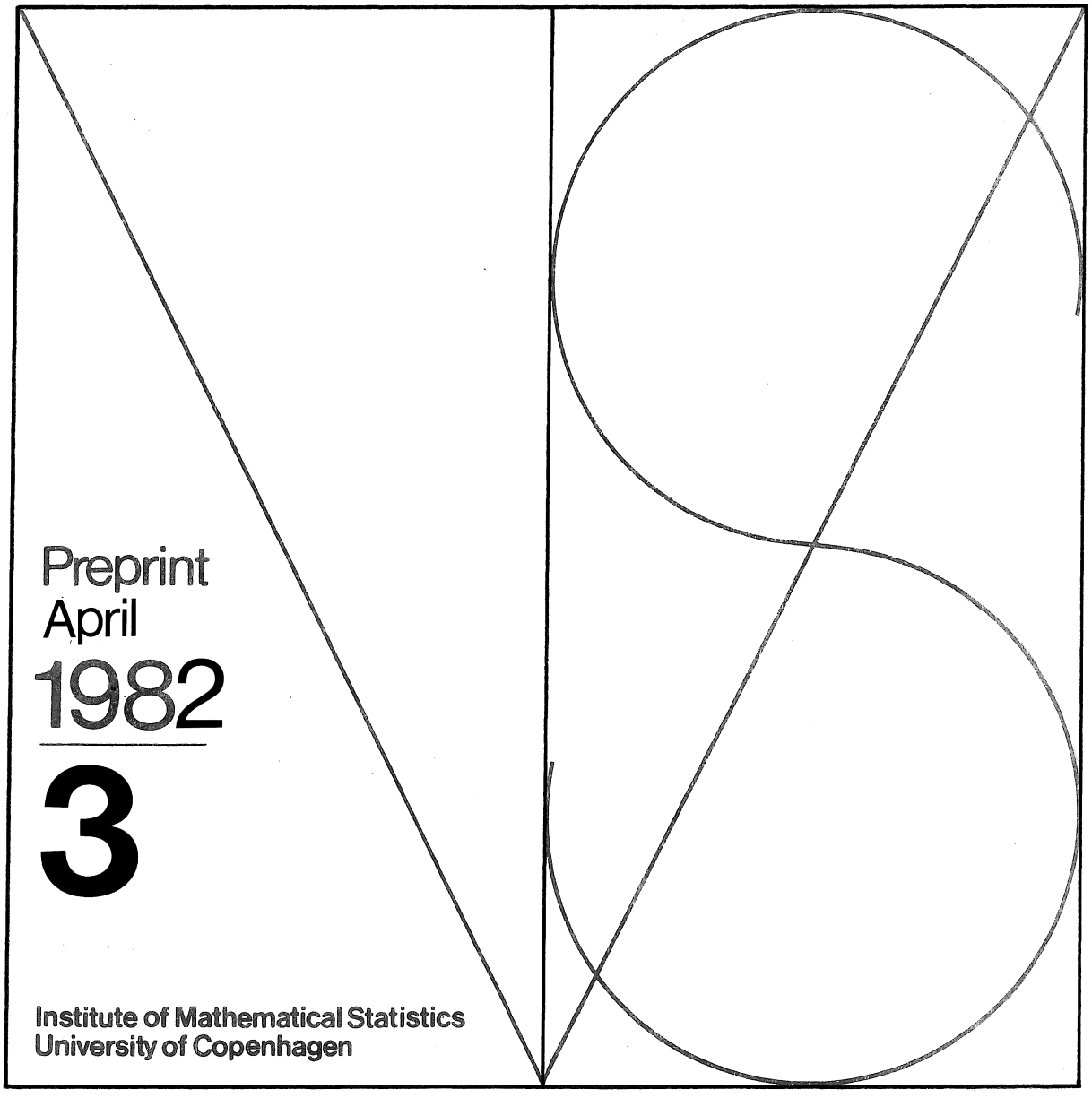
Tue Tjur

# Variance Component Models in Orthogonal Designs



Preprint
April
1982
3

Institute of Mathematical Statistics
University of Copenhagen

Tue Tjur

# VARIANCE COMPONENT MODELS IN ORTHOGONAL DESIGNS

Preprint 1982 No. 3

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN

APRIL 1982

Summary:

This paper presents a unified approach to ANOVA-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 relevant 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. The class of solvable variance component models, stated in terms of factors in a given design, is characterized, and the solutions to problems of estimation and hypothesis testing are given in terms of the factor structure diagram and the ANOVA table induced by the decomposition.

Key words:

Analysis of variance, othogonal design, variance component model.

## 0. INTRODUCTION.

This paper deals with ANOVA models in designs where all factors are orthogonal. Examples are randomized block designs, split-plot designs, complete balanced k-factor designs (possibly with an orthogonal blocking), latin and graeco-latin squares, fractional replications of complete factorials.

The justification of a new paper on this classical subject is that too little seems to have been done to put these models into a unified framework. In textbooks, they are usually presented as a sequence of very similar models, but not as special cases of the same model, in the sense that the solution comes out by inserting the "model structure" in a set of formulas. The present paper does (almost) give such a solution, and this solution is computationally explicit, in the sense that the formulas derived for degrees of freedom, sums of squares, estimates of variance components etc. coincide with those one would usually apply if the computations were to be done on a desk calculator. These formulas are derived in a simple way from a diagram describing the model structure.

Similar ideas, related to a wider class of ANOVA models (namely those with orthogonal block structure only) have been put forward in two papers by Nelder (1965), which together with Wilkinson (1970) and James and Wilkinson (1971) establish the theoretical background for the GENSTAT 'ANOVA' algorithm. However, our formal description of design structure is considered mathematically simpler than that suggested by Nelder, and the Nelder-Wilkinson approach does not lead to the same degree of explicity of the computational formulas in the case of orthogonal block-treatment structure (accordingly, the GENSTAT 'ANOVA' algorithm would not be suited for desk calculators).

It should not be believed from what is said above, that the present paper suggests a serious competetor to the 'ANOVA' algorithm. For practical purposes, the part of the analysis concerned with the structure of the design is more

conveniently done by the statistician than by a computer. The present paper should be regarded as a support to this. The computational work can, of course, be left to a computer, and for this GENSTAT can certainly be recommended. The specification of a model to be analysed by 'ANOVA' requires an understanding of the design on about the same level as given by the factor structure diagrams introduced in the present paper. In particular, it should be noticed by GENSTAT users that 'ANOVA' is unable to detect missing minima and nestedness relations (see section 1) which can not be read directly off the model formula. This may lead to erroneous degrees of freedom, if not taken into account by the programmer.

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 formalisation- does not correspond to a classical statistical concept (though related to concepts like partial aliasing or partial confounding). The concept of a minimum plays a crucial role for the results obtained in later sections.

Readers familiar with Nelder's 1965-papers may be confused by the fact that our arrows between factors (like $F \to G$) are essentially the reverse of those found in Nelder's paper. This inconsistency is difficult to avoid, because our arrows indicate mappings between sets.

## 1. FACTORS : COMBINATORIAL STRUCTURE.

Let $y = (y_i \mid i \in I) \in \mathbb{R}^I$ denote our data set. The (finite) indexing set $I$ is the set of experimental units. A factor

$$\overline{F} : I \to F$$

is a mapping $\overline{F}$ from $I$ into some other finite set $F$. The elements $f \in F$ are called factor levels. Usually, we shall refer to "the factor $F$" (rather than "the factor $\overline{F}: I \to F$"), thus subsuming the mapping as given by the context.

Two factors play a special role as "extremes", the trivial factor $O$, correponding to a constant mapping $\overline{O}: I \to O$ where $O$ is an arbitrary set with one element, and the units factor $I$, corresponding to the identity $\overline{I}: I \to I$.

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

## 1.1. Balanced factors.

As a standard notation in the following, we let $n_f$ denote the size of the class corresponding to $f \in F$, i.e.

$$n_f = \# \overline{F}^{-1}(f)$$

(where $\#$ is used for "number of elements in"). By $|F|$ we denote the number of non-empty classes, i.e.

$$|F| = \#\{ f \in F \mid n_f > 0 \} .$$

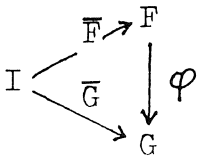
Notice that $|F| = \# F$ if and only if the mapping $\overline{F}$ is surjective. A factor $F$ is called balanced if all the classes are of the same size, which is then denoted

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

## 1.2. Nested factors.

Let two factors $F$ and $G$ be given. Suppose that there exists a mapping $\varphi : F \to G$ such that

$$\varphi \circ \overline{F} = \overline{G}$$



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 $G \subseteq F$ or (in diagrams) $F \rightarrow G$.

Notice that we have $G \subseteq F$ if and only if any of the G-classes can be written as a union of some of the F-classes, namely

$$\overline{G}^{-1}(g) = \bigcup_{f \in \varphi^{-1}(g)} \overline{F}^{-1}(f)$$

## 1.3. Equivalent factors.

Two factors $F$ and $G$ will be called equivalent, written $F \sim G$, if both $F \subseteq G$ and $G \subseteq F$. Roughly speaking, equivalent factors are factors which induce the same partitioning of $I$. Only the labeling of the classes and the number of "empty classes" (corresponding to level names 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 welldefined "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 $\subseteq$ is a partial ordering of the set of factors. We have the maximal element $I$ and the minimal element $0$.

## 1.4. Crossclassifications.

The product $F \times G$ of two factors $F$ and $G$ (or the cross-classification induced by $F$ and $G$) is defined as

$$\overline{F \times G} : I \rightarrow F \times G ,$$

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

$$\overline{F \times G}(i) = (\overline{F}(i), \overline{G}(i)) .$$

The formation of products can be regarded as an operation on equivalence classes, in the sense that

$$F \sim F' \text{ and } G \sim G' \Longrightarrow F \times G \sim F' \times G' \ .$$

Under the partial ordering $\subseteq$, 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,

(1) $F \subseteq F \times G$ and $G \subseteq F \times G$

(2) 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 <u>maxima</u>, and the maximum operation (which we would otherwise denote $\vee$ or $\cup$) coincides with the formation of products ("$F \vee G = F \times G$").

1.5. <u>Minima of factors</u>.

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

(1) $F \wedge G \subseteq F$ and $F \wedge G \subseteq G$

(2) 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 $\mathcal{a}_F$ denote the set of subsets of $I$ of the form $\overline{F}^{-1}(M)$, where $M$ is a subset of the set $F$. Then $\mathcal{a}_F$ is obviously an <u>algebra</u> of subsets of $I$, i.e. $\emptyset \in \mathcal{a}_F$, $I \in \mathcal{a}_F$, and $\mathcal{a}_F$ is closed under the formation of intersections, unions and complements. $\mathcal{a}_F$ can be characterized as the algebra spanned by the clsses $\overline{F}^{-1}(f)$ in the partitioning induced by $F$. Conversely, let $\mathcal{a}$ be an arbitrary algebra of subsets of $I$. A factor $F$ with $\mathcal{a} = \mathcal{a}_F$ can be constructed as follows; consider the atoms of $\mathcal{a}$, i.e. the minimal nonempty sets in $\mathcal{a}$. These constitute a partitioning, and a factor $F$ constructed

by suitable labeling of the classes will obviously have $a_F = a$ .

Hence, we have the one-to-one correspondence

$$F \leftrightarrow a_F$$

between equivalence classes of factors and algebras on $I$ . This correspondence is order preserving , in the sense that

$$F \subseteq G \iff a_F \subseteq a_G \ .$$

This means that the problem of constructing the 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 last problem is straightforward, since the intersection of two algebras is again an algebra. Thus, the minimum of two factors exists and is determined by

$$a_{F \wedge G} = a_F \cap a_G \ .$$

1.6. An example.

The above proof of existence does not give much intuitive feeling for what the operation $\wedge$ really does. This is more conveniently explained by an example. Put $I = \{1, 2, \ldots, 9\}$ , and suppose we have two factors $R$ and $C$ (rows and columns) on $3$ and $4$ levels, respectively. Suppose that the allocation of units to $R \times C$ - levels is as follows.

| | $c_1$ | $c_2$ | $c_3$ | $c_4$ |
|---|---|---|---|---|
| $r_1$ | 1,2 | 3 | | |
| $r_2$ | 4,5 | 6 | | 9 |
| $r_3$ | | | | 7,8 |

$R = \{r_1, r_2, r_3\}$

$C = \{c_1, c_2, c_3, c_4\}$

Put $H = R \wedge C$ . The relation $H \subseteq R$ means that $\bar{H}$ is constant in rows, for example (for the first row)

$$\bar{H}(1) = \bar{H}(2) = \bar{H}(3) \ .$$

Similarly, $H \subseteq C$ means that $\bar{H}$ is constant in columns, for example

$$\bar{H}(3) = \bar{H}(6) \quad .$$

Continuing like this, we can easily deduce that $\bar{H}$ must be constant, i.e. $R \wedge C = 0$ . 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 non-empty cell to any other in a finite sequence of jumps between non-empty cells within the same row or the same column. More generally, this relation between units -that the two corresponding cells can be connected by such a sequence of vertical and horizontal movements- is an equivalence relation, and the partitioning of I into equivalence classes under this relation is exactly the partitioning corresponding to the minimum $R \wedge C$ . Example: If the last element 9 of I above is removed, we will get a non-trivial minimum on two levels, corresponding to the partitioning into $\{1,2,3,4,5,6\}$ and $\{7,8\}$ .

## 2. FACTORS : LINEAR STRUCTURE.

In this and the following section, we shall study the structure imposed by one or several factors on the observation space $\mathbb{R}^I$ . Vectors in $\mathbb{R}^I$ are regarded as $I \times 1$ -matrices (column vectors), and linear mappings are identified with their matrices as usual. $\mathbb{R}^I$ is equipped with the usual inner product

$$(x|y) = x'y = \sum_{i \in I} x_i y_i \quad ,$$

and the corresponding Euclidean norm is denoted

$$\|y\| = \sqrt{(y|y)} \quad .$$

The $I \times I$ identity matrix is denoted by $I$ (without danger of confusion, we hope).

2.1. $X_F$ , $L_F$ and $P_F$ .

A factor $\bar{F} : I \to F$ induces a linear mapping

$$X_F : \mathbb{R}^F \to \mathbb{R}^I$$

defined by

$$X_F((\alpha_f \mid f \in F)) = (\alpha_{\bar{F}(i)} \mid i \in I) .$$

As a matrix, $X_F$ is the $I \times F$ - matrix with elements

$$(X_F)_{if} = \begin{cases} 1 & \text{for } \bar{F}(i) = f \\ 0 & \text{otherwise .} \end{cases}$$

$X_F$ is called the <u>design matrix</u> corresponding to the factor $F$ . The image of the linear mapping $X_F$ , which can also be characterized as the space of functions that are constant on the classes induced by $F$ , is denoted by

$$L_F = \{ (\alpha_{\bar{F}(i)} \mid i \in I) \mid (\alpha_f) \in \mathbb{R}^F \} .$$

Notice that $\dim L_F = |F|$ . By

$$P_F : \mathbb{R}^I \to \mathbb{R}^I$$

we denote the orthogonal projection on $L_F$ . According to well known rules for estimation in a one-way ANOVA model, $P_F$ transforms a vector $y$ by replacement of each coordinate $y_i$ with the average $\bar{y}_f$ over the corresponding class. Hence, the $I \times I$ - matrix $P_F$ has elements

$$(P_F)_{i_1 i_2} = \begin{cases} 1/n_f & \text{for } \bar{F}(i_1) = \bar{F}(i_2) = f \\ 0 & \text{for } \bar{F}(i_1) \neq \bar{F}(i_2) . \end{cases}$$

Notice the relation

$$P_F = \frac{1}{n_F} X_F X_F' ,$$

which holds for any <u>balanced</u> factor $F$ .

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

$$F \subseteq G \implies L_F \subseteq L_G ,$$

and it preserves minima according to the rule

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

(this formula, which is easy to prove, can be regarded as an alternative definition of the minimum). Maxima are not preserved. We have the inclusion

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

which is usually sharp (and leads to the concept of interaction).

## 2.2. Orthogonal factors.

As usual, two linear subspaces $L_1$ and $L_2$ of $\mathbb{R}^I$ are said to be orthogonal if any vector in $L_1$ is orthogonal to any vector in $L_2$. $L_1$ 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 $V_1 \perp V_2$.

The term "geometrically" is motivated by the fact that this kind of orthogonality is the one known from ordinary geometry, where two planes in $\mathbb{R}^3$ may be orthogonal in exactly this sense. The following lemma characterizes the concept in a way which is more convenient in the ANOVA context.

**2.2.1. Lemma.** $L_1$ and $L_2$ are geometrically orthogonal if and only if the corresponding orthogonal projections $P_1$ and $P_2$ commute, i.e. $P_1P_2 = P_2P_1$.

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

2.2.2. Lemma. Let $L_1, \ldots, L_k$ be geometrically orthogonal, and let $P_1, \ldots, P_k$ denote the corresponding orthogonal projections. Then

$$P = P_1 \ldots P_k$$

is the orthogonal projection on $L = L_1 \cap \ldots \cap L_k$ .

Two factors $F$ and $G$ are called orthogonal , and we write $F \perp G$ , if the corresponding subspaces $L_F$ and $L_G$ are geometrically orthogonal; or, equivalently, if

$$P_F P_G = P_G P_F \ .$$

The justification of the concept of orthogonality in relation to ANOVA models lies in the above formula which, among other things, leads to simple expressions for orthogonal projections on sums of subspaces generated by orthogonal factors. However, as a criterion for orthogonality in concrete situations we need a more explicit condition on the cell counts $n_{fg}$ . To this end, notice that $F$ and $G$ are orthogonal if and only if

$$P_F P_G = P_{F \wedge G} \ .$$

Indeed, if $L_F$ and $L_G$ are geometrically orthogonal this formula follows immediately from lemma 2.2.2 above, since $L_F \cap L_G = L_{F \wedge G}$ . Conversely, if the above formula holds we have in particular

$$P_F P_G = (P_F P_G)' = P_G' P_F' = P_G P_F \ ,$$

i.e. $F \perp G$ .

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 $P_F P_G$ and $P_H$ . As to the first one, put $f = \overline{F}(i_1)$ and $g = \overline{G}(i_2)$ . Then, with an obvious notation for indicator functions, we have

$$(P_F P_G)_{i_1 i_2} = \sum_i \frac{1}{n_f} 1\{\overline{F}(i) = f\} \frac{1}{n_g} 1\{\overline{G}(i) = g\} = \frac{n_{fg}}{n_f n_g} \ .$$

For the second matrix, we have

$$(P_H)_{i_1 i_2} = \begin{cases} 1/n_h & \text{for } \bar{H}(i_1) = \bar{H}(i_2) = h \\ 0 & \text{for } \bar{H}(i_1) \neq \bar{H}(i_2) \end{cases} .$$

Now, in the case $\bar{H}(i_1) \neq \bar{H}(i_2)$ we have obviously (cfr. the construction of the minimum) $n_{fg} = 0$, which means that the $(i_1,i_2)$'th element of both matrices is 0 in this case. Hence, our condition for orthogonality boils down to

2.2.3. Proposition. F and G are orthogonal if and only if the relation

$$\frac{n_{fg}}{n_f n_g} = \frac{1}{n_h}$$

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 (i.e. $\bar{F}^{-1}(f) \subseteq \bar{H}^{-1}(h)$ and $\bar{G}^{-1}(g) \subseteq \bar{H}^{-1}(h)$ ).

In the case $F \wedge G = 0$, this criterion simplifies to

$$n_{fg} = n_f n_g / |I| ,$$

which is the well known condition of proportional cell counts. The condition in the general case states that such a proportionality condition should hold within each of the subtables of the table $(n_{fg})$ determined by the levels of $H = F \wedge G$.

## 2.3. The balanced k-way design.

It follows in particular from the above criterion that F and G are orthogonal if $F \times G$ is balanced. In this case $F \wedge G = 0$, obviously. More generally, let k factors $F_1, \ldots, F_k$ be given such 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 2.2.3 above, noticing that the rule for formation of minima of such factors is

$$\left(\prod_{j\in M_1} F_j\right) \wedge \left(\prod_{j\in M_2} F_j\right) = \prod_{j\in M_1\cap M_2} F_j \ ,$$

$M_1 \ , \ M_2 \subseteq \{1 \ , \ \ldots \ , \ k\}$ .

## 3. DECOMPOSITION OF $\mathbb{R}^I$ WITH RESPECT TO AN ORTHOGONAL DESIGN.

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

### 3.1. Orthogonal designs.

Throughout this paper, we make the following three assumptions.

($\mathfrak{D}$ 1) $I \in \mathfrak{D}$ .

($\mathfrak{D}$ 2) Any two factors in $\mathfrak{D}$ are orthogonal.

($\mathfrak{D}$ 3) $\mathfrak{D}$ is closed under the formation of minima.

Notice that only ($\mathfrak{D}$ 2) is really restrictive. The satisfaction of ($\mathfrak{D}$ 1) and ($\mathfrak{D}$ 3) is a matter of including the factor I and some missing minima, and this can be done without destroying the orthogonality.

### 3.2. The decomposition induced by $\mathfrak{D}$ .

Our approach relies on the following main result.

3.2.1. Theorem. Under the assumptions ($\mathfrak{D}$ 1), ($\mathfrak{D}$ 2) and ($\mathfrak{D}$ 3) above, we have a unique decomposition

$$\mathbb{R}^I = \bigoplus_{G \in \mathfrak{D}} V_G$$

of $\mathbb{R}^I$ as a direct sum of pairwise orthogonal subspaces $V_G$, $G \in \mathfrak{D}$, such that for any $F \in \mathfrak{D}$

$$L_F = \bigoplus_{\substack{G \in \mathfrak{D} \\ G \subseteq F}} V_G \ .$$

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

Since the proof of the theorem is somewhat technical, a few remarks may help. It is not surprising that a set $\mathfrak{D}$ of pairwise orthogonal factors induces a decomposition of $\mathbb{R}^I$. Indeed, take all sums and intersections of the geometrically orthogonal spaces $L_F$ with each other and their orthogonal complements. This gives what one might call an algebra of geometrically orthogonal subspaces (closed under $\cap$, $+$ and $^\perp$, containing $\mathbb{R}^I$ and $\{0\}$). The minimal, nontrivial subspaces of this algebra constitute a decomposition of $\mathbb{R}^I$. The orthogonal projections on the components of this decomposition can be regarded as the generators of the commutative algebra spanned by the projections $P_F$, $F \in \mathfrak{D}$. However, the theorem says a little more than this, namely that there is a canonical way of labeling these subspaces, in such a way that the original projections $P_F$ can be recovered as "cumulated" projections on the components, according to the ordering of $\mathfrak{D}$. This result depends strongly on the fact that $\mathfrak{D}$ is closed under the formation of minima.

Proof. Consider the trivial identity

$$I = \prod_{F \in \mathfrak{D}} (P_F + (I - P_F)) \quad .$$

Rewriting this as a sum of products of terms $P_F$ and $I - P_F$, we get

$$I = \sum_{M \subseteq \mathfrak{D}} Q_M \quad ,$$

where for each subset $M$ of $\mathfrak{D}$ the operator $Q_M$ is defined as

$$Q_M = \left( \prod_{F \in M} P_F \right) \left( \prod_{F \notin M} (I - P_F) \right) \quad .$$

As products of commuting orthogonal projections, these operators $Q_M$ are again orthogonal projections. The corresponding subspaces $V_M$ are pairwise orthogonal, since (obviously)

$$M_1 \neq M_2 \implies Q_{M_1} Q_{M_2} = 0 \quad .$$

Hence, the formula $I = \sum_{M \subseteq \mathfrak{D}} Q_M$ corresponds to a decomposition of $\mathbb{R}^I$ as a direct sum of $2^{\#\mathfrak{D}}$ subspaces $V_M$, $M \subseteq \mathfrak{D}$. However, some of these subspaces $V_M$ are trivial. Indeed, $Q_M$ can only be different from 0 if

(1) $F \in M$ and $F \subseteq F' \implies F' \in M$ ,

because otherwise $Q_M = 0$ follows from the fact that the product of $P_F$ and $I - P_{F'}$ is 0 for $F \subseteq F'$. Moreover, if $Q_M$ is to be non-zero, we must assume that the minimum $G$ of all factors in $M$ is itself an element of $M$, i.e.

(2) $G = \bigwedge_{F \in M} F \in M$ ,

because otherwise the product of $(I - P_G)$ and $\prod_{F \in M} P_F = P_G$ gives 0. Thus, if we restrict our attention to non-vanishing terms of the decomposition, we need only take into account the subsets $M$ of $\mathfrak{D}$ satisfying (1) and (2) above. However, these sets $M$ can be characterized in a simpler way. A set $M$ satisfying (2) contains its own minimum, and if (1) is satisfied too, $M$ must contain any factor finer than the minimum. Hence, the only sets $M$

we need to consider are those of the form

$$M = M_G = \{ F \in \mathfrak{D} \mid F \supseteq G \} ,$$

which means that we have reduced to a decomposition of $\mathbb{R}^I$ as a direct sum of $\#\mathfrak{D}$ subspaces,

$$\mathbb{R}^I = \bigoplus_{G \in \mathfrak{D}} V_{M_G} .$$

For simplicity of notation, we write $V_G$ and $Q_G$ instead of $V_{M_G}$ and $Q_{M_G}$ in the following.

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

$$P_F Q_G = \begin{cases} Q_G & \text{for } G \subseteq F \\ 0 & \text{otherwise.} \end{cases}$$

Hence,

$$P_F = P_F\Big(\sum_{G \in \mathfrak{D}} Q_G\Big) = \sum_G P_F Q_G = \sum_{G : G \subseteq F} Q_G ,$$

or, equivalently,

$$L_F = \bigoplus_{G : G \subseteq F} V_G .$$

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

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

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

$$\sum_{\substack{G \subseteq F \\ G \neq F}} Q_G = P_F - Q_F \neq P_F - Q'_F = \sum_{\substack{G \subseteq F \\ G \neq F}} Q_G$$

that there must be some $G$ strictly coarser than $F$ for which $Q_G \neq Q'_G$. Thus, the set $\mathfrak{D}_0$ has the property

that for any element which is not equal to $F_o = \min \mathfrak{D}$ it has an element which is strictly coarser. It is easy to conclude from this that $\mathfrak{D}_o$ must either be empty or contain the minimal element $F_o$ of $\mathfrak{D}$. But $F_o \in \mathfrak{D}_o$ is impossible, since

$$Q_{F_o} = P_{F_o} = Q'_{F_o} .$$

Thus, $\mathfrak{D}_o$ is empty, and the theorem is proved.

The name $Q_G$ for the orthogonal projection onto $V_G$ will be standard notation in the following. Notice that we have now <u>two</u> families of orthogonal projections indexed by $\mathfrak{D}$, the canonical projections $P_F$ on the factor spaces $L_F$ (which depend on $F$ only, not on the remaining factors of the design), and the projections $Q_G$ given by the theorem. The connection between these families is

$$P_F = \sum_{G \subseteq F} Q_G .$$

Conversely, $Q_G$ can be expressed by the projections $P_F$ as follows.

3.2.2. Corollary.

$$Q_G = P_G \prod_{\substack{F : F \subseteq G, \\ F \neq G}} (P_G - P_F) .$$

Proof. In the proof of the theorem, $Q_G$ was constructed as

$$Q_G = Q_{M_G} = \Big(\prod_{F : G \subseteq F} P_F\Big)\Big(\prod_{F : G \not\subseteq F} (I - P_F)\Big)$$

$$= P_G \prod_{F : G \not\subseteq F} (I - P_F) = P_G \prod_{F : G \not\subseteq F} (P_G - P_{F \wedge G})$$

$$= P_G \prod_{\substack{F : F \subseteq G, \\ F \neq G}} (P_G - P_F) .$$

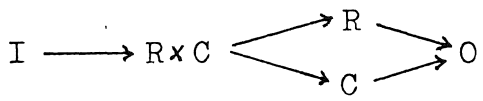
If desirable, an expression of $Q_G$ as a linear combination

$$Q_G = \sum_F a_G^F P_F$$

can be obtained from the formula of the corollary by straightforward application of the distributive law and the rule $P_F P_{F'} = P_{F \wedge F'}$. 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 $\prod (P_G - P_F)$ ; indeed, terms $(P_G - P_{F'})$ with $F' \subseteq F \subset G$ are absorbed by $(P_G - P_F)$ anyway. Similarly, the first term $P_G$ can be omitted if the product is non-empty, i.e. if $G \neq F_o$ ( = min $\mathfrak{D}$ ).

3.3. Example.

Suppose we have two factors $R$ and $C$ such that $R \perp C$ and $R \wedge C = O$. Put $\mathfrak{D} = \{I, R \times C, R, C, O\}$. Obviously, $\mathfrak{D}$ satisfies the conditions ($\mathfrak{D}$ 1), ($\mathfrak{D}$ 2) and ($\mathfrak{D}$ 3) (page 12). The ordering structure of $\mathfrak{D}$ is




Applying corollary 3.2.2 and the above mentioned rules for omission of terms in the product, we obtain the formulas

$$Q_{R \times C} = (P_{R \times C} - P_R)(P_{R \times C} - P_C) = P_{R \times C} - P_R - P_C + P_O$$
$$Q_I = I - P_{R \times C}$$
$$Q_R = P_R - P_O$$
$$Q_C = P_C - P_O$$
$$Q_O = P_O \ .$$

4. THE ANOVA TABLE.

By the ANOVA table for the data set $y$ in the design $\mathfrak{D}$ we mean a table which for each $G \in \mathfrak{D}$ gives the quantities

$$d_G = \dim V_G \quad \text{and} \quad SSD_G = \|Q_G y\|^2 .$$

It should be emphasized that the ANOVA table is not related to a single model. We are not talking about models yet. The ANOVA table is a computational tool, containing the quantities relevant for hypothesis testing and variance estimation in all ANOVA models that can be stated in terms of factors from $\mathfrak{D}$ .

4.1. Construction of the ANOVA table.

Computationally, the ANOVA table can be obtained as follows. Put

$$SS_F = \|P_F y\|^2$$

(and notice that we distinguish between SS (sum of squares) and SSD (sum of squares of deviations)). The square sums $SS_F$ are obtained as

$$SS_F = \sum_{f \in F} S_f^2 / n_f ,$$

where $S_f$ denotes the sum of all observations on the level $f$ of $F$ . Let $Q_G = \sum_F a_G^F P_F$ be the expression of $Q_G$ as a linear combination of the projections $P_F$ . Then

$$SSD_G = \|Q_G y\|^2 = y' Q_G y = \sum_F a_G^F y' P_F y$$

$$= \sum_F a_G^F \|P_F y\|^2 = \sum_F a_G^F SS_F .$$

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 $Q_G$ are obtained as linear combinations of the projections $P_F$ . Similarly, the formula

$$d_G = \operatorname{tr}(Q_G) = \operatorname{tr}\left(\sum_F a_G^F P_F\right) = \sum_F a_G^F \operatorname{tr}(P_F)$$

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

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 formulas for the $SSD_G$'s given above constitute the solution to the equations

$$SS_F = \sum_{G \subseteq F} SSD_G \quad ,$$

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

$$|F| = \sum_{G \subseteq F} d_G \quad .$$

This can be done by means of a factor structure diagram (see example 3.3) containing the information about the nestedness relations between factors in $\mathfrak{D}$. The following four examples illustrate this method.

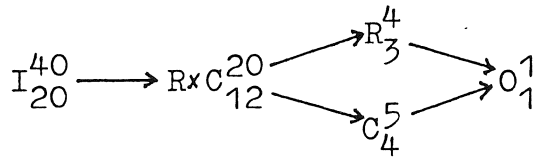
## 4.2. Examples.

### 4.2.1. The balanced two-way table.

Suppose we have a two-way scheme $R \times C$ with all cell counts $n_{rc}$ equal and $\geq 2$. Let $\mathfrak{D}$ consist of the factors $I$, $R \times C$, $R$, $C$ and $0$. The factor structure diagram of example 3.3 (page 17) gives the nestednes relations and the rules for formation of minima in an obvious way. Notice that "composed" arrows, like $I \to 0$, need not be drawn.

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 $V_F$ (like this: $F_{d_F}^{|F|}$). Filling in the superscripts $|F|$ first, it is easy to obtain the numbers $d_F$ recursively, in each step computing $d_F$ as the difference between the corresponding superscript $|F|$ and the sum of all subscripts $d_G$ of factors $G$ strictly coarser

than $F$ . For example, for $|R| = 4$ , $|C| = 5$ and $|I| = 2 \cdot 4 \cdot 5 = 40$ , we get the following picture.



The sums of squares of deviations are obtained similarly (using superscripts $SS_F$ and subscripts $SSD_F$ ), and we obtain the ANOVA table

| $F$ | $d_F$ | $SSD_F$ |
|---|---|---|
| $I$ | 20 | $SS_I - SS_{R \times C}$ |
| $R \times C$ | 12 | $SS_{R \times C} - SS_R - SS_C + SS_0$ |
| $R$ | 3 | $SS_R - SS_0$ |
| $C$ | 4 | $SS_C - SS_0$ |
| $0$ | 1 | $SS_0$ |
| sum | 40 | $SS_I$ |

Since we have no data, formulas for $SSD$'s are given instead of the concrete figures, but the general idea is to compute the numbers $SSD_G$ recursively from the numbers $SS_F$ , as illustrated above for the degrees of freedom. This would certainly be the simplest way of doing it in a computer implementation of this algorithm, where the factor structure diagram could be stored as a binary $\mathfrak{D} \times \mathfrak{D}$ -matrix; in more complicated designs, this may even be the simplest way of doing it manually.

We give three additional examples. In all cases, the degrees of freedom and formulas for the $SSD$'s are immediately obtained from the factor structure diagram.

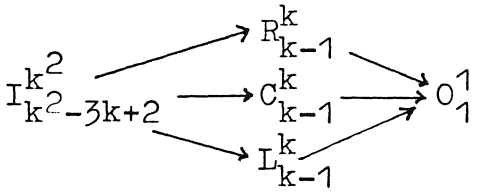
4.2.2. One-way ANOVA (arbitrary group sizes $n_f$).

$$I^n_{n-k} \longrightarrow F^k_{k-1} \longrightarrow 0^1_1 \qquad (|I| = n \,,\ |F| = k).$$

ANOVA table:

| factor | d.f. | SSD |
|---|---|---|
| I | n-k | $SS_I - SS_F$ |
| F | k-1 | $SS_F - SS_0$ |
| 0 | 1 | $SS_0$ |
| sum | n | $SS_I$ |

4.2.3. Latin square of order k .




$|R| = |C| = |L| = k$ ,

$|I| = k^2$

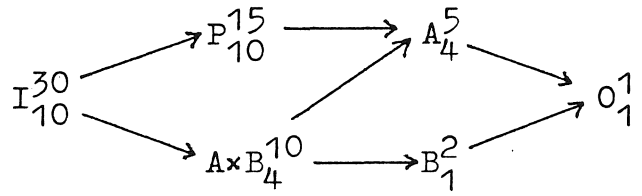
(R for rows , C for columns, L for latin letter) .

| factor | d.f. | SSD |
|---|---|---|
| I | $k^2-3k+2$ | $SS_I - SS_R - SS_C - SS_L + 2\,SS_0$ |
| R | k-1 | $SS_R - SS_0$ |
| C | k-1 | $SS_C - SS_0$ |
| L | k-1 | $SS_L - SS_0$ |
| 0 | 1 | $SS_0$ |
| sum | $k^2$ | $SS_I$ |

4.2.4. Split-plot design. From Cochran and Cox (1957), p. 293. Suppose that 5 treatments $A = \{a_1 , \ldots , a_5\}$ are applied to 15 plots $P = \{1 , \ldots , 15\}$ , each A-treatment being applied to 3 plots. Each plot is divided into two subplots, to which the two treatments $B = \{b_1, b_2\}$ are applied. Thus, the relevant factors are

I on 30 levels (subplots)
P on 15 levels (plots)
A on 5 levels (A-treatments)
B on 2 levels (B-treatments)
AxB on 10 levels ( "interaction " )

Adding $0$, we obtain a set $\mathfrak{D}$ of factors satisfying $(\mathfrak{D}\,1)$, $(\mathfrak{D}\,2)$ and $(\mathfrak{D}\,3)$. The factor structure (with degrees of freedom computed) is




and the ANOVA table, accordingly,

| factor | d.f. | SSD |
|---|---|---|
| I | 10 | $SS_I - SS_P - SS_{A\times B} + SS_A$ |
| P | 10 | $SS_P - SS_A$ |
| A×B | 4 | $SS_{A\times B} - SS_A - SS_B + SS_0$ |
| A | 4 | $SS_A - SS_0$ |
| B | 1 | $SS_B - SS_0$ |
| 0 | 1 | $SS_0$ |
| sum | 30 | $SS_I$ |

(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.)

## 5. LINEAR MODELS.

By a <u>linear model</u> we mean a model assuming that the data set $y$ is (the realization of) a normally distributed random vector with covariance matrix $\sigma^2 \cdot I$ and mean vector $\mu$ in a specified linear subspace $L$ of $\mathbb{R}^I$. We shall restrict our attention to the case where $L$ is given as a sum of subspaces corresponding to factors of our orthogonal design $\mathfrak{D}$, i.e.

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

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 example, in the two-way scheme $R \times C$, the two sets $\{R, C\}$ and $\{R, C, O\}$ represent the same model because $L_R + L_C = L_R + L_C + L_O$.

## 5.1. Model formulas.

We shall refer to the subset $\mathcal{T}$ of $\mathcal{D}$ as the <u>model formula</u>. This is merely a notational convention according to which we list the elements of $\mathcal{T}$ separated by plusses instead of commaes and without the parentheses $\{\ \}$. For example, we talk about the additivity model $R + C$ in a two-way scheme. Notice, however, that we do not adopt the more sophisticated aspects of the GENSTAT model formula conventions (see Wilkinson and 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 a slightly unusual way.

## 5.2. Parameterisations of the mean.

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

$$Ey_i = \alpha_r + \beta_c$$

(subsuming $r = \bar{R}(i)$, $c = \bar{C}(i)$); or, in vector notation, with $\alpha = (\alpha_r) \in \mathbb{R}^R$ and $\beta = (\beta_c) \in \mathbb{R}^C$,

$$Ey = X_R\alpha + X_C\beta \quad .$$

More generally, there is an immediate one-to-one correspondence between model formulas and parameterisations, given by

$$\mathcal{T} \leftrightarrow \left( Ey = \sum_{T \in \mathcal{T}} X_T \alpha^T \right) \quad .$$

Notice that these parameterisations are usually not one-to-one. In fact, as soon as more than one factor is involved, there is a non-identifiability 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 distribution of $y$. We shall not discuss restrictions imposed on the parameters in order to make the parameterisation one-to-one. The non-identifiabilities are usually well justified by the applied context. For example, in a two way additive model, the parameter functions of interest are typically differences between row parameters, while absolute row levels are only meaningful in situations where random (or vanishing) 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 (such as the usual assumption stating that summation of any model term over any of its indices should give zero) should be regarded merely as computational tools, misleading as they are if the constrained parameters are considered "canonical", in some sense. In orthogonal designs, such constraints are not even needed as computational tools. 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 scheme can be written $R \times C$, $R \times C + R$, $R \times C + R + C + 0$, etc. These model formulas correspond to different parameterisations, each possessing its own rules for identifiability of contrasts. Some of these parameterisations may be rather useless, but there seems to be no a priori reason for excluding them.

## 5.3. Estimation of $\sigma^2$.

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 3.2.1, we have

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

From this we conclude that the orthogonal projection $P_L$ onto $L$ is given by

$$P_L = \sum_{G \in \mathcal{T}^*} Q_G \ ,$$

while the residual operator is

$$I - P_L = \sum_{G \notin \mathcal{T}^*} Q_G \ .$$

Accordingly, the residual sum of squares can be obtained from the ANOVA table and the factor structure diagram as the sum of all SSD's corresponding to factors which are not marginal to factors in the model formula,

$$SSD = \| y - P_L y \|^2 = \sum_{G \notin \mathcal{T}^*} \| Q_G y \|^2 = \sum_{G \notin \mathcal{T}^*} SSD_G \ .$$

The degrees of freedom for the residual sum of squares are, similarly,

$$d = \sum_{G \notin \mathcal{T}^*} d_G \ .$$

By standard linear model theory, this gives the variance estimate

$$\hat{\sigma}^2 = SSD/d = \frac{\sum_{G \notin \mathcal{T}^*} SSD_G}{\sum_{G \notin \mathcal{T}^*} d_G} \ .$$

## 5.4. 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 $SSD_0$ and $d_0$ denote the residual sum of squares 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

$$F(d_0 - d \ , \ d) = \frac{(SSD_0 - SSD)/(d_0 - d)}{SSD \ / \ d} \ .$$

All quantities in this expression are easily obtained from the ANOVA table and the factor structure diagram. In particular, in the most important case where $\mathcal{T}_0^*$ is obtained from $\mathcal{T}^*$ by removal of a single factor $T$, the test for "no $T$-effect" becomes

$$F(d_T, d) = \frac{SSD_T / d_T}{SSD / d} .$$

5.5. Estimation of the linear parameters.

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

5.5.1. Theorem. Consider the model

$$E\,y = \mu = \sum_{T \in \mathcal{T}} X_T \alpha^T \qquad (\alpha^T \in \mathbb{R}^T).$$

For $t'_o$ and $t''_o \in T_o$, $T_o \in \mathcal{T}$, the following two conditions are equivalent:

(1) The parameter function $\alpha^{T_o}_{t'_o} - \alpha^{T_o}_{t''_o}$ is estimable (i.e. it can be written as a function of $E\,y$).

(2) For any other factor $T \in \mathcal{T}$, $t'_o$ and $t''_o$ are nested in the same level of $T_o \wedge T$.

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

$$\hat{\alpha}^{T_o}_{t'_o} - \hat{\alpha}^{T_o}_{t''_o} = \bar{y}_{t'_o} - \bar{y}_{t''_o} ,$$

where $\bar{y}_{t_o} = S_{t_o} / n_{t_o}$ denotes the average of all observations on the level $t_o$ of $T_o$. The variance on this estimate is $\sigma^2 \cdot (1/n_{t'_o} + 1/n_{t''_o})$.

Remark. The phrase " $t_o'$ and $t_o''$ are nested in the same level of $T_o \wedge T$ " should, of course, be read as "the classes $\bar{T}_o^{-1}(t_o')$ and $\bar{T}_o^{-1}(t_o'')$ are contained in the same class among those determined by $T_o \wedge T$ ". We illustrate by an example: In a balanced three way scheme $A \times B \times C$, consider the model $A \times B + B \times C$, coordinatewise parameterised as $Ey_i = \alpha_{ab} + \beta_{bc}$. A contrast of the form $\alpha_{a'b'} - \alpha_{a''b''}$ is estimable if and only if $b' = b''$; or, equivalently, if and only if $(a', b')$ and $(a'', b'')$ are nested in the same level of $(A \times B) \wedge (B \times C) = B$.

Proof. First assume that (2) is satisfied, and define $l \in \mathbb{R}^I$ by

$$l_i = \begin{cases} 1/n_{t_o'} & \text{for } \bar{T}_o(i) = t_o' \\ -1/n_{t_o''} & \text{for } \bar{T}_o(i) = t_o'' \\ 0 & \text{otherwise.} \end{cases}$$

Notice that $l'y = \bar{y}_{t_o'} - \bar{y}_{t_o''}$. The vector $l$ belongs to $L_{T_o}$, because $l_i$ is constant on the classes determined by $T_o$. For any other factor $T \in \mathcal{T}$, we have

$$P_T l = P_T P_{T_o} l = P_{T \wedge T_o} l = 0,$$

because condition (2) implies that the two classes $\bar{T}_o^{-1}(t_o')$ and $\bar{T}_o^{-1}(t_o'')$ are contained in the same $T_o \wedge T$-class, from which it follows that averaging of $l_i$ over an arbitrary $T_o \wedge T$-class gives 0. Hence, the linear functional $l'$ vanishes on the subspaces $L_T$ for $T \neq T_o$. From this we conclude that

$$l'(\mu) = l'\Big(\sum_{T \in \mathcal{T}} X_T \alpha^T\Big) = l' X_{T_o} \alpha^{T_o} = \alpha^{T_o}_{t_o'} - \alpha^{T_o}_{t_o''}.$$

This means that our contrast is a function of $\mu$, and the maximum likelihood estimate is given by

$$l'(\hat{\mu}) = l' P_L y = (P_{T_o} l)' P_L y = l' P_{T_o} P_L y =$$

$$l' P_{T_o} y = (P_{T_o} l)' y = l' y = \bar{y}_{t_o'} - \bar{y}_{t_o''}.$$

Evaluation of the variance on this estimate is straightforward. It remains to be shown that (1) implies (2). Suppose that (2) is not satisfied, i.e. there exists a factor $T \in \mathcal{T}$ such that $t_o'$ and $t_o''$ are on different levels of $T_o \wedge T$. Put $H = T_o \wedge T$, and let $h'$ and $h''$ denote the corresponding levels of $H$. Let $M'$ and $M_o'$ denote the subsets of the sets $T$ and $T_o$, respectively, of factor levels nested in the level $h'$ of $H$. Now, suppose that the corresponding parameter vectors $\alpha^T$ and $\alpha^{T_o}$ are modified by addition of a constant $\lambda \neq 0$ to $\alpha_t^T$ for $t \in M'$, and subtraction of that same constant from $\alpha_{t_o}^{T_o}$ for $t_o \in M_o'$. This will leave the mean $\mu$ unchanged, while the contrast $\alpha_{t_o'}^T \rho - \alpha_{t_o''}^T \rho$ decreases by $\lambda$. From this we conclude that the contrast is not estimable.

## 6. VARIANCE COMPONENT MODELS.

By a variance component model we mean a model of the form

$$y = \sum_{T \in \mathcal{T}} X_T \alpha^T + \sum_{B \in \mathcal{B}} \sigma_B X_B u^B$$

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

Coordinatewise, 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 $t = \overline{T}(i)$, $b = \overline{B}(i)$). The idea is that the observation $y_i$ is assumed to be a sum of the fixed effects $\alpha_t^T$ $(T \in \mathcal{T})$ and the random effects $\sigma_B u_b^B$. The variance on a single observation $y_i$ is

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

and the parameters $\sigma_B^2$ are, accordingly, called variance

components.

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

$$Ey = \mu = \sum_{T \in \mathcal{T}} X_T \alpha^T ,$$

$$\mathrm{cov}(y) = \sum_{B \in \mathcal{B}} \sigma_B^2 X_B X_B' .$$

## 6.1. An example.

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

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

Coordinatewise, this model can be written

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

where $\alpha_r$ $(r \in R)$ and $\beta_c$ $(c \in C)$ are the row and column parameters, respectively, $\omega^2$ and $\sigma^2$ are the variance components, and $v_{rc}$ $((r,c) \in R \times C)$ and $u_i$ $(i \in I)$ are independent, normalized normally distributed. This is a two way additive model with random interaction, frequently referred to in applications as a justification for fitting an additive model to the cell averages in situations where the interaction is too large to be ignored against the intra-cell variation.

## 6.2. Model formulas.

A variance component model is specified by the two subsets $\mathcal{T}$ and $\mathcal{B}$ of $\mathcal{D}$. 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 " $\mathcal{T} + [\mathcal{B}]$ " . Notice that the linear models are variance component models with $\mathcal{B} = \{I\}$ , and our conventions for model formulas are consistent with those introduced in section 5.1 (page 23) when an error term $+[I]$ is subsumed. These ideas will be familiar to GENSTAT users (the model formulas $\mathcal{T}$ and $\mathcal{B}$ are simply those occuring in the 'TREATMENT' and 'BLOCKS' directives).

## 6.3. Assumptions.

As in our treatment of linear models, $\mathcal{T}$ is allowed to be an arbitrary subset of $\mathcal{D}$ . However, $\mathcal{B}$ is assumed to satisfy the following conditions:

($\mathcal{B}$1) $I \in \mathcal{B}$ .

($\mathcal{B}$2) All factors in $\mathcal{B}$ are balanced.

($\mathcal{B}$3) $\mathcal{B}$ is closed under the formation of minima.

($\mathcal{B}$4) The matrices $X_B X_B'$ are linearly independent.

Condition ($\mathcal{B}$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 ($\mathcal{B}$2) is restrictive, of course, but necessary for an explicit solution. It is wellknown that even the simplest variance component model, one way analysis of variance with random variation between groups, can not be solved explicitely if the group sizes are unequal.

Condition ($\mathcal{B}$3) is necessary for an algebraically nice solution, and somewhat restrictive in practice. The treatment of variance component models not satisfying ($\mathcal{B}$3) can, to some extend, be based on addition of the missing minima to $\mathcal{B}$ as random "pseudo" factors. The simplest example is the two way model $0 + [R + C + I]$ . Extension to $0 + [0 + R + C + I]$ gives rather satisfactory estimates of the variance components, which, however, may correspond to a covariance matrix (of the original model) which is not positively definite; in particular, we may

obtain a negative estimate of the variance on the grand mean $\bar{y}$ .

Condition ($\mathcal{B}$ 4) ensures identifiability of the variance components, cfr. the parameterisation of cov(y) (page 29). 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 factors).

Notice that we do not make explicit assumptions against non-estimability of variance components due to confounding with fixed effects. Obviously, a variance component $\sigma_B^2$ can not be estimated if $\mathcal{T}$ contains a factor finer than B . The similar problem for linear models occurs when $L = \mathbb{R}^I$ , with zero degrees of freedom left for the residual. Formally, it is an advantage not to exclude models with such non-estimable variance components, cfr. the above remarks on the model $0 + [0 + R + C + I]$ , where $\sigma_0^2$ is non-estimable in exactly this sense. However, our later results on estimation and hypothesis testing are obviously based on the (subsumed) assumption that the degrees of freedom involved are strictly positive.

## 6.4. The null analysis of variance.

The set $\mathcal{B}$ of random factors satisfies our conditions ($\mathcal{D}$ 1) , ($\mathcal{D}$ 2) and ($\mathcal{D}$ 3) (section 3.1, page 12) for the total set $\mathcal{D}$ of factors considered. Hence, by theorem 3.2.1 (page 13), $\mathcal{B}$ induces a decomposition of $\mathbb{R}^I$ similar to that induced by $\mathcal{D}$ . In order to distinguish, components of this new decomposition will be equipped with the superscript o . Thus,

$$\mathbb{R}^I = \bigoplus_{B \in \mathcal{B}} V_B^o$$

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

$$I = \sum_{B \in \mathcal{B}} Q_B^o \ .$$

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

$$SSD_B^o = \| Q_B^o y \|^2 \quad \text{and} \quad d_B^o = \dim V_B^o = \operatorname{tr}(Q_B^o) \ .$$

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

Obviously, the decomposition induced by $\mathfrak{B}$ is coarser than that induced by the whole design $\mathfrak{D}$, in the sense that each $V_B^o$ is a direct sum of some of the subspaces $V_G$ $(G \in \mathfrak{D})$. We say that the factor $G$ belongs to $B$-stratum if $V_G \subseteq V_B^o$. The rule for allocation of factors to strata by means of the factor structure diagram is as follows:

6.4.1. Proposition.

$$V_B^o = \bigoplus_{G \in \mathfrak{D}_B} V_G \ ,$$

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

$$\mathfrak{D}_B = \{ G \in \mathfrak{D} \mid B = \bigwedge_{\substack{B' \in \mathfrak{B} \\ G \subseteq B'}} B' \} \ .$$

Remark. In the decomposition after $\mathfrak{D}$ there may be components of dimension $d_G = 0$. Such factors $G$ are uniquely assigned to a stratum by the proposition, while the criterion $V_G \subseteq V_B^o$ assigns $G$ to any stratum. However, the allocation of such factors to strata is irrelevant for the analysis. The corresponding lines of the ANOVA table can simply be deleted.

Proof. Define a mapping $S: \mathfrak{D} \to \mathfrak{B}$ (S for stratum) by

$$S(G) = \bigwedge_{\substack{B' \in \mathcal{B} \\ G \subseteq B'}} B' \ .$$

Then $\mathcal{D}_B = S^{-1}(B)$ . From this it follows in particular that the sets $\mathcal{D}_B$ constitute a partitioning of $\mathcal{D}$ . Moreover, for any fixed $B_o \in \mathcal{B}$ we have

$$G \subseteq B_o \iff S(G) \subseteq B_o \ .$$

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

$$\{ G \in \mathcal{D} \mid S(G) \subseteq B_o \} = S^{-1}(\{B \in \mathcal{B} \mid B \subseteq B_o\})$$

$$= \bigcup_{\substack{B \in \mathcal{B} \\ B \subseteq B_o}} S^{-1}(B) = \bigcup_{\substack{B \in \mathcal{B} \\ B \subseteq B_o}} \mathcal{D}_B \ .$$

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

$$W_B^o = \bigoplus_{G \in \mathcal{D}_B} V_G \ .$$

Obviously, these subspaces constitute a decomposition

$$\mathbb{R}^I = \bigoplus_{B \in \mathcal{B}} W_B^o$$

of $\mathbb{R}^I$ as a direct sum of orthogonal subspaces, formed by collapse of the subspaces $V_G$ according to the partitioning $\mathcal{D} = \bigcup_B \mathcal{D}_B$ . From what was shown above, we conclude that

$$L_{B_o} = \bigoplus_{\substack{G \in \mathcal{D} \\ G \subseteq B_o}} V_G = \bigoplus_{\substack{B \in \mathcal{B} \\ B \subseteq B_o}} \Big( \bigoplus_{G \in \mathcal{D}_B} V_G \Big) = \bigoplus_{\substack{B \in \mathcal{B} \\ B \subseteq B_o}} W_B^o \ .$$

Hence, the decomposition $\mathbb{R}^I = \bigoplus W_B^o$ satisfies the condition of theorem 3.2.1 (page 13) for the decomposition with respect to $\mathcal{B}$ . Since this condition was shown to

characterize the decomposition uniquely, we must have

$$V_B^o = W_B^o = \bigoplus_{G \in \mathcal{G}_B} V_G \quad ,$$

which concludes the proof.

## 6.5. The canonical covariance parameters.

**6.5.1. Proposition.** The two sets of matrices

$$\{ X_B X_B' \mid B \in \mathcal{B} \} \quad \text{and} \quad \{ Q_B^o \mid B \in \mathcal{B} \}$$

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

*Proof.* It suffices to show that any of the matrices $X_B X_B'$ can be written as a linear combination of the matrices $Q_B^o$ , and vice versa. Since the random factors are assumed to be balanced, we have (cfr. page 8)

$$X_B X_B' = n_B P_B = n_B \sum_{\substack{B' \in \mathcal{B} \\ B' \subseteq B}} Q_{B'}^o \quad .$$

Conversely, by the remarks following the proof of corollary 3.2.2 (page 17), we have an expression

$$Q_B^o = \sum_{B' \in \mathcal{B}} b_B^{B'} P_{B'} = \sum_{B' \in \mathcal{B}} b_B^{B'} \frac{1}{n_{B'}} X_{B'} X_{B'}' \quad .$$

It follows in particular from the above proposition that we have an alternative parameterisation of the covariance matrix as

$$\mathrm{cov}(y) = \sum_{B \in \mathcal{B}} \lambda_B Q_B^o \quad .$$

The quantities $\lambda_B$ will be referred to as the canonical covariance parameters. The explicit solution of the variance component model is based on this parameterisation, which relies heavily on our assumptions $(\mathcal{B}1)$ - $(\mathcal{B}4)$ (page 30). In particular, the assumption $(\mathcal{B}3)$ , which

in the case of balanced k-factor designs was noticed by S.T. Jensen (1979), is essential. Szatrowski and Miller (1980) give the same condition in a less explicit form (they give the criterion for existence of explicit maximum likelihood estimates that the set of all coordinatewise products of columns of the $\mathcal{B} \times \{0, F_1, \ldots, F_k\}$-matrix $((v_{BF})) = ((1_{\{F \subseteq B\}}))$ has exactly $\#\mathcal{B}$ elements, but they do not notice the (equivalent) condition that the set of rows is closed under coordinatewise multiplication).

The connection between the two parameterisations is obtained as follows:

$$\sum_{B' \in \mathcal{B}} \lambda_{B'} Q^o_{B'} = \sum_{B' \in \mathcal{B}} \lambda_{B'} \Big( \sum_{B \in \mathcal{B}} b^B_{B'} \frac{1}{n_B} X_B X_B' \Big)$$

$$= \sum_{B \in \mathcal{B}} \frac{1}{n_B} \Big( \sum_{B' \in \mathcal{B}} b^B_{B'} \lambda_{B'} \Big) X_B X_B' \quad ,$$

i.e.

$$(6.5.2) \qquad \boxed{\sigma^2_B = \frac{1}{n_B} \sum_{B' \in \mathcal{B}} b^B_{B'} \lambda_{B'}}$$

where the coefficients $b^B_{B'}$ are determined by $Q^o_{B'} = \sum_B b^B_{B'} P_B$. And

$$\sum_{B \in \mathcal{B}} \sigma^2_B X_B X_B' = \sum_{B \in \mathcal{B}} \sigma^2_B n_B P_B$$

$$= \sum_{B \in \mathcal{B}} \sigma^2_B n_B \sum_{\substack{B' \in \mathcal{B} \\ B' \subseteq B}} Q^o_{B'} = \sum_{B' \in \mathcal{B}} \Big( \sum_{\substack{B \in \mathcal{B} \\ B' \subseteq B}} n_B \sigma^2_B \Big) Q^o_{B'} \quad ,$$

i.e.

$$(6.5.3) \qquad \boxed{\lambda_{B'} = \sum_{\substack{B \in \mathcal{B} \\ B' \subseteq B}} n_B \sigma^2_B}$$

## 6.6. Negative variance components.

The above discussion of the canonical parameterisation

ignores the problem of specifying the domain of variation for the new parameters $\lambda_B$ , $B \in \mathcal{B}$ . Proposition 6.5.1 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. The formula (6.5.3) expressing $\lambda_B$ as a linear combination of the variance components shows that nonnegative variance components imply nonnegative canonical parameters, but the converse is not always true. This leads to the wellknown problem of negative variance components, which can be explained as follows. A nice solution of the model under the canonical parameterisation is only possible when the parameters are allowed to vary freely in their "natural" domain, which is obviously given by $\lambda_B \geq 0$ . This means that calculation of estimates $\hat{\sigma}_B^2$ from estimates $\hat{\lambda}_B$ may lead to negative estimates for some of the variance components. We shall not, in this paper, discuss formal procedures for estimation of the variance components in their original domain $\sigma_B^2 \geq 0$ . In practice, this seems to be a minor problem. The interpretation of a negative variance component $\sigma_B^2$ is that the correlation between observations in the same B-class is smaller than the correlation 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. Very often, the occurrence of a negative variance component estimate can be taken as a wellcome opportunity to simplify the model by removal of the corresponding factor B from the model. Of course, a <u>significantly</u> negative estimate of a variance component, which ought to 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 the problem and work with the extended model given by $\lambda_B \geq 0$ .

## 6.7. Solving the variance component model.

The basic observation behind the explicit solution to the classical variance component model is that the decomposition according to $\mathcal{B}$ decomposes the data vector $y$ as a sum of stochastically independent components $Q_B^o y$, one for each stratum, and each of these components is described by its own linear model. Indeed, the data components $Q_B^o y$ are easily seen to be independent, normally distributed with mean

$$\mu_B = E(Q_B^o y) = Q_B^o(Ey) = Q_B^o(\mu)$$

and covariance matrix

$$\mathrm{cov}(Q_B^o y) = \lambda_B Q_B^o .$$

The parameters $\mu_B$ and $\lambda_B$ of the distribution of $Q_B^o y$ are functionally independent of those describing the distributions of the remaining data components. Thus, estimation in the original model boils down to estimation in each stratum of the parameters $\mu_B$ and $\lambda_B$ from $Q_B^o y$. This is straightforward, because the model for $Q_B^o y$ is essentially (apart from the lack of a coordinate system) an ordinary linear model with data space $V_B^o$. The covariance matrix $\lambda_B Q_B^o$ is a constant $\lambda_B$ times the "identity" $Q_B^o$ on $V_B^o$, and $\mu_B$ varies in the linear subspace $L \cap V_B^o$ ( $L = \sum_{T \in \mathcal{T}} L_T$ ). The orthogonal projection onto this space is $P_L Q_B^o$, since $L$ and $V_B^o$ are geometrically orthogonal. Estimating as usual in a linear model, we obtain the estimates

$$\hat{\mu}_B = P_L Q_B^o y$$

and (provided that $\dim V_B^o > \dim(L \cap V_B^o)$ )

$$\hat{\lambda}_B = \frac{\| Q_B^o y - P_L Q_B^o y \|^2}{\dim V_B^o - \dim(L \cap V_B^o)} .$$

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

$$\hat{\mu} = \sum_{B \in \mathcal{B}} \hat{\mu}_B = P_L y ,$$

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

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

$$P_L = \sum_{G \in \mathcal{T}^*} Q_G$$

( $\mathcal{T}^*$ is defined in section 5.3, page 24), and

$$Q_B^o = \sum_{G \in \mathcal{D}_B} Q_G$$

( $\mathcal{D}_B$ is defined in proposition 6.4.1, page 32). Thus, the residual operator for our linear model in B - stratum is

$$Q_B^o - P_L Q_B^o = \sum_{G \in \mathcal{D}_B \setminus \mathcal{T}^*} Q_G \quad ,$$

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

$$\| Q_B^o y - P_L Q_B^o y \|^2 = \sum_{G \in \mathcal{D}_B \setminus \mathcal{T}^*} SSD_G \quad .$$

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

$$\hat{\lambda}_B = \frac{\sum_G SSD_G}{\sum_G d_G} \quad ,$$

where both sums are to be taken over $G \in \mathcal{D}_B \setminus \mathcal{T}^*$, i.e. the set of all factors in B - stratum which are not systematic factors or marginal to systematic factors. 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

$$\hat{\lambda}_B = SSD_B / d_B \quad .$$

6.8. Estimation of the variance components.

Estimates $\hat{\sigma}_B^2$ are immediately obtained from the estimates $\hat{\lambda}_B$ by means of formula (6.5.2) (page 35). Notice that these estimates are not $\chi^2$-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, see e.g. Scheffé (1959), Searle (1971).

## 6.9. Hypothesis testing, linear structure.

Let $\mathcal{T}_o$ be a subset of $\mathcal{T}$ specifying a reduced model $\mathcal{T}_o + [\beta]$ . We assume $\mathcal{T}_o^* \subseteq \mathcal{T}^*$ and, accordingly, $L_o \subseteq L$ (where $L_o = \sum_{T \in \mathcal{T}_o} L_T$ ). In order to give an explicit statistical test for the model reduction

$$\mathcal{T} + [\beta] \rightarrow \mathcal{T}_o + [\beta]$$

we must assume that the corresponding square sum

$$\| P_L y - P_{L_o} y \|^2 = \sum_{G \in \mathcal{T}^* \setminus \mathcal{T}_o^*} SSD_G$$

consists of contributions from a single stratum $B_o$ only, i.e.

$$\mathcal{T}^* \setminus \mathcal{T}_o^* \subseteq \mathcal{T}_{B_o} .$$

Notice that this condition is in particular satisfied in the frequently occuring case where $\mathcal{T}_o^*$ is obtained from $\mathcal{T}^*$ by removal of a single factor.

Under the above condition, the model reduction can be regarded as a reduction of the linear model for the data component $Q_{B_o}^o y$ , 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 $B_o$ - stratum,

$$F(d_1, d_2) = \frac{SSD_1 / d_1}{SSD_2 / d_2}$$

where $SSD_1 = \sum_{G \in \mathcal{T}^* \setminus \mathcal{T}_o^*} SSD_G$ , $d_1 = \sum_{G \in \mathcal{T}^* \setminus \mathcal{T}_o^*} d_G$ ,

$$SSD_2 = \sum_{G \in \mathcal{T}_{B_o} \setminus \mathcal{T}^*} SSD_G \quad , \quad d_2 = \sum_{G \in \mathcal{T}_{B_o} \setminus \mathcal{T}^*} d_G .$$

The rules for inclusion of terms $SSD_G$ (and $d_G$) in nominator and denominator are exactly as in the test for reduction of a linear model, cfr. section 5.4 (page 25), except that only factors from $B_0$ - stratum should be taken into account. Usually, when forming the ANOVA table to analyze a variance component model, it is convenient to arrange the lines in such a way that strata are collected as subtables. 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 stratum. Formally, this corresponds to simultaneous reduction of linear models for separate data sets, and there seems to be no simple way of doing it. The best we can do is to carry out the relevant F - test in each of the strata involved (usually not more than two). If no clear conclusion comes out of this, some kind of weighted test statistic, summarizing the information from the strata involved, may be considered. A discussion of this is beyond the scope of the present paper.

## 6.10. Hypothesis testing, covariance structure.

We shall restrict our attention to reductions of the form $\mathcal{T} + [\mathcal{B}] \rightarrow \mathcal{T} + [\mathcal{B}_0]$, 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 the conditions $(\mathcal{B}1) - (\mathcal{B}4)$ of section 6.3 (page 30). This means that the "measurement error" $I$ must not be removed (of course), and that $\mathcal{B}_0$ should again be closed under the formation of minima. (conditions $(\mathcal{B}2)$ and $(\mathcal{B}4)$ 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 $B$, i.e.

$$(6.10.1) \qquad B_o = \bigwedge_{\substack{B' \in \mathcal{B}_o \\ B \subseteq B'}} B' \neq B \ .$$

Indeed, if this condition is satisfied we can obviously not have $B' \wedge B'' = B$ for factors $B'$ and $B'' \in \mathcal{B}_o$, which means that $B' \wedge B''$ must be among the factors left in $\mathcal{B}_o$. Conversely, if the above condition was not satisfied we would have a collection of factors from $\mathcal{B}_o$ (namely those finer than $B$) possessing a minimum not in $\mathcal{B}_o$.

Under the above condition (6.10.1), we have

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

This follows from (6.5.3) (page 35), noticing that the expression for $\lambda_B$ differs from the expression for $\lambda_{B_o}$ only by the occurrence of the term $n_B \sigma_B^2$. Hence, the hypothesis $\sigma_B^2 = 0$ is equivalent to $\lambda_B = \lambda_{B_o}$. Recalling our interpretation of the model as a product of linear models for the data components $Q_{B'}^o y$, $B' \in \mathcal{B}$, 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 a comparison of the two variance estimates by a two-sided $F$-test on their ratio, i.e.

$$F(d, d_o) = \hat{\lambda}_B / \hat{\lambda}_{B_o} \ ,$$

where $d$ and $d_o$ are the degrees of freedom occurring in the denominators of the expressions for $\hat{\lambda}_B$ and $\hat{\lambda}_{B_o}$, respectively. Large values of this $F$-statistic indicates $\lambda_B > \lambda_{B_o}$, or $\sigma_B^2 > 0$. Small values indicate that $\sigma_B^2$ is negative. Thus, the test should be carried out as two-sided when negative values of $\sigma_B^2$ are taken into account. Apart from this, the test is formally equivalent to a test in a linear model, namely the test for the reduction of the linear model in $B_o$-stratum (in the decomposition after $\mathcal{B}_o$) given by

$$\mathcal{T} + B + [\mathcal{B}_o] \rightarrow \mathcal{T} + [\mathcal{B}_o] \ .$$

## 6.11. Estimation of contrasts.

It was shown in section 6.7 (page 37-38) that the maximum likelihood estimate for the mean $\mu = Ey$ coincides with that of the linear model specified by $\mathcal{T}$. In particular, a contrast of the form $\alpha^T_{t'} - \alpha^T_{t''}$ ( $T \in \mathcal{T}$, $t', t'' \in T$ ) should be estimated as in the linear model, by the difference between the corresponding averages. Since the rules for identifiability of contrasts are obviously the same as in the linear model, it remains only to give the formula for the variances on estimated contrasts.

**6.11.1. Proposition.** Let $\alpha^T_{t'_o} - \alpha^T_{t''_o}$ be an estimable contrast (i.e. $t'_o$ and $t''_o$ are nested in the same level of $T \wedge T_o$ for any other $T \in \mathcal{T}$ ). Then the estimate $\bar{y}_{t'_o} - \bar{y}_{t''_o}$ has the variance

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

where

$$c_B = \begin{cases} 0 & \text{for } t'_o \text{ and } t''_o \text{ on the same level of } T_o \wedge B \\ \frac{1}{n_{h'}} + \frac{1}{n_{h''}} & \text{for } t'_o \text{ and } t''_o \text{ on the distinct levels } h' \text{ and } h'' \text{ of } H = T_o \wedge B \, . \end{cases}$$

*Proof.* Let $l \in \mathbb{R}^I$ be defined as in the proof of theorem 5.5.1 (page 27). Then $l'y = \bar{y}_{t'_o} - \bar{y}_{t''_o}$, and since $l \in L_{T_o}$ we have

$$\operatorname{var}(l'y) = l'\operatorname{cov}(y)l = l'\Big(\sum_{B \in \mathcal{B}} \sigma_B^2 X_B X_B'\Big) l$$

$$= \sum_{B \in \mathcal{B}} n_B \sigma_B^2 l' P_B l = \sum_{B \in \mathcal{B}} n_B \sigma_B^2 l' P_{T_o} P_B P_{T_o} l$$

$$= \sum_{B \in \mathcal{B}} n_B \sigma_B^2 \| P_{T_o \wedge B} l \|^2 \, .$$

The proposition follows if we can show that $\|P_{T_o \wedge B} 1\|^2 = c_B$. Noticing that the operator $P_{T_o \wedge B}$ replaces each $l_i$ by the average over the corresponding $T_o \wedge B$-class, this is a matter of straightforward computations which are left to the reader.

6.11.2. Corollary. Suppose that $T_o$ belongs to $B_o$-stratum and that $t'_o$ and $t''_o$ are on the same level of $B \wedge T_o$ for any random factor $B$ which is strictly coarser than $B_o$. Then

$$\operatorname{var}(\bar{y}_{t'_o} - \bar{y}_{t''_o}) = \lambda_{B_o}\left(\frac{1}{n_{t'_o}} + \frac{1}{n_{t''_o}}\right) .$$

Proof. Under the above assumptions we have

$$c_B = \begin{cases} \frac{1}{n_{t'_o}} + \frac{1}{n_{t''_o}} & \text{for } B_o \subseteq B \\ 0 & \text{otherwise.} \end{cases}$$

Indeed, if $B$ is not finer than $B_o$, the minimum $B \wedge B_o$ is strictly coarser than $B_o$, and it follows that

$$c_B = \|P_{B \wedge T_o} 1\|^2 = \|P_{B \wedge (B_o \wedge T_o)} 1\|^2 = \|P_{(B \wedge B_o) \wedge T_o} 1\|^2 = c_{B \wedge B_o} = 0 .$$

For $B_o \subseteq B$ we have

$$c_B = \|P_{B \wedge T_o} 1\|^2 = \|P_{T_o} 1\|^2 = \|1\|^2 = \frac{1}{n_{t'_o}} + \frac{1}{n_{t''_o}} .$$

Hence, by formula (6.5.3) (page 35),

$$\operatorname{var}(\bar{y}_{t'_o} - \bar{y}_{t''_o}) = \sum_{\substack{B \in \mathcal{B} \\ B_o \subseteq B}} n_B \sigma_B^2 \left(\frac{1}{n_{t'_o}} + \frac{1}{n_{t''_o}}\right) = \left(\frac{1}{n_{t'_o}} + \frac{1}{n_{t''_o}}\right) \lambda_{B_o} .$$

Contrasts satisfying the condition of corollary 6.11.2 are said to be estimable in a single stratum (namely $B_o$ - stratum). For such contrasts -and only for such contrasts- can the pairwise comparison of the levels $t'_o$ and $t''_o$ be performed as an exact t-test, since the estimated variance on $\bar{y}_{t'_o} - \bar{y}_{t''_o}$ is $\chi^2$-distributed.

6.12. An example.

In the split-plot experiment of example 4.2.4 (page 21-22), consider the model $A \times B + [P + I]$ ; or, in parametric terms,

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

Proposition 6.4.1 and the factor structure diagram (page 22) gives the allocation of factors to strata:

$$\Omega_I = \{ I , A \times B , B \} ,$$

$$\Omega_P = \{ P , A , 0 \} ,$$

reflecting the obvious fact that A-contrasts (in the additive model $A + B + [P + I]$ ) are estimated from plot totals, while the estimation of B-contrasts and test for $A \times B$ -interaction is based on differences within plots. The ANOVA table, arranged by strata, is

| stratum | effect | d.f. | SSD |
|---|---|---|---|
| I | I | 10 | $SSD_I$ |
| | A×B | 4 | $SSD_{A \times B}$ |
| | B | 1 | $SSD_B$ |
| P | P | 10 | $SSD_P$ |
| | A | 4 | $SSD_A$ |
| | 0 | 1 | $SSD_0$ |

The test for $\omega^2 = 0$ is

$$F(10,10) = \frac{SSD_P/10}{SSD_I/10} \ .$$

If this is accepted, the two strata collapse, and we are left with an ordinary $5 \times 2$ - scheme with 3 observations per cell. However, we shall assume that this hypothesis is rejected (or not considered at all), i.e. that the division into plots is relevant. In this case, the test for interaction (i.e. the reduction $A \times B + [P + I] \rightarrow A + B + [P + I]$ ) is performed in $I$ - stratum,

$$F(4,10) = \frac{SSD_{A \times B} / 4}{SSD_I / 10} .$$

If this is accepted, we have 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

$$F(1,14) = \frac{SSD_B / 1}{(SSD_I + SSD_{A \times B})/14} ,$$

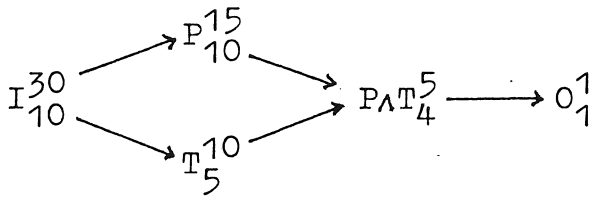
$$F(4,10) = \frac{SSD_A / 4}{SSD_P / 10} .$$

Estimation of variances on contrast estimates is straightforward in this case, by corollary 6.11.2 (page 43).

However, suppose that additivity can not 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 would be $I$, $P$, $T$ and $0$, but the condition ($\mathfrak{A}$ 3) (section 3.1, page 12) forces us to include the "pseudo" factor $P \wedge T$ ( = $A$ above) on 5 levels, reflecting the arrangement of treatments in pairs. Thus, we take

$$\mathfrak{A} = \{ I , P , T , P \wedge T , 0 \}$$

and obtain the factor structure diagram on top of page 46. This diagram is the same as that referred to above, except that the factor $B$ is removed and the degrees of freedom changed accordingly.




The ANOVA table becomes

| stratum | effect | d.f. | SSD |
|---|---|---|---|
| I | I | 10 | $SS_I - SS_P - SS_T + SS_{P\wedge T}$ |
| | T | 5 | $SS_T - SS_{P\wedge T}$ |
| P | P | 10 | $SS_P - SS_{P\wedge T}$ |
| | P∧T | 4 | $SS_{P\wedge T} - SS_0$ |
| | 0 | 1 | $SS_0$ |

The test for overall treatment effect provides an example of a test which does not take place in a single stratum. Indeed, the model reduction $T + [P + I] \rightarrow 0 + [P + I]$ corresponds to removal of both $T$ and $P \wedge T$ from the maximal model formula $\mathcal{F}^* = \{T, T\wedge P, 0\}$, and these two factors are not in the same stratum. All we can do is to perform the two F-tests

$$F(5,10) = \frac{SSD_T / 5}{SSD_I / 10}$$

testing for differences between treatments within pairs, and

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

testing for differences between pair totals.

Similarly, the variances on certain contrast estimates are not estimable within a single stratum. For treatments $t'$ and $t''$ in the same pair, we have (by the corollary, since $t'$ and $t''$ are on the same level of $P \wedge T$ in this case)

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

but for $t'$ and $t''$ not in the same pair we must apply proposition 6.11.1 (page 42), obtaining

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

References.

Cochran, William G. and Cox, Gertrude M. (1957). Experimental designs. Wiley.

James, A. T. and Wilkinson, Graham N. (1971). Factorization of the residual operator and canonical decomposition of nonorthogonal factors in the analysis of variance. Biometrika 58, 279-294.

Jensen, S. Tolver (1979). Varianskomponentmodeller for fuldstændigt balancerede forsøg. Inst. Math. Stat., Un. Copenhagen.

Nelder, John A. (1954). The interpretation of negative components of variance. Biometrika 41, 544-548.

Nelder, John A. (1965). The analysis of randomized experiments with orthogonal block structure, I and II. Proc. Roy. Soc. Lond. A 283, 147-178.

Nelder, John A. (1969). The description of statistical data structures for statistical computing. In Statistical Computation, ed. Roy C. Milton and John A. Nelder. Academic Press.

Nelder, John A. (1977). A reformulation of linear models (with discussion). JRSS A 140, 48-76.

Nelder, John A. et al. The GENSTAT manual. Rothamsted Experimental Station.

Scheffé, Henry (1959). The Analysis of Variance. Wiley.

Searle, S. R. (1971). Topics in variance component estimation. Biometrics 27, 1-76.

Szatrowski, Ted H. and Miller, John J. (1980). Explicit maximum likelihood estimates from balanced data in the mixed model of the analysis of variance. Ann. Stat. 8, 811-819.

Wilkinson, Graham N. (1969). Facilities in a statistical program system for analysis of multiply-indexed data. In Statistical Computing, ed. Roy C. Milton and John A. Nelder. Academic Press.

Wilkinson, Graham N. (1970). A general recursive procedure for analysis of variance. Biometrika 57, 19-46.

Wilkinson, Graham N. and Rogers, C. E. (1973). Symbolic description of factorial models for analysis of variance. Appl. Statist. 22, 392-399.

# PREPRINTS 1981

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK.

No. 1 Johansen, Søren: Asymptotic Inference in Random Coefficient Regression Models.

No. 2 Asmussen, Søren: On the Role of a Certain Eigenvalue in Estimating the Growth Rate of a Branching Process.

No. 3 Lauritzen, Steffen.: Time Series Analysis in 1880. A Discussion of Contributions made by T.N. Thiele.

No. 4 Asmussen, Søren: Conditioned Limit Theorems Relating a Random Walk to its Associate, with Applications to Risk Reserve Processes and the GI/G/1 Queue.

No. 5 Johansen, Søren: The Statistical Analysis of a Markov Branching Process.

No. 6 Asmussen, Søren: Time - Dependent Approximations in some Queueing Systems with Imbedded Markov Chains Related to Random Walks.

No. 7 Skovgaard, Ib M.: A Second - Order Investigation of Asymptotic Ancillarity.

No. 8 Rootzén, Holger: The Rate of Extremes of Stationary Normal Sequences.

No. 9 Skovgaard, Ib M.: Large Deviation Approximations for Maximum Likelihood Estimators.

No. 10 Jensen, Ulla Funck: A Stochastic Projection Model with Implications for Multistate Demography and Manpower Analysis.

No. 11 Johansen, Søren: An Extension of Cox's Regression Model.

No. 12 Skovgaard, Ib M.: Edgeworth Expansions in Statistics.

PREPRINTS 1982

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK.

No. 1 Holmgaard, Simon and Yu, Song Yu: Gaussian Markov Random Fields Applied to Image Segmentation.

No. 2 Andersson, Steen A., Brøns, Hans K. and Jensen, Søren Tolver: Distribution of Eigenvalues in Multivariate Statistical Analysis.

No. 3 Tjur, Tue: Variance Component Models in Orthogonal Designs.