International Statistical Review, (1984), **52**, 1, pp. 33–81. Printed in Great Britain

Analysis of Variance Models in Orthogonal Designs

Tue Tjur

Institute of Mathematical Statistics, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark

Summary

This paper presents an approach to analysis of variance modelling in designs where all factors are orthogonal, based on formal mathematical definitions of concepts related to factors and experimental designs. The structure of an orthogonal design is described by a factor structure diagram containing the information about nestedness relations between the factors. An orthogonal design determines a unique decomposition of the observation space as a direct sum of orthogonal subspaces, one for each factor of the design. A class of well-behaved variance component models, stated in terms of fixed and random effects of factors from a given design, is characterized, and the solutions to problems of estimation and hypothesis testing within this class are given in terms of the factor structure diagram and the analysis of variance table induced by the decomposition.

Key words: Analysis of variance; ANOVA; Mixed models; Orthogonal designs; Variance component models.

1 Introduction

This paper deals with analysis of variance (ANOVA) models in experimental designs where all factors (treatment factors as well as blockings) are orthogonal. Examples are randomized block designs, split-plot designs, completely balanced k-factor designs (with or without an orthogonal blocking), Latin and Graeco-Latin squares, fractional replicates of complete factorials (but not balanced incomplete block designs, etc.).

Mathematically, this is a rather exclusive class of experimental designs. Statistically, however, it is very important, and more or less standardized methods for the handling of analysis of variance models in such designs are given by most books on experimental design.

The main tool of the modern approach to analysis of variance is matrix calculus based on concepts from Euclidean geometry (orthogonal projections, etc.). However, in the case of orthogonal designs it is generally recognized that the matrix calculations involved are purely formal, in the sense that the interpretation of the symbols as matrices plays a secondary role. The final results (e.g. formulae for sums of squares of deviations) are not stated in terms of matrices anyway, and the intermediate matrix manipulations can be more or less replaced by similar operations on other algebraic objects, like symbolic expressions for contrasts and interactions in a 2^k-design (Fisher, 1935), terms of a degrees-of-freedom identity (Nelder, 1965), Kroenecker products of certain types of matrices (Nelder, 1965), subsets of the finite set indexing the factors of a k-way table (Jensen, 1979), or equivalence relations on the set of experimental units (Speed & Bailey, 1982). Apart from the classical origin of this, the tedious algebra of quadratic forms is expressed by summations, bars, dots and a lot of indices

There is a clear line of development here, from the algebra of quadratic forms, via matrices and Euclidean geometry, towards approaches which are more and more directly related to the only thing which is really unavoidable in the statistical context: the combinatorial structure of the design. The aim of the present paper is to give an exposition which is based entirely on this combinatorial structure. Design structures will be represented by *factor structure diagrams*, holding the information about nestedness relations between the factors. It will be shown how degrees of freedom, sums of squares of deviations, etc. can be derived in a simple way from this diagram. The sums of squares of deviations come out as integer linear combinations of square sums of the form $ss_F = \sum n_f \bar{y}_f^2$, where the summation is over the levels of a factor F, n_f denoting the number of observations on level f, $\bar{y}_f$ the average of these. Rules for estimation of variance components, allocation of treatment effects to strata, tests for model reductions, estimation of contrasts, etc. will also be given in terms of the factor structure diagram and the analysis of variance table constructed from it.

The theory in the following is based on formal mathematical definitions of basic concepts related to experimental designs. Some of these concepts (e.g. the concept of a factor) are so simple and well-known that their mathematical meaning is usually subsumed or ignored. It should be noticed, however, that the concept of a *minimum* of two factors, which is a standard mathematical construction and an almost unavoidable part of the mathematical formalism, does not correspond directly to a classical statistical concept, though related in an obvious way to partial aliasing or partial confounding. The concept of a minimum plays a crucial role for the results obtained in the present paper.

However, these mathematical concepts are not new, nor is their relation to analysis of variance models. There is, in particular, an overlap with ideas put forward by Nelder (1965), Zedeler (1970), Jensen (1979), Speed & Bailey (1982). More detailed statements about this follow below. It is hoped that the present paper will contribute to the propagation of the ideas put forward by these authors, and that these ideas are not too obscured by the simplifying assumptions and omissions of many other aspects, which are made here.

Related papers. Nelder (1965) developed a theory for a class of analysis of variance models (including finite randomization models, which are not discussed in the present paper), usually referred to as the *generally balanced models*. This theory has been developed further by Wilkinson (1970) and James & Wilkinson (1971), and some of the computational aspects of this theory have been implemented in the GENSTAT 'ANOVA' algorithm. For a very rough description, the power of this algorithm lies in the fact that it is able to recognize and make use of symmetries in a very large class of analysis of variance models. Computationally, as far as linear (single stratum) models are concerned, the 'ANOVA' algorithm is situated somewhere between the general linear regression solution (involving the representation of the factor structure by a suitable design matrix, and the numerical inversion of a $k \times k$ matrix for k = the rank of this design matrix) and the explicit 'matrix free' solutions required for desk calculators. The 'ANOVA' algorithm performs a sequence of 'sweeps' to the data vector. This sequence can be divided into subsequences, each of which corresponds to the addition of a term to the model formula. The procedure ends up with the vector of residuals in the specified model. A 'sweep' can be described as the subtraction of level averages (with respect to a factor of the model) from data, possibly with a simple correction for nonorthogonality.

A more detailed discussion of the generally balanced models is beyond the scope of the present paper. See Houtman & Speed (1983) for a mathematically very clear exposition of these models and their properties.

However, our concept of an orthogonal design can be viewed as a generalization of the orthogonal block structures discussed by Nelder (1965, I). In the terminology of the present paper, Nelder's block structures are orthogonal designs, built up from balanced factors by repeated use of two operations, 'crossing' and 'nesting'. These operations can be regarded as operations on designs. Both give a new design with a set of experimental units that can be identified with the Cartesian product of the units sets of the two designs operated on. Crossing leads to a design containing all 'direct products' of factors from the two designs, while nesting corresponds to the replacement of each experimental unit of the first design with a copy of the second.

It is easy to show that the so constructed designs are orthogonal designs in our sense. Conversely, as far as only the random factors (our $\mathcal{B}$ in § 7) are concerned, it is roughly correct to say that our block structures $\mathcal{B}$ coincide with Nelder's block structures, though there are some counter-examples (e.g. a Latin square with random effects of row, column *and* Latin letter, which is included in our definition).

Nelder emphasizes three different forms of the covariance matrix. The two of these correspond to the two parameterizations discussed in § 7 of the present paper, and the third is a description in terms of distinct correlations between pairs of experimental units. Recursive rules for the expansion of these three forms under nesting and crossing are given.

Note that the arrows $\rightarrow$ occurring in the present paper are essentially the reverse of those used by Nelder (1965) to indicate nesting. This inconsistency is difficult to avoid, because our arrows indicate mappings between finite sets.

Zedeler (1970) discusses the structure of two-way designs in a mathematical framework similar to that of the present paper. The minimum (in our terminology) of two factors is characterized as their *cointersection* in the category of mappings between finite sets, and the combinatorial condition for orthogonality (our Proposition 1) is given. The relation of these concepts to the analysis of linear models in a two-way design is discussed.

Speed & Bailey (1982) clarify and develop further some of Nelder's ideas by means of a formal definition of factor structures similar to the one given here. A factor is identified by the equivalence relation which it imposes on the set of experimental units. The operations *conjunction* and *composition* on relations are discussed. Conjunction corresponds to our formation of products $F \times G$. Compositions are related to our minima $F \wedge G$. In terms of these operations, a characterization in combinatorial terms of balanced orthogonal factors with balanced minima (in our terminology) is given, and a version of our Theorem 1, is proved. Speed & Bailey relate their exposition to the combinatorial theory of finite lattices, see e.g. Aigner (1979), and show the relation to the Möbius function; see § 4.2. Nelder's concepts of crossing and nesting are discussed in this mathematical framework.

Jensen (1979) discusses variance component models in completely balanced k-factor designs. A covariance structure is determined by a set $\mathcal{A}$ of subsets of $\{1, \ldots, k\}$, each subset determining (in our terminology) a random factor $B \in \mathcal{B}$. The condition that $\mathcal{A}$ is closed under the formation of intersections (slightly generalized here to the condition that $\mathcal{B}$ should be closed under the formation of minima) is imposed, and the consequences of a second condition, closedness under the formation of unions, are discussed. Three parameterizations, corresponding to Nelder's three forms of the covariance matrix, are studied, and the six relations connecting these are derived. It is noticed that the representation by distinct correlations between pairs of observations (which is not discussed in the present paper) is only meaningful under the second condition (closedness under unions, which holds also for Nelder's orthogonal block structures). The rules for estimation and hypothesis testing under orthogonal block-treatment structure are derived.

2 Factors: Combinatorial structure

Let $\mathbf{y} = (y_i \mid i \in I) \in \mathbf{R}^I$ denote the data set. The elements i of the (finite) index set I are called *experimental units.*

A *factor* $\varphi_F : I \to F$ is a mapping φ_F from I into some other finite set F. The elements f of F are called *factor levels.* Usually, we shall talk about 'the factor F', rather than 'the factor $\varphi_F : I \to F$', thus subsuming the mapping φ_F as given by the context.

Two factors play a special role as 'extremes', the *trivial factor* 0, corresponding to a constant mapping $\varphi_0 : I \to 0$, where 0 is an arbitrary set with a single element, and the *units factor* I, corresponding to the identity $\varphi_I : I \to I$.

Intuitively, a factor F should be thought of as a partitioning of I into classes $\varphi_F^{-1}(f)$, each equipped with a label $f \in F$. Thus, the trivial factor 0 is the 'partitioning' into a single class, and I is the partitioning into single units.

2.1 Balanced factors

As a standard notation in the following, we let n_f denote the number of experimental units on the level f of the factor F,

$$n_f = \text{card } \varphi_F^{-1}(f),$$

where card denotes 'number of elements in'. By $|F|$ we denote the number of nonempty classes,

$$|F| = \text{card}\{f \mid n_f > 0\}.$$

A factor F is called *balanced* if all the classes are of the same size, which is then denoted

$$n_F = n_f = |I|/|F|.$$

2.2 Nested factors

Let two factors F and G be given. Suppose that there exists a mapping $\varphi_{FG} : F \to G$ such that $\varphi_{FG} \circ \varphi_F = \varphi_G$. In this case we say that F is *nested in* or *finer than* G, or that G is *marginal to* or *coarser than* F. We write $G \leqslant F$ or (in diagrams) $F \to G$. Notice that we have $G \leqslant F$ if and only if any of the G-classes can be written as a union of some of the F-classes, namely

$$\varphi_G^{-1}(g) = \bigcup_{f \in \varphi_{FG}^{-1}(g)} \varphi_F^{-1}(f).$$

We say that *the level f is nested in the level g* if $\varphi_F^{-1}(f) \subseteq \varphi_G^{-1}(g)$, or, equivalently, $\varphi_{FG}(f) = g$.

2.3 Equivalent factors

Two factors F and G will be called *equivalent* if both $F \leqslant G$ and $G \leqslant F$. Thus, equivalent factors are factors which induce the same partitioning of I. Only the labels on these classes and the number of empty classes (i.e. factor levels which are not used) may be different. For many purposes, the properties of a factor are sufficiently described by its equivalence class, and most of the concepts discussed in the following are well defined 'up to equivalence', (notice, however, that balancedness is *not* a property of the equivalence class).

Under the subsumed convention of not distinguishing between equivalent factors, the

relation $\leqslant$ is a partial ordering of the set of factors. We have the maximal element I and the minimal element 0.

The symbol $<$ is used for 'coarser than and nonequivalent to', that is $F<G \Leftrightarrow F \leqslant G$ but not $F \geqslant G$.

2.4 *Cross-classifications*

The *product* $F \times G$ of two factors F and G, or the *cross-classification* induced by F and G, is the factor

$$\varphi_{F\times G} : I \to F \times G,$$

where the set $F \times G$ is the ordinary Cartesian product of F and G, and

$$\varphi_{F\times G}(i) = (\varphi_F(i), \varphi_G(i)).$$

The formation of products can be regarded as an operation on equivalence classes, in the sense that replacement of F and G with equivalent factors F' and G' leads to an equivalent product $F' \times G'$. Under the partial ordering $\leqslant$, the product $F \times G$ can be characterized as the coarsest factor which is finer than both F and G. Indeed, the following two properties are easily seen to characterize $F \times G$ up to equivalence:

(i) $F \leqslant F \times G$ and $G \leqslant F \times G$;
(ii) any factor H finer than both F and G is also finer than $F \times G$.

In this sense, the partially ordered set of (equivalence classes of) factors possesses *maxima*, and the maximum operation, which we would otherwise denote by $\vee$, coincides with the formation of products ('$F \vee G = F \times G$').

2.5 *Minima of factors*

The dual concept, the *minimum* of two factors, is defined as follows. For two factors F and G, their minimum $F \wedge G$ is a factor with the properties:

(i) $F \wedge G \leqslant F$ and $F \wedge G \leqslant G$;
(ii) any factor H coarser than both F and G is also coarser than $F \wedge G$.

It is easy to show that the minimum $F \wedge G$, if it exists, is uniquely determined up to equivalence. Existence of the minimum can be proved as follows. For any factor F, let $\mathscr{A}_F$ denote the set of subsets of I of the form $\varphi_F^{-1}(M)$, where M is a subset of the set F. Then $\mathscr{A}_F$ is obviously an *algebra* of subsets of I, that is $\mathscr{A}_F$ includes the empty set and the set I, and is closed under the formation of intersections, unions and complements; $\mathscr{A}_F$ can be characterized as the algebra spanned by the classes $\varphi_F^{-1}(f)$ in the partitioning induced by F. Conversely, let $\mathscr{A}$ be an arbitrary algebra of subsets of I. A factor F with $\mathscr{A} = \mathscr{A}_F$ can be constructed as follows; consider the atoms of $\mathscr{A}$, that is the minimal nonempty sets in $\mathscr{A}$. These constitute a partitioning, and a factor F constructed by suitable labelling of the classes will obviously have $\mathscr{A}_F = \mathscr{A}$.

Hence, we have the one-to-one correspondence $F \leftrightarrow \mathscr{A}_F$ between equivalence classes of factors and algebras of subsets of I. This correspondence is order preserving, in the sense that

$$F \leqslant G \Leftrightarrow \mathscr{A}_F \subseteq \mathscr{A}_G.$$

This means that the problem of constructing a minimum of two factors is equivalent to the problem of constructing the 'minimum' of two algebras under the usual ordering by

inclusion. But the solution to the latter problem is straightforward, since the intersection of two algebras is again an algebra. Hence, the minimum of two factors exists and is determined by

$$\mathcal{A}_{F\wedge G} = \mathcal{A}_F \cap \mathcal{A}_G.$$

2.6 *Two examples*

The proof of existence does not give much intuitive feeling for what the operation $\wedge$ really does. This is illustrated by the following two examples.

Example 1. Let $I=\{1, 2, \ldots, 9\}$, and suppose we have two factors $R=\{r_1, r_2, r_3\}$ and $C=\{c_1, c_2, c_3, c_4\}$, i.e. rows and columns, on 3 and 4 levels, respectively. Suppose that the allocation of units to $R\times C$ levels is as indicated by Table 1. Put $H=R\wedge C$. From the relation $H\leqslant R$ we conclude that φ_H is constant on rows; for example, for the first row,

$$\varphi_H(1)=\varphi_H(2)=\varphi_H(3).$$

Similarly, $H\leqslant C$ means that φ_H is constant on columns, for example

$$\varphi_H(3)=\varphi_H(6).$$

Continuing like this, we can easily show that φ_H must be constant, that is $R\wedge C=0$. Thus, a criterion for the property $R\wedge C=0$, which is sometimes called *connectedness* of the two-way table, is that we can move from any nonempty cell to any other in a finite sequence of jumps between nonempty cells within the same row or the same column. More generally, this relation between experimental units, that the two corresponding cells can be connected by such a sequence of vertical and horizontal jumps, is obviously an equivalence relation, and the partitioning of I into equivalence classes under this relation coincides with the partitioning induced by the minimum $R\wedge C$. For example, if the last element 9 of I is removed, we will get a nontrivial minimum on two levels, corresponding to the partitioning into $\{1, 2, 3, 4, 5, 6\}$ and $\{7, 8\}$.

Table 1

Allocation of units to $R\times C$ levels

	Columns			
Rows	c_1	c_2	c_3	c_4
r_1	1, 2	3		
r_2	4, 5	6		9
r_3				7, 8

Example 2. Let $F_1, \ldots, F_k$ be factors such that the product $F_1\times\ldots\times F_k$ has $n_{f_1\ldots f_k}>0$ for all $(f_1, \ldots, f_k)\in F_1\times\ldots\times F_k$. This means that we have a k-way table with all cell counts positive. For $M\subseteq\{1, \ldots, k\}$, Let

$$F_M=\prod_{j\in M} F_j$$

denote the cross-classification according to the factors F_j, $j\in M$. It is easy to show, then, that we have the rule

$$F_{M_1}\wedge F_{M_2}=F_{M_1\cap M_2}.$$

For example, in a three-way table $A\times B\times C$ with all cell counts positive,

$$(A\times B)\wedge(B\times C)=B.$$

3 Factors: Linear structure

In this and § 4, we shall study the structure imposed by one or several factors on the observation space $\mathbf{R}^I$. Vectors in $\mathbf{R}^I$ are regarded as $I \times 1$ matrices, i.e. column vectors, and linear mappings are identified with their matrices as usual; $\mathbf{R}^I$ is equipped with the inner product

$$(\mathbf{x} \mid \mathbf{y}) = \mathbf{x}^*\mathbf{y} = \sum_{i \in I} x_i y_i,$$

the asterisk denoting transposition, and the corresponding Euclidean norm is denoted

$$\|\mathbf{y}\| = (\mathbf{y} \mid \mathbf{y})^{\frac{1}{2}}.$$

The $I \times I$ identity matrix is denoted by $\mathbf{I}$.

3.1 Notation and some basic relations

A factor $\varphi_F : I \to F$ induces a linear mapping $\mathbf{X}_F : \mathbf{R}^F \to \mathbf{R}^I$ defined by

$$\mathbf{X}_F((\alpha_f)) = (\alpha_{\varphi_F(i)} \mid i \in I).$$

As a matrix, $\mathbf{X}_F$ is the $I \times F$ matrix with elements

$$(\mathbf{X}_F)_{if} = \begin{cases} 1 & \text{for} \quad \varphi_F(i) = f, \\ 0 & \text{otherwise.} \end{cases}$$

The image of the linear mapping $\mathbf{X}_F$, which can also be characterized as the space of functions $I \to \mathbf{R}$ that are constant on the classes induced by F, is denoted

$$L_F = \{\mathbf{X}_F\boldsymbol{\alpha} \mid \boldsymbol{\alpha} \in \mathbf{R}^F\}.$$

Notice that $\dim L_F = |F|$. By $\mathbf{P}_F : \mathbf{R}^I \to \mathbf{R}^I$ we denote the orthogonal projection onto L_F. According to well-known rules for least squares estimation in a one-way analysis of variance model, $\mathbf{P}_F$ transforms a vector $\mathbf{y}$ by replacement of each co-ordinate y_i with the average $\bar{y}_f$ of all observations on the same level $f = \varphi_F(i)$. Hence, the $I \times I$ matrix $\mathbf{P}_F$ has elements

$$(\mathbf{P}_F)_{i_1 i_2} = \begin{cases} 1/n_f & \text{for} \quad \varphi_F(i_1) = \varphi_F(i_2) = f, \\ 0 & \text{for} \quad \varphi_F(i_1) \neq \varphi_F(i_2). \end{cases}$$

Notice the relation

$$\mathbf{P}_F = n_F^{-1}\mathbf{X}_F\mathbf{X}_F^*, \tag{3.1}$$

which holds for any *balanced* factor F.

The mapping $F \to L_F$ is order preserving in the sense that

$$F \leqslant G \Rightarrow L_F \subseteq L_G,$$

and it preserves minima according to the rule

$$L_{F \wedge G} = L_F \cap L_G.$$

Indeed, the inclusion $L_{F \wedge G} \subseteq L_F \cap L_G$ follows immediately from the order preservation property. The opposite inclusion follows if one thinks of a vector $\mathbf{v} = (v_i) \in L_F \cap L_G$ as a factor with real numbers as levels; $\mathbf{v} \in L_F \cap L_G$ means that this factor is coarser than both F and G, hence coarser than $F \wedge G$, by definition of the minimum.

Maxima are not preserved. It is easy to show that

$$L_{F \times G} \supseteq L_F + L_G,$$

but this inclusion is usually sharp. In a strictly mathematical sense, this is where the statistical concept of interaction comes in.

3.2 *Orthogonal factors*

As usual, two linear subspaces L_1 and L_2 of $\mathbf{R}^I$ are said to be orthogonal if any vector in L_1 is orthogonal to any vector in L_2. Here L_1 and L_2 are called *geometrically orthogonal* if they satisfy the following (weaker) condition. Let $L_1 = V \oplus V_1$ and $L_2 = V \oplus V_2$ be the decompositions of L_1 and L_2 as direct orthogonal sums of $V = L_1 \cap L_2$ and 'remainders' $V_1 = L_1 \cap V^\perp$ and $V_2 = L_2 \cap V^\perp$; then V_1 and V_2 are orthogonal.

The term 'geometrically' is motivated by the fact that this kind of orthogonality is the one known from Euclidean geometry, where two planes in $\mathbf{R}^3$ may be 'orthogonal' in exactly this sense. Roughly speaking, geometric orthogonality means that the subspaces are orthogonal except that they may have a nontrivial intersection. The following lemma characterizes the concept in a way which is more convenient in the analysis of variance context.

LEMMA 1. *Subspaces L_1 and L_2 are geometrically orthogonal if and only if the corresponding orthogonal projections $\mathbf{P}_1$ and $\mathbf{P}_2$ commute, that is $\mathbf{P}_1\mathbf{P}_2 = \mathbf{P}_2\mathbf{P}_1$.*

The proof is left to the reader, and so is the proof of the following useful result.

LEMMA 2. *Let $L_1, \ldots, L_k$ be pairwise geometrically orthogonal, and let $\mathbf{P}_1, \ldots, \mathbf{P}_k$ denote the corresponding orthogonal projections. Then $\mathbf{P} = \mathbf{P}_1 \ldots \mathbf{P}_k$ is the orthogonal projection on $L = L_1 \cap \ldots \cap L_k$.*

Two factors F and G are said to be orthogonal, written $F \perp G$, if the corresponding subspaces L_F and L_G are geometrically orthogonal. Or, equivalently, if

$$\mathbf{P}_F\mathbf{P}_G = \mathbf{P}_G\mathbf{P}_F. \tag{3.2}$$

The justification of the concept of orthogonality in relation to analysis of variance lies in formula (3.2), which, among other things, yields simple expressions for orthogonal projections on sums of subspaces generated by orthogonal factors. For the benefit of GENSTAT users, this can be illustrated by the way linear models with orthogonal terms are handled by the 'ANOVA' algorithm. Let $F_1, \ldots, F_k$ be orthogonal factors, and consider the model specified by $E\mathbf{y} \in L_{F_1} + \ldots + L_{F_k}$. According to Lemma 2, the vector of residuals in this model, i.e. the orthogonal projection of the data vector $\mathbf{y}$ on

$$(L_{F_1} + \ldots + L_{F_k})^\perp = L_{F_1}^\perp \cap \ldots \cap L_{F_k}^\perp,$$

is given by

$$\mathbf{r} = (\mathbf{I} - \mathbf{P}_{F_1}) \ldots (\mathbf{I} - \mathbf{P}_{F_k})\mathbf{y},$$

where each term $(\mathbf{I} - \mathbf{P}_{F_i})$ can be regarded as the residual operator in a one-way analysis of variance model, performing the subtraction of averages over F_i-levels from the vector operated on; i.e. a 'sweep'.

As a criterion for orthogonality in concrete situations, we need an explicit condition on the cell counts n_{fg}. To this end, notice that F and G are orthogonal if and only if

$$\mathbf{P}_F\mathbf{P}_G = \mathbf{P}_{F \wedge G}. \tag{3.3}$$

Indeed, if L_F and L_G are geometrically orthogonal, this formula follows immediately from

Lemma 2, since $L_F \cap L_G = L_{F\wedge G}$. Conversely, if (3.3) holds, we have in particular, since an orthogonal projection is symmetric,

$$\mathbf{P}_F\mathbf{P}_G = (\mathbf{P}_F\mathbf{P}_G)^* = \mathbf{P}_G^*\mathbf{P}_F^* = \mathbf{P}_G\mathbf{P}_F.$$

Now, put $H = F \wedge G$. For $i_1, i_2 \in I$, we shall compute the (i_1, i_2)th element of the two matrices $\mathbf{P}_F\mathbf{P}_G$ and $\mathbf{P}_H$. As to the first one, let $f = \varphi_F(i_1)$ and $g = \varphi_G(i_2)$; then, with an obvious notation for indicator functions, we have

$$(\mathbf{P}_F\mathbf{P}_G)_{i_1 i_2} = \sum_{i\in I} n_f^{-1} 1_{\{\varphi_F(i)=f\}} n_g^{-1} 1_{\{\varphi_G(i)=g\}} = n_{fg}/(n_f n_g).$$

For the second matrix, we have

$$(\mathbf{P}_H)_{i_1 i_2} = \begin{cases} n_h^{-1} & \text{for} \quad \varphi_H(i_1) = \varphi_H(i_2) = h, \\ 0 & \text{for} \quad \varphi_H(i_1) \neq \varphi_H(i_2). \end{cases}$$

Now, in the case $\varphi_H(i_1) \neq \varphi_H(i_2)$ we have obviously, as in the construction of the minimum in Example 1 of § 2.6, $n_{fg} = 0$, which means that the (i_1, i_2)th element of both matrices is 0 in this case. Hence, our condition for orthogonality results in Proposition 1.

PROPOSITION 1. *Factors F and G are orthogonal if and only if the relation*

$$n_{fg} n_h = n_f n_g$$

holds for all $f \in F$, $g \in G$ and $h \in H = F \wedge G$ such that f and g are nested in h.

For $F \wedge G = 0$, this criterion simplifies to $n_{fg}|I| = n_f n_g$, which is the well-known condition of *proportional cell counts*. The condition in the general case states that this proportionality condition should hold for each of the subtables of the table (n_{fg}) determined by the levels of $F \wedge G$.

3.3 Example: The balanced k-way table

It follows from Proposition 1 that F and G are orthogonal if $F \times G$ is balanced. In that case, $F \wedge G = 0$, obviously. More generally, let k factors $F_1, \ldots, F_k$ be given and assume that the cell counts $n_{f_1 \ldots f_k}$ of the k-way table $F_1 \times \ldots \times F_k$ are all equal. It is then easy to show that any two factors, formed as products of some of the k factors, are orthogonal. This follows from Proposition 1; the rule for formation of minima of such factors is given in Example 2 of § 2.6.

4 Decomposition of the observation space with respect to an orthogonal design

4.1 Orthogonal designs

Our 'universe', when we make analysis of variance modelling for a given data set, is a set $\mathcal{D}$ of nonequivalent factors, which we shall refer to as the *design*. The idea is that $\mathcal{D}$ should include *all* factors relevant for the model building, also the cross-classifications which are to occur as interaction terms in our models. For example, if data are arranged in a balanced k-way table, $\mathcal{D}$ will typically consist of all (or almost all) possible products of the k 'main factors'.

Throughout this paper, we make the following three assumptions.

Assumption 1. The factor $I \in \mathcal{D}$.

Assumption 2. Any two factors in $\mathcal{D}$ are orthogonal.

Assumption 3. Set $\mathcal{D}$ is closed under the formation of minima.

Notice that only Assumption 2 is really restrictive. The satisfaction of Assumptions 1 and 3 is a matter of including the factor I and the missing minima, if any, and this extension will not destroy the orthogonality. However, in more complicted designs, the extension may introduce new factors without any statistical meaning. Such factors will be called *pseudofactors*.

4.2 *The decomposition induced by* $\mathcal{D}$

Our approach relies on the following main result.

THEOREM 1. *Under Assumptions* 1, 2 *and* 3, *we have a unique decomposition*

$$\mathbf{R}^I = \bigoplus_{G \in \mathcal{D}} V_G$$

of the observation space as a direct sum of orthogonal subspaces V_G $(G \in \mathcal{D})$, *such that for any* $F \in \mathcal{D}$,

$$L_F = \bigoplus_{\substack{G \in \mathcal{D} \\ G \leqslant F}} V_G.$$

Remarks. In the following, we shall frequently consider sums or direct sums taken over subsets of $\mathcal{D}$. For notational convenience, we omit the specification $G \in \mathcal{D}$, writing, for example, the last identity of the theorem as $L_F = \oplus_{G \leqslant F} V_G$.

Since the proof of the theorem is somewhat technical, a few remarks may be of help here. It is well known that a set of orthogonal factors induces a decomposition of the observation space into orthogonal components. In Nelder's (1965, I) discussion of orthogonal block structures, these components are referred to as the *strata* subspaces. They can also be characterized (in this case) as the eigenspaces for the covariance matrix, and the corresponding orthogonal projections (constituting what Nelder calls a *complete binary set*) are the generators in the spectral representation of the *relationship algebra* of the design (James, 1957).

However, the theorem says a little more; namely that there is a canonical way of labelling these subspaces by factors of the design, in such a way that the original projections $\mathbf{P}_F$ are recovered as 'cumulated' projections on the components, and that this property characterizes the decomposition. This result, which generalizes results given by Nelder (1965) and Speed & Bailey (1982), depends crucially on the fact that $\mathcal{D}$ is closed under the formation of minima.

Proof. Consider the trivial identity

$$\mathbf{I} = \prod_{F \in \mathcal{D}} (\mathbf{P}_F + (\mathbf{I} - \mathbf{P}_F)).$$

Expanding this as a sum of products of terms $\mathbf{P}_F$ and $(\mathbf{I} - \mathbf{P}_F)$, we get

$$\mathbf{I} = \sum_{M \subseteq \mathcal{D}} \mathbf{Q}_M,$$

where for each subset M of $\mathcal{D}$ the operator $\mathbf{Q}_M$ is defined as

$$\mathbf{Q}_M = \Big(\prod_{F \in M} \mathbf{P}_F\Big)\Big(\prod_{F \notin M} (\mathbf{I} - \mathbf{P}_F)\Big).$$

As products of commuting orthogonal projections, these operators $\mathbf{Q}_M$ are again orthogonal projections. The corresponding subspaces V_M are pairwise orthogonal, since (obviously)

$$M_1 \neq M_2 \Rightarrow \mathbf{Q}_{M_1}\mathbf{Q}_{M_2} = 0.$$

Hence, the formula $\mathbf{I} = \sum_{M \subseteq \mathcal{D}} \mathbf{Q}_M$ corresponds to a decomposition of $\mathbf{R}^I$ as a direct orthogonal sum of $2^{\operatorname{card} \mathcal{D}}$ subspaces V_M, one for each subset M of $\mathcal{D}$. However, many of these subspaces are trivial. Indeed, $\mathbf{Q}_M$ can only be different from 0 if (a) and (b) are satisfied.

(a) If $F \in M$ and $F \leqslant F'$, then

$$F' \in M.$$

This follows because otherwise $\mathbf{Q}_M = 0$ follows from the fact that the product of $\mathbf{P}_F$ and $\mathbf{I} - \mathbf{P}_{F'}$ is 0 for $F \leqslant F'$. Moreover, if $\mathbf{Q}_M$ is to be nonzero, we must assume that the minimum G of all factors in M is itself an element of M, that is as follows.

(b) Factor

$$G = \min_{F \in M} F$$

belongs to M.

This follows because otherwise the expression for $\mathbf{Q}_M$ contains the product of $(\mathbf{I} - \mathbf{P}_G)$ and $\prod_{F \in M} \mathbf{P}_F = \mathbf{P}_G$, which is 0. Thus, if we restrict our attention to nonvanishing terms of the decomposition, we need only take into account the subsets M of $\mathcal{D}$ satisfying (a) and (b). However, these sets M can be characterized in a simpler way. A set M satisfying (b) contains its own minimum, and if (a) is also satisfied, M must contain any factor finer than that minimum. Hence, the only sets M to be considered are those of the form

$$M_G = \{F \in \mathcal{D} \mid G \leqslant F\},$$

which means that we have now reduced to a decomposition into at most card $\mathcal{D}$ nontrivial subspaces,

$$\mathbf{R}^I = \bigoplus_{G \in \mathcal{D}} V_{M_G}.$$

For simplicity of notation, we write V_G and $\mathbf{Q}_G$ instead of V_{M_G} and $\mathbf{Q}_{M_G}$ in the following.

Our next step is to show that this decomposition satisfies the condition of the theorem. We notice that

$$\mathbf{P}_F\mathbf{Q}_G = \begin{cases} \mathbf{Q}_G & \text{for } G \leqslant F, \\ 0 & \text{otherwise.} \end{cases}$$

Hence,

$$\mathbf{P}_F = \mathbf{P}_F\Big(\sum_{G \in \mathcal{D}} \mathbf{Q}_G\Big) = \sum_G \mathbf{P}_F\mathbf{Q}_G = \sum_{G \leqslant F} \mathbf{Q}_G,$$

or, equivalently,

$$L_F = \bigoplus_{G \leqslant F} V_G.$$

Finally, we must show that the decomposition constructed here is the only one satisfying the condition of the theorem. Suppose we had another, say

$$\mathbf{R}^I = \bigoplus_G V'_G \quad \text{or} \quad \mathbf{I} = \sum_G \mathbf{Q}'_G.$$

Let $\mathcal{D}_0 \subseteq \mathcal{D}$ denote the set of factors G for which $\mathbf{Q}'_G \neq \mathbf{Q}_G$. We intend to show that $\mathcal{D}_0$ is empty. Suppose that $\mathcal{D}_0$ has an element G_0. If G_0 is not the minimal element of $\mathcal{D}$, we conclude from

$$\sum_{G<G_0} \mathbf{Q}_G = \mathbf{P}_{G_0} - \mathbf{Q}_{G_0} \neq \mathbf{P}_{G_0} - \mathbf{Q}'_{G_0} = \sum_{G<G_0} \mathbf{Q}'_G$$

that there must be some G strictly coarser than G_0 for which $\mathbf{Q}_G \neq \mathbf{Q}'_G$. Thus, the set $\mathcal{D}_0$ has the property that for any element which is not equal to $F_0 = \min \mathcal{D}$, it has an element which is strictly coarser. From this, it is easy to conclude that $\mathcal{D}_0$ must either be empty or contain the minimal element F_0. But $F_0 \in \mathcal{D}_0$ is impossible, since $\mathbf{Q}_{F_0} = \mathbf{P}_{F_0}$ and $\mathbf{Q}'_{F_0} = \mathbf{P}_{F_0}$. Thus, $\mathcal{D}_0$ is empty, and the theorem is proved.

The names $\mathbf{Q}_G$ and V_G will be standard notation in the following. Notice that we have now *two* families of orthogonal projections indexed by $\mathcal{D}$, the canonical projections $\mathbf{P}_F$ on the factor spaces L_F (depending on F only, not on the remaining factors of the design), and the projections $\mathbf{Q}_G$ on the subspaces V_G given by the theorem. The connection between these families is

$$\mathbf{P}_F = \sum_{G \leqslant F} \mathbf{Q}_G.$$

Conversely, $\mathbf{Q}_G$ can be expressed by the projections $\mathbf{P}_F$ as follows.

COROLLARY 1. *The projection* $\mathbf{Q}_G$ *can be expressed as*

$$\mathbf{Q}_G = \mathbf{P}_G \prod_{F<G} (\mathbf{P}_G - \mathbf{P}_F).$$

Proof. In the proof of the theorem, $\mathbf{Q}_G$ was constructed as

$$\mathbf{Q}_G = \mathbf{Q}_{M_G} = \Big(\prod_{F \geqslant G} \mathbf{P}_F\Big)\Big(\prod_{F \geqslant G} (\mathbf{I} - \mathbf{P}_F)\Big) = P_G \prod_{F \neq G} (\mathbf{I} - \mathbf{P}_F)$$

$$= \mathbf{P}_G \prod_{F \neq G} (\mathbf{P}_G - \mathbf{P}_{F \wedge G}) = \mathbf{P}_G \prod_{H<G} (\mathbf{P}_G - \mathbf{P}_H).$$

If desirable, an expression of $\mathbf{Q}_G$ as a linear combination

$$\mathbf{Q}_G = \sum_F a_G^F \mathbf{P}_F$$

can be obtained from Corollary 1 by straightforward applications of the distributive law and the rule $\mathbf{P}_F \mathbf{P}_{F'} = \mathbf{P}_{F \wedge F'}$. These computations are simplified if it is noticed that only the maximal factors F among those strictly coarser than G need to be included in the product $\prod (\mathbf{P}_G - \mathbf{P}_F)$, since terms $(\mathbf{P}_G - \mathbf{P}_{F'})$ with $F' < F < G$ are absorbed by $(\mathbf{P}_G - \mathbf{P}_F)$ anyway. Similarly, the first term P_G can be omitted if the product is nonempty, i.e. if $G \neq \min \mathcal{D}$. Notice also that $a_G^F \neq 0$ only for $F \leqslant G$.

It is easy to show that the $\mathcal{D} \times \mathcal{D}$ matrix (a_G^F) of integer coefficients can be characterized as the (transposed) inverse of the $\mathcal{D} \times \mathcal{D}$ matrix $(1_{\{G \leqslant F\}})$. In the combinatorial theory of finite lattices, the function a_G^F on $\mathcal{D} \times \mathcal{D}$ is known as the *Möbius function* (Speed & Bailey, 1982; Aigner, 1979). However, this characterization plays a secondary role in the present exposition. It is usually simpler to work directly with the expression of the $\mathbf{P}_F$'s as sums of $\mathbf{Q}_G$'s.

4.3 Example

Consider a connected two-way table with proportional cell counts, that is $R \wedge C = 0$ and $R \perp C$. We assume $n_{rc} > 1$ for some (r, c), that is I and $R \times C$ are not equivalent. Put $\mathcal{D} = \{I, R \times C, R, C, 0\}$. Then $\mathcal{D}$ satisfies the Assumptions 1, 2 and 3. The ordering

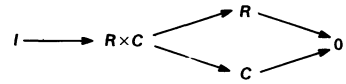

Figure 1. *Nestedness relations between the factors of example of* § 4.3.

structure of $\mathcal{D}$ is given by Fig. 1. Applying Corollary 1 or solving the equations $\mathbf{P}_F = \sum_{G \leq F} \mathbf{Q}_G$ directly, we obtain the following well-known formulae:

$$\begin{aligned}
\mathbf{Q}_I &= \mathbf{I} - \mathbf{P}_{R\times C},\\
\mathbf{Q}_{R\times C} &= \mathbf{P}_{R\times C} - \mathbf{P}_R - \mathbf{P}_C + \mathbf{P}_0,\\
\mathbf{Q}_R &= \mathbf{P}_R - \mathbf{P}_0,\\
\mathbf{Q}_C &= \mathbf{P}_C - \mathbf{P}_0,\\
\mathbf{Q}_0 &= \mathbf{P}_0.
\end{aligned}$$

Accordingly, the matrix

$$(a_G^F) = \begin{bmatrix} 1 & -1 & 0 & 0 & 0\\ 0 & 1 & -1 & -1 & 1\\ 0 & 0 & 1 & 0 & -1\\ 0 & 0 & 0 & 1 & -1\\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

is the transposed inverse of

$$(1_{\{G \leq F\}}) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0\\ 1 & 1 & 0 & 0 & 0\\ 1 & 1 & 1 & 0 & 0\\ 1 & 1 & 0 & 1 & 0\\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

5 The analysis of variance table and the factor structure diagram

By the analysis of variance table for the data set $\mathbf{y}$ in the design $\mathcal{D}$ we mean a table which for each $G \in \mathcal{D}$ gives the quantities $d_G = \dim V_G$ and $\mathrm{SSD}_G = \| \mathbf{Q}_G \mathbf{y} \|^2$. It should be emphasized that the analysis of variance table is not related to a particular statistical model. We are not talking about models yet. The analysis of variance table is a computational tool, containing the quantities relevant for hypothesis testing and variance component estimation in all analysis of variance models that can be stated in terms of factors from $\mathcal{D}$.

5.1 Construction of the analysis of variance table

Computationally, the analysis of variance table can be obtained as follows. Put $\mathrm{SS}_F = \|\mathbf{P}_F \mathbf{y}\|^2$ (and notice that we distinguish between SS = sum of squares and SSD = sum of

squares of deviations). The quantities SS_F are obtained as

$$\text{SS}_F = \sum_{f \in F} S_f^2/n_f,$$

where S_f denotes the sum of all observations on the level f of F. Let $\mathbf{Q}_G = \sum_F a_G^F \mathbf{P}_F$ be the expression of $\mathbf{Q}_G$ as an integer linear combination of the projections $\mathbf{P}_F$. Then

$$\text{SSD}_G = \|\mathbf{Q}_G \mathbf{y}\|^2 = \mathbf{y}^* \mathbf{Q}_G \mathbf{y} = \sum_F a_G^F \mathbf{y}^* \mathbf{P}_F \mathbf{y}$$

$$= \sum_F a_G^F \|\mathbf{P}_F \mathbf{y}\|^2 = \sum_F a_G^F \text{SS}_F.$$

Thus, the sums of squares of deviations SSD_G are obtained as linear combinations of the sums of squares SS_F in exactly the same way as the projections $\mathbf{Q}_G$ are obtained as linear combinations of the projections $\mathbf{P}_F$. Similarly, the formula

$$d_G = \operatorname{tr} \mathbf{Q}_G = \operatorname{tr} \sum_F a_G^F \mathbf{P}_F = \sum_F a_G^F \operatorname{tr} \mathbf{P}_F$$

$$= \sum_F a_G^F \dim L_F = \sum_F a_G^F |F|$$

shows that the degrees of freedom d_G can be obtained as linear combinations of the integers $|F|$ in exactly the same manner.

In concrete situations, it is not even necessary to compute the coefficients a_G^F. The formulae for the SSD_G's constitute the solution to the equations

$$\text{SS}_F = \sum_{G \leqslant F} \text{SSD}_G,$$

and it is just as simple to work directly with these equations, solving them recursively as F varies from the coarsest factor (usually 0) to the finest (I). Similarly, degrees of freedom are obtained by solving

$$|F| = \sum_{G \leqslant F} d_G.$$

This can be done by means of a *factor structure diagram*, like Fig. 1, holding the information about the nestedness relations between factors in $\mathscr{D}$. The following examples illustrate this method.

5.2 *Four examples*

The balanced two-way table. Suppose we have a two-way scheme $R \times C$, for simplicity assumed to be balanced with cell counts $n_{R \times C} \geqslant 2$. Let $\mathscr{D}$ consist of the factors occurring in Fig. 1. In order to compute the degrees of freedom d_G, we write as a superscript to each factor the number of (effectively used) factor levels $|F|$ and as a subscript the dimension d_F of V_F. Filling in the superscripts $|F|$ first, it is easy to obtain the subscripts d_F recursively, in each step computing d_F as the difference between the corresponding superscript $|F|$ and the sum of all subscripts to factors G strictly coarser than F. For example, for $|R| = 4$, $|C| = 5$ and $|I| = 2 \times 4 \times 5 = 40$, we get the picture given by Fig. 2a. The sums of squares of deviations are obtained similarly, using superscripts SS_F and subscripts SSD_F, and we obtain the analysis of variance table given as Table 2a.

One-way classification (arbitrary group sizes). For an arbitrary factor F, we take $\mathscr{D} = \{0, F, I\}$. The ordering structure is linear in this case, given by

$$I_{n-k}^n \to F_{k-1}^k \to 0_1^1 \quad (n = |I|, k = |F|).$$

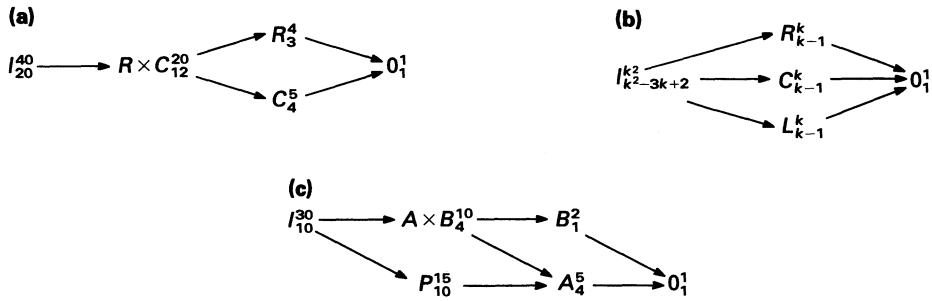

Figure 2. *Factor structure diagrams*: (a) *balanced two-way table*; (b) *Latin square of order k*; (c) *split-plot design.*

The analysis of variance table is given as Table 2b.

Latin square of order k. A Latin square of order k can be described as a design of the form $\mathfrak{D} = \{0, R, C, L, I\}$, where R (rows), C (columns) and L (latin letter) are factors on k levels such that the three cross-classifications $R \times C$, $R \times L$ and $C \times L$ are balanced and equivalent to I. The factor structure diagram is given as Fig. 2b, the analysis of variance table as Table 2c.

Split-plot design (*Cochran & Cox*, 1957, *p.* 293). Suppose that five treatments $a_1, a_2, \ldots, a_5$ are applied to 15 plots, each treatment being applied to 3 plots. Each plot is divided into two subplots, to which the two further treatments b_1 and b_2 are applied. Thus, the relevant factors are: I on 30 levels (subplots); P on 15 levels (plots); $A = \{a_1, a_2, \ldots, a_5\}$; $B = \{b_1, b_2\}$; $A \times B$ on 10 levels. Adding $0 = A \wedge B$, we obtain a set $\mathfrak{D}$ of factors satisfying our conditions for an orthogonal design. The factor structure with

Table 2

Analysis of variance table

(a) Balanced two-way table

Factor	Degrees of freedom	Sum of squares of deviations
0	1	ss_0
C	4	$ss_C - ss_0$
R	3	$ss_R - ss_0$
$R \times C$	12	$ss_{R\times C} - ss_R - ss_C + ss_0$
I	20	$ss_I - ss_{R\times C}$
Sum	40	ss_I

(b) One-way classification

Factor	Degrees of freedom	Sum of squares of deviations
0	1	ss_0
F	$k-1$	$ss_F - ss_0$
I	$n-k$	$ss_I - ss_F$
Sum	n	ss_I

(c) Latin square of order k

Factor	Degrees of freedom	Sum of squares of deviations
0	1	ss_0
L	$k-1$	$ss_L - ss_0$
C	$k-1$	$ss_C - ss_0$
R	$k-1$	$ss_R - ss_0$
I	k^2-3k+2	$ss_I - ss_R - ss_C - ss_L + 2ss_0$
Sum	k^2	ss_I

(d) Split-plot design

Factor	Degrees of freedom	Sum of squares of deviations
0	1	ss_0
A	4	$ss_A - ss_0$
B	1	$ss_B - ss_0$
P	10	$ss_P - ss_A$
$A \times B$	4	$ss_{A\times B} - ss_A - ss_B + ss_0$
I	10	$ss_I - ss_P - ss_{A\times B} + ss_A$
Sum	30	ss_I

degrees of freedom is given by Fig. 2c, and the analysis of variance table as Table 2d.

We have ignored a factor 'replicate', dividing the 15 plots into 3 groups of 5. Depending on the concrete circumstances, this factor may or may not be relevant as a third level of blocking. The inclusion of it is left to the reader as an exercise.

6 Linear models

By a *linear model* we mean a model assuming that the data vector is (the realization of) a normally distributed random vector with covariance matrix $\sigma^2\mathbf{I}$ and mean vector $\boldsymbol{\mu}$ in a specified linear subspace L of $\mathbf{R}^I$. We shall restrict our attention to cases where the linear structure is given as an additive effect of factors from our design, that is

$$L = \sum_{T \in \mathcal{T}} L_T, \quad \mathcal{T} \subseteq \mathcal{D}.$$

Thus, a linear model is specified by a subset $\mathcal{T}$ of $\mathcal{D}$. However, different subsets of $\mathcal{D}$ may specify the same model. For instance, in the two-way design, the two sets $\{R, C\}$ and $\{R, C, 0\}$ represent the same model because $L_R + L_C = L_R + L_C + L_0$.

6.1 *Model formulae*

We shall refer to the subset $\mathcal{T}$ of $\mathcal{D}$ as the *model formula.* This is merely a notational convention, according to which we list the elements of $\mathcal{T}$ separated by pluses instead of commas and without the braces, {}. For example, we talk about the additivity model $R + C$ in a two-way design. Notice, however, that we do not adopt more advanced model formula conventions (Wilkinson & Rogers, 1973), like distributivity of $\times$ over $+$, nesting operations, etc. A model formula in this text is nothing but a set of factors, written in this special way.

6.2 *Parameterizations of the mean*

The intuitive appeal of the model formula notation lies in the fact that it reflects the parametric representation of linear models. For example, the additivity model $R + C$ can be stated as

$$Ey_i = \mu_i = \alpha_r + \beta_c,$$

if we subsume $r = \varphi_R(i)$, $c = \varphi_C(i)$. Or, in vector notation, with $\boldsymbol{\alpha} = (\alpha_r) \in \mathbf{R}^R$, $\boldsymbol{\beta} = (\beta_c) \in \mathbf{R}^C$,

$$\boldsymbol{\mu} = \mathbf{X}_R\boldsymbol{\alpha} + \mathbf{X}_C\boldsymbol{\beta}.$$

More generally, there is an immediate one-to-one correspondence between model formulae and parameterizations of linear models, given by

$$\mathcal{T} \leftrightarrow \left(\boldsymbol{\mu} = \sum_{T \in \mathcal{T}} \mathbf{X}_T\boldsymbol{\alpha}^T\right).$$

Notice that these parameterizations of the mean are usually not one-to-one. In fact, as soon as more than one term is involved, we have a nonidentifiability of parameters, since a constant may be added to the linear parameters of a first factor and subtracted from

those of a second, without changing the mean. We shall *not* discuss restrictions imposed on the parameters in order to make such parameterizations one-to-one. The nonidentifiabilities are usually well justified in the applied context. For instance, in a two-way additive model $R+C$, the parameter functions of interest are typically differences between row parameters, whereas absolute row levels are only meaningful in situations where vanishing (or random) column effect can be assumed. Similarly, estimation of a main effect in a two-way table is usually not meaningful in the presence of interaction. Constraints on parameters (like the usual assumption that summation of any model term over any of its indices should give zero) should be regarded merely as computational tools, if they have to be considered at all. We seem to be in agreement with Nelder (1977) on these matters.

Notice also that we do not impose restrictions on the set $\mathcal{T}$ of terms in a model formula. For example, the interaction model in a two-way table can be written $R\times C$, $R\times C+R$, $R\times C+R+C+0$, etc. These model formulae correspond to different parameterizations of the same model, each possessing its own rules for identifiability of contrasts. Some of these parameterizations may be of limited interest, but there seems to be no *a priori* reason for excluding them.

6.3 *Estimation of the variance*

By $\mathcal{T}^*$ we denote the set of factors $F\in\mathcal{D}$ such that F is marginal to (or equal to) some factor in $\mathcal{T}$. One may think of $\mathcal{T}^*$ as the *maximal model formula*, in the sense that $\mathcal{T}^*$ specifies the same model as $\mathcal{T}$, but with the greatest possible number of redundant terms. By Theorem 1, we have

$$L=\sum_{T\in\mathcal{T}} L_T=\sum_{T\in\mathcal{T}}\Big(\bigoplus_{G\leqslant T} V_G\Big)=\sum_{G\in\mathcal{T}^*} V_G=\bigoplus_{G\in\mathcal{T}^*} V_G.$$

From this we conclude that the orthogonal projection $\mathbf{P}$ on L is given by

$$\mathbf{P}=\sum_{G\in\mathcal{T}^*}\mathbf{Q}_G,$$

while the residual operator, the orthogonal projection on $L^{\perp}$, is given by

$$\mathbf{I}-\mathbf{P}=\sum_{G\notin\mathcal{T}^*}\mathbf{Q}_G.$$

Accordingly, the residual sum of squares can be obtained from the analysis of variance table and the factor structure diagram as the sum of the SSD_G for factors G which are not marginal to factors occurring in the model formula,

$$\mathrm{SSD}^{\mathrm{res}}=\|\mathbf{y}-\mathbf{P}\mathbf{y}\|^2=\sum_{G\notin\mathcal{T}^*}\mathrm{SSD}_G.$$

The degrees of freedom for this residual sum of squares of deviations are obtained similarly as

$$d^{\mathrm{res}}=\sum_{G\in\mathcal{T}^*} d_G.$$

By standard linear model theory, the variance σ^2 should then be estimated by $\hat{\sigma}^2=\mathrm{SSD}^{\mathrm{res}}/d^{\mathrm{res}}$.

6.4 *Test for model reduction*

Let $\mathcal{T}_0^* \subseteq \mathcal{T}^*$ be the maximal model formula for a reduced model $\mathcal{T}_0$, and let $\text{SSD}_0^{\text{res}}$ and d_0^{res} denote the residual sum of squares of deviations and its degrees of freedom in this reduced model. By standard linear model theory, the likelihood ratio test for $\mathcal{T}_0$ against $\mathcal{T}$ is equivalent to the F test

$$F(d_0^{\text{res}} - d^{\text{res}}, d^{\text{res}}) = \frac{(\text{SSD}_0^{\text{res}} - \text{SSD}^{\text{res}})/(d_0^{\text{res}} - d^{\text{res}})}{\text{SSD}^{\text{res}}/d^{\text{res}}}.$$

All quantities in this expression are easily obtained from the analysis of variance table and the factor structure diagram. In particular, in the most important case where $\mathcal{T}_0^*$ is obtained from $\mathcal{T}^*$ be removal of a single factor T, the test for 'no T-effect' becomes

$$F(d_T, d^{\text{res}}) = \frac{\text{SSD}_T/d_T}{\text{SSD}^{\text{res}}/d^{\text{res}}}.$$

6.5 *Estimation of the linear parameters*

We shall restrict our attention to estimation of contrasts of the form $\alpha_{t'}^T - \alpha_{t''}^T$, $t', t'' \in T$, $T \in \mathcal{T}$. The first question posing itself is, of course, whether or not a given contrast of this form can be estimated at all. The answer to this question and the rule for estimation is given by the following theorem.

THEOREM 2. *Consider the model*

$$E\mathbf{y} = \boldsymbol{\mu} = \sum_{T \in \mathcal{T}} \mathbf{X}_T \boldsymbol{\alpha}^T \quad (\boldsymbol{\alpha}^T \in \mathbf{R}^T).$$

For t_0' and $t_0'' \in T_0$, $T_0 \in \mathcal{T}$, the following two conditions are equivalent:

(i) *the parameter function $\alpha_{t_0'}^{T_0} - \alpha_{t_0''}^{T_0}$ is estimable, that is it can be written as a function of $\boldsymbol{\mu}$;*

(ii) *for any other factor $T \in \mathcal{T}$, t_0' and t_0'' are nested in the same level of $T_0 \wedge T$.*

In case of estimability, the maximum likelihood estimate of this contrast is

$$\hat{\alpha}_{t_0'}^{T_0} - \hat{\alpha}_{t_0''}^{T_0} = \bar{y}_{t_0'} - \bar{y}_{t_0''},$$

where $\bar{y}_{t_0} = S_{t_0}/n_{t_0}$ denotes the average of all observations on the level t_0 of T_0. The variance on this estimate is $(n_{t_0'}^{-1} + n_{t_0''}^{-1})\sigma^2$.

We illustrate by an example. In a balanced three-way scheme $A \times B \times C$, consider the model $A \times B + B \times C$, co-ordinatewise parameterized as $\mu_i = \alpha_{ab} + \beta_{bc}$. A contrast of the form $\alpha_{a'b'} - \alpha_{a''b''}$ is estimable if and only if (a', b') and (a'', b'') are nested in the same level of $(A \times B) \wedge (B \times C) = B$; this means that the contrast is estimable if and only if $b' = b''$.

Proof. First assume that (ii) is satisfied, and define $\mathbf{v} \in \mathbf{R}^I$ by

$$v_i = \begin{cases} 1/n_{t_0'} & \text{for } \varphi_{T_0}(i) = t_0', \\ -1/n_{t_0''} & \text{for } \varphi_{T_0}(i) = t_0'', \\ 0 & \text{otherwise.} \end{cases}$$

Notice that $\mathbf{v}^*\mathbf{y} = \bar{y}_{t_0'} - \bar{y}_{t_0''}$. The vector $\mathbf{v}$ belongs to L_{T_0}, because v_i as a function of i is

constant on the classes determined by T_0. For any other factor $T \in \mathcal{T}$, we have

$$\mathbf{P}_T \mathbf{v} = \mathbf{P}_T \mathbf{P}_{T_0} \mathbf{v} = \mathbf{P}_{T \wedge T_0} \mathbf{v} = \mathbf{0},$$

because condition (ii) implies that the two classes $\varphi_{T_0}^{-1}(t_0')$ and $\varphi_{T_0}^{-1}(t_0'')$ are contained in *the same* $T_0 \wedge T$ class, from which it follows that averaging v_i over an arbitrary $T_0 \wedge T$ class gives 0. Hence, the linear functional $\mathbf{v}^*$ vanishes on the subspace L_T for $T \neq T_0$. From this we conclude that

$$\mathbf{v}^*(\boldsymbol{\mu}) = \mathbf{v}^*\left(\sum_{T \in \mathcal{T}} \mathbf{X}_T \boldsymbol{\alpha}^T\right) = \mathbf{v}^* \mathbf{X}_{T_0} \boldsymbol{\alpha}^{T_0} = \alpha_{t_0'}^{T_0} - \alpha_{t_0''}^{T_0}.$$

This means that our contrast is a function of $\boldsymbol{\mu}$. The maximum likelihood estimate is given by

$$\mathbf{v}^* \hat{\boldsymbol{\mu}} = \mathbf{v}^* \mathbf{P} \mathbf{y} = (\mathbf{P}\mathbf{v})^* \mathbf{y} = \mathbf{v}^* \mathbf{y} = \bar{y}_{t_0'} - \bar{y}_{t_0''}.$$

The calculation of the variance on this estimate is straightforward. It remains to be shown that (i) implies (ii). Suppose that (ii) is not satisfied, that is, there exists a factor $T \in \mathcal{T}$ such that t_0' and t_0'' are on different levels of $T_0 \wedge T$. Put $H = T_0 \wedge T$, and let h' and h'' denote the corresponding levels of H. Let M' and M_0' denote the subsets of T and T_0, respectively, of factor levels nested in the level h' of H. Now, suppose that the corresponding parameter vectors $\boldsymbol{\alpha}^T$ and $\boldsymbol{\alpha}^{T_0}$ are modified by the addition of a constant $\lambda \neq 0$ to α_t^T for $t \in M'$ and the subtraction of that same constant from $\alpha_{t_0}^{T_0}$ for $t_0 \in M_0'$. This will leave the mean $\boldsymbol{\mu}$ unchanged, while the contrast $\alpha_{t_0'}^{T_0} - \alpha_{t_0''}^{T_0}$ decreases by λ. From this we conclude that the contrast is not estimable.

7 Variance component models

By a *variance component model* in the design $\mathcal{D}$ we mean a model of the form

$$\mathbf{y} = \sum_{T \in \mathcal{T}} \mathbf{X}_T \boldsymbol{\alpha}^T + \sum_{B \in \mathcal{B}} \sigma_B \mathbf{X}_B \mathbf{u}^B,$$

where $\mathcal{T}$ and $\mathcal{B}$ are subsets of $\mathcal{D}$, $\boldsymbol{\alpha}^T = (\alpha_t^T) \in \mathbf{R}^T$ $(T \in \mathcal{T})$ and $\sigma_B \geqslant 0$ $(B \in \mathcal{B})$ are unknown parameters, and $\mathbf{u}^B = (u_b^B) \in \mathbf{R}^B$ $(B \in \mathcal{B})$ denote independent, normalized normally distributed vectors.

Co-ordinatewise, we can write this model as

$$y_i = \sum_{T \in \mathcal{T}} \alpha_t^T + \sum_{B \in \mathcal{B}} \sigma_B u_b^B,$$

subsuming $t = \varphi_T(i)$ and $b = \varphi_B(i)$. The idea is that the observation y_i is assumed to come out as a sum of *fixed effects* α_t^T $(T \in \mathcal{T})$ and *random effects* $\sigma_B u_b^B$ $(B \in \mathcal{B})$. The variance on a single observation y_i is

$$\operatorname{var}(y_i) = \sum_{B \in \mathcal{B}} \sigma_B^2,$$

and the parameters σ_B^2 are, accordingly, called *variance components.*

Alternatively, we may specify the model by mean and covariance matrix of the data set $\mathbf{y}$. We have

$$E\mathbf{y} = \boldsymbol{\mu} = \sum_{T \in \mathcal{T}} \mathbf{X}_T \boldsymbol{\alpha}^T, \quad \operatorname{cov}(\mathbf{y}) = \sum_{B \in \mathcal{B}} \sigma_B^2 \mathbf{X}_B \mathbf{X}_B^*.$$

7.1 *An example*

Suppose we have a balanced two-way table $R \times C$ with $n_{R\times C} \geqslant 2$, and put $\mathscr{D} = \{0, R, C, R \times C, I\}$. Consider the variance component model given by

$$\mathscr{T} = \{R, C\}, \quad \mathscr{B} = \{R \times C, I\}.$$

Co-ordinatewise, this model can be written

$$y_i = \alpha_r + \beta_c + \omega v_{rc} + \sigma u_i,$$

where α_r and β_c are the row and column parameters, respectively, ω^2 and σ^2 the variance components, and v_{rc} and u_i $((r, c) \in R \times C, i \in I)$ are independent random variables, normally distributed with mean 0 and variance 1. This is the *two-way additive model with random interaction*, frequently referred to as the justification for fitting an additive model to the cell averages in situations where the interaction is too large to be ignored against the intracell variation.

7.2 *Model formulae*

A variance component model is specified by the two subsets $\mathscr{T}$ and $\mathscr{B}$ of $\mathscr{D}$. We shall condense this information in a single model formula, adopting the convention that random factors should be in brackets. Thus, the two-way additive model with random interaction is written

$$R + C + [R \times C + I],$$

and the general idea is to write $\mathscr{T} + [\mathscr{B}]$. Notice that linear models are variance component models with $\mathscr{B} = \{I\}$, and that our conventions for model formulae are consistent with those introduced in § 6.1 if an error term '$+[I]$' is subsumed. These ideas will be familiar to GENSTAT users (the model formulae $\mathscr{T}$ and $\mathscr{B}$ are simply those occurring in the 'TREATMENTS' and 'BLOCKS' directives).

7.3 *Assumptions*

As in our discussion of linear models, $\mathscr{T}$ is an arbitrary subset of our orthogonal design $\mathscr{D}$. However, $\mathscr{B}$ is assumed to satisfy the following conditions.

Condition 1. Assume $I \in \mathscr{B}$.

Condition 2. All factors in $\mathscr{B}$ are balanced.

Condition 3. Assume $\mathscr{B}$ is closed under the formation of minima.

Condition 4. The matrices $\mathbf{X}_B \mathbf{X}_B^*$ are linearly independent.

Condition 1 means that an 'error term', taking care of the random variation between experimental units, should be present in the model. In practice, this condition seems to be unrestrictive.

Condition 2 is obviously restrictive, but indispensable. It is well known that even for the simplest model, a one-way model with random variation between groups, a satisfactory solution of the unbalanced case requires a technique going far beyond classical analysis of variance methods. The immediate reason for this is that formula (3.1), which, as we shall see, establishes the link between the variance components and the spectral decomposition of the covariance matrix, is only valid for balanced factors.

Similarly, Condition 3 is necessary for an algebraically nice solution, and somewhat

restrictive. The treatment of variance component models not satisfying Condition 3 can, to some extent, be based on the extension to a larger model (including some random 'pseudo' factors) satisfying Condition 3. The simplest example is the two-way model $0+[R+C+I]$ with both main effects random. Extension to $0+[0+R+C+I]$ yields simple and relatively well-behaved estimates of the variance components. However, these estimates may correspond to a covariance matrix (in the original model) which is not positively definite. In particular, the estimated variance on the grand mean $\bar{y}$ may come out negative.

Condition 4 ensures identifiability of the variance components; compare with the parameterization of $\operatorname{cov}(\mathbf{y})$. Linear dependence seems to occur only in pathological situations (the simplest example is a Latin square of order 2, with the three 'main' factors and I as random).

Notice that we do not make explicit assumptions against nonestimability of variance components due to confounding with fixed effects. Obviously, a variance component σ_B^2 can not be estimated if $\mathcal{T}$ contains a factor finer than B. The similar problem for linear models occurs when $L=\mathbf{R}^I$, with zero degrees of freedom left for the residual. Formally, it is an advantage not to exclude models with such nonestimable variance components, see the above remarks on the model $0+[0+R+C+I]$, where σ_0^2 is nonestimable in exactly this sense. However, some of our later results on estimation and hypothesis testing are based on the (subsumed) assumption that the degrees of freedom involved are strictly positive.

7.4 *The null analysis of variance*

The set $\mathcal{B}$ of random factors satisfies, in particular, Assumptions 1, 2 and 3 in § 4.1 for an orthogonal design. Hence, by Theorem 1, $\mathcal{B}$ induces a decomposition of $\mathbf{R}^I$ similar to that induced by $\mathcal{D}$. In order to distinguish, components of this new decomposition will be equipped with the superscript 0. Thus,

$$\mathbf{R}^I = \bigoplus_{B\in\mathcal{B}} V_B^0$$

is the decomposition induced by $\mathcal{B}$. Or, in terms of orthogonal projections,

$$\mathbf{I} = \sum_{B\in\mathcal{B}} \mathbf{Q}_B^0.$$

Sums of squares of deviations and their degrees of freedom are similarly denoted by

$$\mathrm{SSD}_B^0 = \|\mathbf{Q}_B^0\mathbf{y}\|^2, \quad d_B^0 = \dim V_B^0.$$

The condensed analysis of variance table, giving for each $B\in\mathcal{B}$ the quantities SSD_B^0 and d_B^0, corresponds to what Nelder (1965) calls the *null analysis of variance*, the analysis without treatment structure. The components of the decomposition after $\mathcal{B}$ are called *error strata.*

The decomposition induced by $\mathcal{B}$ is coarser than that induced by the whole design $\mathcal{D}$, in the sense that each V_B^0 is a direct sum of some of the subspaces V_G ($G\in\mathcal{D}$). We say that the factor G *belongs to* B*-stratum* if $V_G \subseteq V_B^0$. The rule for allocation of factors to strata follows from the following proposition

PROPOSITION 2. *We have*

$$V_B^0 = \bigoplus_{G\in\mathcal{D}_B} V_G,$$

where $\mathscr{D}_B$ consists of those factors $G \in \mathscr{D}$ for which B is the coarsest random factor finer than G, that is

$$\mathscr{D}_B = \left\{ G \in \mathscr{D} \mid B = \min_{\substack{B' \in \mathscr{B} \\ G \leqslant B'}} B' \right\}.$$

Thus, the rule for allocation to strata by means of the factor structure diagram is that a factor G belongs to the stratum of the coarsest random factor finer than G. Notice that the decomposition after $\mathscr{D}$ may have components V_G of dimension zero; such factors G are assigned to any stratum according to the rule $V_G \subseteq V_B^0$, while the proposition assigns G uniquely to a single stratum. However, the allocation of such factors to strata is irrelevant for the analysis. The corresponding lines of the analysis of variance table can simply be deleted.

Proof. Define a mapping $S : \mathscr{D} \to \mathscr{B}$ by

$$S(G) = \min_{\substack{B' \in \mathscr{B} \\ G \leqslant B'}} B'.$$

Then $\mathscr{D}_B = S^{-1}(B)$, and the sets $\mathscr{D}_B$ are seen to form a partitioning of $\mathscr{D}$. Moreover, for any fixed $B_0 \in \mathscr{B}$, we have

$$G \leqslant B_0 \Leftrightarrow S(G) \leqslant B_0.$$

Indeed, B_0 is finer than G if and only if B_0 is finer than the coarsest factor in $\mathscr{B}$ which is finer than G. Now, the set of factors G satisfying this can be rewritten as follows:

$$\begin{aligned} \{G \in \mathscr{D} \mid S(G) \leqslant B_0\} &= S^{-1}(\{B \in \mathscr{B} \mid B \leqslant B_0\}) \\ &= \bigcup_{\substack{B \in \mathscr{B} \\ B \leqslant B_0}} S^{-1}(B) = \bigcup_{\substack{B \in \mathscr{B} \\ B \leqslant B_0}} \mathscr{D}_B. \end{aligned}$$

This means that the set of factors $G \in \mathscr{D}$ coarser than a given random factor B_0 equals the (disjoint) union of the sets $\mathscr{D}_B$ for $B \leqslant B_0$. Now define

$$W_B^0 = \bigoplus_{G \in \mathscr{D}_B} V_G.$$

Obviously, these subspaces constitute a decomposition of $\mathbf{R}^I$ as a direct sum of orthogonal subspaces, formed by collapse of the subspaces V_G according to the partitioning $\mathscr{D} = \cup\, \mathscr{D}_B$. From what was shown above, we conclude that

$$L_{B_0} = \bigoplus_{\substack{G \in \mathscr{D} \\ G \leqslant B_0}} V_G = \bigoplus_{\substack{B \in \mathscr{B} \\ B \leqslant B_0}} \left(\bigoplus_{G \in \mathscr{D}_B} V_G \right) = \bigoplus_{\substack{B \in \mathscr{B} \\ B \leqslant B_0}} W_B^0.$$

Hence, the decomposition $\mathbf{R}^I = \oplus W_B^0$ satisfies the condition of Theorem 1, for the decomposition with respect to $\mathscr{B}$. Since this condition was shown to characterize the decomposition uniquely, we must have

$$V_B^0 = W_B^0 = \bigoplus_{G \in \mathscr{D}_B} V_G,$$

which concludes the proof.

7.5 The spectral decomposition of the covariance matrix

PROPOSITION 3. *The two sets of matrices*

$$\{\mathbf{X}_B\mathbf{X}_B^* \mid B \in \mathcal{B}\}, \quad \{\mathbf{Q}_B^0 \mid B \in \mathcal{B}\}$$

span the same linear subspace of $\mathbf{R}^{I\times I}$.

Proof. It sufficies to show that any of the matrices $\mathbf{X}_B\mathbf{X}_B^*$ can be written as a linear combination of the matrices $\mathbf{Q}_B^0$, and vice versa. Since the random factors are assumed to be balanced, we have, see formula (3.1)

$$\mathbf{X}_B\mathbf{X}_B^* = n_B\mathbf{P}_B = n_B \sum_{\substack{B'\in\mathcal{B}\\ B'\leqslant B}} \mathbf{Q}_{B'}^0.$$

Conversely, by the remarks following Corollary 1 in § 4.2, we have an expression

$$\mathbf{Q}_B^0 = \sum_{B'\in\mathcal{B}} b_B^{B'}\mathbf{P}_{B'} = \sum_{B'\in\mathcal{B}} b_B^{B'} n_{B'}^{-1}\mathbf{X}_{B'}\mathbf{X}_{B'}^*,$$

with integer coefficients $b_B^{B'}$, corresponding to the coefficients a_G^F of § 4.2.

It follows from Proposition 3 that we have an alternative parametrization of the covariance matrix as

$$\operatorname{cov}(\mathbf{y}) = \sum_{B\in\mathcal{B}} \lambda_B\mathbf{Q}_B^0,$$

where the new parameters λ_B $(B\in\mathcal{B})$ are the eigenvalues of the covariance matrix corresponding to the eigenspaces V_B^0. The explicit solution of the variance component model is based on this parameterization, which relies heavily on our Conditions 1–4. In particular, Condition 3, which was noticed by Jensen (1979) in the case of balanced k-factor designs, is essential. Szatrowski & Miller (1980) give a similar condition. There they give the criterion for existence of explicit maximum likelihood estimates that the set of all co-ordinatewise products of columns of the $\mathcal{B}\times\{0, F_1, \ldots, F_k\}$ matrix $((1_{\{F\leqslant B\}}))$ has exactly card $\mathcal{B}$ elements, but that they do not acknowledge Jensen's (equivalent) condition that the set of rows of this matrix is closed under co-ordinatewise multiplication.

The connection between the two parameterizations is obtained as follows:

$$\sum_{B'\in\mathcal{B}} \lambda_{B'}\mathbf{Q}_{B'}^0 = \sum_{B'\in\mathcal{B}} \lambda_{B'}\Big(\sum_{B\in\mathcal{B}} b_{B'}^B n_B^{-1}\mathbf{X}_B\mathbf{X}_B^*\Big) = \sum_{B\in\mathcal{B}} n_B^{-1}\Big(\sum_{B'\in\mathcal{B}} b_{B'}^B\lambda_{B'}\Big)\mathbf{X}_B\mathbf{X}_B^*;$$

that is

$$\sigma_B^2 = n_B^{-1}\sum_{B'\in\mathcal{B}} b_{B'}^B\lambda_{B'}, \tag{7.1}$$

where the coefficents $b_{B'}^B$ are determined by $\mathbf{Q}_{B'}^0 = \sum_B b_{B'}^B\mathbf{P}_B$. And

$$\sum_{B\in\mathcal{B}} \sigma_B^2\mathbf{X}_B\mathbf{X}_B^* = \sum_{B\in\mathcal{B}} \sigma_B^2 n_B\mathbf{P}_B = \sum_{B\in\mathcal{B}} \sigma_B^2 n_B \sum_{\substack{B'\in\mathcal{B}\\ B'\leqslant B}} \mathbf{Q}_{B'}^0 = \sum_{B'\in\mathcal{B}} \Big(\sum_{\substack{B\in\mathcal{B}\\ B'\leqslant B}} n_B\sigma_B^2\Big)\mathbf{Q}_{B'}^0;$$

that is

$$\lambda_{B'} = \sum_{\substack{B\in\mathcal{B}\\ B'\leqslant B}} n_B\sigma_B^2. \tag{7.2}$$

7.6 *Negative variance components*

Our discussion of the parameterization of the covariance matrix by its eigenvalues λ_B ignores the problem of specifying the domain of variation for these new parameters. Proposition 3 gives an identity between the linear spaces spanned by two sets of matrices, but the corresponding cones of nonnegative linear combinations are usually not identical. Formula (7.2) expressing λ_B as a linear combination of the variance components shows that nonnegative variance components imply nonnegative eigenvalues, but the converse is not always true. This leads to a well-known problem of 'negative variance components', which can be explained as follows. A nice solution to the model is only available when the parameters are allowed to vary freely in their maximal domain, given by $\lambda_B \geqslant 0$. But this may lead to negative estimates for some of the variance components. We shall not discuss formal procedures for estimation of the variance components in their original domain $\sigma_B^2 \geqslant 0$. In practice, this seems to be a secondary problem. The interpretation of a negative variance component σ_B^2 is that correlations between observations in the same B class are smaller than correlations between observations in different B classes, all other random factor levels kept fixed. This phenomenon is explainable in some applied contexts, and in some it is not. Quite often, the occurrence of a negative variance component estimate can be taken as a welcome opportunity to simplify the model by removal of the corresponding variance component. Of course, a *significantly* negative estimate of a variance component, which should be positive, will always be a problem. But the immediate conclusion in this case seems to be that the model fails to describe data, rather than that a more sophisticated estimation procedure is required. See Nelder (1954) and Searle (1971) for more careful discussions of these matters. In the following, we will simply ignore this problem and work with the extended model given by $\lambda_B \geqslant 0$.

7.7 *Solving the variance component model*

The basic, and classical, observation behind the solution to the variance component model in the 'balanced' case (i.e. under the conditions assumed here) is that the decomposition according to $\mathscr{B}$ decomposes the data vector $\mathbf{y}$ as a sum of stochastically independent components $\mathbf{Q}_B^0\mathbf{y}$, one in each error stratum, and that each of these components is described by its own *linear* model. Indeed, the data components $\mathbf{Q}_B^0\mathbf{y}$ are easily seen to be independent, normally distributed with means,

$$\boldsymbol{\mu}_B = E(\mathbf{Q}_B^0\mathbf{y}) = \mathbf{Q}_B^0(E\mathbf{y}) = \mathbf{Q}_B^0(\boldsymbol{\mu}),$$

and covariance matrices,

$$\operatorname{cov}(\mathbf{Q}_B^0\mathbf{y}) = \lambda_B\mathbf{Q}_B^0.$$

The parameters $\boldsymbol{\mu}_B$ and λ_B of the distribution of $\mathbf{Q}_B^0\mathbf{y}$ are functionally independent of those describing the distributions of the remaining data components. Thus, estimation in the original model reduces to estimation in each stratum of the parameters $\boldsymbol{\mu}_B$ and λ_B. This is straightforward, because the model for $\mathbf{Q}_B^0\mathbf{y}$ is essentially, i.e. in a co-ordinate-free sense, an ordinary linear model with data space V_B^0. The covariance matrix $\lambda_B\mathbf{Q}_B^0$ is a constant λ_B times the 'identity' $\mathbf{Q}_B^0$ on V_B^0, and $\boldsymbol{\mu}_B$ varies in the linear subspace $L \cap V_B^0$, for $L = \sum_{T\in\mathscr{T}} L_T$. The orthogonal projection onto this space is $\mathbf{PQ}_B^0$, for $\mathbf{P}$ = the orthogonal projection on L, as usual, since L and V_B^0 are geomtrically orthogonal. Estimating as usual in a linear model, we obtain the estimates $\hat{\boldsymbol{\mu}}_B = \mathbf{PQ}_B^0\mathbf{y}$ and, provided that $\dim V_B^0 > \dim(L \cap V_B^0)$,

$$\hat{\lambda}_B = \|\mathbf{Q}_B^0\mathbf{y} - \mathbf{PQ}_B^0\mathbf{y}\|^2/(\dim V_B^0 - \dim(L \cap V_B^0)).$$

The estimates $\hat{\boldsymbol{\mu}}_B$ are recombined to

$$\hat{\boldsymbol{\mu}} = \sum_{B \in \mathscr{B}} \hat{\boldsymbol{\mu}}_B = \mathbf{P}\mathbf{y},$$

which is recognized as the estimate for the mean in a *linear* model specified by $\mathscr{T}$.

The estimates $\hat{\lambda}_B$ can be computed from the analysis of variance table and the factor structure diagram as follows. We have

$$\mathbf{P} = \sum_{G \in \mathscr{T}^*} \mathbf{Q}_G,$$

where $\mathscr{T}^*$ is defined in § 6.3, and

$$\mathbf{Q}_B^0 = \sum_{G \in \mathscr{D}_B} \mathbf{Q}_G$$

where $\mathscr{D}_B$ is defined in Proposition 2. Thus, the residual operator for our linear model in B stratum is

$$\mathbf{Q}_B^0 - \mathbf{P}\mathbf{Q}_B^0 = \sum_{G \in \mathscr{D}_B \setminus \mathscr{T}^*} \mathbf{Q}_G,$$

and the residual sum of squares in B stratum is, accordingly,

$$\|\mathbf{Q}_B^0\mathbf{y} - \mathbf{P}\mathbf{Q}_B^0\mathbf{y}\|^2 = \sum_{G \in \mathscr{D}_B \setminus \mathscr{T}^*} \mathrm{SSD}_G.$$

Applying the analogous rules for computation of degrees of freedom, we get

$$\hat{\lambda}_B = \sum \mathrm{SSD}_G \Big/ \sum d_G,$$

where both sums are to be taken over $G \in \mathscr{D}_B \setminus \mathscr{T}^*$, that is over all factors G in B stratum which are not in $\mathscr{T}$ and not marginal to factors in $\mathscr{T}$. Very often, at least for the initial model in a statistical analysis, this set consists of B only, in which case we have the simpler formula

$$\hat{\lambda}_B = \mathrm{SSD}_B / d_B.$$

7.8 *Estimation of the variance components*

Estimates $\hat{\sigma}_B^2$ are immediately obtained from the estimates $\hat{\lambda}_B$ by means of (7.1) or (7.2). With reference to the literature on general variance component models, these are the restricted maximum likelihood (REML) estimates, *without* the variance component constraint $\sigma_B^2 \geqslant 0$. Notice that these estimates are usually *not* χ^2 distributed, except $\hat{\sigma}_I^2$, which is always equal to $\hat{\lambda}_I$. In particular, some of them may be negative. However, the moments of $\hat{\sigma}_B^2$ are not difficult to obtain, and various methods for construction of confidence limits exist (Scheffé, 1959; Searle, 1971).

7.9 *Hypothesis testing: Treatment structure*

Let $\mathscr{T}_0$ be a subset of $\mathscr{T}$ specifying a reduced model $\mathscr{T}_0 + [\mathscr{B}]$. we assume $\mathscr{T}_0^* \subseteq \mathscr{T}^*$ and, accordingly, $L_0 \subseteq L$, where $L_0 = \sum_{T \in \mathscr{T}_0} L_T$. In order to obtain an ordinary F test for the model reduction

$$\mathscr{T} + [\mathscr{B}] \to \mathscr{T}_0 + [\mathscr{B}]$$

we must assume that the corresponding square sum of deviations

$$\|\mathbf{Py}-\mathbf{P}_0\mathbf{y}\|^2=\sum_{G\in\mathcal{T}^*\setminus\mathcal{T}_0^*}\mathrm{SSD}_G,$$

for $\mathbf{P}_0$ equal to the orthogonal projection on L_0, consists of contributions from a single stratum B_0 only, that is $\mathcal{T}^*\setminus\mathcal{T}_0^*\subseteq\mathcal{D}_{B_0}$. Notice that this condition is automatically satisfied in the frequently occuring case where $\mathcal{T}_0^*$ is obtained from $\mathcal{T}^*$ by removal of a single factor.

Under this condition, the model reduction can be regarded as a reduction of the linear model for the data component $\mathbf{Q}_{B_0}^0\mathbf{y}$, while the models for the remaining data components are left unchanged. Accordingly, the likelihood ratio test takes the form of an ordinary F test for reduction of the linear model in B_0 stratum,

$$F(\textstyle\sum_1 d_G,\sum_2 d_G)=\dfrac{\sum_1 \mathrm{SSD}_G/\sum_1 d_G}{\sum_2 \mathrm{SSD}_G/\sum_2 d_G},$$

where $\sum_1$ stands for summation over $G\in\mathcal{T}^*\setminus\mathcal{T}_0^*$ and $\sum_2$ for summation over $G\in\mathcal{D}_{B_0}\setminus\mathcal{T}^*$. The rules for inclusion of terms in nominator and denominator are exactly as in the test for reduction of a linear model, see § 6.4, except that only factors from B_0 stratum should be taken into account. Usually, when forming the analysis of variance table to analyse a variance component model, it is convenient to arrange the lines in such a way that strata are collected as subtables, as in output from the GENSTAT 'ANOVA' algorithm. Under this convention, tests for reduction of linear structure are carried out exactly as in the case of a linear model, on the basis of the relevant subtable and the factor structure diagram.

In more complicated situations, it is sometimes desirable to test reductions of linear structure which do *not* take place in a single error stratum. This happens, typically, when a partial confounding of a treatment factor T with a blocking B induces a nontrivial minimum $B\wedge T$. Removal of T from the model formula in this case implies removal of T *and* the pseudofactor $B\wedge T$ from the maximal model formula $\mathcal{T}^*$. Formally, this corresponds to simultaneous reduction of linear models for separate data sets, and there seems to be no standard way of doing it. The immediate thing to do is to perform the relevant F test in each of the strata involved. If no decisive conclusion comes out of this, some kind of weighted test statistic (e.g. the likelihood ratio), summarizing the information from different strata, may be considered.

7.10 Hypothesis testing: Block structure

We shall restrict our attention to model reductions of the form $\mathcal{T}+[\mathcal{B}]\to\mathcal{T}+[\mathcal{B}_0]$, where $\mathcal{B}_0$ is obtained from $\mathcal{B}$ by removal of a single factor B. Thus, in parametric terms, we are considering the hypothesis $\sigma_B^2=0$.

In order to derive an explicit test, we must assume that $\mathcal{B}_0$, as well as $\mathcal{B}$, satisfies Conditions 1–4 of § 7.3. This means that the 'measurement error' I must not be removed, and that $\mathcal{B}_0$ should again be closed under the formation of minima; Conditions 2 and 4 are automatically carried over from $\mathcal{B}$ to $\mathcal{B}_0$. Closedness under minima is satisfied by $\mathcal{B}_0$ if and only if the minimum B_0 of all factors $B'\in\mathcal{B}_0$ which are finer than B is distinct from B, that is

$$B_0=\min_{\substack{B'\in\mathcal{B}_0\\ B\leq B'}}B'\neq B. \tag{7.3}$$

Indeed, if this condition is satisfied, we can obviously not have $B'\wedge B''=B$ for factors B'

and $B'' \in \mathscr{B}_0$, which means that $B' \wedge B''$ must be among the factors left in $\mathscr{B}_0$. Conversely, if (7.3) was not satisfied, we would have a collection of factors from $\mathscr{B}_0$, namely those finer than B, possessing a minimum not in $\mathscr{B}_0$.

Under this condition (7.3), we have

$$\lambda_B = n_B \sigma_B^2 + \lambda_{B_0}.$$

This follows from (7.2), if we note that the expression for λ_B differs from the expression for λ_{B_0} only by the occurrence of the term $n_B \sigma_B^2$. Hence, the hypothesis $\sigma_B^2 = 0$ is equivalent to $\lambda_B = \lambda_{B_0}$. Recalling our interpretation of the model as a product of models for the data components $\mathbf{Q}_B^0 \mathbf{y}$, this hypothesis is formally equivalent to a hypothesis stating that two linear models for separate, independent data sets have the same variance. The usual procedure for test of this is the comparison of the two variance estimates by a two-sided F test on their ratio, that is

$$F(d, d_0) = \hat{\lambda}_B / \hat{\lambda}_{B_0},$$

where d and d_0 are the degrees of freedom occurring in the denominators of the expressions for $\hat{\lambda}_B$ and $\hat{\lambda}_{B_0}$. Large values of this test statistic indicate $\lambda_B > \lambda_{B_0}$, or $\sigma_B^2 > 0$. Small values indicate that σ_B^2 is negative. Hence, the test should be carried out as two sided when negative values of σ_B^2 are to be taken into account. Apart from this, the test is formally equivalent to a test for reduction of treatment structure, namely the test in B_0 stratum for

$$\mathscr{T} + B + [\mathscr{B}_0] \to \mathscr{T} + [\mathscr{B}_0].$$

7.11 *Estimation of contrasts*

It was shown in § 7.7 that the maximum likelihood estimate of the mean coincides with that in the linear model specified by $\mathscr{T}$. In particular, a contrast of the form $\alpha_{t'}^T - \alpha_{t''}^T$ ($T \in \mathscr{T}$, $t', t'' \in T$) should be estimated as in the linear model, by the difference between the corresponding averages. Obviously, the rules for estimability of contrasts are also as in the linear model, so it remains only to give the formula for the variance of an estimated contrast as follows.

PROPOSITION 4. *Let $\alpha_{t_0'}^{T_0} - \alpha_{t_0''}^{T_0}$ be an estimable contrast, that is t_0' and t_0'' are nested in the same level of $T \wedge T_0$ for any other $T \in \mathscr{T}$. Then the estimate $\bar{y}_{t_0'} - \bar{y}_{t_0''}$ has the variance*

$$\sum_{B \in \mathscr{B}} c_B n_B \sigma_B^2,$$

where

$$c_B = \begin{cases} 0 & \text{for } t_0', t_0'' \text{ nested in same level of } T_0 \wedge B \\ n_{h'}^{-1} + n_{h''}^{-1} & \text{for } t_0', t_0'' \text{ nested in distinct levels } h', h'' \text{ of } H = T_0 \wedge B. \end{cases}$$

Proof. Let $\mathbf{v} \in \mathbf{R}^I$ be defined, as in the proof of Theorem 2, by $\mathbf{v}^* \mathbf{y} = \bar{y}_{t_0'} - \bar{y}_{t_0''}$. Since $\mathbf{v} \in L_{T_0}$, we have

$$\begin{aligned} \operatorname{var}(\mathbf{v}^* \mathbf{y}) &= \mathbf{v}^* \operatorname{cov}(\mathbf{y}) \mathbf{v} = \mathbf{v}^* \Big(\sum_{B \in \mathscr{B}} \sigma_B^2 \mathbf{X}_B \mathbf{X}_B^* \Big) \mathbf{v} \\ &= \sum_{B \in \mathscr{B}} n_B \sigma_B^2 \mathbf{v}^* \mathbf{P}_B \mathbf{v} = \sum_{B \in \mathscr{B}} n_B \sigma_B^2 \mathbf{v}^* \mathbf{P}_{T_0} \mathbf{P}_B \mathbf{P}_{T_0} \mathbf{v} = \sum_{B \in \mathscr{B}} n_B \sigma_B^2 \| \mathbf{P}_{T_0 \wedge B} \mathbf{v} \|^2. \end{aligned}$$

The proposition follows if we can show that $\| \mathbf{P}_{T_0 \wedge B} \mathbf{v} \|^2 = c_B$. If we note that the operator

$\mathbf{P}_{T_0 \wedge B}$ replaces each v_i by the average over the corresponding $T_0 \wedge B$ class, this is a matter of straightforward computations which are left to the reader.

PROPOSITION 5. *Suppose that T_0 belongs to B_0 stratum and that t_0' and t_0'' are nested in the same level of $T_0 \wedge B$ for any random factor B which is strictly coarser than B_0. Then*

$$\text{var}\,(\bar{y}_{t_0'} - \bar{y}_{t_0''}) = (n_{t_0'}^{-1} + n_{t_0''}^{-1})\lambda_{B_0}.$$

Proof. Under these assumptions, we obviously have $\mathbf{v} \in V_{B_0}^0$, see the proof of Proposition 4, since $\mathbf{v} \in L_{B_0}$ but $\mathbf{P}_B \mathbf{v} = 0$ for $B < B_0$, $B \in \mathcal{B}$. Hence

$$\text{var}\,(\bar{y}_{t_0'} - \bar{y}_{t_0''}) = \mathbf{v}^* \,\text{cov}\,(\mathbf{y})\mathbf{v} = \mathbf{v}^* \Big(\sum_{B \in \mathcal{B}} \lambda_B \mathbf{Q}_B^0 \Big) \mathbf{v}$$
$$= \lambda_{B_0} \mathbf{v}^* \mathbf{Q}_{B_0}^0 \mathbf{v} = \lambda_{B_0} \|\mathbf{v}\|^2 = \lambda_{B_0}(n_{t_0'}^{-1} + n_{t_0''}^{-1}).$$

Contrasts satisfying the condition of Proposition 5 are called *estimable in a single stratum*, namely B_0 stratum. For such contrasts, and only for such contrasts, can the pairwise comparison of the levels t_0' and t_0'' be performed as an exact t test, since the estimated variance on $\bar{y}_{t_0'} - \bar{y}_{t_0''}$ is χ^2 distributed.

8 Two examples

8.1 *The split-plot experiment*

In the split-plot design of the example in § 5.2, consider the model $A \times B + [P + I]$; or, in parametric terms,

$$y_i = \gamma_{ab} + \omega v_p + \sigma u_i.$$

Proposition 2 in § 7.4 and the factor structure diagram, Fig. 2c, gives the allocation of factors to strata:

$$\mathcal{D}_I = \{I, A \times B, B\}, \quad \mathcal{D}_P = \{P, A, 0\},$$

reflecting the facts that A contrasts in an additive model $A + B + [P + I]$ should be estimated from plot totals, while the estimation of B contrasts and the test for $A \times B$ interaction is based on differences within plots. The analysis of variance table, arranged by strata, is given as Table 3.

The estimated eigenvalues of the covariance matrix are, see § 7.7,

$$\hat{\lambda}_I = \text{SSD}_I/10, \quad \hat{\lambda}_P = \text{SSD}_P/10.$$

Estimates for the two variance components are obtained as the solutions to (7.2) with the

Table 3

Analysis of variance table for the split-plot example

Stratum	Factor	Degrees of freedom	Sums of squares of deviations
P	0	1	SS_0
	A	4	$\text{SS}_A - \text{SS}_0$
	P	10	$\text{SS}_P - \text{SS}_A$
I	B	1	$\text{SS}_B - \text{SS}_0$
	$A \times B$	4	$\text{SS}_{A\times B} - \text{SS}_A - \text{SS}_B + \text{SS}_0$
	I	10	$\text{SS}_I - \text{SS}_P - \text{SS}_{A\times B} + \text{SS}_A$

estimated eigenvalues inserted:

$$\hat{\lambda}_I = \hat{\sigma}^2, \quad \hat{\lambda}_P = 2\hat{\omega}^2 + \hat{\sigma}^2,$$

giving

$$\hat{\sigma}^2 = \hat{\lambda}_I = \text{SSD}_I/10, \quad \hat{\omega}^2 = (\hat{\lambda}_P - \hat{\lambda}_I)/2 = (\text{SSD}_P - \text{SSD}_I)/20.$$

Notice that $\hat{\omega}^2$ may be negative.

The test for $\omega^2 = 0$ is, see § 7.10,

$$F(10, 10) = (\text{SSD}_P/10)/(\text{SSD}_I/10).$$

If $\omega^2 = 0$, the two strata collapse, and we are left with an ordinary 5×2 scheme with 3 observations per cell. However, we shall assume that this hypothesis is rejected or not considered at all, i.e. the division into plots is assumed to be relevant. In this case, the test for interaction, i.e. the reduction

$$A \times B + [P + I] \to A + B + [P + I],$$

is performed in I stratum, see § 7.9, by

$$F(4, 10) = (\text{SSD}_{A \times B}/4)/(\text{SSD}_I/10).$$

If this model reduction is accepted, we are left with an additive model, $\gamma_{ab} = \alpha_a + \beta_b$, with the main effect of B in I stratum and the main effect of A in P stratum. The tests for main effects are

$$F(1, 14) = (\text{SSD}_B/1)/((\text{SSD}_I + \text{SSD}_{A \times B})/14)),$$

$$F(4, 10) = (\text{SSD}_A/4)/(\text{SSD}_P/10).$$

The estimation of contrast variances is straightforward in this case, by Proposition 5.

However, suppose that additivity cannot be accepted. For illustrative purposes, we may even consider the situation where the product structure of the treatment factor $T = A \times B$ is irrelevant, the experiment being designed for comparison of $|T| = 10$ different treatments, arbitrarily arranged in 5 pairs, each pair being applied to the pair of subplots of 3 plots. The relevant factors in this case are I, P, T and 0, but Assumption 3 in § 4.1 forces us to include the pseudofactor $P \wedge T$ ($= A$ above) on 5 levels, reflecting the partial confounding of treatments with plots. Thus, we take

$$\mathcal{D} = \{I, P, T, P \wedge T, 0\}$$

and obtain the factor structure diagram given as Fig. 3. This diagram is the same as that of Fig. 2c except that the factor B has been removed and the degrees of freedom have been changed accordingly. Random factors are in brackets to simplify the allocation of factors to strata. The analysis of variance table is given as Table 4.

The test for overall treatment effect provides an example of a test which does not take place in a single stratum, see § 7.9. Indeed, the model reduction $T + [P + I] \to 0 + [P + I]$ corresponds to the removal of *two* factors T and $P \wedge T$ from the maximal model formula

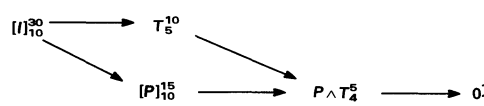

Figure 3. *Factor structure diagram of the modified split-plot example.*

Table 4

Modified analysis of variance table for the split-plot example

Stratum	Factor	Degrees of freedom	Sum of squares of deviations
P	0	1	SS_0
	$P \wedge T$	4	$SS_{P\wedge T} - SS_0$
	P	10	$SS_P - SS_{P\wedge T}$
I	T	5	$SS_T - SS_{P\wedge T}$
	I	10	$SS_I - SS_P - SS_T + SS_{P\wedge T}$

$\mathcal{T}^* = T + P \wedge T + 0$, and these are not in the same stratum. The two F tests are

$$F(5, 10) = (SSD_T/5)/(SSD_I/10)$$

testing in I stratum for differences between treatments within pairs, and

$$F(4, 10) = (SSD_{P\wedge T}/4)/(SSD_P/10)$$

testing in P stratum for differences between pair totals.

Accordingly, certain contrasts are not estimable in a single stratum. For treatments in the same pair, we have, by Proposition 5 since t' and t'' are on the same level of $P \wedge T$ in this case,

$$\text{var}\,(\bar{y}_{t'} - \bar{y}_{t''}) = \lambda_I(3^{-1} + 3^{-1}) = 2\sigma^2/3;$$

but for t' and t'' not in the same pair, we must apply Proposition 4, obtaining

$$\text{var}\,(\bar{y}_{t'} - \bar{y}_{t''}) = 2(6^{-1} + 6^{-1})\omega^2 + (3^{-1} + 3^{-1})\sigma^2 = 2(\omega^2 + \sigma^2)/3.$$

8.2 Three replicates of a complete 2^3 factorial in blocks of four (Cochran & Cox, 1957, p. 183; Federer, 1955, p. 233)

The eight combinations of three dichotomous treatment factors

$$A = \{a_0, a_1\}, \quad B = \{b_0, b_1\}, \quad C = \{c_0, c_1\}$$

are applied to 24 plots, divided into six blocks of four plots, as follows, after randomized labelling of blocks and plots within blocks:

Blocks 1, 3 and 5: $a_0b_0c_1$ $a_0b_1c_0$ $a_1b_0c_0$ $a_1b_1c_1$

Blocks 2, 4 and 6: $a_0b_0c_0$ $a_0b_1c_1$ $a_1b_0c_1$ $a_1b_1c_0$.

The rule is that treatment combinations with an odd/even sum of indices occur in odd/even numbered blocks. We denote the blocking by $P = \{1, \ldots, 6\}$, since the symbol B is reserved for one of the treatment factors. The relevant factors are: I (plots) on 24 levels; P (blocks) on 6 levels; $A \times B \times C$ on 8 levels; $A \times B$, $A \times C$ and $B \times C$ on 4 levels; A, B and C on 2 levels; 0 on 1 level. There is also a pseudofactor $P \wedge (A \times B \times C)$ on 2 levels, reflecting the classification of experimental units by parity of the block number.

The factor structure diagram is given as Fig. 4 and the analysis of variance table as Table 5, arranged according to the covariance structure $\mathcal{B} = \{I, P\}$. We shall not discuss the statistical analysis in detail. Notice, however, that $SSD_{A\times B\times C} = 0$, since $d_{A\times B\times C} = 0$, while the usual square sum for three-factor interaction in a 2^3 table occurs in P stratum as

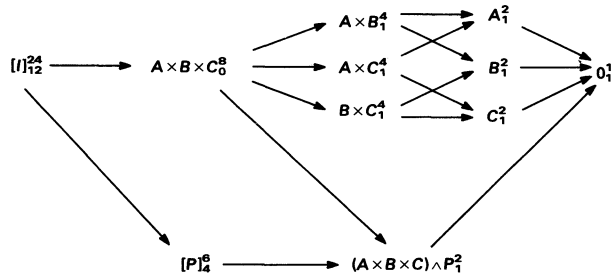

Figure 4. *Factor structure diagram of the* 2^3 *example.*

$\mathrm{SSD}_{P\wedge(A\times B\times C)}$. This represents a characteristic feature of 2^n designs with orthogonal blockings, where the role of the pseudofactors is merely to keep track of the allocation of main effects and interactions to strata. Model reductions taking place in several strata simultaneously do not occur here, since all the relevant model reductions are of dimension one. In the present example, the occurrence of the nontrivial minimum $P\wedge(A\times B\times C)$ reflects the confounding of the three-factor interaction with blocks. Thus, the test for three-factor interaction takes place in P stratum, while two-factor interactions and main effects are tested for in I stratum.

Table 5

Analysis of variance table for the 2^3 *design*

Stratum	Factor	Degrees of freedom	Sum of squares of deviations
P	0	1	SS_0
	$P\wedge(A\times B\times C)$	1	$\mathrm{SS}_{P\wedge(A\times B\times C)}-\mathrm{SS}_0$
	P	4	$\mathrm{SS}_P-\mathrm{SS}_{P\wedge(A\times B\times C)}$
I	A	1	$\mathrm{SS}_A-\mathrm{SS}_0$
	B	1	$\mathrm{SS}_B-\mathrm{SS}_0$
	C	1	$\mathrm{SS}_C-\mathrm{SS}_0$
	$A\times B$	1	$\mathrm{SS}_{A\times B}-\mathrm{SS}_A-\mathrm{SS}_B+\mathrm{SS}_0$
	$A\times C$	1	$\mathrm{SS}_{A\times C}-\mathrm{SS}_A-\mathrm{SS}_C+\mathrm{SS}_0$
	$B\times C$	1	$\mathrm{SS}_{B\times C}-\mathrm{SS}_B-\mathrm{SS}_C+\mathrm{SS}_0$
	$A\times B\times C$	0	$0=(\mathrm{SS}_{A\times B\times C}-\mathrm{SS}_{A\times B}-\mathrm{SS}_{A\times C}-\mathrm{SS}_{B\times C}+\mathrm{SS}_A+\mathrm{SS}_B+\mathrm{SS}_C-\mathrm{SS}_0)-(\mathrm{SS}_{P\wedge(A\times B\times C)}-\mathrm{SS}_0)$
	I	12	$\mathrm{SS}_I-\mathrm{SS}_{A\times B\times C}-\mathrm{SS}_P+\mathrm{SS}_{P\wedge(A\times B\times C)}$

9 Some suggestions for further research

9.1 Computational aspects

As a computational tool, the theory of the present paper can be regarded as a unified way of deriving the formulae needed for classical desk calculator methods in the orthogonal case. It is tempting (at least for the author) to consider its potential ability as a computer algorithm also. Factor structure diagrams could easily be stored as binary $\mathfrak{D}\times\mathfrak{D}$ matrices, and optional checks for orthogonality, nestedness and missing minima should also be within the limits of relatively straightforward combinatorial programming. In practice, however, this work is not likely to be worth the trouble. The gain in space and computing time over the GENSTAT 'ANOVA' algorithm would hardly be significant, and the need for auxiliary facilities (such as replacement of missing values, covariance analysis and

general data structure manipulations) makes the idea of an isolated 'orthogonal ANOVA package' rather unrealistic.

9.2 *Group generated designs*

Most orthogonal designs are constructed, or can be constructed, in such a way that the set I of experimental units has a commutative group structure, under which the factors of the design can be regarded as homomorphisms φ_F into other groups F. Such factors are balanced if and only if they are surjective as mappings. Any two such factors are orthogonal, and there is a simple interpretation of their minimum. A detailed study of these designs in relation to the theory presented here might be of some interest. A potential aspect of this would be the algebraic rules for aliasing and confounding in fractionally replicated 2^n designs, and their more complicated counterparts for 3^n designs, etc.; see the 2^3 example of § 8.

9.3 *Generally balanced designs*

We have called our approach 'combinatorial' because linear estimates and sums of squares of deviations come out as linear combinations of sums S_f or square sums SS_F with *combinatorial* coefficients, i.e. coefficients given in terms of the factor structure diagram, the cell counts n_f and the level counts $|F|$. A similar approach for the wider class of generally balanced designs (Nelder, 1965) is not available. The class is known to include balanced incomplete block designs, certain partially balanced incomplete block designs (provided that suitable 'pseudofactors' are included in the design) and some other designs with nice combinatorial properties. But an exhaustive characterization of these designs by their combinatorial properties seems not to exist. Such a characterization could throw new light on the ideas behind the GENSTAT 'ANOVA' algorithm, and might even yield a simpler 'desk calculator algorithm' for these designs. However, this theory is bound to be much more complicated than the theory for orthogonal designs. Even the rules for computation of degrees of freedom would have to be modified. Three nonorthogonal factors F, G and H with $F \wedge G = G \wedge H = F \wedge H = 0$ may be 'confounded' in the sense that

$$\dim (L_F + L_G + L_H) < |F| + |G| + |H| - 2,$$

which means that the algorithm of § 5 for computation of degrees of freedom does not always hold for nonorthogonal designs. The step from orthogonal designs to generally balanced designs corresponds, in a sense, to a generalization from commutative algebras to certain noncommutative algebras. The counterpart to our Theorem 1 should, in principle, be something like the theory presented by James & Wilkinson (1971). It is not *a priori* given, that a 'combinatorial version' of this theory will be simple enough to be of any use.

Acknowledgements

Thanks are due to a referee for very detailed comments on an earlier version of this paper, and to a second referee for requesting the relation to the work of Speed and others. I am also grateful to S.A. Andersson and S.T. Jensen for valuable comments, and to T.P. Speed, who really initiated this work by drawing the attention here to the GENSTAT programming language.

References

Aigner, M. (1979). *Combinatorial Theory*. New York: Springer.
Cochran, W.G. & Cox, G.M. (1957). *Experimental Designs*. New York: Wiley.
Federer, W.T. (1955). *Experimental Design, Theory and Application*. New York: MacMillan.

Fisher, R.A. (1935). *The Design of Experiments.* London: Oliver & Boyd.
Houtman, A.M. & Speed, T.P. (1983). Balance in designed experiments with orthogonal block structure. *Ann. Statist.* **11**. To appear.
James, A.T. (1957). The relationship algebra of an experimental design. *Ann. Math. Statist.* **28,** 993–1002.
James. A.T. & Wilkinson, G.N. (1971). Factorization of the residual operator and canonical decomposition of nonorthogonal factors in the analysis of variance. *Biometrika* **58,** 279–294.
Jensen, S.T. (1979). Varianskomponentmodeller i fuldstændigt balancerede forsøg. Inst. Math. Statist., University of Copenhagen, manuscript.
Nelder, J. A.(1954). The interpretation of negative components of variance. *Biometrika* **41,** 544–548.
Nelder, J.A. (1965). The analysis of randomized experiments with orthogonal block structure, I and II. *Proc. R. Soc.* Lond. A **283,** 147–178.
Nelder, J.A. (1969). The description of statistical data structures for statistical computing. In *Statistical Computation,* Ed. R.C. Milton and J.A. Nelder, pp. 13–36. New York: Academic Press.
Nelder, J.A. (1977). A reformulation of linear models (with discussion). *J. R. Statist. Soc.* A **140,** 48–76.
Nelder, J.A. *et al.* (1977). *The GENSTAT Manual.* Harpenden: Rothamsted Experimental Station.
Scheffé, H. (1959). *The Analysis of Variance.* New York: Wiley.
Searle, S.R. (1971). Topics in variance component estimation. *Biometrics* **27,** 1–76.
Speed, T.P. & Bailey, R.A. (1982). On a class of association schemes derived from lattices of equivalence relations. In *Algebraic Structures and Applications,* Ed. P. Schultz, C.E. Praeger and R.P. Sullivan. New York: Marcel Dekker.
Szatrowski, T.H. & Miller, J.J. (1980). Explicit maximum likelihood estimates from balanced data in the mixed model of the analysis of variance. *Ann. Statist.* **8,** 811–819.
Wilkinson, G.N. (1969). Facilities in a statistical program system for analysis of multiply-indexed data. In *Statistical Computing,* Ed. R.C. Milton and J.A. Nelder, pp. 201–228. Academic Press.
Wilkinson, G.N. (1970). A general recursive procedure for analysis of variance. *Biometrika* **57,** 19–46.
Wilkinson, G.N. & Rogers, C.E. (1973). Symbolic description of factorial models for analysis of variance. *Appl. Statist.* **22,** 392–399.
Zedeler, K. (1970). Algebraisk behandling af tosidet variansanalyse. Graduate dissertation, supervised by H. Brøns. Inst. Math. Stat., University of Copenhagen.

Résumé

L'objet de cet article est de présenter une théorie de modèles de l'analyse de la variance dans plans d'expérience composé de facteurs orthogonaux. La structure d'un plan d'expérience orthogonale est résumée par un diagramme de structure des facteurs, contenant l'information sur les relations de subordination entre les facteurs. Un plan d'expérience orthogonal détermine une décomposition unique de l'espace des données, comme une somme directe de sous-espaces orthogonaux. Une classe de modèles mixtes, exprimés par des facteurs fixes et des facteurs aléatoires, est caracterisée. Décrirons l'estimation des composantes de la variance et des tests de signification dans cette classe au moyen de la diagramme de structure des facteurs et du tableau d'analyse de la variance correspondant à la décomposition de l'espace des donneés.

[*Paper received April* 1982, *revised March* 1983]

Discussion of paper by T. Tjur

R.A. Bailey

Statistics Department, Rothamsted Experimental Station, Harpenden, Hertfordshire, AL5 2JQ, UK

I am very pleased both to read and to be invited to comment on Tjur's paper, which is a careful coherent account of a subject whose edges have been chipped at by many people. I hope that these ideas become widely accepted. I have assembled my comments under subject headings. A major theme is a plea for standardization of helpful terminology and notation.

D1 Balance

This word has been over-used in the statistical literature: see Preece (1982). So I hope an alternative can be found for 'balanced' in 'balanced factor' (§ 2.1). Speed & Bailey