normal populations. Adaptation to variance components in the balanced 1-way classification is indicated. Savings appear to be small in terms of numbers of observations needed in order to have efficiency equivalent to that of the usual F-test. Sequential probability ratio tests of hypotheses have also been suggested by Ghosh [1965, 1967]. Industrial environments may well be suited to the use of such tests, where balanced data may be readily obtainable sequentially as outflow of one or more machine processes. However, this situation seems less likely to occur in a biological environment.

### 21.4 Other models.

Most comparative studies of variance component estimators are made on the basis of normality assumptions. The relative intractability of results even in this case discourages extensive studies of non-normal models. Nevertheless, in this situation Hammersley [1949] and Tukey [1957a] have looked at the 1-way classification in some detail, and Harville [1969a] discusses the form of quadrmissible estimators. Atiqullah [1962] considers variances of variance components estimators for the mixed model having one random factor (balanced data), and Kelleher and Robinson [1958] report comparative studies of estimates obtained from corn breeding experiments over a period of years. They conclude that their results show no departure from the underlying normality assumptions.

Another departure from the usual model is dealt with by Harville [1967b, 1968] who uses models in which the numbers of observations are taken as random

variables. Marginal distributions of the random effects are assumed normal, with the conditional distribution of the  $n$ 's, given the effects, being a Poisson distribution. These models are appropriate to situations where the number of observations in a class depends on the magnitude of that class effect; e.g., in the artificial breeding of dairy cattle, bulls that sire high-yielding daughters are apt to have more daughters than those who leave low-yielding daughters. The absolute value of the bias of equal-weights estimators in such cases is found by Harville to be generally less than that of analysis of variance estimators.

#### 21.5 Harmonic components.

Variance components for factors of a model that are largely time periods have been considered by Lawton and Jackson [1969] and Jackson and Lawton [1969]. These authors consider nested classification random models, with balanced data, as asymptotic stationary random processes for which they derive the asymptotic spectrum and hence what they call harmonic variance components. These are described as "representing the amount of variability due to cyclic patterns of the various periods," and they appear to be best suited to such time-defined factors as, for example, weeks, days within weeks, shifts within days and hours within shifts, in an industrial situation. The linear relationship of such components to those defined in the usual analysis of variance context is shown, and numerical examples given.

#### 21.6 Dispensing with unbiasedness.

Most current methods of variance component estimation yield estimators that are mean unbiased. This property has been righteously adopted from linear estimation, - but for the variance components situation it can be questioned, or at least amended. As Kempthorne [1968] points out, unbiasedness in estimating

the fixed effects of a model "...leads to residuals which do not contain systematic effects and is therefore valuable...and is fertile mathematically in that it reduces the class of candidate statistics (or estimates)." But "...in the variance component problem, it does not lead to a fertile smaller class of statistics." Furthermore, in linear estimation, we usually conceive of repetitions of data and associated repetitions of estimates. But with unbalanced data from random models this is often not the case - more data, yes, but not necessarily with the same pattern of unbalancedness. Any replications of data cannot, therefore, altogether be thought of as mere re-samplings of data already available. The concept of mean unbiasedness is therefore not necessarily pertinent. Furthermore, random model data are often so voluminous in extent that additional data may effectively be unavailable or, if available they may only be obtainable from different populations. This demands contentment with the data already available, and conceiving of repetitions of data is no longer permitted. It may therefore be reasonable to dispense with the mean unbiasedness property of present estimation procedures, perhaps replacing it by one of modal unbiasedness, as hinted at in Searle [1968, discussion]. It would mean developing a variance component estimator for which the probability distribution had its (only) mode at the population value of the parameter. Such a development would maximize the probability of the estimator being close to its parameter, rather than minimizing the probability of being far away from it, as Eisenhart [1968] points out. Harville [1969b] doubts if estimators satisfying the modal unbiasedness criterion exist and questions, on decision-theoretic grounds, the justification of such a criterion preferring to use an appropriate loss function.

If abandonment of traditional methods of estimation were to be considered, the use of range by Ghosh [1965] in developing tests of hypotheses concerning ratios of variances could prompt speculation on using range statistics for developing estimation methods. Whilst resulting estimators might be expected to have less attractive properties than those of current estimators they might, if not very much less attractive, be worthwhile, since current estimators are difficult to compute from unbalanced data, and have few optimal properties. Of course, the development of such estimation methods, even if appropriate, might well be no easy task.

References

- Ahrens, H. [1965]. Standardfehler geschätzter Varianzkomponenten eines unbalanzierten Versuchplanes in r-stufiger hierarchischer Klassifikation. Monatsb. Deutsch Akad. Wiss. Berlin 7, 89-94.
- Anderson, R. L. [1960a]. Some remarks on the design and analysis of factorial experiments. In Contributions to Probability and Statistics (Ed. I. Olkin et al) p.p. 35-56. Stanford University Press, Stanford, California.
- Anderson, R. L. [1960b]. Uses of variance components analysis in the interpretation of biological experiments. International Statistical Institute Bulletin 37, Vol. 3, 71-90.
- Anderson, R. L. [1962]. Designs for estimating variance components. University of N. Carolina, Mimeo Series No. 310, Raleigh, North Carolina.
- Anderson, R. L. [1965]. Negative variance estimates. Query No. 9, Technometrics, 7, 75-76.
- Anderson, R. L. and Bancroft, T. A. [1952]. Statistical Theory in Research. McGraw-Hill, New York.
- Anderson, R. L. and Crump, P. P. [1967]. Comparisons of designs and estimation procedures for estimating parameters in a two-stage nested process. Technometrics, 9, 499-516.
- Atiqullah, M. [1962]. On the effect of non-normality on the estimation of variance components. J. Roy. Stat. Soc. (B), 24, 140-146.
- Bainbridge, T. R. [1963]. Staggered, nested designs for estimating variance components. Am. Soc. Qual. Control 17-th Annual Conference Transactions, 93-103.
- Bennett, C. A. and Franklin, N. L. [1954]. Statistical Analysis in Chemistry and the Chemical Industry. John Wiley and Sons, New York.
- Birnbaum, Allan [1958]. Sequential tests for variance ratios and components of variance. Ann. Math. Stat., 29, 506-515.
- Blischke, W. R. [1966]. Variances of estimates of variance components in a three-way classification. Biometrics, 22, 553-565.
- Blischke, W. R. [1968]. Variances of moment estimators of variance components in the unbalanced r-way classification. Biometrics, 24, 527-540.
- Broemeling, L. D. [1969a]. Confidence regions for variance ratios of random models. J. Am. Stat. Ass. 64, 660-664.

- Broemeling, L. D. [1969b]. Confidence intervals for measures of heritability. Biometrics, 25, 424-427.
- Bross, Irwin [1950]. Fiducial intervals for variance components. Biometrics, 6, 136-144.
- Bulmer, M. G. [1957]. Approximate confidence limits for components of variance. Biometrika, 44, 159-167.
- Bush, N. and Anderson, R. L. [1963]. A comparison of three different procedures for estimating variance components. Technometrics, 5, 421-440.
- Cochran, W. G. [1951]. Testing a linear relation among variances. Biometrics, 7, 17-32.
- Cornfield, J. and Tukey, J. W. [1956]. Average values of mean squares in factorials. Ann. Math. Stat., 27, 907-949.
- Crump, S. L. [1946]. The estimation of variance components in analysis of variance. Biometrics (Bulletin), 2, 7-11.
- Crump, S. L. [1947]. The estimation of components of variance in multiple classifications. Ph.D. Thesis, Iowa State College Library, Ames, Iowa.
- Crump, S. L. [1951]. Present status of variance component analysis. Biometrics, 7, 1-16.
- Cunningham, E. P. [1969]. A note on the estimation of variance components by the method of fitting constants. Biometrika, 56, 683-684.
- Cunningham, E. P. and Henderson, C. R. [1968]. An iterative procedure for estimating fixed effects and variance components in mixed model situations. Biometrics, 24, 13-25. Corrigenda, Biometrics, 25, 777-778, 1969.
- Daniels, H. E. [1939]. The estimation of components of variance. J. Roy. Stat. Soc. (Supplement), 6, 186-197.
- Eisenhart, C. [1947]. The assumptions underlying the analysis of variance. Biometrics, 3, 1-21.
- Eisenhart, C. [1968]. Discussion of Searle [1968]. Biometrics, 24, 784.
- Evans, D. A. [1968]. Personal communication.
- Federer, W. T. [1968]. Non-negative estimators for components of variance. App. Stat. 17, 171-174.
- Fryer, H. C. [1966]. Concepts and Methods of Experimental Statistics. Allyn and Bacon, Boston, Mass.

- Furukawa, N. [1959]. The point estimation of the variance components in random effect model. Kunamoto J. of Sci., (A), 4, 111-133.
- Ganguli, M. [1941]. A note on nested sampling. Sankhyā, 5, 449-452.
- Gates, C. E. and Shiue C. [1962]. The analysis of variance of the S-stage hierarchical classification. Biometrics, 18, 529-536.
- Gaylor, D. W. and Hartwell, T. D. [1969]. Expected mean squares for nested classifications. Biometrics, 25, 427-430.
- Gaylor, D. W. and Hopper, F. N. [1969]. Estimating the degrees of freedom for linear combinations of mean squares by Satterthwaite's formula. Technometrics, 11, 691-706.
- Gaylor, D. W., Lucas, H. L. and Anderson, R. L. [1970]. Calculation of expected mean squares by the abbreviated Doolittle and square root methods, Biometrics, 26, (in press).
- Ghosh, B. K. [1965]. Sequential range tests for components of variance. J. Am. Stat. Ass., 60, 826-836.
- Ghosh, B. K. [1967]. Sequential analysis of variance under random and mixed models. J. Am. Stat. Ass., 62, 1401-1417.
- Gower, J. C. [1962]. Variance component estimation for unbalanced hierarchical classifications. Biometrics, 18, 537-542.
- Graybill, F. A. [1954]. On quadratic estimates of variance components. Ann. Math. Stat., 25, 367-372.
- Graybill, F. A. [1961]. An Introduction to Linear Statistical Models. Vol. I. McGraw-Hill, New York.
- Graybill, F. A. and Hultquist, R. A. [1961]. Theorems concerning Eisenhart's Model II. Ann. Math. Stat., 32, 261-269.
- Graybill, F. A. and Weeks, D. L. [1959]. Combining inter-block and intra-block information in balanced incomplete blocks. Ann. Math. Stat., 30, 799-805.
- Graybill, F. A. and Wortham, A. W. [1956]. A note on uniformly best unbiased estimators for variance components. J. Am. Stat. Assoc., 51, 266-268.
- Hammersley, J. M. [1949]. The unbiased estimate and standard error of the interclass variance. Metron, 15, 189-205.
- Hartley, H. O. [1967]. Expectation, variances and covariances of ANOVA mean squares by 'synthesis'. Biometrics, 23, 105-114, and Corrigenda, 853.
- Hartley, H. O. and Rao, J. N. K. [1967]. Maximum likelihood estimation for the mixed analysis of variance model. Biometrika, 54, 93-108.



- Hartley, H. O. and Searle, S. R. [1969]. On interaction variance components in mixed models. Biometrics, 25, 573-576.
- Harville, D. A. [1967a]. Estimability of variance components for the 2-way classification with interaction. Ann. Math. Stat., 38, 1508-1519.
- Harville, D. A. [1967b]. Statistical dependence between random effects and the numbers of observations on the effects for the unbalanced one-way classification. J. Am. Stat. Ass., 62, 1375-1386.
- Harville, D. A. [1968]. Statistical dependence between subclass means and the numbers of observations in the subclasses for the two-way completely-random classification. J. Am. Stat. Ass., 63, 1484-1494.
- Harville, D. A. [1969a]. Quadratic unbiased estimation of variance components for the one-way classification. Biometrika, 56, 313-326.
- Harville, D. A. [1969b]. Variance component estimation for the unbalanced one-way random classification - a critique. Aerospace Research Laboratories, ARL-69-0180, Wright-Patterson Air Force Base, Ohio.
- Harville, D. A. [1969c]. Variances of variance-component estimators for the unbalanced 2-way cross-classification with application to balanced incomplete block designs. Ann. Math. Stat. 40, 408-416.
- Harville, D. A. [1969d]. Expression of variance component estimators as linear combinations of independent non-central chi-square variates. Ann. Math. Stat., 40, 2189-2194.
- Henderson, C. R. [1953]. Estimation of variance and covariance components. Biometrics, 9, 226-252.
- Henderson, C. R. [1959]. Design and analysis of animal husbandry experiments. Chapter 1 of Techniques and Procedures in Animal Production Research. American Society of Animal Production Monograph.
- Henderson, C. R., Kempthorne, O., Searle, S. R., and Von Krosigk, C. N. [1959]. Estimation of environmental and genetic trends from records subject to culling. Biometrics, 15, 192-218.
- Henderson, C. R. and Searle, S. R. [1970]. On mixed models. Paper number BU-274-M in the mimeograph series of the Biometrics Unit, Cornell University, Ithaca, N.Y.
- Herbach, L. H. [1959]. Properties of Model II type analysis of variance tests. Ann. Math. Stat., 30, 939-959.
- Herzberg, A. M., and Cox, D. R. [1969]. Recent work on the design of experiments: a bibliography and a review. J. Roy. Stat. Soc. (A), 132, 29-67.

- Hill, B. M. [1965]. Inference about variance components in the one-way model. J. Am. Stat. Assoc., 60, 806-825.
- Hill, B. M. [1967]. Correlated errors in the random model. J. Am. Stat. Assoc., 62, 1387-1400.
- Hooke, R. [1956]. Some applications of bipolykays to the estimation of variance components and their moments. Ann. Math. Stat., 27, 80-98.
- Jackson, J. E., and Lawton, W. H. [1969]. Comparison of ANOVA and harmonic components of variance. Technometrics, 11, 75-90.
- Kapadia, C. H. and Weeks, D. L. [1963]. Variance components in two-way classification models with interaction. Biometrika, 50, 327-336.
- Kempthorne, O. [1968]. Discussion of Searle [1968]. Biometrics, 24, 782-784.
- Kimball, A. W. [1951]. On dependent tests of significance in the analysis of variance. Ann. Math. Stat., 22, 600-602.
- Kirk, R. E. [1968]. Experimental Design: Procedures for the Behavioural Science. Brooks/Cole, Belmont, California.
- Klotz, J. H. and Milton, R. C. [1967]. Estimators of components of variance with small mean square error. Tech. Report No. 136, Dept. of Statistics, University of Wisconsin, Madison, Wisconsin.
- Klotz, J. H., Milton, R. C., and Zacks, S. [1969]. Mean square efficiency of estimators of variance components. J. Am. Stat. Ass., 64, 1383-1402.
- Koch, G. G. [1967a]. A general approach to the estimation of variance components. Technometrics, 9, 93-118.
- Koch, G. G. [1967b]. A procedure to estimate the population mean in random effects models. Technometrics, 9, 577-586.
- Koch, G. G. [1968]. Some further remarks on "A general approach to the estimation of variance components". Technometrics, 10, 551-558.
- Kussmaul, Keith and Anderson, R. L. [1967]. Estimation of variance components in two-stage nested designs with composite samples. Technometrics, 9, 373-390.
- Lawton, W. H. and Jackson, J. E. [1969]. The spectrum of a Model II nested ANOVA and its applications. Technometrics, 11, 91-102.
- Leone, F. C. and Nelson, L. S. [1966]. Sampling distributions of variance components. I. Empirical studies of balanced nested designs. Technometrics, 8, 457-468.

- Leone, F. C., Nelson, L. S., Johnson, N. L., and Eisenstat, S. [1968]. Sampling distributions of variance components. II Empirical studies of unbalanced nested designs. Technometrics, 10, 719-738.
- Low, L. Y. [1964]. Sampling variances of estimates of components of variance from a non-orthogonal two-way classification. Biometrika, 51, 491-494.
- Low, L. Y. [1969]. Estimators of variance components in the balanced incomplete block. J. Am. Stat. Ass., 64, 1014-1030.
- Lum, Mary D. [1954]. Rules for determining error terms in hierarchal and partially hierarchal models. Wright Air Development Centre, Dayton, Ohio.
- Mahamunulu, D. M. [1963]. Sampling variances of the estimates of variance components in the unbalanced 3-way nested classification. Ann. Math. Stat., 34, 521-527.
- McHugh, R. B. and Mielke, P. W., Jr. [1968]. Negative variance estimates and statistical dependence in nested sampling. J. Am. Stat. Assoc., 63, 1000-1003.
- Millman, J. and Glass, G. V. [1967]. Rules of thumb for writing the anova table. J. of Educational Measurement, 4, 41-51.
- Mood, A. M. [1950]. Introduction to the Theory of Statistics. McGraw-Hill, New York.
- Mood, A. M. and Graybill, F. A. [1963]. Introduction to the Theory of Statistics. 2nd Ed., McGraw-Hill, New York.
- Mostafa, M. G. [1967]. Designs for the simultaneous estimation of functions of variance components from 2-way crossed classifications. Biometrika, 54, 127-131.
- Nelder, J. A. [1954]. The interpretation of negative components of variance. Biometrika, 41, 544-548.
- Oktaba, W. [1968]. A note on estimating the variance components by the method two of Henderson. Biometrische Zeitschrift, 10, 97-108.
- Plackett, R. L. [1960]. Models in the analysis of variance. J. Roy. Stat. Soc. (B), 22, 195-217.
- Rao, J. N. K. [1968]. On expectations, variances and covariances of anova mean squares by 'synthesis'. Biometrics, 24, 963-978.
- Robertson, A. [1962]. Weighting in the estimation of variance components in the unbalanced single classification. Biometrics, 18, 413-417.
- Robinson, J. [1965]. The distribution of a general quadratic form in normal variables. Australian J. of Stats., 7, 110-114.

- Robson, D. S. [1957]. Personal communication.
- Robson, D. S. [1965]. On the admissability of non-negative maximum likelihood estimators of the "between groups" variance component. Paper No. BU-179-M in the mimeo series, Biometrics Unit, Cornell University, Ithaca, New York.
- Rohde, C. A. and Tallis, G. M. [1969]. Exact first- and second-order moments of estimates of components of covariance. Biometrika, 56, 517-526.
- Roy, S. N. and Whitfield, Cobb. [1960]. Mixed model variance analysis with normal error and possibly non-normal other random effects. I. The univariate case. II. The multivariate case. Ann. Math. Stat., 31, 939-957 and 958-968.
- Satterthwaite, F. E. [1946]. An approximate distribution of estimates of variance components. Biometrics Bulletin, 2, 110-114.
- Scheffé, H. [1959]. The Analysis of Variance. Wiley, New York, (Chapters 7 and 8).
- Schultz, E. F., Jr. [1955]. Rules of thumb for determining expectations of mean squares in analysis of variance. Biometrics, 11, 123-135.
- Searle, S. R. [1956]. Matrix methods in variance and covariance components analysis. Ann. Math. Stat., 27, 737-748.
- Searle, S. R. [1958]. Sampling variances of estimates of components of variance. Ann. Math. Stat., 29, 167-178.
- Searle, S. R. [1961a]. Variance components in the unbalanced 2-way nested classification. Ann. Math. Stat., 32, 1161-1166.
- Searle, S. R. [1961b]. Estimating the heritability of butterfat production. J. Agr. Sci., 57, 289-294.
- Searle, S. R. [1966]. Matrix Algebra for the Biological Sciences. John Wiley and Sons, Inc., New York.
- Searle, S. R. [1968]. Another look at Henderson's methods of estimating variance components. Biometrics, 24, 749-788.
- Searle, S. R. [1969]. Variance component estimation : proof of equivalence of alternative method 4 to Henderson's method 3. Paper No. BU-260-M in the Mimeo Series, Biometrics Unit, Cornell University.
- Searle, S. R. [1970]. Large sample variances of maximum likelihood estimators of variance components. Biometrics, 26, (in press).
- Searle, S. R. and Fawcett, R. F. [1970]. Expected mean squares in variance components models having finite populations. Biometrics, 26, (in press).

- Searle, S. R. and Henderson, C. R. [1960]. Judging the effectiveness of age correction factors. J. Dairy Sci., 43, 966-974.
- Searle, S. R. and Henderson, C. R. [1961]. Computing procedures for estimating components of variance in the two-way classification, mixed model. Biometrics, 17, 607-616. Corrigenda, Biometrics, 23, 852, 1967.
- Snedecor, G. W. and Cochran, W. G. [1967]. Statistical Methods. 6th Ed., Iowa State University Press, Ames, Iowa.
- Steele, R. G. D. and Torrie, J. H. [1960]. Principles and Procedures of Statistics. McGraw-Hill, New York.
- Thompson, R. [1969]. Iterative estimation of variance components for non-orthogonal data. Biometrics, 25, 767-773.
- Thompson, W. A., Jr. [1961]. Negative estimates of variance components: an introduction. Bulletin, International Institute of Statistics, 34, 1-4.
- Thompson, W. A., Jr. [1962]. The problem of negative estimates of variance components. Ann. Math. Stat., 33, 273-289.
- Thompson, W. A., Jr. and Moore, James, R. [1963]. Non-negative estimates of variance components. Technometrics, 5, 441-450.
- Tiao, G. C. and Box, G. E. P. [1967]. Bayesian analysis of a three-component hierarchical design model. Biometrika, 54, 109-125.
- Tiao, G. C. and Tan, W. Y. [1965]. Bayesian analysis of random-effects models in the analysis of variance I. Posterior distribution of variance components. Biometrika, 52, 37-53.
- Tiao, G. C. and Tan, W. Y. [1966]. Bayesian analysis of random-effects models in the analysis of variance. II. Effect of autocorrelated errors. Biometrika, 53, 477-495.
- Townsend, C. E. [1968]. Unbiased estimators of variance components in simple unbalanced designs. Unpublished Ph.D. Thesis, Cornell University, Ithaca, New York.
- Tukey, J. W. [1956]. Variances of variance components: I. Balanced designs. Ann. Math. Stat., 27, 722-736.
- Tukey, J. W. [1957a]. Variances of variance components: II. The unbalanced single classification. Ann. Math. Stat., 28, 43-56.
- Tukey, J. W. [1957b]. Variances of variance components: III. Third moments in a balanced single classification. Ann. Math. Stat., 28, 378-386.
- Wald, A. [1943]. Tests of statistical hypotheses concerning several parameters when the number of variables is large. Trans. Amer. Math. Soc. 54, 426-482.

- Wang, Y. Y. [1967]. A comparison of several variance component estimators. Biometrika, 54, 301-305.
- Weeks, D. L. and Graybill, F. A. [1961]. Minimal sufficient statistics for the balanced incomplete block designs under an Eisenhart Model II. Sankhyā, (A), 23, 261-268.
- Weeks, D. L. and Graybill, F. A. [1962]. A minimal sufficient statistic for a general class of designs. Sankhyā, (A), 24, 339-354.
- Welch, B. L. [1956]. On linear combinations of several variances. J. Am. Stat. Assoc., 51, 132-148.
- Wilk, M. B. [1955]. The randomized analysis of a generalized randomized block design. Biometrika, 42, 70-79.
- Wilk, M. B. and Kempthorne, O. [1955]. Fixed, mixed and random models. J. Am. Stat. Assoc., 50, 1144-1167.
- Wilk, M. B. and Kempthorne, O. [1956]. Some aspects of the analysis of factorial experiments in a completely randomized design. Ann. Math. Stat., 27, 950-985.
- Wilk, M. B. and Kempthorne, O. [1957]. Non-additivities in a Latin Square design. J. Am. Stat. Ass., 52, 218-236.
- Williams, J. S. [1962]. A confidence interval for variance components. Biometrika, 49, 278-281.
- Winsor, C. P. and Clarke, G. L. [1940]. A statistical study of variation in the catch of plankton nets. Sears Foundation J. of Marine Res., 3, 1.
- Yates, F. [1934]. The analysis of multiple classifications with unequal numbers in the different classes. J. Am. Stat. Assoc., 29, 51-66.
- Yates, F. [1967]. A fresh look at the basic principles of the design and analysis of experiments. Proceedings of the 5'th Berkeley Symposium on Mathematical Statistics and Probability. Ed. L. Lecam and J. Neyman, University of California Press, Vol IV, 777-790.
- Young, C. W., Legates, J. E. and Farthing, B. R. [1965]. Pre- and post-natal influences on growth, prolificacy and maternal performance in mice. Genetics, 52, 553-561.
- Zacks, S. [1967]. More efficient estimators of variance components. Tech. Report No. 4. Dept. of Statistics, Kansas State University, Manhattan, Kansas.
- Zelen, M. [1968]. Discussion of Searle [1968]. Biometrics, 24, 779-780.