A Reformulation of Linear Models

J. A. Nelder

Journal of the Royal Statistical Society. Series A (General), Vol. 140, No. 1 (1977), 48-77.

Stable URL:
http://links.jstor.org/sici?sici=0035-9238%281977%29140%3A1%3C48%3AAROLM%3E2.0.CO%3B2-8

Journal of the Royal Statistical Society. Series A (General) is currently published by Royal Statistical Society.

J. R. Statist. Soc. A,
(1977), **140**, *Part* 1, p. 48

A Reformulation of Linear Models

By J. A. NELDER

Rothamsted Experimental Station, Harpenden, Herts, U.K.

[Read before the ROYAL STATISTICAL SOCIETY on Wednesday, November 10th, 1976, the President, Miss STELLA V. CUNLIFFE, in the Chair]

SUMMARY

Dissatisfaction is expressed with aspects of the current exposition of linear models, including the neglect of marginality, unnecessary differences between models for finite and infinite populations, failure to distinguish different kinds of random terms, impositon of unnecessary and inconsistent constraints on parameters, and lack of an adequate notation for negative components of variance. The reformulation, exemplified for crossed and nested classifications of balanced data, and for simple orthogonal designed experiments, is designed to integrate finite and infinite populations, random and fixed effects, excess and deficit of variance, to avoid unnecessary constraints on parameters, and to lead naturally to interesting hypotheses about the model terms.

Keywords: LINEAR MODEL; FIXED EFFECT; RANDOM EFFECT; VARIANCE COMPONENT; CANONICAL COMPONENT; MARGINALITY; CONSTRAINTS; INTRINSIC ALIASING; EXTRINSIC ALIASING

1. INTRODUCTION

THIS paper has arisen from a personal dissatisfaction with several aspects of the current exposition of linear models. Such models, rooted in classical least-squares theory, and more recently extended to include multiple error terms and random effects, form a fundamental tool, perhaps *the* fundamental tool, for the practising statistician. It is, therefore, vital that we have a clear and agreed exposition of the subject. In spite of the enormous literature I do not believe that we have such an agreed exposition, for reasons which will be enlarged upon in Section 2. There follows in subsequent sections a development of linear models which I believe to be free of inconsistencies, and one which is tied as closely as possible to the structure of data as collected in experiments and surveys.

1.1 *Notation*

We shall use capitals $A, B, C, \ldots$ to denote *factors* classifying the populations under discussion. A factor divides the elements of a population into disjunct sets conventionally indexed by $1, 2, 3, \ldots$. We shall associate the indices $i, j, k, \ldots$ with $A, B, C, \ldots$ respectively. Factor A has N_A *levels* in the population, and n_A levels in a sample from that population, so that $n_A \leqslant N_A$. *Structures* defined by the factors are written using the notation of Wilkinson and Rogers (1973). A structure implies a linear model; thus the nested structure A/B gives the model formula

$$A + A.B,$$

where $A.B$ means "B within A" because it is accompanied by one main effect A only. The corresponding crossed structure $A * B$ gives

$$A + B + A.B,$$

where $A.B$ is now the interaction of A and B because it is accompanied by both main effects.

A simple term such as B in a model formula corresponds to a one-dimensional set of *parameters* β_j, and a compound term such as $A.B$ exemplifies a multi-dimensional set $(\alpha\beta)_{ij}$.

Estimates of the parameters will be written with the corresponding Latin letters, e.g. b_j. The population of elements will be denoted by **x** with suitable subscripts, and a set of data values obtained by sampling **x** will be denoted by **y**.

We shall consider sampling schemes defined by the selection of levels of factors, and shall distinguish *complete* from *incomplete sampling*. Incomplete sampling has $n<N$ and implies a random sample of n levels from the population of N. The corresponding parameters in the linear model are random variables. Complete sampling implies that $n=N$, so that N is necessarily finite. The idea of randomness now disappears, and the corresponding parameters in the linear model correspond to what are usually called *fixed effects*. Note that if we select a subset n *non-randomly* from a population of size N', then our inferences are necessarily conditioned by the n actually selected, so that N is reduced from N' to n.

2. Reasons for Dissatisfaction

2.1. *The Neglect of Marginality*

When two factors, A and B, say, index some data by rows and columns and the model $A+B+A.B$ is to be fitted by least squares, there is no disagreement about the *arithmetic* of the analysis of variance when the data are orthogonal, though there is disagreement about the underlying algebra and its interpretation, a disagreement which is the subject of this paper. When the data are non-orthogonal the algebraic disagreement leads to divergencies in the arithmetic also, as Francis (1973) showed. When quantitative covariates (explanatory variates) are involved in a linear model, the test for the contribution of any one, given that the others are to be in the model, is obtained by fitting first all the covariates and then omitting the one to be tested, deriving the appropriate sum of squares (S.S.) with 1 d.f. as the difference in residual sums of squares. There is usually no special ordering among the covariates, so that it makes sense to fit them in any order. However, with the sets of qualitative variates generated by factors, as exemplified in A, B and $A.B$, i.e. two main effects and their interaction, a special ordering does exist. $A.B$ represents a two-way array of parameters $(\alpha\beta)_{ij}$; in the space of all linear contrasts between these parameters are the two sub-spaces representing contrasts between row totals, and between column totals. The two sub-spaces underlie, of course, the two main effects A and B. Note that A and B are sub-spaces of the space of $A.B$ *independently of the design matrix of the observations*; thus their relationship to $A.B$ is different in kind from the sort of collinearity (corresponding to confounding) which depends upon the particular form of the design matrix. A and B will be said to be *marginal* to $A.B$; similarly A, B, C, $A.B$, $A.C$ and $B.C$ are all marginal to $A.B.C$. The term for the grand mean is marginal to all terms involving factors.

A consequence of marginality is that the interpretation of, say, $A.B$ will, as noted above, depend on which terms marginal to it precede it in a model formula. Consider the three formulae

$$A.B$$

$$A+A.B$$

$$A+B+A.B$$

The first describes a two-way array of parameters indexed by combinations of levels of A and B (e.g. the expected values of the cell means); the second extracts parameters representing the average effect of A so that $A.B$ represents parameters for residual B-within-A effects (such a model is relevant when B is nested within A); the third has both main effects removed so that $A.B$ represents an interaction term and a cross-classification of A and B is implicit. These apparent ambiguities in the interpretation of $A.B$ may seem at first sight disconcerting, but they will be found to mirror what happens numerically when an algorithm is used for fitting the models. If, for example, the term A is fitted first then the subsequent fitting of $A.B$

will reveal explicitly that there is now no further information about those contrasts in $A.B$ that represent contrasts between rows. There is thus no alternative to omitting these contrasts when fitting $A.B$, and this changes the interpretation of $A.B$ accordingly.

If the marginality relations are ignored and A is fitted after $A.B$, then of course it will be found that there is no information about A once $A.B$ has been fitted, because A is a sub-space of $A.B$. Thus A must be omitted from the model, and the effect is *algebraically* equivalent to setting the A effects to zero; it follows that, as Searle (1971) has maintained, the only possible value for the S.S. for A eliminating $A.B$ is zero. However, in a very widely used program for fitting linear models, BMDX64 (Dixon, 1970), non-zero S.S. for A eliminating $A.B$, etc. are given. How has this come about? The answer is that, in the initial fitting of $A.B$, constraints are imposed on the parameters; the undesirable effects of imposing constraints will be considered in more detail in Section 2.4. Two points are relevant here: one is that it is not always obvious what the relevant constraints should be, particularly when unequal cell numbers occur in the data; the other, much more fundamental, is that it leads to models where an interaction, $A.B$ say, is postulated to exist whose marginal effects, A or B, are to be null. *I contend that such models are of no practical interest.* Thus a S.S. for A eliminating $A.B$, though calculable given certain assumptions, makes no practical sense. I assert that the general rules for calculating S.S. must take account of marginality, the meaningful S.S. in this example being those for $A+B$, A eliminating B, B eliminating A and $A.B$ eliminating A and B.

2.2. *Finite and Infinite Populations*

Linear models have been developed both with infinite and finite populations in mind. Infinite populations are classical; finite populations have been mostly explored in the context of randomization, first by Fisher, followed by many others, see e.g. Eden and Yates (1933), Welch (1937) and Pitman (1938) for early references. In textbooks these two developments usually exist as uneasy bedfellows, with exact distribution theory for one and approximate theory for the other. In the literature the present distinction, between finite and infinite populations, has not always been clearly separated from the two other, logically quite separate, distinctions, namely those between fixed and random effects, and between sample and population. Thus Cornfield and Tukey (1956), in a paper that gives many valuable insights, attempt to equate finite populations with fixed effects, and infinite with random effects. However, if this equation is accepted it makes impossible the development of a framework that includes randomization models (which use finite populations and random effects) as a special case. The point is that though fixed effects must correspond to a finite population of levels, and in fact comprise the whole of that population, random effects can be defined by sampling from a population of levels which may be either finite or infinite. We shall equate fixed and random effects, not with finite and infinite populations, but with complete and incomplete sampling. This gives a unified development for both finite and infinite populations and, in particular, displays the often awkward "corrections for finite populations" in variances as just a special case of a general formula.

2.3. *Two aspects of a Random Effect*

Classical least squares has a model containing only one random term, the residual. Remaining terms correspond to what are now called fixed effects. Developments in experimental design produced linear models, e.g. for split-plot designs, which contained more than one random term, together with treatment effects regarded as fixed. Then more general models appeared with arbitrary combinations of fixed and random terms—note that again marginality relations must be recognized, because a model having A or B or both random but their interaction fixed makes no sense. The classical model became known as a fixed model (illogically, because the last term is random), the model with all terms random became a random model, and all other types of model became mixed models (Eisenhart, 1947). This generalization, however, fails to make an important distinction between two kinds of random term.

One kind is a component of error; no interest attaches to individual values, only to their mass behaviour. The other kind of random term represents an effect of interest (a "treatment" term) which is nevertheless specified in terms of a sample from a population; an example might be a set of measuring instruments in a measurement-error trial, chosen at random from a larger set. Individual values may be of interest as well as the variance component that describes the population. Random terms of the first kind describe the error structure of the data, that is they classify the experimental units by assigning contrasts between the observations to various *error strata* (Nelder, 1965). Sums of squares deriving from them form the denominators in F-ratios for which the variation of either fixed terms or random terms of the second kind form the numerators. Again a suitable theory must embody this distinction.

2.4. *Constraints*

Hartley and Searle (1969) drew attention to what they called a "discontinuity" in that an expected mean square in a nested classification sometimes contained a term with a certain variance component and sometimes did not, according to the assumptions made about the effects. This difference is a simple example of a much wider problem about the definition of variance components arising from random terms in a model. It is common (though undesirable) in setting up a model with fixed effects, say $Y_{ij} = \mu + \alpha_i + \beta_j$, to constrain both the αs and βs to sum to zero, a property borrowed from their estimates, a_i and b_j. When such constraints are carried over to random effects, however, inconsistencies immediately arise. Thus Kempthorne and Folks (1971) describe the mixed model for the two-way classification with interactions (A fixed and B random) in which the interaction effects γ_{ij} are first defined as independent $N(0, \sigma^2)$ variables and then have their B margin $\gamma_{.j}$ constrained to be zero. These mutually inconsistent properties are confusing to the beginner and indeed to the expert as well. They lead to unnecessary complexity in the rules for deriving expected mean squares (see e.g. Bennett and Franklin, 1954). They also lead to unrealistic hypotheses, already touched on in Section 2.1, whereby the variation in a margin is hypothesized to be zero though real interactions are supposed to be present in the body of the table. The theory we shall develop contains no such constraints, leads to simpler rules for expected mean squares and to realistic hypotheses rather than unrealistic ones.

2.5. *Negative Components of Variance*

With infinite populations the model for a nested classification A/B implies that the expected mean square (EMS) for A cannot be less than that for $A.B$ (B within A). However, with randomization models based on finite populations this is not so. If A/B is now the block structure of a randomized block experiment (A = blocks, B = plots within blocks) and we unfortunately lay down the blocks to lie along a trend instead of across it, then the EMS for A under randomization will be less than that for $A.B$. The notation developed for infinite populations is inadequate to express this, except by allowing a negative component of variance σ^2_A, which many find unhelpful. However σ^2_A, as a variance component, is in fact not a variance but an *excess of variance* in the A margin of the $A.B$ table over that internal to the table, and this "excess" can be negative even in the population, implying simply that the margin is less variable than would be expected from the internal variation. Components of excess of variance have been called Σ-quantities by Wilk and Kempthorne (1956), f-quantities by Nelder (1965) and canonical components by Fairfield Smith (1955) in his unjustly neglected monograph. We shall develop a notation *ab initio* which allows these quantities to appear in a form which does not exclude their being negative.

3. THE REFORMULATION

3.1. *Basic Scope*

In reformulating linear models we shall restrict the development considerably in order not to obscure the basic points by algebraic complexity. We shall restrict ourselves to first- and

second-order moments only and deal only with balanced structures having equal numbers in all sub-classes. We distinguish observational data (as collected in surveys) from experimental data (as collected in experiments). The investigator may choose how to clas sify observational data, i.e. define the factors and their levels, but he has no control over the value of a factor level for any particular observational unit. The factors and levels chosen will be used both to characterize the population and to draw samples from it. When, however, experiments are done on a set of units, treatment factors and levels are defined by the investigator, who can now control the values of these factor levels for each unit. There may exist, of course, prior classifications by observational-type factors (block factors), and balancing of the treatment-factor levels with respect to the block-factor levels can lead to very complex sampling schemes. The form of two such sampling schemes for finite populations will be discussed in Sections 7.1 and 7.2. We discuss first models for observational data with simple nested and crossed structures, and begin by describing the underlying populations.

4. Simple Nested Structure

4.1. *The Analysis of Populations*

We are concerned with the structure A/B where the levels of A index a set of sub-populations and those of B the elements within a sub-population. We shall be concerned with those characteristics of the sub-populations which do not depend upon knowing which particular level number has been assigned to which sub-population. The assumption is that if some subset of the sub-populations, or of elements within them, requires to be distinguished a new factor would be set up to make the distinction. Both the set of sub-populations, and of elements within a sub-population, can then be treated as defining a probability distribution. If either set is finite then the members are treated as equi-probable. An infinite set is assumed to have a probability distribution of the scale-and-location type, i.e. $f\{(x-\mu)/\sigma\}$, with finite variance. We shall assume that the scale parameter associated with a B sub-population does not depend on the level of A (though of course the location parameter may).

Write x_{ij} for an element taken at random from the structure. If both A and B have finite numbers of levels then i is taken at random from $1, 2, \ldots, N_A$ and j from $1, 2, \ldots, N_B$. The mean of the ith sub-population is $x_{i.} = E_j(x_{ij})$ and the grand mean is $x_{..} = E_{ij}(x_{ij})$. The suffices of the expectation operators denote the extent of the averaging.

4.1.1. *The population linear identity*

For the A/B structure this takes the form

$$x_{ij} \equiv x_{..} + (x_{i.} - x_{..}) + (x_{ij} - x_{i.}) \tag{4.1}$$

and expresses an element as the sum of three components, the grand mean, the deviation of the ith population mean from the grand mean and the deviation of the element from the ith population mean. For finite N_A and N_B the population linear identity can also be regarded as a linear transformation of the $N_A N_B$ elements to three sets of *contrasts* with numbers in each set

$$1, \quad (N_A - 1) \quad \text{and} \quad N_A(N_B - 1)$$

respectively. (We include $x_{..}$, conventionally, as a contrast to complete the set of linear combinations.) We define the *variance components* of the structure by

$$\left.\begin{aligned} \sigma_A^2 &= \frac{\sum_i (x_{i.} - x_{..})^2}{N_A - 1}, \\ \sigma_{AB}^2 &= \frac{\sum_{ij} (x_{ij} - x_{i.})^2}{N_A(N_B - 1)}, \end{aligned}\right\} \tag{4.2}$$

when N_A and N_B are finite, replacing averaging by integration over the corresponding suffix when either becomes infinite. Note that $x_{..}$ is by definition a constant and so generates no variance component.

4.1.2. *The population quadratic identity*

This is derived by squaring the linear identity and taking expectations to give

$$E_{ij}(x_{ij}-x_{..})^2 = E_{ij}(x_{i.}-x_{..})^2+E_{ij}(x_{ij}-x_{i.})^2. \tag{4.3}$$

For finite N_A and N_B the r.h.s. becomes, in terms of the variance components,

$$\text{var}\,(x_{ij}) \equiv \sigma_A^2\left(1-\frac{1}{N_A}\right)+\sigma_{AB}^2\left(1-\frac{1}{N_B}\right). \tag{4.4}$$

(Daniels (1939) argued that the variance components should be defined with divisors N_A and $N_A N_B$ respectively, so that (4.4) becomes

$$\text{var}\,(x_{ij}) = \sigma_A^2+\sigma_{AB}^2.$$

The choice is to some extent a matter of convention and convenience. We shall argue below that the fundamental quadratic quantities are the canonical components rather than the variance components, and that for this development the definitions given are more convenient.)

We now develop two alternative representations of the second-order statistics of the structure, and then derive the relationships between all three.

4.1.3. *The covariance structure*

Two elements, x_{ij} and $x_{i'j'}$, from the structure A/B may have one of three possible relations: either they are identical ($i=i', j=j'$), or they belong to the same A sub-population but are not identical ($i=i', j\neq j'$), or they belong to different A sub-populations ($i\neq i'$). There are therefore just three distinct covariances between elements, which we write in the form

$$\text{cov}\,(x_{ij}, x_{ij}) = \text{var}\,(x_{ij}) = \rho_{12}\sigma^2, \quad \text{cov}\,(x_{ij}, x_{ij'}) = \rho_1\sigma^2, \quad \text{cov}\,(x_{ij}, x_{i'j'}) = \rho\sigma^2. \tag{4.5}$$

4.1.4. *The population linear model*

The first-order statistics of the population are summed up by the $x_{i.}$, the sub-population means, and $x_{..}$, the grand mean, and any expectations of linear combinations of the x_{ij}s can be simply expressed in terms of them. We set up a population linear model to provide an analogous facility for quadratic combinations. We express x_{ij} in the form

$$x_{ij} = \mu+\alpha_i+\beta_{ij}, \tag{4.6}$$

where $\mu, \alpha_i, \beta_{ij}$ are all random variables in a certain formal sense to be explained. It is assumed that the components on the r.h.s. are all uncorrelated (e.g. not only μ and α_i but also α_i and $\alpha_{i'}$, etc). The components μ, α, β of any pair of elements x_{ij} and $x_{ij'}$ are then either identical or uncorrelated. It is easy to show that if we assign to the components the following "variances" then the covariance structure (4.5) is reproduced:

$$\phi = \text{var}\,\mu = \rho\sigma^2; \quad \phi_A = \text{var}\,\alpha_i = (\rho_1-\rho)\sigma^2; \quad \phi_{AB} = \text{var}\,\beta_{ij} = (\rho_{12}-\rho_1)\sigma^2. \tag{4.7}$$

The ϕ quantities will be termed *canonical components*, and their interpretation will be discussed further below. First we establish relations between the three forms of the second-order statistics, namely the variance components (4.2), the covariances (4.5) and the canonical components (4.7).

4.1.5. *Second-order statistics; the three forms*

The derivation is most easily done for finite N_A and N_B, results for infinite populations following by letting $N\to\infty$. Consider first the relation between $\phi = \rho\sigma^2$ and the variance

components. Now

$$\rho\sigma^2 = \mathrm{cov}(x_{ij}, x_{i'j'})$$
$$= \sum_{i\neq i'}\sum_{j}\sum_{j'}(x_{ij}-x_{..})(x_{i'j'}-x_{..})/\{N_A(N_A-1)N_B^2\}.$$

If the summation is written in the form

$$\sum_{i}\sum_{i'}\sum_{j}\sum_{j'} - \sum_{i=i'}\sum_{j}\sum_{j'}$$

the first term vanishes identically and the second gives

$$-\sum_{i}(x_{i.}-x_{..})^2/\{N_A(N_A-1)\}.$$

Thus $$\phi = \rho\sigma^2 = -\sigma_A^2/N_A.$$

Similarly we find $$\phi_A = (\rho_1-\rho)\sigma^2 = \sigma_A^2 - \frac{\sigma_{AB}^2}{N_B}, \quad \phi_{AB} = (\rho_{12}-\rho_1)\sigma^2 = \sigma_{AB}^2. \qquad (4.8)$$

Note also the relations

$$\rho\sigma^2 = \phi; \quad \rho_1\sigma^2 = \phi+\phi_A; \quad \rho_{12}\sigma^2 = \phi+\phi_A+\phi_{AB}. \qquad (4.9)$$

4.1.6. *The canonical components*

The canonical components are the formal variances of the components of the population linear model, formal because $\mathrm{var}\,\mu$ is necessarily non-positive and $\mathrm{var}\,\alpha_i$ will also be so if $\sigma_A^2 < \sigma_{AB}^2/N_B$. However any calculation of (co)variances of linear functions of the xs, or of expectations of quadratic functions is facilitated by using these formal quantities. The ϕs will coincide with the variance components if N_A and N_B both tend to infinity, when we recover the standard infinite model with

$$\mathrm{var}\,\mu = 0, \quad \mathrm{var}\,\alpha_i = \sigma_A^2, \quad \mathrm{var}\,\beta_{ij} = \sigma_{AB}^2.$$

The canonical components can be interpreted either via the covariance representation (4.5), or via the variance representation (4.2). In covariance terms, $\mathrm{var}\,\mu$ is the basic covariance of "unrelated" xs, $\mathrm{var}\,\alpha_i$ is the excess covariance of elements in the same A sub-population over the basic, and $\mathrm{var}\,\beta_{ij}$ is the excess of the covariance of identical elements over that of different elements in the same sub-population. In terms of variances, $\mathrm{var}\,\beta_{ij}$ is the variance within A sub-populations, $\mathrm{var}\,\alpha_i$ is the excess variance in the means of the A sub-populations over the corresponding variance within those sub-populations, and $\mathrm{var}\,\mu$ can be interpreted in the form

$$E(\mu^2) = x_{..}^2 - \sigma_A^2/N_A,$$

where the r.h.s. gives the excess of the square of the grand mean over the variance derived from the A sub-populations. The term "canonical component", adopted from Fairfield Smith (1955), is thus a convenient abbreviation for "canonical component of excess variation".

4.1.7. *The difference between the linear identity and the linear model*

The similarity of the structures of the linear identity and the linear model makes it tempting to equate corresponding terms. This temptation should be resisted, because, for example, $x_{..}$ in the linear identity is not the same as μ in the linear model. In fact

$$\left.\begin{aligned} x_{..} &= \mu+\alpha_.+\beta_{..}, \\ x_{i.}-x_{..} &= \alpha_i-\alpha_.+\beta_{i.}-\beta_{..}, \\ x_{ij}-x_{i.} &= \beta_{ij}-\beta_{i.}, \end{aligned}\right\} \qquad (4.10)$$

so that only the last terms are similar. The others can be made similar by constraining $\alpha_.$ and $\beta_{i.}$ to be zero, a procedure often followed. However, such a procedure induces correlations between the αs and βs when the populations are finite, and so spoils the simplicity of the linear model. Without constraints there are, of course, more parameters than elements, and the nature of the subsequent non-uniqueness of the parameters is seen in equations (4.10). The set $(\mu, \alpha_., \beta_.)$ and the sets $(\alpha_i, \beta_{i.})$ are *intrinsically aliased*; aliased, because there is no way in which the contributions of the elements within a set can be distinguished and intrinsically aliased, because the aliasing is independent of the way in which the elements may be sampled to produce estimates of the parameters. Intrinsic aliasing is closely analogous to non-identifiability in econometric models.

4.2. *The Analysis of Samples*

Suppose we generate a sample of values y_{ij} from our A/B structure, where $i = 1, \ldots, n_A$ and $j = 1, \ldots, n_B$. If N_A and N_B are finite this amounts to choosing a sample of n_A levels of A and n_B levels of B at random without replacement. If either becomes infinite it amounts to choosing a sample of n from the appropriate distribution. Note that $n \leqslant N$, with equality implying that the sample is complete with respect to that factor.

4.2.1. *The linear and quadratic identities*

The sample gives rise to linear and quadratic identities

$$y_{ij} = y_{..} + (y_{i.} - y_{..}) + (y_{ij} - y_{i.}), \tag{4.11}$$

$$\sum_{ij} y_{ij}^2 = \sum_{ij} y_{..}^2 + \sum_{ij} (y_{i.} - y_{..})^2 + \sum_{ij} (y_{ij} - y_{i.})^2, \tag{4.12}$$

where dots denote averaging over the sample, which now takes the place of expectations.

4.2.2. *The analysis of variance*

The sample quadratic identity is of course the sample ANOVA, and we can derive the expected mean squares (EMS) of its components by substituting the population linear model and using (4.7). Thus, for example,

$$\begin{aligned} E_{ij} \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} (y_{i.} - y_{..})^2 &= E_{ij} \sum_i \sum_j (\mu + \alpha_i + \beta_{i.} - \mu - \alpha_. - \beta_.)^2 \\ &= n_B E_i \sum_i \{(\alpha_i - \alpha_.)^2 + (\beta_{i.} - \beta_.)^2\} \\ &= n_B[(n_A - 1)\phi_A + (n_A - 1)\phi_{AB}/n_B] \\ &= (n_A - 1)[\phi_{AB} + n_B \phi_A]. \end{aligned}$$

This derivation illustrates the usefulness of assuming that the α_i and β_{ij} are all uncorrelated random variables. The resulting table is thus

Component	*d.f.*	*EMS*
Mean	1	$\phi_{AB} + n_B \phi_A + n_A n_B(\phi + \mu^2)$
A	$n_A - 1$	$\phi_{AB} + n_B \phi_A$
$A.B$	$n_A(n_B - 1)$	ϕ_{AB}

The EMS show the central role of the ϕs. Note the separation of the ϕs, which are population quantities, from the multipliers, which are sample quantities; also that the EMS for the mean and A contain not only the canonical component corresponding to the model term in question

but also *all* those corresponding to terms to which they are marginal. This reflects the fact that $y_{..}$ is $\mu+\alpha_.+\beta_.$ and $y_{i.}-y_{..}$ is $\alpha_i-\alpha_.+\beta_{i.}-\beta_.$ (averaging now being over the sample). As noted above, ϕ_A appears as the expected *excess* variation between A means over that between B within A; however, because ϕ_A can be negative, this excess can be a deficit, indicating that variation between levels of A is less than would be predicted from variation between levels of B within levels of A. Such deficits are not just a theoretical oddity, but represent actual occurrences. Consider, for example, litters produced by females of a genetically uniform inbred line. Competition for food may generate negative correlations within litters and may give $\rho_1<\rho$, i.e. $\phi_A<0$. Note that while we might, for the theoretical population, make $N_A\to\infty$, i.e. envisage a hypothetically infinite population of females, it would be pointless to make $N_B\to\infty$, i.e. to regard a litter as a sample from a hypothetically infinite litter. Thus an infinite population for the elements **x** does *not* imply an infinite population of levels for all the factors in the structure. Note the two special cases for the EMS (see (4.8)):

(i) If $N_B\to\infty$, with ϕ_{AB} remaining finite, ϕ_A becomes the variance component σ^2_A, so that the excess variation, as measured by the canonical component, must be non-negative.

(ii) If $n_B = N_B$, i.e. B corresponds to a fixed effect, then $\phi_{AB}+N_B\phi_A = N_B\sigma^2_A$ and it is common in many accounts to consider the hypothesis that σ_A is zero, which involves comparing the mean square for A not with $A.B$, but with some measure of replicate error. The present approach, in which the mean square for A is compared with that for $A.B$ is a test not for σ^2_A, but for ϕ_A, and answers the question "are there consistent effects of A over and above the variation described by ϕ_{AB}?"

Note that if $A.B$ is postulated to exist then A effects will exist in consequence. It is thus pointless to formulate the hypothesis that A is zero; the relevant question is "does the variation in A exceed that expected from that of $A.B$?" and this requires the comparison of the A mean square with that for $A.B$.

4.3. *Models with unknown Ns*

Sometimes the numbers of levels N_A and N_B in the population are clearly defined, e.g. as would be so if litter size in the above example was fixed (clutch size in pigeons, which is always 2, would be a better example). Consider, however, a field experiment laid out in randomized blocks; the definition of units in terms of blocks and plots is to a large extent arbitrary and the values assigned to N_A and N_B are artifacts of the particular layout adopted. In such circumstances it is better to define the ϕs in terms of the (co)variances σ^2, ρ_1 and ρ. The previous arguments are otherwise unaffected, and the ϕs remain the central quantities.

5. Simple Crossed Structures

The structure is now $A*B$, where the levels of A index a *row* population and those of B a *column* population. Elements are defined for each intersection of a row and column.

5.1. *The Analysis of Populations*

Both rows and columns define probability distributions as before, and these may be finite or infinite. The population of elements in a row (column) has its mean dependent on the row (column) index. The scale parameter is assumed not to depend on the indexing. We define σ^2_A and σ^2_B as the variance components of rows and columns respectively and σ^2_{AB} as the corresponding interaction component. Again we write x_{ij} for an element taken at random from the structure.

5.2. *The Linear and Quadratic Identities*

These are

$$x_{ij}\equiv x_{..}+(x_{i.}-x_{..})+(x_{.j}-x_{..})+(x_{ij}-x_{i.}-x_{.j}+x_{..}) \tag{5.1}$$

and

$$Ex_{ij}^2 \equiv Ex_{..}^2 + E(x_{i.}-x_{..})^2 + E(x_{.j}-x_{..})^2 + E(x_{ij}-x_{i.}-x_{.j}+x_{..})^2 \tag{5.2}$$

which for finite N_A and N_B can be written

$$\text{var}(x_{ij}) = \sigma_A^2\left(1-\frac{1}{N_A}\right)+\sigma_B^2\left(1-\frac{1}{N_B}\right)+\sigma_{AB}^2\left(1-\frac{1}{N_A}\right)\left(1-\frac{1}{N_B}\right). \tag{5.3}$$

5.3. *The Covariance Structure*

There are now four distinct kinds of pairs of elements; identical, non-identical in same row, non-identical in same column, not in same row or column. The four covariances will be written:

$$\text{cov}(x_{ij}, x_{ij}) = \text{var}(x_{ij}) = \rho_{12}\sigma^2, \quad \text{cov}(x_{ij}, x_{i'j}) = \rho_2\sigma^2$$
$$\text{cov}(x_{ij}, x_{ij'}) = \rho_1\sigma^2, \quad \text{cov}(x_{ij}, x_{i'j'}) = \rho\sigma^2.$$

The four canonical components are related to the variance components by

$$\phi = -\frac{\sigma_A^2}{N_A}-\frac{\sigma_B^2}{N_B}+\frac{\sigma_{AB}^2}{N_AN_B}, \qquad \phi_B = \sigma_B^2-\frac{\sigma_{AB}^2}{N_A},$$

$$\phi_A = \sigma_A^2-\frac{\sigma_{AB}^2}{N_B}, \qquad \phi_{AB} = \sigma_{AB}^2$$

and to the covariances by

$$\rho_{12}\sigma^2 = \phi+\phi_A+\phi_B+\phi_{AB}, \quad \rho_2\sigma^2 = \phi+\phi_B,$$
$$\rho_1\sigma^2 = \phi+\phi_A, \qquad \rho\sigma^2 = \phi.$$

5.4. *The Population Linear Model*

This takes the form

$$x_{ij} = \mu+\alpha_i+\beta_j+(\alpha\beta)_{ij}$$

with the components independent and assigned variances

$$\text{var}\,\mu = \rho\sigma^2 = \phi, \qquad \text{var}\,\beta_j = (\rho_2-\rho)\sigma^2 = \phi_B,$$
$$\text{var}\,\alpha_i = (\rho_1-\rho)\sigma^2 = \phi_A, \qquad \text{var}\,(\alpha\beta)_{ij} = (\rho_{12}-\rho_1-\rho_2+\rho)\sigma^2 = \phi_{AB}.$$

Thus $\text{var}\,(\alpha\beta)_{ij}$ is the interaction variance, $\text{var}\,\beta_j$ is the excess variance in the B margin over the interaction variance, $\text{var}\,\alpha_i$ similarly for the A margin, while $\text{var}\,\mu+x_{..}^2$ gives the excess of the square of the grand mean over the sum of the interaction variance and the two marginal excess variances. Again the terms in the population linear model are *not* to be equated to those in the population linear identity.

5.5. *The Analysis of Samples*

If we generate a sample of values, y_{ij}, from an $A*B$ structure, with $i = 1,\ldots,n_A$ and $j = 1,\ldots,n_B$, then we have analogous linear and quadratic identities and expected mean squares as follows:

Component	*d.f.*	*EMS*
Mean	1	$\phi_{AB}+n_B\phi_A+n_A\phi_B+n_An_B(\phi+\mu^2)$
A	n_A-1	$\phi_{AB}+n_B\phi_A$
B	n_B-1	$\phi_{AB}+n_A\phi_B$
$A.B$	$(n_A-1)(n_B-1)$	ϕ_{AB}

Again each EMS contains the ϕ with corresponding suffices, and also contributions from ϕs of all terms to which that term is marginal. The forms of the EMS hold for finite or infinite populations and for complete or incomplete sampling, i.e. for fixed or random effects. Compare this uniformity with the diversity produced by a commonly used parametrization. The following table is adapted from Tables 16.9, 16.10 and 16.11 of Kempthorne and Folks (1971):

	EMS			
Component	*A fixed B fixed*	*A fixed B random*	*A random B fixed*	*A random B random*
A	$n_B \sum_A$	$\sigma^2_{AB}+n_B \sum_A$	$n_B \sigma^2_A$	$\sigma^2_{AB}+n_B \sigma^2_A$
B	$n_A \sum_B$	$n_A \sigma^2_B$	$\sigma^2_{AB}+n_A \sum_B$	$\sigma^2_{AB}+n_A \sigma^2_B$
$A.B$	$\sum_{AB}$	σ^2_{AB}	σ^2_{AB}	σ^2_{AB}

In this table

$$\sum_A = \sum a_i^2/(n_A-1),\ \sum_B = \sum b_j^2/(n_B-1),\ \sum_{AB} = \sum (ab)_{ij}^2/(n_A-1)(n_B-1)$$

and the σ's are variance components.

The apparent differences in the forms of the EMS according as the factors are treated as fixed or random is a source of considerable confusion to students. These differences arise from the imposition of constraints on the terms in the linear model which have the effect of shifting variation from one term to another. The above exposition shows that, properly parametrized, the EMS are completely regular; furthermore, because the ϕs above the lowest level (i.e. ϕ, ϕ_A and ϕ_B) can be negative with finite populations, implying margins of less than expected variability, hypotheses that this is so can be stated naturally, something that is impossible when the standard notation of variance components is used. This development also exposes clearly the marginal relations between the main effects and their interaction, and shows that if the interaction is assumed to be present then the only interesting question to be asked about the marginal differences is whether they are larger (or smaller) than that predicted by the interaction, not whether they are larger than zero.

6. Estimation Procedures

The confusions and inconsistencies that can arise from defining linear models with constraints on the parameters have been discussed above. It remains to present the fitting of linear models without constraints in a way that makes the process simple and straightforward. The idea of sequential fitting of parameters is vital; although the mathematical formulae can be written down as though the fitting of a model yields simultaneous estimates, an algorithm for the process will in fact proceed parameter by parameter. Consider, for example, the Gauss–Jordan method of elimination of the SSP-matrix; for each parameter we sweep out the corresponding row-column of the matrix. Or rather we do so *if there is any information about the parameter*, given the other parameters already in the model. If there is no information, the pivot (diagonal element) will be zero (we ignore problems of rounding error); in the absence of information we have no alternative but to omit the term involving this parameter from the model. Omission is algebraically equivalent to setting the parameter value to zero. Absence of information is equivalent to aliasing of the parameter with some linear combination of parameters already fitted. The important point is that the residual S.S., the fitted values and the estimators of any estimable contrast are invariant with respect to the order of fitting of the parameters. It is useful to distinguish two kinds of aliasing, extrinsic and intrinsic. Intrinsic aliasing occurs, for example, when we consider the fitting of the interaction $A.B$ after the fitting of main effects A and B. As pointed out in Section 2.1, the $(n_A n_B-1)$-dimensional space of the contrasts in the two-way array of effects $(\alpha\beta)_{ij}$ contains as sub-spaces the (n_A-1) and

(n_B-1)-dimensional spaces of contrasts for the main effects α_i and β_j respectively, and this inclusion of the sub-spaces in the $A.B$ space is independent of the design matrix, so that the corresponding aliasing will *always* occur. Extrinsic aliasing is a function of the design matrix; thus for the unbalanced design consisting of the three points marked with a cross in the diagram

		B		
		1	2	3
	1	×		
A	2			×
	3		×	

$\alpha_1-\alpha_2$ is aliased with $\beta_1-\beta_3$ and $\alpha_1+\alpha_2-2\alpha_3$ with $\beta_1-2\beta_2+\beta_3$. Implementations of algorithms for the sequential fitting of linear models need not provide a row-column in the SSP matrix for intrinsically aliased parameters or comment on their existence, which can be deduced from the form of the model without knowing the design matrix. Extrinsically aliased parameters need to be pointed out as a warning to the user. Note that the arguments about aliasing carry over unchanged to generalized linear models (Nelder and Wedderburn, 1972), because the pattern of aliasing is a function solely of the design matrix and the model for the linear predictor, and does not depend on the error distribution or the link function. It thus applies, in particular, to log-linear models with Poisson errors, and logit models with Binomial errors.

7. Linear Models for Randomized Experiments

We now turn from linear models for observational data to consider the corresponding models for designed experiments. The general approach will be that of Nelder (1965a,b) whereby a treatment structure is imposed on an existing block structure. We shall not assume unit-treatment additivity, however, and will restrict discussion to the two simplest designs, the completely randomized and randomized block experiment.

7.1. *Completely Randomized Experiment*

Experiments imply a set of experimental units (plots, animals, plants, etc.) to which are applied a set of *treatments*. Consider the simplest type of experiment in which t treatments are applied to nt units, each treatment occurring n times. We consider a population consisting of the conceptual nt^2 yields obtainable from applying any of the treatments to any of the units. Unlike populations of survey data, which may be, at least in theory, wholly ascertainable, such populations for experimental data are necessarily non-ascertainable, and only the sample population of items is available. We now consider the aliasing that follows from this.

The diagram shows the theoretical 6×3 table of yields for 6 plots and 3 treatments, where an actual experiment produces yields in the starred positions:

		Treatment		
		1	2	3
	1	× *	×	×
	2	×	×	× *
Plot	3	×	× *	×
	4	×	×	× *
	5	× *	×	×
	6	×	× *	×

There must be, of course, one sample point per row and two per column. The population analysis of variance is that of a treatment × plot structure, and we indicate the presence of a

canonical component in the EMS by a cross.

Component	*d.f.*	*Canonical components in EMS* *P*	*T*	*PT*
Mean	1	×	×	×
P	$nt-1$	×		×
T	$t-1$		×	×
P.T	$(nt-1)(t-1)$			×

The population linear model can be written as

$$\mu+\alpha_i+\beta_j+\gamma_{ij}$$

and contrasts between the nt^2 elements γ_{ij} of *PT* appear in all four EMS; $(t-1)$ of these contrasts are aliased with those of *T*, $(nt-1)$ contrasts are aliased with those of *P* and one "contrast" with the mean term. All aliasing is intrinsic.

Consider now the sample analysis of variance. Further (extrinsic) aliasing appears, caused by the incompleteness of the $P\times T$ table. In this table a fraction $1/t$ only of the cells is occupied, so that the *P.T* "margin" (actually the whole table) is *incomplete*. By contrast the *P* and *T* margins are *complete* because sample values occur in each row and column. However the sets of contrasts measuring *P* and *T* are partially aliased; for the example in the diagram $\alpha_1-\alpha_2-\alpha_4+\alpha_5$ is aliased with $\beta_1-\beta_3$. Generally $(t-1)$ d.f. of the plot contrasts are aliased with treatment contrasts. If a margin is complete the corresponding canonical component will occur in one or more of the EMS, and the total degrees of freedom for these MS will equal the number of parameters of the term in the population linear model. Thus since the sample ANOVA has *nt* d.f. in all, the component ϕ_P must occur in all the EMS.

The incomplete table *P.T* is now identical with its *P* margin and hence the corresponding contrasts are wholly aliased with each other. In agreement with this the total d.f. for the EMS of terms containing the component ϕ_{PT} is now $1/t$ of nt^2, i.e. *nt*. Thus the canonical components occurring in the sample ANOVA are as follows, where starred components denote those arising from extrinsic aliasing.

Component	*d.f.*	*Canonical components in EMS* *P*	*T*	*PT*
Mean	1	×	×	×
T	$t-1$	×*	×	×
P.T (≡*P*)	$(n-1)t$	×*		×

The identifiable quantities are thus ϕ_T and $\phi_P+\phi_{PT}$.

7.2. *The Randomized Block Experiment*

Let there be *bt* plots arranged in *b* blocks of *t* plots with *t* treatments. The population has bt^2 values with structure $(B/P)*T$, the factors being B = blocks, P = plots (within blocks) and T = treatments. The *population* ANOVA, with the canonical components occurring in the EMS for each line, thus has the form:

		Canonical components				
Component	*d.f.*	*B*	*BP*	*T*	*BT*	*BPT*
(1)	1	×	×	×	×	×
B	$b-1$	×	×		×	×
B.P	$b(t-1)$		×			×
T	$t-1$			×	×	×
B.T	$(b-1)(t-1)$				×	×
B.P.T	$b(t-1)^2$					×

For the form of the sample ANOVA, consider first a simple example with $b = 2$ and $t = 3$ and a layout given by

		T		
B	*P*	1	2	3
1	1	×*	×	×
1	2	×	×	×*
1	3	×	×*	×
2	1	×	×	×*
2	2	×*	×	×
2	3	×	×*	×

The *sample* is again a $1/t$ replicate. All one-way and two-way margins are full, hence all contrasts of the corresponding terms occur in full in the EMS. The three-way "margin" is present, however, in fraction $1/t$ and the three margins *BPT*, *BP* and *BT* are identical. The canonical components are thus present in the sample ANOVA with the following pattern, where a starred entry denotes occurrence by virtue of extrinsic aliasing.

		Canonical components				
Component	*d.f.*	*B*	*BP*	*T*	*BT*	*BPT*
(1)	1	×	×	×	×	×
B	$b-1$	×	×		×	×
T	$t-1$		×*	×	×	×
B.P.T	$(b-1)(t-1)$		×*		×*	×

The three identifiable quantities among the canonical components are thus ϕ_B, ϕ_T, and $\phi_{BP}+\phi_{BT}+\phi_{BPT}$. Valid randomization means making the *B.P* contrasts in the *T* line a random sample of all *B.P* contrasts (Grundy and Healy, 1950).

7.3. *Conditions for a Valid Error*

If the strong assumption of unit-treatment additivity is made, then $\phi_{BT} = \phi_{BPT} = 0$. The *B.P* contrasts remain and these are homogeneous w.r.t. treatments by definition. However, although a valid randomization procedure involves making the *B.P* contrasts which become the treatment contrasts a random sample of all contrasts, there remains the possibility that the *total* set of *B.P* contrasts does not look like a sample from a Normal distribution. We return to this point later.

If unit-treatment additivity does not hold, we need to consider carefully what null hypothesis to entertain about the treatment effects. Neyman (1935) and Wilk and Kempthorne

(1956) both considered the hypothesis that the mean treatment difference over all plots was zero, and both concluded that for Latin squares the standard form of variance would be biased for testing this hypothesis. Cox (1958) pointed out that such a hypothesis was unlikely to be of interest, and that for a more realistic null hypothesis the usual error was satisfactory. In our terminology the Neyman hypothesis is equivalent to assuming the presence of interaction in the unit-treatment table with the absence of effect in one of the margins. This is thus another example of the class of hypothesis that I assert to be without practical interest. Cox's hypothesis is equivalent to saying that the treatment means do not vary by more than the variation implied by the interaction, i.e. that the interaction variation rather than zero provides the baseline from which to measure the treatment effects. If the treatment differences were small compared to the interaction variation, then rather than regard them as little greater than zero, we should be surprised that they are so much *less* than the interaction variation.

However, even if it is accepted that the usual error line is not biased on the average there remains a problem of homogeneity. As Yates (1967) has pointed out, once an interaction with many d.f. has been found to be appreciable, doubts must immediately arise about its homogeneity. If it is heterogeneous then it cannot act as a suitable denominator for testing all marginal contrasts, but needs to be split into homogeneous pieces. When we are dealing with an interaction term which is to act as an error, the problem is to specify what we mean by homogeneity. If we consider the $P \times T$ table then each column gives the (hypothetical) yields for a treatment applied to all plots. The columns might be, among many other possibilities, (1) identical except for origin shifts, (2) independent samples from the same distribution or (3) some mixture of (1) and (2). Unit-treatment additivity is implied by (1), but all would produce homogeneity at least in so far that any two treatments would produce exchangeable sets of values as regards the pattern about the mean. In the context of finite populations it is not obvious how this definition can be made rigorous; attempts to make it so give rise to fascinating problems, which cannot be pursued here.

7.4. *More Complex Designs*

All experimental designs can be regarded as producing a fraction of the non-ascertainable population of yields classified by units × treatments. Nelder (1965b) contains a development for generally balanced designs with unit-treatment additivity. The more general approach developed here for the two simplest designs, whereby the intrinsic aliasing inherent in the population linear model is augmented by further extrinsic aliasing caused by the particular experimental sampling scheme which constitutes the design, can be extended without difficulty to more complex balanced designs with one or more errors. Also the restricted randomization schemes of Grundy and Healy (1950), whereby contrasts of treatments are applied to contrasts of units, rather than individual treatments to individual units, cause no extra difficulty.

8. CONCLUSION

The objections to existing formulations of linear models made in the course of this reformulation are not that they are mathematically incorrect; the objections fall into two categories, notational deficiencies and the generation of unsatisfactory hypotheses. The reformulation has been designed to integrate finite and infinite populations, random and fixed effects, and excess and deficit of variance, to avoid the unnecessary introduction of constraints on parameters, and to lead naturally to interesting hypotheses about the terms in a model. On its success in doing this the reformulation should be judged.

ACKNOWLEDGEMENTS

It is a pleasure to record the clarification I have received from many helpful discussions with Mr G. N. Wilkinson; he was co-author of an earlier unpublished paper from which this one has developed. In particular I owe the word "marginality" to him and also the idea that a transfer of variance results from the imposition of constraints.

I have also benefited from the comments of Dr F. Yates, Professor D. R. Cox, Mr. M. J. R. Healy (who made detailed criticisms of early drafts), Mrs. Pamela Morse, Professor Murray Aitkin and numerous colleagues at Rothamsted.

References

BENNETT, C. A. and FRANKLIN, N. L. (1954). *Statistical Analysis in Chemistry and the Chemical Industry.* New York: Wiley.

CORNFIELD, J. and TUKEY, J. W. (1956). Average values of mean squares in factorials. *Ann. Math. Statist.*, **27**, 907–949.

COX, D. R. (1958). The interpretation of the effects of non-additivity in the Latin square. *Biometrika*, **45**, 69–73.

DANIELS, H. E. (1939). The estimation of components of variance. *J. R. Statist. Soc.* Suppl., **4**, 186–197.

DIXON, W. J. (ed.) (1970). *Biomedical Computer Programs, X-Series Supplement.* Los Angeles: University of California Press.

EDEN, T. and YATES, F. (1933). On the validity of Fisher's z-test when applied to an actual example of non-normal data. *J. Agric. Sci.*, **23**, 6–16.

EISENHART, C. (1947). The assumptions underlying the analysis of variance. *Biometrics*, **3**, 1–21.

FAIRFIELD SMITH, H. (1955). Variance components, finite populations, and experimental inference. *North Carolina Institute of Statistics, Mimeo Series No.* 135.

FRANCIS, I. (1973). Comparison of several analysis of variance programs. *J. Amer. Statist. Ass.*, **68**, 860–865.

GRUNDY, P. M. and HEALY, M. J. R. (1950). Restricted randomization and quasi-latin squares. *J. R. Statist. Soc.* B, **12**, 286–291.

HARTLEY, H. O. and SEARLE, S. R. (1969). A discontinuity in mixed model analyses. *Biometrics*, **25**, 572–576.

KEMPTHORNE, O. and FOLKS, L. (1971). *Probability, Statistics and Data Analysis.* Ames: Iowa State University Press.

NELDER, J. A. (1965a). The analysis of randomized experiments with orthogonal block structure. I. Block structure and the null analysis of variance. *Proc. Roy. Soc.* A, **283**, 147–162.

—— (1965b). The analysis of randomized experiments with orthogonal block structure. II. Treatment structure and the general analysis of variance. *Proc. Roy. Soc.* A, **283**, 163–178.

NELDER, J. A. and WEDDERBURN, R. W. M. (1972). Generalized linear models. *J. R. Statist. Soc.* A, **135**, 370–384.

NEYMAN, J. (1935). Statistical problems in agricultural experimentation. *J. R. Statist. Soc. Suppl.*, **2**, 107–154.

PITMAN, E. J. G. (1938). Significance tests which may be applied to samples from any population. III. The analysis of variance test. *Biometrika*, **29**, 322–335.

SEARLE, S. E. (1971). *Linear Models.* New York: Wiley.

WELCH, B. L. (1937). On the z-test in randomized blocks and Latin squares. *Biometrika*, **29**, 21–52.

WILK, M. B. and KEMPTHORNE, O. (1956). Some aspects of the analysis of factorial experiments in a completely randomized design. *Ann. Math. Statist.*, **27**, 950–985.

WILKINSON, G. N. and ROGERS, C. E. (1973). Symbolic description of factorial models for analysis of variance. *Appl. Statist.*, **22**, 392–399.

YATES, F. (1967). A fresh look at the basic principles of the design and analysis of experiments. *Proc. 5th Berkeley Symp. on Math. Statist. and Prob.*, **4**, 777–790.

Discussion on Dr Nelder's Paper

Mr M. J. R. HEALY (Medical Research Council): It is a particular pleasure to propose a vote of thanks to Dr Nelder for his interesting paper which I first encountered when it was entitled "The Great Mixed Model Muddle". It seems to me to be a paper well in the Rothamsted tradition, in which long practical experience of analysing data has led to the clarification of a situation which had been rendered confused by ill-directed theoretical work. It also may serve as a long-delayed tribute to the late Hugh Fairfield Smith whose unpublished report is another important contribution of the same kind.

As Dr Nelder remarks, it is strange that one of our most basic techniques, one in constant use by the majority of statisticians, should give rise to such confusion. While it could be argued that the users (who seem largely unconcerned with the theoretical complications) are simply too naive to understand fully what they are doing, it would seem likely that serious defects in such a widespread method as commonly applied would have come to light when it led to erroneous conclusions. That this has not occurred to any great extent suggests either that the complexities do not really exist or that they occur only infrequently in practical situations.

The forthright practitioner, impatient with theoretical niceties, is tempted to adopt the first of these alternatives, but I think in doing so he would be wrong. Dr Nelder's development of linear model theory seems to me an excellent one, leading to simple and unified formulations of the main results, but there is one price which not everybody will be willing to pay; this is the radical re-interpretation of the standard linear model (4.6) to include items with possibly negative variances. I have (perhaps I should add "as yet") no experience of teaching this to students, but I can see the potential difficulties. One way out would be to explain that the "independent random variables" in 4.6 are no more than a formal device to simplify the formulae and that their "variances" are just short-hand for the expressions involving correlations in 4.7; but this leaves us with the question of just what it is that we are estimating in the usual analysis. In random models, it may well be the ϕ's, but in fixed models it is usually thought to be the α's and β's.

This usual view is of course wrong unless qualified; as 4.10 shows and as is well known the α's and β's are not well defined unless constrained in some manner. Dr Nelder has stressed the vital point that such constraints are undesirable in theory and he could perhaps have laid more emphasis on the parallel finding that they are unnecessary in practice; I believe C. R. Rao was the first to set out clearly the theory showing that certain properties of a linear model (such as the residual mean square and the estimable contrasts) can be estimated without reference to constraints, and that the latter are only needed to undo the mischief that we have gratuitously introduced by over-parametrizing the model in the first place. Indeed, thinking in terms of constraints can divert attention from some real problems. Suppose I measure serum iron on 1,000 blood donors in order to determine the normal range. Subject to a suitable sampling design, I would have little hesitation in quoting the sample mean as an estimate of the population mean. But now suppose that I cross-classify the donors by age and sex and fit a linear model; in what sense have the sub-group means now become non-estimable?

Apart from the variance component story, Dr Nelder's distinction between complete and incomplete sampling seems to me to cut an immense amount of cackle. But here I must raise my other alternative—if this distinction has remained confused, is it because it almost never matters? The only experiment I can think of with one treatment factor an incomplete sample from a finite population (the only really tricky case) is one reported by Daniels (1939) and mentioned by Fairfield Smith and in the present paper. Dr Nelder gives an entirely realistic example involving clutch size in birds, but it seems to me that this is best considered either as complete sampling (as he seems to imply) or more likely as incomplete sampling from an effectively infinite population of gametes. That inference about finite populations is difficult is evident from the stream of papers on the topic which is still flowing strongly, but I am not sure that the matter under discussion is one in which the difficulties genuinely arise.

I hope that other discussants will take up Section 7 of the paper which in many ways I found the most interesting. It could act as something of a shibboleth to distinguish between those who find the idea of the actual experiment as a fractional replicate with consequent aliasing natural and helpful and those for whom it adds to the obscurity. For myself, I would like to end by congratulating Dr Nelder on an important piece of work, and by proposing that we offer him a hearty vote of thanks.

PROFESSOR R. L. PLACKETT (University of Newcastle-upon-Tyne): An opportunity to discuss linear models is welcome. As John Nelder says, they form a fundamental tool for the practising statistician and so a clear exposition is most important. The developments of the 1950s were reviewed here not long afterwards (Plackett, 1960), and Dr Nelder's contribution to that discussion seems already to contain the seeds of dissatisfaction. My comments on his paper are those of a teacher of statistics who has a professional interest in clear exposition.

Three sorts of exposition are mentioned here: in computer manuals, in statistical textbooks, and in teaching departments. Computer manuals seem generally to be disregarded by the review sections of statistical journals. This is perhaps natural, on the grounds that they appear to be concerned with computers, instructions for the use of. In fact, such manuals provide direct access to computers for those who have data to analyse and who wish, for one reason or another, to short-cut statistical courses and consultants. Computer manuals are the statistical cookbooks of the computer revolution. They must be reviewed with care, in the hope that errors committed in earlier editions will be corrected subsequently.

Textbooks are more influential at all levels of organized teaching. But any teacher of statistics ought to strive for an individual viewpoint, and many will prefer to do so. Neither the elegant perspectives of textbooks nor their regrettable deficiencies are necessarily reproduced in teaching. Therefore the most influential method of exposition now available is that by the teachers themselves.

For an example of what happens at the grass roots, consider the matter of constraints for fixed effects, which Dr Nelder describes as undesirable. The reason why constraints of the form $\sum \alpha_i = 0$ were introduced is that the equations of estimation would otherwise have a singular matrix. This method is perfectly valid, and useful at introductory and intermediate level. It forms a basic part of the classical analysis for balanced incomplete block designs. At an advanced level, some of the necessary equipment is provided by the theory of generalized inverse matrices and methods for computing them. The last sentence of the monograph on this topic by Pringle and Rayner (1971) reads as follows.

"The discomfort of statisticians of an earlier generation who used the method of imposed linear restrictions ... as a sort of *deus ex machina* should now be completely banished ...."

Many teachers of statistics will agree that constraints play a much smaller role than in the past, but their usefulness has not disappeared.

Dr Nelder criticizes the standard approach for notational deficiencies. But such deficiencies are also present in the reformulation. A minor example concerns the meaning of $A.B$, which changes with the circumstances in a way which he grants may seem at first sight disconcerting. The mathematical behaviour is unconventional and some explanation is called for. We find that the notation was devised for the control language of a statistical program. This is perhaps taking the computer revolution too far. But the main results are derived using a standard notation, so there is no need to press the point.

A more serious example arises from the concept of canonical components. Although defined throughout as variances, they can be negative. Teachers of statistics will recall the many occasions in lectures and practical classes when the fact that a variance is non-negative has been stressed: as a consequence of the definition, as a means of showing that $-1 \leqslant \rho \leqslant 1$, and as a proof that some calculation has gone astray. The problem of explaining canonical components is obviously not insuperable in an environment where many amazing ideas find unquestioning acceptance, but there is likely to be confusion nevertheless.

A general formulation of the type described in Sections 4 and 5 is certainly attractive, and merits serious consideration. The normal linear model is equally general, and widely taught, but usually after the ground has been prepared. My preference is to consider fixed effects first, and introduce random effects later. There are two reasons. First, a general model needs a careful introduction, with examples of special cases, and may be more difficult to grasp than simpler models taken separately. In Section 5.5, not all the four combinations from fixed and random effects need arise: two are the same anyway, and one can be reserved for the exercises or examinations. Secondly my assessment of the current state of the art for mixed models in the analysis of variance is as follows. Interesting developments are still in progress, which include the maximum likelihood estimation of variance components (e.g. Patterson and Thompson, 1971) and the use of minimum norm quadratic unbiased estimation (e.g. Rao, 1971). Methods of teaching will be strongly influenced by the fact that these advances are geared to the existing framework.

The material in Section 7 presents a teacher of statistics with problems of a rather different kind. There was indeed a time when linear models formed *the* fundamental tool for the practising statistician, and most experiments within his remit were randomized field experiments. Detailed consideration could have been given to matters as technical as restricted randomization and unit-treatment interactions, except of course that the opportunities for advanced teaching were much smaller then than now. Under present conditions, linear models have to compete for their slice of the syllabus with many other topics of practical importance, some of which have come on remarkably as a result of the computer revolution. They include multivariate analysis, time series, optimization algorithms, response surfaces, and spatial analysis. The expanding frontier is now a serious limiting factor on advanced teaching.

I therefore predict that Dr Nelder's reformulation is unlikely to be taken over as an agreed exposition but that his thoughful criticisms and constructive proposals will affect the teaching of linear models at several crucial points. His paper shows once again that fundamental topics in statistics can, with advantage, be illuminated from a variety of different angles. I have much pleasure in seconding the vote of thanks.

The vote of thanks was passed by acclamation.

Professor M. AITKIN (University of Lancaster): Dr Nelder raises the important question of marginality in his paper. The ignoring of marginality relationships in unbalanced survey cross-classifications can lead to quite incorrect conclusions. The paper by Francis referenced by Dr Nelder gives a good example.

Francis discussed a consulting problem in which an unbalanced 2 × 5 (Sex × Religion) classification was analysed by four different ANOVA programs, all of which gave different answers. These are reproduced as columns (a) to (d) in Table 1, slightly rearranged from Francis' paper.

TABLE 1

Analyses from Francis

Source	*d.f.*	*SS* (*a*)	*SS* (*b*)	*SS* (*c*)	*SS* (*d*)	*SS* (*e*)
Mean	1	Not given	7,305.78	12,982.73	12,982.73	12,982.73
Sex	1	43.58	11.17	43.58	28.71	28.71
Religion	4	69.36	57.74	54.49	69.36	54.49
Sex × Religion	4	−4.61	10.25	10.25	10.25	10.25
Error	1,300	2,988.95	2,988.95	2,988.95	2,989.00	2,988.95

Column (*a*) arises from a computing method which ignored the non-orthogonality, subtracting the unadjusted main effects *SS* from the among cell *SS*. Columns (*c*) and (*d*) are hierarchical analyses, giving repectively sex, religion adjusted for sex and interaction adjusted for sex and religion, and religion, sex adjusted for religion, and interaction adjusted for religion and sex. Column (*b*) arises from a "regression" analysis in which each effect is adjusted for all others in the model: thus sex is adjusted for religion and the interaction, and religion is adjusted for sex and the interaction (and the mean is adjusted for all effects!).

We need not consider column (*a*) further. Columns (*c*) and (*d*) *together* lead to the conclusion that a final model must incorporate both sex and religion terms, but there is no interaction; both main effect mean squares are significant at well beyond the 1 per cent level, regardless of the order in which the effects are fitted.

Column (*b*) leads to the conclusion that only the religion effect is significant, for the sex effect is not significant even at the 2½ per cent level. The adjustment of the sex effect for the interaction, which is not included in the final model, results in a considerable reduction of the sex sum of squares, and the choice of the wrong model.

It is surprising to find Francis strongly recommending method (*b*) as the kind of test that "most people would really want to do..." "because it is precisely what is done in multiple regression in testing whether particular coefficients are zero...". In fact, in choosing a suitable multiple regression model, the *t*-tests on individual coefficients in the full (*p*-variable) model are of very little use, precisely because such tests compare the full model with all (*p*-1) variable models. Unless the final model is itself a (*p*-1) variable model, such a battery of tests may be quite misleading, for it gives no information about the usefulness of the remaining variables when more than one variable is omitted.

A very astute research worker might re-run the analysis omitting the sex × religion interaction. Columns (*c*) and (*d*) remain unchanged except for the deletion of the interaction row, and the pooling of its *SS* and d.f. with error. However, column (*b*) changes quite dramatically, its first three rows becoming those of column (*e*). It will now be found that both sex and religion are necessary. How many research workers would appreciate the necessity of fitting this second model, given the appearance of column (*b*)? Francis clams that "testing in this way [as in column (*b*)] the sex effect... would provide a test of the pure main sex effect, uncompromised and unadulterated by the sex–religion impact". Unfortunately, the reverse is true.

Yet another composite ANOVA table is shown in column (*e*). Here the main effects are adjusted for each other as in column (*b*), but are not adjusted for the interaction. This table can also lead to the wrong conclusion if the sex and religion effects are highly correlated: each adjusted *SS* may be non-significant while each unadjusted *SS* may be highly significant.

The presentation of *both* hierarchical analyses leads to the choice of the appropriate model(s), regardless of the degree of unbalance of the design.

In conclusion, most textbooks warn against interpreting main effect tests when significant interactions are present in balanced factorial designs. In unbalanced designs, the use of ANOVA tables equivalent to t-tests on the regression coefficients is unreasonable on two counts: "averaged" main effects are being tested when interactions may be present, and the non-orthogonality may produce substantial changes in adjusted sums of squares when non-significant high-order interactions are subsequently omitted. The hierarchical analysis, in association with a simultaneous test (Aitkin, 1974) and a small number of permutations of retained terms, ensures that correct final models are chosen.

Dr F. YATES (Rothamsted Experimental Station): I am very glad that Dr Nelder has joined me in urging a reconsideration of the fixed and random effects models of the analysis of variance. As most of you know, I have always regarded these concepts as unnecessary and confusing. I expressed my doubts publicly some ten years ago in a general paper on experimental design. This paper was presented at the Fifth Berkeley Symposium, and a further note on the subject was added when the paper was reprinted in *Experimental Design* (1970).

I fully expected there would be protests at the meeting, but none came. Nor, so far as I am aware, have the issues been argued in print. I did expect to find some discussion in a recent paper by Kempthorne in *Biometrics* entitled "Fixed and random models in the analysis of variance", but he was only concerned to warn his readers against my heresies:

> In spite of highly critical (and I believe unfounded) remarks of Yates [1965-6] (which should, however, be read with care *after* a perusal of the intervening literature), I am of the opinion (and I am not alone) that Eisenhart's paper was of great importance.

Kempthorne followed up this paper with a circular letter, beginning with the erroneous statement that "workers at Rothamsted, including Yates and Nelder, are of the opinion that the ideas on interpretation of the analysis of variance with a mixed two-way table... that have been expounded by essentially everyone who has written on the topic are wrong"—erroneous in that I never said the concepts were wrong, merely that they were unnecessary and confusing and could lead to error. (I quoted an awful example in my Berkeley paper.)

With this letter he enclosed a copy of his paper, and 18 mimeographed pages of turgid algebra on "sampling" a mixed two-factor model. This latter, he said, he inflicts on classes at Iowa State University, but again the issues were not discussed.

In view of all this I am naturally pleased, though not surprised, that Nelder's reformulation confirms my contentions. The canonical components, the ϕs, are merely the variance components of the random effects model, as the tables of Sections 4.2.2 and 5.5 show. Nelder's final conclusion, therefore, must be that this formulation, which is that adopted by Fisher, is appropriate for indicating the correct tests of significance, whether or not the levels of a factor are regarded as a random selection of a larger population of levels.

My only reservation on Nelder's paper is that it gives the impression, at least to the casual reader, that the distinction between fixed, random and mixed effects models is of some importance. My own approach, set out in the Berkeley paper, is more basic. I there wrote:

> What are the facts? The first and crucial point to recognise is that whether the factor levels are a random selection from some defined set (as might be the case with, say, varieties), or are deliberately chosen by the experimenter, does not affect the logical basis of the formal analysis of variance or the derivation of variance components. Once the selection or choice has been made the levels are known, and the two cases are indistinguishable as far as the actual experiment is concerned....
>
> There is an analogy here with the classical problem of determining the error of a linear regression coefficient, in which, it may be remembered, it was sometimes claimed that allowance had to be made for the fact that the observed x were a sample from some population of x, whereas Fisher rightly insisted they could be taken as known.

It is worth remembering that the concepts of random and fixed effects were only introduced by Eisenhart in 1947, by which time the theory of experimental design and analysis had been very fully developed and successfully applied in practice for many years. The concept of components of variance was included by Fisher in *Statistical Methods for Research Workers* (1925)

as an alternative to intraclass correlation; this, as Fisher said, was "a very great simplification". Fisher did not there mention that the second component might be negative, but this must obviously have been apparent to him (and indeed to his more perceptive readers) as he had just been discussing at length the minimum negative value of the intraclass correlation coefficient.

With a negative component Fisher's statement that the values can be regarded as the sum of two independent normal variates no longer holds—instead, in some form or other, deviates from the sample class means are required to obtain the necessary negative correlations. Had Fisher discussed this point, and had the later inappropriate σ^2 notation for variance components not been adopted, it is possible that the present confusion might never have arisen.

May I conclude by reiterating that in my opinion the fixed-effects model, in the form it is customarily expounded, is a source of confusion rather than enlightenment, and should be dropped. With the rapid accretion of theory in all branches of statistics it is most important that only what is really useful is retained and taught to students.

Professor O. KEMPTHORNE (Iowa State University): The Secretary wrote me to limit my remarks and the statement I shall make cannot therefore contain a detailed analysis of the paper. I must confine myself to general remarks. The title is pretentious in view of the contents. Notations for handling factorial structures have been developed by myself, M. B. Wilk, G. Zyskind, and undoubtedly others. One would get no intimation of this from the paper. I suggest to the profession that Dr Nelder has been professionally unethical in not making reference to work which covered essentially everything he does with a mere difference of notation. Who has neglected "marginality"? Perhaps Dr Nelder in the past. Who have exposited differences between finite and infinite populations? The mere reference to a one-page reference by Kempthorne in Kempthorne and Folks, stated by them to be "sloppy" or careless, is outright dishonesty in view of the published record, and what I know Dr Nelder knows but chooses to ignore. It will be apparent to real students of the area that Dr Nelder has had the habit for more than a decade of failing to read the literature or failing to acknowledge its existence (cf. his 1965 paper). The reformulation is not at all new, I assert. Population linear identities were given by me in my 1952 book. They were developed extensively by M. B. Wilk and G. Zyskind and others. I hope that the profession will take note of the facts.

On the matter of linear models for randomized experiments, surely Dr Nelder has read and indeed had read by 1965 my own book. There is nothing new in Section 7.

There are philosophical questions in the area but I have used the space available here.

I assert that there is nothing new and original in this paper.

I suggest that Dr Nelder owes deep apologies to many professional workers individually and to the profession as a whole. Finally, his offence is one against science and research in general. I ask the judgement of the profession.

Professor D. V. LINDLEY (University College London). Although Dr Nelder and I have radically different approaches to the linear model, it is interesting how much agreement there is between us when it comes to the "mechanics" of the situation and consequently how illuminating his algebraic manipulations are for the Bayesian view. He starts with the concept of a population, undefined in tonight's paper, which, in the Bayesian framework, is a collection of random quantities judged exchangeable. The linear model structure can be interpreted in terms of assumptions of partial or complete exchangeability. Thus if $X_1, X_2, \ldots X_n$ are exchangeable, $E(X_i) = \mu$, $\text{var}(X_i) = \omega^2$, $\text{cov}(X_i, X_j) = \rho\omega^2$ for all $i, j, i \neq j$. If we write $X_i = \theta + \varepsilon_i$ with

$$E(\theta) = \mu, \text{var}(\theta) = \tau^2, E(\varepsilon_i) = 0, \text{var}(\varepsilon_i) = \sigma^2,$$

all $n+1$ quantities being independent, we have a linear model with the same first- and second-order structure, the identification being $\omega^2 = \sigma^2 + \tau^2$ and $\rho = \tau^2/(\sigma^2 + \tau^2)$. Unfortunately the linear model implies $\rho \geqslant 0$ whereas if the exchangeability is only finite we may have $(n-1)^{-1} \leqslant \rho$. Nelder manages to embrace this case by using a linear model with negative "variances". In the Bayesian view this is not essential, but nevertheless the understanding gained by his device is clearly useful to the personalistic approach and tonight's paper is to be welcomed.

If we agree on the inner workings of the black box, we appear to part company when we look at the casing and consider its relation to the environment. Why are we engaged in all this algebra?

What is to be done with the mean squares or the F-tests? How is the practitioner to interpret the linear models? (Incidentally I thought the speaker was a "likelihood man"—if so, what are F-ratios doing?). Let me illustrate some of the issues by taking remarks from the paper.

In Section 2.2 fixed and random effects are respectively associated with complete and incomplete sampling. Suppose that treatment 27 is included in an experiment. If it is desired to make inferences about treatment 27 then the effect is fixed. If inferences are required about a a treatment not included in the experiment but exchangeable with 27 the effect is random. Sampling does not enter into the matter; the distinction is concerned with the inference, a topic not mentioned in the paper.

I suppose that with the mud of Rothamsted on his boots, Dr Nelder is in a better position than an academic like me to comment on practical matters, but his assertion that an interaction with no main effects is of no practical interest seems rash. A main effect can be turned into an interaction and vice versa. Consider two factors each at levels 0 and 1 with yields as follows

	0	1
0	3	5
1	3	5

with one main effect and no interaction. Let one factor remain unaltered but let the other be at level 0 if the original two factors were at the same level, and 1 otherwise. The new factors have no main effects, only an interaction. The important point is what inferences are to be made?

In the Bayesian view the analysis cannot be separated from the input or the output. We are used to no input in the classical model, but to have no output either seems carrying things too far. This is a useful—and I choose the word deliberately, for I shall use it—paper but one that is incomplete because of its failure to consider what all the activity is about. Once this is included, inference and associated probability ideas are essential to an appreciation of the black box's role.

Mr A. P. Dawid (University College London). To a very large extent my comments will echo those of Professor Lindley, although they were prepared independently. Together with Professor Nelder, I feel that there is more to analysis of variance than the normal linear model, and that most current formulations do not emphasize the distinctive features. Consequently a re-formulation is to be welcomed, and I find Nelder's approach valuable. In fact, my principal criticism of this paper is the length of time we have had to wait for it since some of the ideas were presented in Nelder (1954).

I do have some difficulty, however, with Nelder's concept of probability, which, as I understand it, is similar to Fisher's idea of probability as a proportion in a possibly infinite population. This seems to extend straightforwardly to cover nested structures (populations of populations of ...), but its application to cross-classifications baffles me, and I find a related subjectivist approach more satisfying.

Suppose the statistician has a joint (subjective) distribution for the unobserved values (x_{ij}), where x_{ij} is the value associated with levels i and j of factors A and B respectively. If these factors are crossed, then this distribution might be unaltered if we permute *either* the levels of A *or* those of B (or both). This leads immediately to the covariance structure of Section 5.3, and hence, following Nelder's arguments, to the population linear model of Section 5.4 to describe the subjective distribution. The implied variance components may be negative, but usually will not be.

Nelder distinguishes, rightly, between two kinds of random term: error and effect. But this distinction cannot be absolute; it is related to the use to which the experimental data are to be put. Thus for A/B, with population linear model $x_{ij} = \mu + \alpha_i + \beta_{ij}$, the α's are effects if we wish to predict x's associated with further levels of B within levels of A occurring in the experiment; and we should want to estimate the $(\mu + \alpha_i)$. But if we wish to predict for *new* levels of A, then the (α_i) become errors, and the relevant effect, to be estimated, is simply μ.

The above ideas are developed, in the subjectivist framework, in Dawid (1977).

Dr D. A. Preece (University of Kent at Canterbury): I regard tonight's paper as outstandingly important and elegant. Many present-day textbooks, particularly some written for non-mathematicians, give the impression that the analysis for even the simplest experimental designs is a task

of terrifying and tedious mathematical complexity. Dr Nelder's attempt to do something about this deserves a warm welcome.

I suspect that a prime cause of the trouble, and a reason for many bad analyses of experimental data, is that people have tried to express, in mathematical notation or terminology, things that they have failed to express first in ordinary everyday language.

I am sorry that Dr Nelder's detailed discussion in Section 7 had to stop before reaching split-plot experiments. The crazy idea has been spread abroad that the analysis of even simple split-plot experiments is dangerous and difficult. Dr Yates dealt with an aspect of this in a paper already referred to tonight (Yates, 1967, pp. 785–786). But, more recently, Anderson and McLean (1974, p. 145) have stated that, when a student is first "exposed to the split-plot design, ... there is utter chaos". They seek to deal with this unnecessary or supposed chaos by introducing into the analysis of variance extra lines, each with zero degrees of freedom. I myself see more future in explaining and discussing, in everyday language, terms in the *standard* analysis, than in introducing and explaining terms *that are not really there anyway*. I hope Dr Nelder will have time and space to touch on this.

In practical experimentation, there is one situation that may seem to contradict Dr Nelder's remarks about postulating the existence of interactions in the absence of corresponding main effects. This arises when an experimenter sets out to study a summit on a response surface. The experimenter may say that he knows roughly what levels of his individual factors will produce a maximum response, and that he specifically wishes to study *interactions* in the region around the summit. If, for example, he sets up a 2^n experiment that happens to straddle the summit neatly, his data may well contain negligible main effects but large interactions. To ask "Are the main effects zero or nearly so?" is to ask "Are we working around a summit (or trough)?"—and this seems to me to be a question of practical interest.

Many of us were taught to use linear constraints for the parameters in the analyses of simple experimental designs. Now that we have heard Dr Nelder's advice about not using these constraints, some of us may fear that this will make teaching more difficult, just because there will be all those extra terms in the normal equations, and solution of these equations will be so much more difficult. (I am here thinking of teaching done without the use of matrices.) I had such a fear when I heard earlier versions of tonight's paper. However, I cut all constraints out of my teaching last year and found that, with a little care, things could proceed very smoothly. Indeed, it seems to me that the idea of an estimable contrast can be the more readily understood when there are no constraints to confuse the issue.

Dr H. P. WYNN (Imperial College, London): Dr Nelder has an important constraint running through his models. This is the constraint of *balance*. His very brief statement of the condition (Section 3.1) hides this severe limitation of the classical theory of random effect models.

Under normality and balance assumptions the usual estimators of variance are minimum variance unbiased estimators. This follows from the Lehmann–Scheffé theory of complete sufficient statistics combined with a suitable partitioning of the sum of squares using a generalized Fisher–Cochran theorem. There has been much recent work in weakening the normality assumptions to distributions whose third and fourth moments mimic those of the normal distribution. (For example, Graybill and Hultquist, 1961; Kleffe and Pincus, 1974). The estimators are proved to be uniformly minimum variance *quadratic* unbiased (MINQUE).

When balance is lost MINQUE estimators can be shown not to exist. Full necessary and sufficient conditions still seem to be lacking, however. The problems of balance also go over to the search for simple admissible estimators of covariance matrices when the fixed effects (means) are unknown. This is true both in the Stein and Bayes frameworks. I should welcome some comments from Professor Nelder on the unbalanced case.

Dr J. A. JOHN (University of Southampton): I would like to comment on marginality. Suppose we have two factors A and B at a and b levels respectively. Consider the two-way non-additive cross-classification model

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

I agree with Dr Nelder that since the γ_{ij} parameters span an $(ab-1)$-dimensional space of contrasts the effects of A and B are marginal to $A.B$. It makes no sense to test this model against, for

instance, the model

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

Dr Nelder, therefore, concludes that the meaningful sum of squares are those for A eliminating *B*, *B* eliminating *A* and *A*.*B* eliminating *A* and *B*. By this I assume he means that he has the following sequence of models:

$$\begin{aligned}
&(1) \quad y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk},\\
&(2) \quad y_{ijk} = \mu + \alpha_i + \beta_j \qquad + e_{ijk},\\
&(3) \quad y_{ijk} = \mu \qquad + \beta_j \qquad + e_{ijk},\\
&(4) \quad y_{ijk} = \mu + \alpha_i \qquad\quad + e_{ijk}.
\end{aligned}$$

A.*B* eliminating *A* and *B* involves testing (1) against (2). *A* eliminating *B* is (2) against (3) and is only valid if interactions are not significant in the first test. *B* eliminating *A* is (2) against (4).

However, one can approach the problem in another way. Suppose we write the model as follows

$$\begin{aligned}
y_{ijk} &= (\mu - \bar{\gamma}) + (\alpha_i + \bar{\gamma}_{i.}) + (\beta_j + \bar{\gamma}_{.j}) + (\gamma_{ij} - \bar{\gamma}_{i.} - \bar{\gamma}_{.j} + \bar{\gamma}) + e_{ijk}\\
&= \mu^* + \alpha_i^* + \beta_j^* + \gamma_{ij}^* + e_{ijk}.
\end{aligned}$$

Now the γ_{ij}^* parameters span an $(a-1)(b-1)$-dimensional space of contrasts of which the two spaces spanned by the α_i^* and by the β_j^* parameters are not subspaces. It seems perfectly valid to define the main effects of *A* and *B* in terms of the α_i^* and β_j^* respectively so that, in this model, *A* and *B* are not marginal to *A*.*B*. In other words, contrasts in the α_i^* and β_j^* are estimable and it is meaningful to test main effects by testing the equality of the α_i^* or β_j^*. Yates (1934) indicates that this is the appropriate test to use when it cannot be assumed that the interactions are non-existent. He refers to this analysis as the "method of weighted squares of means".

It is interesting, although it possibly adds to the confusion, to look further at the models involved when testing the equality of, say, the α_i^*. We are now testing

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

against

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

or, if we re-arrange the models to involve the original parameters, we test

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

against

$$y_{ijk} = \mu + \beta_j + (\gamma_{ij} - \bar{\gamma}_{i.}) + e_{ijk}.$$

I am not familiar with the BIOMED program but I wonder whether, with its choice of constraints, its sum of squares for *A* eliminating *A*.*B* does in fact test this hypothesis. If so then it seems perfectly valid.

I am also uneasy about adopting the sequential approach for more than two factors. Suppose in testing five factors the five-factor interaction is statistically significant. I might still be interested in testing main effects and other interactions since the statistical significance may be spurious. It may happen, for instance, that no other interaction is significant which would throw considerable doubt on the importance of the five-factor interaction. I would be interested to hear what Dr Nelder would do in such a situation.

Professor D. R. Cox (Imperial College): A simple example where a zero main effect occurs in conjunction with interaction can be based on the textile problems studied by Daniels (1939). Consider a rectangle of paper of constant density cut very accurately into 1 cm strips parallel to one side. Suppose now that each of these strips is cut by scissors into 10 parts, the aim being to judge by eye a division into 10 parts as nearly equal as possible. If this is done for m such strips and the pieces weighed, an $m \times 10$ data array is generated the row sums of which are constant, the interaction mean square providing a measure of the variability of the scissor cutting. The moral seems to be that in each application we have to consider carefully what are the physically

meaningful parameters. I agree with Dr Nelder that in the great majority of cases the measures of excess variability are appropriate, but feel that he has gone too far in suggesting that this is always so.

The following contributions were received in writing, after the meeting.

Dr JOHN W. TUKEY (Bell Laboratories, Murray Hill, N.J., U.S.A.): I join with the speaker in hoping for an eventual unified and agreed-upon description. I hope the present paper will help us approach this ideal state, but I must say that it has not brought us there. In particular, as I explain below, I cannot accept a universal prohibition on asking about $A = 0$ in the presence of $AB \neq 0$.

Three types of variability arise in almost any question about a set of comparative measurements, experimental or not: measurement variability, sampling variability, contextual sampling variability. The general picture is that we have measured a sample of versions (about whose comparison the question is posed) in each of a sample of contexts.

A major point, on which I cannot yet hope for universal agreement, is that our focus must be on questions, not models. Suppose, for maximum simplicity, that we have measured some response for each of a number of versions or treatments under circumstances where we can separate measurement variability from sampling variability. Then, whether or not both are wise, we may correctly ask both of the following questions: (a) What confidence limits can we place on a mean of *exactly* these versions? (b) What confidence limits can we place on the mean over a corresponding population of versions of which these are a reasonable sample? (Each of these is, as is easy to see, really treating versions as contexts.) If our versions are a probability sample of a population, this fact cannot possibly stop us from inquiring about exactly these versions. If our versions are *not* a sample from a well-defined population, we have long known that we should not "cop out" by only considering limits on the mean of the version we have. Granted that our estimate of sampling variability is, perhaps, biased to an unknown degree, we need to produce measures of uncertainty, such as confidence limits, that combine measurement variability and sampling variability. We need to face the larger variability, even though our view of it may be broad. (This example can, of course, be generalized almost without bound.)

One conclusion that I draw from such examples is this: Models can—and will—get us in deep troubles if we expect them to tell us what the unique proper questions are. (We also have to be almost equally careful about models, for quite separate reasons, once we face up to robust/resistant considerations.)

If we know just what versions we are to compare, but are concerned with the contexts in which the comparisons are to be made a similar pair of questions are easy to formulate—one in which we ask about means over the contexts actually used (context-sampling variability excluded) and one in which we ask about means over a population of contexts from which we have used a sample (context-sampling variability included). Again we can almost always ask both questions, though one may be quite silly. In agriculture, of course, we usually want to include the context-sampling variability—even though we rarely, if ever, find our sample of contexts to be a random sample of the contexts that concern us.

In an $A \times B \times C$ analysis, we may well want to include C's context-sampling variability when analysing A and exclude it when analysing B, something no model could tell us to do.

Consider, finally, an experiment comparing two detailed flavours for a standard brand of confectionery. If we do not know sexes for individuals—only that our samples are at least random—we will clearly be very concerned as to which flavour will sell better in total. If we know individual sexes and if we find a large flavour-by-sex interaction, we are still almost sure to want to compare total sales between flavours without regard to whether it is larger than the interaction would suggest. (After all, any such suggestion has to come from regarding the two sexes as a sample from a larger number of sexes!) This example shows clearly, I believe, that it may be important, not just possible, to test A against zero in the presence of even quite large interaction.

Mr J. H. MAINDONALD (Victoria University of Wellington, N.Z.): Computer programs ought, I believe, to do more than check for extrinsic aliasing as discussed in Section 6; details of the linear relation ought to be printed out. Such information is easily obtained as a by-product of any of the standard methods of computation. Suppose that the kth column of the design matrix is a

linear combination of earlier columns. Then with Cholesky or a related method, or with $Q-R$, the necessary information is contained in the $(k-1)\times k$ submatrix of the upper triangle matrix obtained. If, for example, the elimination process reduces the matrix of coefficients in the normal equations to

$$\begin{pmatrix} 6 & 2 & 5 \\ 0 & 4 & -2 \\ 0 & 0 & 0 \end{pmatrix},$$

one will solve

$$\begin{pmatrix} 6 & 2 \\ 0 & 4 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 5 \\ -2 \end{pmatrix},$$

to yield

$$\mathbf{x}_3 = \mathbf{x}_1 - \tfrac{1}{2}\mathbf{x}_2.$$

(Any zero in one of the $k-1$ diagonal positions will be replaced by 1·0.)

With Gauss–Jordan the necessary information is available directly, provided one adopts a suitable convention for dealing with zero or near-zero pivots. In the above case the coefficients 1 and $-\frac{1}{2}$ would occupy positions 1 and 2 of column 3 (or row 3, if the algorithm works with the lower triangle) of the reduced matrix.

Dr Nelder's examples have too simple a structure to provide good illustrations of this suggested procedure.

Drs J. FRANE and R. JENNRICH (UCLA): In this brief space, we can touch only on a few of the issues raised by Dr Nelder. First, we would like to clarify what the computer program BMD10V (né BMDX64) does in terms of a simple 2×2 fixed effects design: $\mathrm{E}Y_{ij} = \mu_{ij}$. The following hypotheses are automatically tested (for both balanced and unbalanced designs).

$$\begin{aligned} &\text{row effects:} && \mu_{11}+\mu_{12} = \mu_{21}+\mu_{22}, \\ &\text{column effects:} && \mu_{11}+\mu_{21} = \mu_{12}+\mu_{22}, \\ &\text{interaction:} && \mu_{11}+\mu_{22} = \mu_{12}+\mu_{21}. \end{aligned}$$

These hypotheses are, of course, the hypotheses that are classically tested. Several authors recommend these for most unbalanced problems (Francis, 1973; Kutner, 1974; Speed and Hocking, 1976). Other schemes for unbalanced data generally test hypotheses that weight the expected cell means by the observed cell sizes, but Kutner (1974), Speed and Hocking (1976) and Searle (1971, p. 317) recommend against these methods. BMD10V is a general linear hypothesis program and as such can test a wide variety of hypotheses besides those automatically tested.

The above test for row effects in the presence of interaction has been questioned. Indeed, we recognize that the interpretation of the test for row effects is generally influenced by the magnitude of the interaction. Consider, however, data on the cost of operation of two chemical analyses used in two laboratories. For each analyser in each laboratory we are give a sample of daily costs. Suppose we have clear-cut interaction, but our problem is to purchase one analyser to be used by the two laboratories. We are interested in minimizing the total cost for both laboratories and in the hypothesis that there is no difference in total cost, namely,

$$\mu_{11}+\mu_{12} = \mu_{21}+\mu_{22},$$

where μ_{ij} is the mean cost of operating analyser i in laboratory j. This is, of course, a test for zero main effect in the presence of significant interaction of the form given by BMD10V.

On the other hand, if we are not constrained to use the same analyser in both laboratories in the future and if interaction is significant then hypotheses of the form (analyser 2 is always at least as good as and is sometimes better than analyser 1) are of interest. If the evidence supports $\mu_{11} < \mu_{12}$ and $\mu_{21} > \mu_{22}$, then we use separate analysers for each laboratory. BMD10V can perform the necessary computations except for modest side calculations.

While certain tests are not always used, it is better to have them reported than not to have them available at all.

Use of analysis of variance for model building is sometimes different from hypothesis testing. An important feature in BMD10V is its ability to consider a sequence of models (each as a separate analysis).

Dr NELDER replied briefly at the meeting, and subsequently in writing as follows.

I had wondered, in advance, whether I should find myself in a minority of one at the end of the discussion; in the event the amount of support has been gratifying. Mr Healy wonders if some of the distinctions and differences have any practical consequences. The ambiguities in the interpretation of variance components caused by differing constraints, not always made explicit, have, I believe, caused confusion. The dangers of making false inferences by neglecting marginality are exemplified by Professor Aitkin and are real. The difficulties encountered by students with the current formulation should not be underestimated.

Professor Plackett raises the important question of how formulations of statistical models reach the target populations. He is surely right in noting the increasing influence of computer manuals (and the Editorial Board of *Applied Statistics* decided some years ago that manuals should be included among books to be reviewed). I do not have to teach linear models and he does; I would be unhappy, however, about introducing constraints for fixed effects, because I would have to leave them out for random effects (otherwise the distributional assumptions are violated), and this would lead to the kind of differences in dealing with the two kinds of term that I am objecting to. Dr Preece's experience in banishing constraints in his teaching is encouraging, and suggests that a unified approach can be put across at the elementary level.

Professor Plackett also questions the adequacy of a notation in which the interpretation of $A.B$ depends upon which marginal terms precede it. Now it is arguable that the interpretation of $\beta_2 x^2$ is different in the two expressions

$$\beta_0+\beta_1 x+\beta_2 x^2 \quad \text{and} \quad \beta_0+\beta_2 x^2$$

yet we do not distinguish notationally between them. It would of course be perfectly possible to write, for example,

$$A.B \quad A.B[A] \quad A.B[B] \quad A.B[A,B]$$

when the marginal terms eliminated are appended in square brackets. Experience with the notation of the paper, which is compact and unambiguous, suggests that users quickly assimilate it and are not confused by the convention adopted. This contrasts with the current use of variance components in E.M.S., where the notation *is* ambiguous unless the constraints are also specified.

Professor Lindley has, as usual, asked some awkward questions. In considering the status of his treatment 27 I regard sampling as a way of ensuring some objectivity in the assumption of exchangeability. It is quite true that by redefinition of factors we can convert a main effect into an interaction and vice versa. The process is analogous to the rotation of axes in multivariate methods. Like models which violate marginality relations, rotation procedures are well defined mathematically but may not make any practical sense. The fact that tables can be constructed with interactions associated with null margins does not mean that the corresponding model is a useful one *a priori*.

Both Professors Lindley and Tukey have commented, rightly, on the incompleteness of the paper. I dealt only with the specifying and fitting of models, whereby the data are effectively replaced by a set of fitted values derived from a model, the model being a parsimonious one, which nevertheless gives fitted values not appreciably discrepant from the data. This may be called the *smoothing* phase. Beyond this is the *prediction* phase, when predictions and inferences are made, and questions (in Tukey's sense) asked. The relationships between questions, the structure of data, and models are both important and complex, but are outside the scope of the paper; very briefly I take the view that models are useful tools in trying to answer questions. The prediction phase I was obliged to leave undiscussed for lack of space; of course the smoothing phase also contains inferences, e.g. whether to infer that an interaction is zero, and in such goodness-of-fit questions I doubt that likelihoods are enough, i.e. data actually obtained do have to be embedded in other data sets that might have been obtained but were not. The strong statement in the paper about practically interesting models in relation to marginality is a statement about the smoothing phase. I agree entirely with Professor Tukey that, having fitted a model containing an interaction, one may wish to predict one of the margins; with balanced experimental data the

sample margins serve both as sufficient statistics in the fitting process and as predictors, and these roles are often not distinguished. However, with survey data the sample and predicted margins are often quite different, with individual elements combined with different sets of weights for the two types. Standard errors may be assigned to the elements of a predicted margin, and hence, in Tukey's example, a significance test could be constructed for the difference in the two kinds of confectionery summed over the factors with which the preferences interacted, though I am not convinced that zero is a specially interesting point on the difference scale.

Drs Frane and Jennrich give a similar example involving a decision about instruments which interact with laboratories. Again the distinction must be made between significance tests for goodness of fit at the model-building stage and tests for predicted margins, given the model. With unbalanced data the use of A eliminating $A.B$ when the interaction is negligible can lead to severe loss of power in the test for A, so that to present this sum of squares at the model-building stage can lead to false inferences. The use of a test of significance in the decision problem—which instrument?—can be questioned. Why not choose the one that minimizes costs?

Professor Cox presents an example of a table with a null margin and a non-null interaction. Such tables certainly exist, and perhaps a simpler example is afforded by a contingency table with a fixed null margin produced by an *a priori* sampling scheme. In the fitting of models, the terms corresponding to such margins belong to a *minimal model* (Nelder and Wedderburn, 1972) which has to be fitted first before exploration of further model terms begins. Two points about the terms in the minimal model should be noted: first, there is no question of a hypothesis being tested about them and, secondly, the terms should be fitted *even if null* so that the d.f. for the residuals after the minimal model has been fitted will be correct. Thus null terms in the minimal model do not contradict the asserted necessity of recognizing marginality relations.

Dr Preece raises a slightly different point, namely the relations between terms in a polynomial model, as might be used in studying a surface near a summit. If the position of the summit is not a special point on the measurement scale (and generally it is not) then the quadratic surface must include the linear one as a subset, i.e.

$$Y = \beta_0 + \beta_1 + \beta_2 x^2$$

includes

$$Y = \beta_0 + \beta_1 x.$$

If interest centres on the curvature then the linear term becomes part of the minimal model and is fitted first; its smallness relative to the quadratic is not under test. It would be wrong, however, to ignore the fact that the linear model is part of the quadratic one, and to fit a term for linear eliminating quadratic (and cubic etc.). This leads to the same danger of making misleading inferences as Professor Aitken demonstrated so clearly for main effects and interactions. The ordering implicit in the models must be respected. I hope Professor Aitkin will expand his remarks for the point he makes is of great importance.

Dr Wynn rightly comments on the complexities of estimation of models with both fixed and random effects when no balance exists. For the estimation of the variance components I would use the modified maximum-likelihood estimation described by, for example, Patterson and Thompson (1971). The problem of estimating the fixed effects, given the estimates of the variances, and of assigning measures of uncertainty to them, seems much more difficult, the still-controversial Fisher–Behrens distribution relating to almost the simplest possible case. I am unsure about the validity of demanding admissibility as a necessary property of an estimator. For unbalanced data with one error I believe the use of canonical components should lead to some simplification in the formulae for standard errors and expected mean squares, though of course much of the simplicity of the balanced case will be lost.

Dr John's reparametrization underlies the type of model which ignores marginality. This model is well defined algebraically, and my quarrel is not with its mathematics, but with its practical relevance. I should emphasize again the distinction between the smoothing and prediction phases. His second point is one of some practical importance, namely whether marginal effects give any information about interior effects. Thus if $A.B.C$ is large, but $A.B$, $A.C$ and $B.C$ are all negligible should we feel the less convinced about the reality of the three-factor interaction? If the $A.B.C$ effect is real, the position is analogous to having accidentally sited an experiment exactly at a col in a response surface. My own practice is to discount large high-order interactions not accompanied

by at least some large marginal effects, but the justification is intuitive and the matter deserves serious study.

Dr Yates believes that I have not gone far enough and that I should have abolished altogether the distinction between fixed and random effects. We are agreed in believing that the arithmetic of the analysis of variance and the form of expected mean squares are not affected by the fixed/random distinction, but I still find it useful to distinguish between a set of, say, varietal effects which can be regarded as a random sample from a population and a set of effects which cannot. I am sure he is right in tracing back negative canonical components to negative intra-class correlation. Mr Dawid rightly draws attention to the two-faced attitude we may adopt towards random terms that occur in the numerator of variance ratios; my main concern, however, was to distinguish this type of term from a random "denominator" term in setting up a model. It is true that even here the distinction is not absolute, because one man's noise may be another man's interesting sound.

It is difficult to know how to respond to Professor Kempthorne's somewhat intemperate outburst. My paper was written from a strong dissatisfaction with the current exposition of linear models in many textbooks. I have spent some considerable time in recent years trying to undo the confusions that have arisen in various people's minds from the exposition they have received. I am not alone in this; a Professor of Statistics, currently in Canada, wrote to me recently (my italics): "The psychologists I teach here are *strongly confused* on all this", i.e. on random and fixed effects. I have given five reasons for my dissatisfaction; if Professor Kempthorne can convince the referees that these points have been previously expounded in exactly the form I give, then I shall be only too happy to acknowledge the necessary priority. In my reformulation it was necessary to go back to first principles and to simple structures in order to make as clear as possible the points at issue; no implication was intended that all the steps in the reformulation were original. Such a notion is manifestly absurd. For instance, I know that Professor Kempthorne, among others, has written about the population linear identity; partly as a result of his writings such an idea can now be taken as given, and he should be pleased that the idea has now reached such a firm place in the generally accepted corpus of statistical thought.

To the melodrama of his last paragraph I have nothing to contribute. My reformulation may be accepted, in whole or in part, or rejected entirely, or replaced by something better as other ideas are put forward. Other statisticians will decide. Meanwhile, Professor Kempthorne might do worse than consider the possibility that there really are substantial points at issue here, points that would be worth while his trying to understand.

References in the Discussion

Aitkin, M. A. (1974). Simultaneous inference and the choice of variable subsets in multiple regression. *Technometrics*, **16**, 221–227.

Anderson, V. L. and McLean, R. A. (1974). Restriction errors: another dimension in teaching experimental statistics. *Amer. Statist'n*, **28**, 145–152.

Dawid, A. P. (1977). Invariant distributions and analysis of variance models. *Biometrika* (in press).

Graybill, F. A. and Hultquist, R. A. (1961). Theorems concerning Eisenhart's Model II. *Ann. Math. Statist.*, **32**, 261–269.

Kempthorne, O. (1952). *The Design and Analysis of Experiments*. New York: Wiley.

Kleffe, J. and Pincus, R. (1974). Bayes and best quadratic unbiased estimators for variance components and heteroscedastic variances in linear models. *Math. Operationsf. & Statistik*, **5**, 147–159.

Kutner, M. H. (1974). Hypothesis testing in linear models (Eisenhart Model I). *Amer. Statist'n*, **28**, 98–100.

Nelder, J. A. (1954). The interpretation of negative components of variance. *Biometrika*, **41**, 544–548.

Patterson, H. D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, **58**, 545–554.

Plackett, R. L. (1960). Models in the analysis of variance (with Discussion). *J. R. Statist. Soc.* B, **22**, 195–217.

Pringle, R. M. and Rayner, A. A. (1971). *Generalised Inverse Matrices with Applications to Statistics*. London: Griffin.

Rao, C. R. (1971). Estimation of variance and covariance components—Minque theory. *J. Multivar. Anal.*, **1**, 257–275.

Speed, F. M. and Hocking, R. R. (1976). The use of the $R(\,)$-notation with unbalanced data. *Amer. Statist'n*, **30**, 30–33.

Yates, F. (1934). The analysis of multiple classifications with unequal numbers in the difference classes. *J. Amer. Statist. Ass.*, **29**, 51–66.

As a result of the ballot held during the meeting, the following were elected Fellows of the Society.

ABRAHAM, Bovas
AITKIN, Murray A.
ATSEM, Francis, K.
BAMIDURO, Timothy A.
BENETT, Evelyn M. M.
BIGNOLD, Peter R.
BIRD, Anthony P.
BREWER, Alan C.
BRUBACHER, Sanford R.
BULL, Roderick P.
BURRIDGE, James
BUTLER, Michael
CAMPBELL, Norman A.
CHALONER, Kathryn M.
CHATER, Robin E. J.
CHONG, Yok C.
CLIFFORD, Peter
DARBY, Sarah C.
DAVIES, John G.
DOLBY, Geoffrey R.
DOUGLAS, Malcolm
EL-ATOUM, S. A. M.
FIELD, David A.
FIELLER, Nicholas R.
FORD, Ashley P.
FRANCIS, Brian J.
GIORGI, Roberto S.
GRANT, Ian H.
HAMAD, Yahia H.
HERMIS, Sabah R.
HOINVILLE, Elizabeth A.
JOHNSTON, Bruce O.
KAMEL, Raymond D.
KELSEY, William F.
KENWARD, Andrew C.
LAM, Lai C.
LANE, John A.
LANGFORD, Roger J.
LOCKWOOD, Barbara M.
MARLOW, Andrea K.
MARRIOTT, John M.
MCCOMB, Robert W.
MEMON, Abdul G.
MOUQUE, Colin V.
MULENGA, Mpafya S.
NAIDU, Sandrasagarren N.
NEMENYI, Peter
NEVILL, Alan M.
OWEN, Jean M.
PANAYIOTOU, George N.
PATEL, Mahendra D.
PETCH, Geoffrey A.
PETSA, Katerina
PNEUMATICOS, Tryphon M.
RICHARDS, Michael J.
RICHARDSON, Robin E.
SCOTT, Kenneth, E.
SHELDON, Neil A.
SHINE, Philip J.
SMITH, Alison M.
SUTTON, Terence W.
SWAN, Robin N.
SYKES, Alan M.
TAPPIN, Michael
TRAXLER, Robert H.
TSE, Yiu K.
TYLER, Michael J.
TZORTZOPOULOS, Panos T.
WADE, Robert E.
WHITE, Robert J.
WILKINSON, Ian R.
WINSLOW, Richard L.
WYBROW, Graham M.