Chapter 5
Last Words on Anova?

Terry Speed

Many people like to say the last words in an academic debate, and I am no exception. I have tried to do this on a few occasions, only to discover that when I came to say my piece, everyone had left the room. The analysis of variance is a case in point, and my comments on Tukey's contributions to anova explain the problem. If – as I believe to be true – people don't care much these days what Tukey thought about anova, they are going to care even less what I think. This is not said with any sense of bitterness. Indeed I regard myself as something of a student of fads, fashions and trends in statistics, so why should I expect otherwise? Nevertheless, I'm very happy to see these articles reprinted, as their easier availability may kindle the interest of someone, somewhere, sometime in what I still believe to be an important part of (the history of) our subject.

My main stimulus for work in this area came from the papers of six people: R.A. Fisher, Frank Yates, and John A. Nelder from the U.K, indeed all from Rothamsted, Alan T. James and Graham Wilkinson from Adelaide, Australia, and John W. Tukey from the U.S.A. Unpublished lecture notes by James were extremely helpful in getting me going. The anova program within GENSTAT, initially created by Wilkinson based on research by James, Wilkinson, James & Wilkinson and Nelder, was enormously influential. It was (and remains) truly brilliant in conception and execution, and I wanted to understand it. For a long time I was interested in – one might say obsessed with – the symmetries underlying much of anova, and that is reflected in some of the papers reprinted (thank you Rosemary Bailey)! But also I wanted to understand how users of anova saw things, including gory details such as the combination of information, the analysis of covariance and dealing with missing values, all topics with wonderful histories. I made one attempt to put it all together for general consumption, but that got rejected, and so I moved on to other things. As explained above, it is not clear how many people now care. I hope you enjoy the papers. There are several more if you do.

T. Speed
Department of Statistics, University of California, Berkeley, and Division of Bioinformatics, Walter and Eliza Hall Institute, Australia
e-mail: terry@stat.berkeley.edu

S. Dudoit (ed.), *Selected Works of Terry Speed*, Selected Works in Probability and Statistics,
DOI 10.1007/978-1-4614-1347-9_5,

References

[1] R. A. Bailey and T. P. Speed. On a class of association schemes derived from lattices of equivalence relations. In P. Schultz, C. E. Praeger, and R. Sullivan, editors, *Algebraic Structures and their Applications*. Marcel Dekker, New York, 1982.

[2] R. A. Bailey and T. P. Speed. A note on rectangular lattices. *Ann. Stat.*, 14: 874–895, 1986.

[3] T. P. Speed and R. A. Bailey. Factorial dispersion models. *Int. Stat. Rev.*, 55: 261–277, 1987.

[4] R. A. Bailey, C. E. Praeger, C. A. Rowley, and T. P. Speed. Generalized wreath products of permutation groups. *Proc. Lond. Math. Soc. (3)*, 47:69–82, 1983.

[5] A. Houtman and T. P. Speed. Balance in designed experiments with orthogonal block structure. *Ann. Stat.*, 11:1069–1085, 1983.

[6] A. Houtman and T. P. Speed. The analysis of multi-stratum designed experiments with incomplete data. *Aust. J. Stat.*, 26:227–246, 1984.

[7] T. P. Speed. John W. Tukey's contributions to analysis of variance. *Ann. Stat.*, 30(6):1649–1665, 2002.

[8] T. P. Speed. ANOVA models with random effects: An approach via symmetry. In J. Gani and M. B. Priestley, editors, *Essays in Time Series and Allied Processes : Papers in Honour of E. J. Hannan*, pages 355–368. Applied Probability Trust, Sheffield, UK, 1986.

[9] T. P. Speed. What is an analysis of variance? *Ann. Stat.*, 15:885–910, 1987.

GENERALIZED WREATH PRODUCTS OF PERMUTATION GROUPS

R. A. BAILEY, CHERYL E. PRAEGER, C. A. ROWLEY, *and* T. P. SPEED

[Received 19 July 1982]

1. *Introduction*

For $i = 1, 2$, let (G_i, Δ_i) be a permutation group. The *permutation direct product* $(G_1, \Delta_1) \times (G_2, \Delta_2)$ is $(G_1 \times G_2, \Delta_1 \times \Delta_2)$ with action defined by

$$(\delta_1, \delta_2)(g_1, g_2) = (\delta_1 g_1, \delta_2 g_2).$$

If $\Delta_1 \times \Delta_2$ is visualized as a rectangular array, an element of $G_1 \times G_2$ may be described as a permutation of rows by an element of G_1 followed by a permutation of columns by an element of G_2.

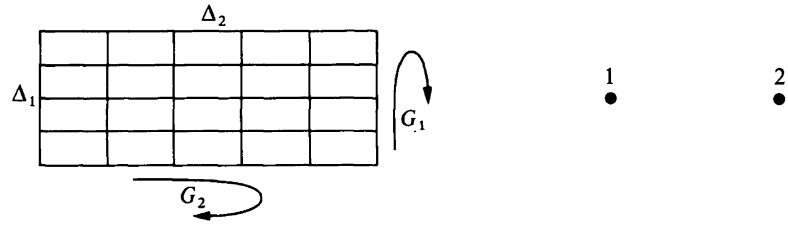

FIG. 1

The *permutation wreath product* $(G_1, \Delta_1) \mathrm{wr} (G_2, \Delta_2)$ is $(G_1^{\Delta_2} \times G_2, \Delta_1 \times \Delta_2)$, with action defined by

$$(\delta_1, \delta_2)(f, g_2) = (\delta_1(\delta_2 f), \delta_2 g_2),$$

where f is a function from Δ_2 to G_1. Thus an informal description of an element of $G_1 \mathrm{wr} G_2$ is 'independent permutations of the points within each column by elements of G_1, followed by a permutation of the columns by an element of G_2'.

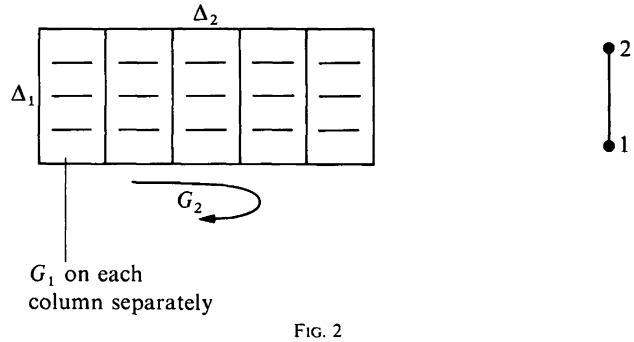

FIG. 2

In the first case, the indexes 1 and 2 play an equal role, and the *rows* of the array are significant. In the second case, the index 2 *dominates* the index 1, and the rows have no significance.

Sets with the two structures described above are frequently used in designed experiments. Nelder [**9**] described a class of structures obtained from these by successive *crossing* (corresponding to the direct product) and *nesting* (corresponding to the wreath product). He developed a large body of theory for these structures and asked whether only these structures satisfied his results.

However, there are many structures that are recognized as tractable by designers of statistical experiments, but which are not in Nelder's class: the simplest such was described by Throckmorton [**14**].

EXAMPLE 1. The set is divided into *rows* (index 1) crossed with *columns* (index 2). Each row is subdivided into *minirows* (index 3), which meet all columns. Within each *square* (row–column intersection), the fragments of minirows are crossed with *microcolumns* (index 4). Thus 1 dominates 3 and 4, while 2 dominates 4.

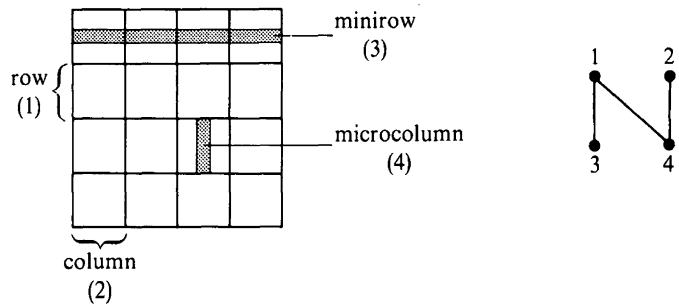

FIG. 3

Although this is not one of Nelder's structures, most statisticians working in the design of experiments could describe its automorphism group, in such terms as 'permute rows, permute columns, within each row separately permute minirows, within each square separately permute microcolumns'.

A more precise discussion of structures such as this is given by Speed and Bailey [**12**], and Bailey [**2**]: however, our concern here is with the associated *permutation groups*. In § 3 of this paper we introduce an explicit description of the elements of both the full automorphism group of such a structure and some of its subgroups. For Nelder's structures these groups can be obtained by successively forming direct products and wreath products of the appropriate permutation groups (group actions), but for our more general class we need to use a construction which Wells [**15**, § 7] described for actions of semigroups: he called it *the wreath product of an ordered set of actions*. The *ordered set* here is a partially ordered set, the partial order being given by the combinatorial structure. (The right-hand part of Figs 1–3 shows the appropriate partially ordered set.) We need to prove that if we start with *group* actions then Wells's wreath product action is also a group action.

Since statistical experiments are, necessarily, finite, our main interest is in structures defined by *finite* partially ordered sets: in this case, as is quite straightforward to check

using the results of §4, the permutation group which we construct is identical with that constructed by Holland [6] and Silcock [11]. This is why we call it a generalized wreath product. However, finiteness is not essential to all of the theory: and so we have not confined our attention to this case.

Throughout this paper we rely heavily on Wells's explicit representation of the elements of the generalized wreath product. We use a slightly modified version based on the notation introduced in §2.

In §§5–6 we address two questions which are relevant to the use of these structures in designed experiments. What are the orbits on pairs of points (§5)? In particular, which subgroups have the 'expected' orbits? What are the irreducible constituents of the permutation linear representation (§6)? In particular, when is the centralizer algebra commutative?

2. *Notation and terminology*

The notation introduced in this section is used without comment in the rest of this paper. Throughout, (I, ρ) denotes a partially ordered set. We shall find it convenient to use both of the symbols ρ and $\leqslant$ for the partial order, the former in descriptive work and the latter in computations (when we shall also use the associated symbols $<$, $>$, and $\geqslant$ with their obvious meanings).

DEFINITIONS. Following Grätzer [4], we define a subset J of I to be

hereditary if, whenever $i < j$ and $j \in J$, then $i \in J$;

ancestral if, whenever $i > j$ and $j \in J$, then $i \in J$;

a *chain* if, whenever $i, j \in J$, then either $i \leqslant j$ or $j \leqslant i$;

an *antichain* if, whenever $i, j \in J$ and $i \neq j$, then neither $i \leqslant j$ nor $j \leqslant i$.

For $i \in I$ we define

$$A(i) = \{j \in I\colon j > i\}, \quad A[i] = \{j \in I\colon j \geqslant i\},$$

$$H(i) = \{j \in I\colon j < i\}, \quad H[i] = \{j \in I\colon j \leqslant i\},$$

and for $J \subseteq I$ we define

$$A(J) = \bigcup_{i \in J} A(i), \quad A[J] = \bigcup_{i \in J} A[i],$$

$$H(J) = \bigcup_{i \in J} H(i), \quad H[J] = \bigcup_{i \in J} H[i].$$

Note that all the A-subsets are ancestral and all the H-subsets are hereditary.

For $i \in I$, let Δ_i be a set with $|\Delta_i| \geqslant 2$ (this restriction is to avoid irritating special cases). For $J \subseteq I$, put $\Delta_J = \prod_{i \in J} \Delta_i$. If $K \subseteq J \subseteq I$, let π_K^J denote the natural projection from Δ_J onto Δ_K. If $K = \{k\}$, we shall often write π_k^J for π_K^J. We shall also abbreviate Δ_I to Δ and π_J^I to π_J. We shall need $\Delta_{A(i)}$ and $\pi_{A(i)}$ so often that we abbreviate them to Δ^i and π^i respectively.

We write elements of Δ as $\delta = (\delta_i)$ with $\delta_i \in \Delta_i$. For $J \subseteq I$ we define the equivalence relation $\underset{J}{\sim}$ on Δ by $\delta \underset{J}{\sim} \varepsilon$ if and only if $\delta\pi_J = \varepsilon\pi_J$.

For each $i \in I$ let G_i be a (faithful) permutation group on Δ_i with identity 1_i, and let F_i be the set of all functions from Δ^i into G_i. For $J \subseteq I$ put $F_J = \prod_{i \in J} F_i$, and let $F = F_I$. We write elements of F as $f = (f_i)$ with $f_i \in F_i$. If $K \subseteq J \subseteq I$, let φ_K^J denote the natural projection from F_J to F_K. The abbreviations φ_k^J and φ_J are used analogously to π_k^J and π_J.

Note that if J is empty then both Δ_J and F_J are singletons, and that if J is infinite then each is the full Cartesian product.

3. *Specification of the generalized wreath product*

Our aim is to identify F with a set of functions from Δ to Δ. We do this by defining an 'action' of F on Δ and showing that this action is, in a natural sense, faithful.

DEFINITION. For each $f = (f_i) \in F$, we define an *action* of f on Δ by the following rule: for each $\delta = (\delta_i) \in \Delta$,

$$\delta f = \varepsilon, \quad \text{where } \varepsilon = (\varepsilon_i) \in \Delta, \text{ and } \varepsilon_i = \delta_i(\delta\pi^i f_i);$$

that is, $(\delta f)_i = \delta_i(\delta\pi^i f_i)$.

LEMMA 1. *The action on Δ is faithful in the sense that if $f, h \in F$ and if $\delta f = \delta h$ for all $\delta \in \Delta$ then $f = h$.*

Proof. Let $i \in I$ and let $\gamma \in \Delta^i$. Put $x = \gamma f_i$ and $y = \gamma h_i$, so that $x, y \in G_i$. Let $\alpha \in \Delta_i$, and choose $\delta \in \Delta$ so that $\delta_i = \alpha$ and $\delta\pi^i = \gamma$. Then $(\delta f)_i = \alpha x$ and $(\delta h)_i = \alpha y$ so $\alpha x = \alpha y$. We can choose such a δ for every $\alpha \in \Delta_i$, and so $x = y$. Thus $\gamma f_i = \gamma h_i$ for all $\gamma \in \Delta^i$, and so $f_i = h_i$.

Lemma 1 shows that the above definition identifies F with a subset of Δ^Δ, and henceforth we shall regard F as this subset. Thus elements of F are 'multiplied' according to the composition of the corresponding functions from Δ to Δ. We wish to prove that this subset F of the semigroup Δ^Δ is a sub*group*. To do this we first need to investigate in some detail the relationship between the action of F on Δ and the partial order on I; these results (Lemmas 2, 3, and 6) are also used in subsequent sections. The proof that F is a submonoid (Lemmas 4 and 5) is immediate but the existence of inverses in F (Lemmas 7 and 8) requires a restriction on the partial order on I.

LEMMA 2. *Let J be an ancestral subset of I; let $\delta, \varepsilon \in \Delta$ and $f \in F$. If $\delta \underset{J}{\sim} \varepsilon$ then $\delta f \underset{J}{\sim} \varepsilon f$.*

Proof. Suppose that $\delta \underset{J}{\sim} \varepsilon$. Let $i \in J$. Then, since J is ancestral, $\delta_j = \varepsilon_j$ for all $j \geqslant i$. Thus $\delta_i = \varepsilon_i$ and $\delta\pi^i = \varepsilon\pi^i$. It follows that $(\delta f)_i = (\varepsilon f)_i$. Hence $\delta f \underset{J}{\sim} \varepsilon f$.

Lemma 2 shows that, if $f \in F$ and J is an ancestral subset of I, then f induces a map $\hat{f}_J : \Delta_J \to \Delta_J$ such that $f\pi_J = \pi_J \hat{f}_J$. It follows directly from the definitions that for J *any* subset of I and $f, g \in F$ we have $f\pi_J = g\pi_J$ if and only if $f\varphi_J = g\varphi_J$. Thus Lemma 2 gives us a way of defining an action of F_J on Δ_J, in the case when J is an ancestral subset, by identifying $f\varphi_J$ with $\hat{f}_J$. It is evident that this definition coincides with that obtained by defining the action of F_J on Δ_J analogously to that of F on Δ. If $J = \{j\}$ then $\Delta_J = \Delta_j$ and $\hat{f}_J = f_j$: there is therefore no ambiguity in writing f_J for $\hat{f}_J$ and f_j for $\hat{f}_{\{j\}}$.

LEMMA 3. *Let J, K be ancestral subsets of I with $K \subseteq J$, and let $f \in F$. Then $f_J \pi_K^J = \pi_K^J f_K$; in particular, $f\pi_J = \pi_J f_J$.*

LEMMA 4. *Let $f, h \in F$. Then $fh = t$, where*

$$t_i = f_i \times f_{A(i)} h_i$$

and the product of the functions f_i and $f_{A(i)}h_i$ from $\Delta_{A(i)}$ to G_i is defined pointwise.

Proof. Let $\delta \in \Delta$. Then

$$\begin{aligned}(\delta fh)_i &= (\delta f)_i(\delta f \pi^i h_i) \\ &= \delta_i(\delta\pi^i f_i)(\delta\pi^i f_{A(i)} h_i) \quad \text{(by Lemma 3)} \\ &= \delta_i(\delta\pi^i t_i).\end{aligned}$$

LEMMA 5. *If, for each $i \in I$, the function $z_i \in F_i$ is defined by $\gamma z_i = 1_i$ for all $\gamma \in \Delta^i$ then $z = (z_i)$ is the identity permutation on Δ.*

LEMMA 6. *Let J and K be ancestral subsets of I with $K \subseteq J$. Then $\varphi_K^J \colon F_J \to F_K$ is a semigroup homomorphism.*

DEFINITION. Let J be an ancestral subset of I and let $f \in F$. Then f is *invertible on J* if f_J has an inverse in F_J.

LEMMA 7. *Let $\mathscr{L}$ be a family of ancestral subsets of I; let $L = \bigcup \mathscr{L}$; let $f \in F$. If f is invertible on J for all J in $\mathscr{L}$ then f is invertible on the ancestral subset L.*

Proof. Since all the projections involved are semigroup homomorphisms, if $J, K \in \mathscr{L}$ and $i \in J \cap K$ then $(f_J)^{-1}\varphi_i^J = (f_K)^{-1}\varphi_i^K$. Hence we may define h in F_L by $h\varphi_i^L = (f_J)^{-1}\varphi_i^J$, using any $J \in \mathscr{L}$ such that $i \in J$. It is straightforward to check that $h = f_L^{-1}$.

LEMMA 8. *Suppose that (I, ρ) satisfies the maximal condition. If $f \in F$ and J is an ancestral subset of I then f is invertible on J.*

Proof. Let $X = \{i \in J : f \text{ is not invertible on } A[i]\}$. If X is not empty, then X contains a maximal element m. Then f is invertible on $A[i]$ for all $i > m$. By Lemma 7, f is invertible on $A(m)$, because $\bigcup_{i>m} A[i] = A(m)$. Define h in $F_{A[m]}$ by

$$h\varphi_i^{A[m]} = f_{A(m)}^{-1}\varphi_i^{A(m)} \quad \text{for all } i \in A(m)$$

and

$$\gamma(h\varphi_m^{A[m]}) = (\gamma f_{A(m)}^{-1} f_m)^{-1} \quad \text{for all } \gamma \in \Delta^m.$$

Then Lemma 4 shows that h is the inverse of $f_{A[m]}$, so f is invertible on $A[m]$. This contradiction shows that X is empty. Since $J = \bigcup_{i \in J} A[i]$, Lemma 7 shows that f is invertible on J.

The following example can be modified to show that, if (I, ρ) is any partially ordered set which contains an infinite ascending chain, then F contains an element which is not invertible on I. Thus the maximal condition in Lemma 8 is necessary.

EXAMPLE 2. Let $\mathbb{N}$ be the natural numbers with the usual ordering. For all $i \in \mathbb{N}$, let Δ_i be an arbitrary fixed 2-element set $\{a, b\}$, and let $G_i = \text{Symm}(\Delta_i)$. Now, for all $i \in \mathbb{N}$, let f_i be the function in F_i which maps every element of Δ^i, with one exception,

to the transposition in G_i: the exception being the element **a** in Δ^i, each of whose co-ordinates is equal to a; this element **a** is mapped to the identity of G_i. Then f is not a bijection on Δ, because its image does not include the element **b** with each co-ordinate equal to b.

Holland [**6**] and Silcock [**11**] avoid this problem by restricting the sets Δ and F; other approaches are worth investigation.

We summarize the results of this section in the following theorem.

THEOREM A. *Let (I, ρ) be a partially ordered set with the maximal condition. Then*

(i) *for all ancestral subsets J of I, (F_J, Δ_J) is a (faithful) permutation group;*

(ii) *if J and K are ancestral subsets of I with $J \supseteq K$ then (φ_K^J, π_K^J) is a permutation homomorphism from (F_J, Δ_J) onto (F_K, Δ_K) with kernel*

$$N_K^J = \{f \in F_J : f_j = z_j \text{ for } j \in K\};$$

(iii) *if J, K and L are ancestral subsets of I and $J \supseteq K \supseteq L$ then*

$$(\varphi_K^J, \pi_K^J)(\varphi_L^K, \pi_L^K) = (\varphi_L^J, \pi_L^J).$$

In particular, (F, Δ) is a permutation group, which we call the generalized wreath product of the permutation groups $(G_i, \Delta_i)_{i \in I}$. More formally, we write $\prod_{(I,\rho)}(G_i, \Delta_i)$ for this generalized wreath product.

We observe that if ρ is the identity relation on I then $\prod_{(I,\rho)}(G_i, \Delta_i)$ is the permutation Cartesian product of the (G_i, Δ_i); if I is finite, this is simply the permutation direct product. Similarly, if I is the disjoint union $I_1 \cup I_2$ and no element of I_1 is comparable with any element of I_2 then $\prod_{(I,\rho)}(G_i, \Delta_i)$ is the permutation direct product of the generalized wreath products $\prod_{(I_1,\rho)}(G_i, \Delta_i)$ and $\prod_{(I_2,\rho)}(G_i, \Delta_i)$.

At the other extreme, if (I, ρ) is the finite chain $1 \leqslant 2 \leqslant \ldots \leqslant n$ then $\prod_{(I,\rho)}(G_i, \Delta_i)$ is the permutation wreath product $(G_1, \Delta_1) \operatorname{wr} (G_2, \Delta_2) \operatorname{wr} \ldots \operatorname{wr} (G_n, \Delta_n)$. More generally, if I is the disjoint union $I_1 \cup I_2$ and, for all $i \in I_1$ and $j \in I_2$, $i \leqslant j$, then $\prod_{(I,\rho)}(G_i, \Delta_i)$ is the permutation wreath product of the generalized wreath products $\prod_{(I_1,\rho)}(G_i, \Delta_i)$ and $\prod_{(I_2,\rho)}(G_i, \Delta_i)$.

Although, as remarked above, we cannot apply our generalized wreath product construction to obtain a group if (I, ρ) is the natural numbers with the usual ordering, we can if we take the opposite ordering $\mathbb{N}^-$. This gives a class of potentially interesting examples of uncountable permutation groups on uncountable sets which are built up by an explicit construction from (possibly) finite permutation groups: for example, for all $i \in \mathbb{N}$ take $|\Delta_i| = 2$ and $G_i = \operatorname{Symm}(\Delta_i) \cong \mathbb{Z}_2$.

Of practical significance to statistics, Example 1 shows that if $|I| \geqslant 4$ then there are partial orders on I which cannot be decomposed into chains and identity relations: thus these generalized wreath products include more than the wreath and direct products and their iterated composites.

4. *Poset block structures and their automorphism groups*

DEFINITION. A *poset block structure* is a pair (Δ, S), where

(i) Δ is the Cartesian product over a partially ordered set (I, ρ) of sets Δ_i $(i \in I)$, with $|\Delta_i| \geqslant 2$,

(ii) S is the following set of equivalence relations on Δ:

$$S = \{ \gamma_J : J \text{ an ancestral subset of } I \}.$$

These structures, and their relationship with association schemes and distributive lattices, are discussed by Speed and Bailey in [12], with the appellation 'distributive lattices of commuting uniform equivalence relations'. Their automorphism groups are described, without proof, by Bailey in [2]; the purpose of this section is to prove Theorem 1 of that paper (our Theorem B).

DEFINITION. An *automorphism* of a poset block structure (Δ, S) is a permutation t of Δ such that, for all $\sigma \in S$,

$$\delta \sigma \varepsilon \quad \text{if and only if} \quad (\delta t)\sigma(\varepsilon t) \quad (\delta, \varepsilon \in \Delta).$$

The following example shows that the weaker condition

$$\delta \sigma \varepsilon \quad \text{implies} \quad (\delta t)\sigma(\varepsilon t)$$

is not sufficient to ensure that the inverse of t also preserves the block structure.

EXAMPLE 3. Let (I, ρ) be the two-element chain $1 \leqslant 2$, let $\Delta_1 = \Delta_2 = \mathbb{N}$, and define t as follows:

$$(n,0)t = (2n,0), \quad (n,1)t = (2n+1,0),$$
$$(n,m)t = (n,m-1) \quad \text{for } m \geqslant 2.$$

THEOREM B. *Let (Δ, S) be a poset block structure with poset (I, ρ). Let F be the generalized wreath product $\prod_{(I,\rho)} \mathrm{Symm}(\Delta_i)$. If (I, ρ) satisfies the maximal condition then F is the group of automorphisms of (Δ, S).*

Proof. Let $f \in F$ and let J be an ancestral subset of I. Lemma 2 shows that

$$\delta \gamma_J \varepsilon \quad \text{implies} \quad \delta f \gamma_J \varepsilon f.$$

But f^{-1} is also in F, and application of Lemma 2 to f^{-1} shows that

$$\delta \gamma_J \varepsilon \quad \text{is implied by} \quad \delta f \gamma_J \varepsilon f.$$

Hence f is an automorphism of (Δ, S).

Now let t be an automorphism of (Δ, S). We need to prove that there are functions $t_i \in F_i$ such that $t = (t_i) \in F$.

Fix $i \in I$, and put $J = A[i]$, $K = A(i)$, so that $\Delta_K = \Delta^i$. Because t is an automorphism, there exist permutations t_J and t_K of Δ_J, Δ_K respectively such that $t\pi_J = \pi_J t_J$ and $t\pi_K = \pi_K t_K$. Also, as in Lemma 3, $\pi_K^J t_K = t_J \pi_K^J$.

Identifying Δ_J with $\Delta_i \times \Delta_K$, for $\beta \in \Delta^i$ define $\beta t_i \colon \Delta_i \to \Delta_i$ by

$$\alpha(\beta t_i) = (\alpha, \beta) t_J \pi_i^J \quad \text{for } \alpha \in \Delta_i.$$

We shall show that

(i) for all β in Δ^i, the function βt_i is a permutation of Δ_i, which shows that $t_i \in F_i$;

(ii) for all $\delta \in \Delta$, $\delta t \pi_i = \delta_i (\delta \pi^i t_i)$.

Since these results hold for all $i \in I$, we can complete the proof as follows: by (i), $(t_i) \in F$, and by (ii), $t = (t_i)$.

Proof of (i). We need a preliminary equality. For all $\alpha \in \Delta_i$ and all $\beta \in \Delta_K$,

$$(\alpha, \beta)t_J = ((\alpha, \beta)t_J\pi_i^J, (\alpha, \beta)t_J\pi_K^J) = (\alpha(\beta t_i), (\alpha, \beta)\pi_K^J t_K) = (\alpha(\beta t_i), \beta t_K).$$

To prove that βt_i is injective, suppose that $\alpha, \alpha' \in \Delta_i$ and $\alpha(\beta t_i) = \alpha'(\beta t_i)$. By the above equality, $(\alpha, \beta)t_J = (\alpha', \beta)t_J$. Since t_J is injective, $\alpha = \alpha'$.

To prove that βt_i is surjective, we observe that, because t_J is surjective, for each $\alpha \in \Delta_i$ there are $\alpha' \in \Delta_i$ and $\beta' \in \Delta_K$ such that $(\alpha', \beta')t_J = (\alpha, \beta t_K)$; then $(\alpha'(\beta' t_i), \beta' t_K) = (\alpha, \beta t_K)$. But t_K is injective, so $\beta = \beta'$ and $\alpha'(\beta t_i) = \alpha$.

Proof of (ii).

$$\delta t \pi_i = \delta t \pi_J \pi_i^J = \delta \pi_J t_J \pi_i^J = (\delta \pi_i, \delta \pi_K) t_J \pi_i^J.$$

By the definition of t_i this is $(\delta \pi_i)(\delta \pi_K t_i)$, which is $\delta_i(\delta \pi^i t_i)$.

5. *Orbits on* $\Delta \times \Delta$

Throughout this section we assume that (I, ρ) satisfies the maximal condition, so that, as we established in Theorem A, (F, Δ) is a permutation group. We are interested only in the case when F is transitive on Δ, so we first state the following lemma.

LEMMA 9. *The generalized wreath product of the permutation groups* $(G_i, \Delta_i)_{i \in I}$ *is transitive if and only if* (G_i, Δ_i) *is transitive for all* $i \in I$.

When F *is* transitive on Δ, we are concerned with the orbits of F on $\Delta \times \Delta$. For subsets Γ_i of $\Delta_i \times \Delta_i$, we denote by $\bigotimes_{i \in I} \Gamma_i$ the subset of $\Delta \times \Delta$ which contains (δ, ε) if and only if $(\delta_i, \varepsilon_i) \in \Gamma_i$ for all $i \in I$. For each $i \in I$ we denote by D_i the diagonal subset $\{(\alpha, \alpha): \alpha \in \Delta_i\}$ of $\Delta_i \times \Delta_i$, and put $E_i = \Delta_i \times \Delta_i$.

DEFINITIONS. Let J be an ancestral subset of I. The *border* of J, denoted $B(J)$, is the set of maximal elements of $I \setminus J$. The subset O_J of $\Delta \times \Delta$ is defined by

$$O_J = (\bigotimes_{i \in J} D_i) \otimes (\bigotimes_{i \in B(J)} (E_i \setminus D_i)) \otimes (\bigotimes_{i \in I \setminus J \setminus B(J)} E_i).$$

LEMMA 10. (i) *If* J *is an ancestral subset of* I *and* $(\delta, \varepsilon) \in O_J$ *then* $\delta \sim_J \varepsilon$ *and* J *is the maximal ancestral subset with this property.*

(ii) *The set* $\Delta \times \Delta$ *is the disjoint union of the subsets* O_J, *taken over all ancestral subsets* J *of* I.

We shall refer to the subsets O_J as *association sets.*

In work on the design of experiments, Δ is taken to be a set of random variables and a covariance model on Δ is specified in terms of the equivalence relations $(\underset{A[i]}{\sim})_{i \in I}$. Many authors (see, for example, John [7]), assume a model in which the value of the covariance of δ and ε depends only on the values of i for which $\delta \underset{A[i]}{\sim} \varepsilon$; that is, only on the association set O_J containing (δ, ε). Thus the decomposition of $\Delta \times \Delta$ into the association sets is a useful one to study. Other authors (see, for example, Nelder [9] and Bailey [1]) assume a model in which the covariance of δ and ε depends only on the orbit of F on $\Delta \times \Delta$ containing (δ, ε), where $F = \prod_{(I, \rho)}(G_i, \Delta_i)$ and the groups G_i are specified transitive subgroups of the Symm(Δ_i). Still other authors (see, for example, Yates [17] and Preece, Pearce, and Kerr [10]), say that *randomization* based

on the group F is valid if and only if the orbits of F on $\Delta \times \Delta$ coincide with the association sets.

By Lemma 2, it is evident that each association set is a union of orbits of F on $\Delta \times \Delta$. Grundy and Healy [5] showed that, for the direct product $(G_1, \Delta_1) \times (G_2, \Delta_2)$, 2-transitivity of each (G_i, Δ_i) ensures that each association set is a single orbit. Bailey [2] proved a similar result for generalized wreath products. Here we prove the following stronger result.

THEOREM C. *The orbits of* $\prod_{(I,\rho)}(G_i, \Delta_i)$ *on* $\Delta \times \Delta$ *are precisely the association sets if and only if, for each* $i \in I$, *the permutation group* (G_i, Δ_i) *is 2-transitive.*

We shall prove Theorem C as a corollary to Theorem D. First we illustrate Theorem C by an example, and comment on its significance.

EXAMPLE 1. When F is transitive on Δ, as it is in this example, the orbits of F on $\Delta \times \Delta$ are more conveniently displayed as the orbits on Δ of the stabilizer F_δ in F of a fixed element $\delta \in \Delta$. In this case the ancestral subsets of I are $\varnothing$, $\{1\}$, $\{2\}$, $\{1,2\}$, $\{1,3\}$, $\{1,2,3\}$, $\{1,2,4\}$, and $\{1,2,3,4\}$. If each of the four groups (G_i, Δ_i) is 2-transitive, the orbits of F_δ on Δ are as shown in Table 1.

TABLE 1

Orbit	Corresponding ancestral set
$\{\delta\}$	$\{1,2,3,4\}$
$\{\varepsilon$: δ and ε are in the same micro-column, and $\delta \neq \varepsilon\}$	$\{1,2,4\}$
$\{\varepsilon$: δ and ε are in the same column and minirow, and $\delta \neq \varepsilon\}$	$\{1,2,3\}$
$\{\varepsilon$: δ and ε are in the same minirow but different columns$\}$	$\{1,3\}$
$\{\varepsilon$: δ and ε are in the same square but different minirows and different microcolumns$\}$	$\{1,2\}$
$\{\varepsilon$: δ and ε are in the same row but different minirows and different columns$\}$	$\{1\}$
$\{\varepsilon$: δ and ε are in the same column but different rows$\}$	$\{2\}$
$\{\varepsilon$: δ and ε are in different rows and different columns$\}$	$\varnothing$

The *significance* of Theorem C is that, by Theorem B, the orbits of the automorphism group of a poset block structure are precisely the combinatorially defined association sets. Nelder [9] showed that this is true for those poset block structures in which the partial order ρ is successively built up from chains and identity relations, and asked which other structures have this property. We have here a wider class of structures with this property, and hence a partial answer to the question. There are still other block structures, not based on posets, for which association schemes can be combinatorially defined (see Speed and Bailey [12]). The association sets of some of these structures are identical to the orbits of their automorphism groups, but only under fairly severe extra conditions (see Bailey [2, 3]). Thus a complete answer to the question does not yet seem to be known.

THEOREM D. *Suppose that, for* $i \in I$, (G_i, Δ_i) *is transitive. For* $i \in I$, *let* $M_{ij(i)}$ *for* $j(i) \in \Lambda_i$ *be the non-diagonal orbits of* G_i *on* $\Delta_i \times \Delta_i$. *Then, if* S *is any antichain in* I, *and,*

for $i \in S, j(i)$ is any element of Λ_i, the following subset M is an orbit of F on $\Delta \times \Delta$:

$$M = (\bigotimes_{i \in I \setminus H[S]} D_i) \otimes (\bigotimes_{i \in S} M_{ij(i)}) \otimes (\bigotimes_{i \in H(S)} E_i).$$

Moreover, each orbit of F on $\Delta \times \Delta$ has a unique representation of this form.

Proof. Since $|\Delta_i| \geqslant 2$, none of the D_i or $M_{ij(i)}$ is E_i, nor is Λ_i ever empty; therefore distinct antichains S and distinct families $(j(i))_{i \in S}$ give distinct subsets of $\Delta \times \Delta$.

Now let $\delta, \varepsilon \in \Delta$. Then $(\delta, \varepsilon) \in M$ if and only if both $\delta_i = \varepsilon_i$ for all $i \in I \setminus H[S]$ and $(\delta_i, \varepsilon_i) \in M_{ij(i)}$ for all $i \in S$. Let $(\delta, \varepsilon) \in M$. If $i \in I \setminus H(S)$ then $A(i) \subseteq I \setminus H[S]$, so $\delta \pi^i = \varepsilon \pi^i$. Thus if $f \in F$ then $\delta \pi^i f_i$ and $\varepsilon \pi^i f_i$ are the *same* elements of G_i for $i \in I \setminus H(S)$, and hence $(\delta f)_i = (\varepsilon f)_i$ for $i \in I \setminus H[S]$ and $((\delta f)_i, (\varepsilon f)_i) \in M_{ij(i)}$ for $i \in S$. Therefore $(\delta f, \varepsilon f) \in M$.

Conversely, suppose that $\alpha, \beta, \gamma, \varepsilon \in \Delta$, that $(\alpha, \beta) \in M$ and $(\gamma, \varepsilon) \in M$. We must show that there is an element $f \in F$ such that $\alpha f = \gamma$ and $\beta f = \varepsilon$. We deal with coordinates in $I \setminus H(S)$ and $H(S)$ separately.

(i) If $i \in I \setminus H(S)$ then there is an element $g_i \in G_i$ such that $\alpha_i g_i = \gamma_i$ and $\beta_i g_i = \varepsilon_i$. Let $f_i \colon \Delta^i \to G_i$ be the constant function with image g_i.

(ii) If $i \in H(S)$ then there is an element $k \in S$ with $i < k$. Since $(\alpha_k, \beta_k) \in M_{kj(k)}$, we have $\alpha_k \neq \beta_k$. By transitivity, there are elements g_i and h_i in G_i such that $\alpha_i g_i = \gamma_i$ and $\beta_i h_i = \varepsilon_i$. Define $s_i \colon \Delta_k \to G_i$ by $\alpha_k s_i = g_i$ and $\omega s_i = h_i$ for all ω in $\Delta_k \setminus \alpha_k$. Let $f_i \colon \Delta^i \to G_i$ be the function $\pi_k^{A(i)} s_i$. Then $\alpha_i(\alpha \pi^i f_i) = \alpha_i g_i = \gamma_i$ and $\beta_i(\beta \pi^i f_i) = \beta_i h_i = \varepsilon_i$.

Now the element $f = (f_i)$ we have constructed maps α to γ and β to ε. Thus M is indeed an orbit of F on $\Delta \times \Delta$.

All that remains to show is that every orbit of F on $\Delta \times \Delta$ arises in this way. Let $\delta, \varepsilon \in \Delta$ and let J be the unique ancestral subset of I such that $(\delta, \varepsilon) \in O_J$. Now, $B(J)$ is an antichain. Moreover, $H(B(J)) = (I \setminus J) \setminus B(J)$, and so $I \setminus H[B(J)] = J$. By definition of O_J, for each i in $B(J)$ there is a unique $j(i)$ in Λ_i such that $(\delta_i, \varepsilon_i) \in M_{ij(i)}$. Thus, if M is defined as above for $S = B(J)$ and these $j(i)$, then $(\delta, \varepsilon) \in M$.

The proof of Theorem D gives the following, alternative, description of the orbits of F on $\Delta \times \Delta$. Each such orbit is specified by

(i) an ancestral subset J of I,

(ii) for each maximal element i of $I \setminus J$, a non-diagonal orbit M_i of G_i on $\Delta_i \times \Delta_i$.

The pair (α, β) is in the corresponding orbit if and only if

(i) $\alpha_i = \beta_i$ for all $i \in J$,

(ii) $(\alpha_i, \beta_i) \in M_i$ for each maximal element i of $I \setminus J$.

Proof of Theorem C. If (G_i, Δ_i) is 2-transitive then the only $M_{ij(i)}$ which occurs is $E_i \setminus D_i$: hence each antichain S gives just one orbit M_S. Let $J = I \setminus H[S]$. Then J is an ancestral subset of I, and $B(J) = S$; the last part of the proof of Theorem D shows that $M_S = O_J$.

To prove the converse, for $k \in I$ let $J = I \setminus H[k]$; then O_J is the union of $|\Lambda_k|$ orbits. If (G_k, Δ_k) is not 2-transitive then $|\Lambda_k| \geqslant 2$.

6. *Characters*

In this section we assume that I is finite and that, for each $i \in I$, (G_i, Δ_i) is finite and transitive. We investigate the permutation characters of the action of F on Δ and on Δ_J, for ancestral subsets J of I. Since we are using the letter π for projections, we

denote the permutation character of F on Δ_J (that is, on equivalence classes of $\sim_J$) by ψ_J, and put $\psi = \psi_I$. Once again, the case in which every (G_i, Δ_i) is 2-transitive is particularly interesting and straightforward.

THEOREM E. *If I is finite and, for each $i \in I$, (G_i, Δ_i) is finite and 2-transitive then $\prod_{(I,\rho)}(G_i, \Delta_i)$ has distinct irreducible characters $\{\chi_J: J \subseteq I, J \text{ ancestral}\}$ such that, if K is any ancestral subset of I, the permutation character ψ_K of $\prod_{(I,\rho)}(G_i, \Delta_i)$ on the equivalence classes of $\sim_K$ is $\sum_{J \subseteq K} \chi_J$: in particular, ψ is the sum $\sum \chi_J$ over all ancestral subsets J. Moreover, if, for each $i \in I$, $|\Delta_i| = n_i$, the degree of χ_J is $\prod_{i \in m(J)}(n_i - 1)\prod_{i \in J \setminus m(J)} n_i$, where $m(J)$ is the set of minimal elements of J.*

We shall prove Theorem E as a corollary to Theorem F. Here we simply note that, since the ψ_J are very easy to compute in practice, so are the (irreducible) χ_J.

The permutation linear representations of (G_i, Δ_i) and (F, Δ) are afforded canonically by the vector spaces $\mathbb{R}^{\Delta_i}$ and $\mathbb{R}^{\Delta}$ respectively. Let $W_i = \mathbb{R}^{\Delta_i}$. We shall identify $\mathbb{R}^{\Delta}$ with $\bigotimes_{i \in I} W_i$ by regarding the tensor product of the functions w_i, with $i \in I$, to be the function which maps $\delta \in \Delta$ to $\prod_{i \in I} \delta_i w_i$. We shall also use the natural inner product on $\mathbb{R}^{\Delta}$ given by $v \cdot w = \sum_{\delta \in \Delta}(\delta v)(\delta w)$.

THEOREM F. *Let $F = \prod_{(I,\rho)}(G_i, \Delta_i)$. Suppose that I is finite and, for $i \in I$, (G_i, Δ_i) is finite and transitive. Let C_i be the subspace of constant functions in W_i, and let $V_{ij(i)}$ for $j(i) \in \Theta_i$ denote the other components of a direct decomposition of W_i into G_i-irreducible subspaces. Let S be an antichain in I, and, for each $i \in S$, let $j(i)$ be an element of Θ_i. Then*

(i) *the following subspace V of $\mathbb{R}^{\Delta}$ is F-irreducible:*

$$V = \Big(\bigotimes_{i \in A(S)} W_i\Big) \otimes \Big(\bigotimes_{i \in S} V_{ij(i)}\Big) \otimes \Big(\bigotimes_{i \in I \setminus A[S]} C_i\Big).$$

(ii) *Moreover, $\mathbb{R}^{\Delta}$ is the direct sum of such subspaces.*

Let

$$W = \Big(\bigotimes_{i \in A(T)} W_i\Big) \otimes \Big(\bigotimes_{i \in T} V_{ik(i)}\Big) \otimes \Big(\bigotimes_{i \in I \setminus A[T]} C_i\Big)$$

for some antichain T and some family $(k(i))_{i \in T}$ such that, for each $i \in T$, $k(i) \in \Theta_i$. Then

(iii) *$V = W$ if and only if $S = T$ and $j(i) = k(i)$ for each $i \in S$.*

(iv) *The subspaces V and W afford equivalent representations of F if and only if $S = T$ and, for each $i \in S$, the spaces $V_{ij(i)}$ and $V_{ik(i)}$ afford equivalent representations of G_i.*

Proof. (iii) Since, for each $i \in I$, $|\Delta_i| \geqslant 2$, none of the subspaces W_i or $V_{ij(i)}$ is equal to C_i and Θ_i is non-empty; therefore $V \cap W$ is the zero subspace unless $S = T$ and, for each $i \in S$, $j(i) = k(i)$.

(i) (a) Now we show that the subspace V given above is F-invariant. For each $\alpha \in \Delta_{A(S)}$ let $w_\alpha: \Delta_{A(S)} \to \mathbb{R}$ be the characteristic function of $\{\alpha\}$. Then V is spanned by the set of all functions $v: \Delta \to \mathbb{R}$ of the following form:

$$v = (\pi_{A(S)} w_\alpha) \prod_{i \in S} \pi_i v_i,$$

where $\alpha \in \Delta_{A(S)}$, and, for $i \in S$, $v_i \in V_{ij(i)}$; the product is pointwise. It is, therefore, sufficient to show that, for each $f \in F$ and each function v of this form, the function

$fv\colon \Delta \to \mathbb{R}$ is in V. Put $\beta = \alpha f_{A(S)}^{-1}$, and let $\delta \in \Delta$. Since $\pi_{A(S)} f_{A(S)} = f\pi_{A(S)}$, we have $\delta f \pi_{A(S)} = \alpha$ if and only if $\delta \pi_{A(S)} = \beta$.

Thus if $\delta \pi_{A(S)} \neq \beta$ then $\delta f v = 0$. On the other hand, if $\delta \pi_{A(S)} = \beta$ then, for each $i \in S$,

$$\delta f \pi_i v_i = \delta_i(\delta \pi^i f_i) v_i = \delta_i(\delta \pi_{A(S)} \pi_{A(i)}^{A(S)} f_i) v_i = \delta_i(\beta \pi_{A(i)}^{A(S)} f_i) v_i = \delta_i g_i v_i = \delta \pi_i g_i v_i,$$

where $g_i = \beta \pi_{A(i)}^{A(S)} f_i \in G_i$ and so $g_i v_i \in V_{ij(i)}$ because $V_{ij(i)}$ is G_i-invariant.

Thus $fv = (\pi_{A(S)} w_\beta) \prod_{i \in S} \pi_i g_i v_i$, which is in V, and so V is F-invariant.

(ii) To see that $\mathbb{R}^\Delta$ is the direct sum of such subspaces, for each $i \in I$ we let H_i be the orthogonal complement of C_i in W_i. Now for each antichain S let

$$Y_S = \Big(\bigotimes_{i \in A(S)} W_i\Big) \otimes \Big(\bigotimes_{i \in S} H_i\Big) \otimes \Big(\bigotimes_{i \in I \setminus A[S]} C_i\Big).$$

Then, since F acts orthogonally on $\mathbb{R}^\Delta$, Y_S is the direct sum

$$\bigoplus_{j \in \Sigma_S} \Big(\Big(\bigotimes_{i \in A(S)} W_i\Big) \otimes \Big(\bigotimes_{i \in S} V_{ij(i)}\Big) \otimes \Big(\bigotimes_{i \in I \setminus A[S]} C_i\Big)\Big), \tag{6.1}$$

where Σ_S is the set of families $(j(i))_{i \in S}$ such that, for all $i \in S$, $j(i) \in \Theta_i$. Moreover, if X_S is the subspace generated by $\{Y_T\colon T \text{ is an antichain, } T \subseteq A[S], \text{ and } T \neq S\}$, then $Y_S \cap X_S$ is the zero subspace. Thus the Y_S generate their direct sum. Now, $\mathbb{R}^\Delta$ is spanned by vectors of the form $w = \bigotimes_{i \in I} w_i$, where, for $i \in I$, either $w_i \in H_i$ or $w_i \in C_i$. Let $S(w) = m(\{i \in I\colon w_i \in H_i\})$. Then $S(w)$ is an antichain and $w \in Y_{S(w)}$. Thus

$$\mathbb{R}^\Delta = \bigoplus_{\substack{\text{antichains} \\ S}} Y_S.$$

(i)(b) and (iv). For each $i \in I$, denote by m_i the sum of the squares of the multiplicities of the distinct inequivalent G_i-irreducible components of H_i. Let m and m_S be the corresponding sums of squares for the F-irreducible components of $\mathbb{R}^\Delta$ and Y_S respectively. Then Proposition 29.2 of Wielandt [**16**] shows that $m_i = |\Lambda_i|$, where Λ_i is as defined in § 5; then, with Theorem D, it shows that

$$m = \sum_{\substack{\text{antichains} \\ S}} \prod_{i \in S} m_i. \tag{6.2}$$

Since $\mathbb{R}^\Delta$ is the direct sum of the F-subspaces Y_S,

$$m \geqslant \sum_{\substack{\text{antichains} \\ S}} m_S, \tag{6.3}$$

with equality if and only if, for $S \neq T$, no F-component of Y_S is equivalent to any F-component of Y_T.

Let $j, k \in \Sigma_S$. If, for all $i \in S$, the spaces $V_{ij(i)}$ and $V_{ik(i)}$ afford equivalent representations of G_i, then the direct summands in (6.1) corresponding to j and k afford equivalent representations of F. The sum of the squares of the numbers of these F-equivalent direct summands is calculated to be $\prod_{i \in S} m_i$. Since (6.1) gives Y_S, we have

$$m_S \geqslant \prod_{i \in S} m_i, \tag{6.4}$$

with equality if and only if *both* each V is F-irreducible *and* there are no more F-equivalences among the direct summands of (6.1) than those just described.

Now Equation 6.2 forces equality in Equations 6.3 and 6.4, and the result follows.

Proof of Theorem E. If (G_i, Δ_i) is 2-transitive then the orthogonal complement H_i of C_i in W_i is G_i-irreducible. Thus, for each antichain S, the set Σ_S contains just one element j, and so, by Theorem F, the subspace Y_S defined above is F-irreducible.

For each antichain S, the subset $A[S]$ is ancestral and $m(A[S]) = S$. For each ancestral subset J, the subset $m(J)$ is an antichain and $A[m(J)] = J$. Thus the maps $J \mapsto m(J)$ and $S \mapsto A[S]$ are mutually inverse bijections. For each ancestral subset J, let χ_J be the irreducible character of F afforded by $Y_{m(J)}$.

If K is any ancestral subset of I, the permutation character ψ_K of F is afforded by the subspace

$$V_K = \mathbb{R}^{\Delta_K} \otimes \bigotimes_{i \notin K} C_i.$$

Applying Theorem F to the partially ordered set K, and then tensoring the result with $\bigotimes_{i \notin K} C_i$, gives

$$V_K = \bigoplus_{\substack{\text{antichains} \\ S \subseteq K}} Y_S = \bigoplus_{\substack{\text{ancestral} \\ J \subseteq K}} Y_{m(J)},$$

and so $\psi_K = \sum_{\text{ancestral } J \subseteq K} \chi_J$.

From the point of view of the statistician, a permutation group (G, Γ) is useful only if the centralizer algebra $\mathscr{A}_G$ of G in $\mathbb{R}^\Gamma$ is commutative, or possibly if only the subset $\mathscr{S}_G$ of symmetric matrices in this centralizer algebra is commutative (and so forms a subalgebra) (see McLaren [**8**], and Speed, Bailey, Praeger, and Taylor [**13**]). Denote the permutation character of (G, Γ) by ψ_G.

LEMMA 11. (i) *$\mathscr{A}_G$ is commutative if and only if ψ_G is multiplicity-free.*

(ii) *$\mathscr{S}_G$ is commutative if and only if all irreducible quaternionic characters in ψ_G have multiplicity 2 and all other irreducible characters in ψ_G have multiplicity 1.*

Proof. Part (i) is well-known (see Wielandt [**16**, Theorem 29.3]), whilst (ii) is a slight modification of (i), and is proved by McLaren [**8**] and Speed et al. [**13**].

COROLLARY TO THEOREM F. *Under the hypotheses of Theorem F:*

(i) *$\mathscr{A}_F$ is commutative if and only if $\mathscr{A}_{G_i}$ is commutative for all $i \in I$;*

(ii) *$\mathscr{S}_F$ is commutative if and only if $\mathscr{S}_{G_i}$ is commutative for all $i \in I$ and there is no two-element antichain $\{i, j\}$ such that ψ_{G_i} includes quaternionic characters and ψ_{G_j} includes either non-real or quaternionic characters.*

For example, if I is the two-element antichain $\{1, 2\}$ and, for $i = 1, 2$, (G_i, Δ_i) is the regular representation of the quaternion group Q_8, then $\mathscr{S}_{G_1}$ and $\mathscr{S}_{G_2}$ are both commutative but $\mathscr{S}_F$ is not.

References

1. R. A. BAILEY, 'A unified approach to design of experiments', *J. Roy. Statist. Soc. Ser. A*, 144 (1981), 214–223.
2. R. A. BAILEY, 'Distributive block structures and their automorphisms', *Combinatorial mathematics VIII* (ed. K. L. McAvaney), Lecture Notes in Mathematics 884 (Springer, Berlin, 1981), pp. 115–124.
3. R. A. BAILEY, 'Latin squares with highly transitive automorphism groups', *J. Austral. Math. Soc. Ser. A*, 32 (1982), 18–22.
4. G. GRÄTZER, *Lattice theory. First concepts and distributive lattices* (W. H. Freeman, San Francisco, 1971).

5. P. M. GRUNDY and M. J. R. HEALY, 'Restricted randomization and quasi-Latin squares', *J. Roy. Statist. Soc. Ser. B*, 12 (1950), 286–291.
6. W. C. HOLLAND, 'The characterization of generalized wreath products', *J. Algebra*, 13 (1969), 152–172.
7. P. W. M. JOHN, *Statistical design and analysis of experiments* (Macmillan, New York, 1971).
8. A. D. MCLAREN, 'On group representations and invariant stochastic processes', *Proc. Cambridge Philos. Soc.*, 59 (1963), 431–450.
9. J. A. NELDER, 'The analysis of randomized experiments with orthogonal block structure', *Proc. Roy. Soc. London Ser. A*, 283 (1965), 147–178.
10. D. A. PREECE, S. C. PEARCE, and J. R. KERR, 'Orthogonal designs for three-dimensional experiments', *Biometrika*, 60 (1973), 349–358.
11. H. L. SILCOCK, 'Generalized wreath products and the lattice of normal subgroups of a group', *Algebra Universalis*, 7 (1977), 361–372.
12. T. P. SPEED and R. A. BAILEY, 'On a class of association schemes derived from lattices of equivalence relations', *Algebraic structures and applications* (eds. P. Schultz, C. E. Praeger, and R. P. Sullivan, Marcel Dekker, New York, 1981), pp. 45–54.
13. T. P. SPEED, R. A. BAILEY, C. E. PRAEGER, and D. E. TAYLOR, *The analysis of variance*, to appear.
14. T. N. THROCKMORTON, 'Structures of classification data', Ph.D. thesis, Iowa State University, 1961.
15. C. WELLS, 'Some applications of the wreath product construction', *Amer. Math. Monthly*, 83 (1976), 317–338.
16. H. WIELANDT, *Finite permutation groups* (Academic Press, New York, 1964).
17. F. YATES, 'The formation of Latin squares for use in field experiments', *Empire J. Exp. Agric.* (1933), 235–244.

R. A. Bailey
Statistics Department
Rothamsted Experimental Station
Harpenden
Herts. AL5 2JQ

C. A. Rowley
Mathematics Faculty
The Open University
Milton Keynes
MK7 6AA

C. E. Praeger *and* T. P. Speed
Mathematics Department
University of Western Australia
Nedlands 6009
Western Australia

The Annals of Statistics,
1983, Vol. 11, No. 4, 1069–1085

BALANCE IN DESIGNED EXPERIMENTS WITH ORTHOGONAL BLOCK STRUCTURE

BY A. M. HOUTMAN AND T. P. SPEED

A. C. Nielsen and CSIRO

The notion of general balance due to Nelder is discussed in relation to the eigenvectors of an information matrix, combinatorial balance and the simple combinability of information from uncorrelated sources in an experiment.

1. Introduction. This paper is about the notion of general balance (GB) introduced by Nelder (1965) in two papers on designed experiments with orthogonal block structure. Nelder defined (GB) as a relationship between the block structure or dispersion model for the data and the treatment structure or model for the expected value of the data. It embodies and unifies three important and apparently unrelated ideas concerning designed experiments: the usefulness of eigenvectors of the associated information matrices, the combinatorial and statistical notions of balance, and the simple combinability of information from different, uncorrelated, sources in the experiment. These ideas have been discussed independently by a number of authors including Yates (1936, 1939, 1940), Sprott (1956), Morley Jones (1959), Pearce (1963), Martin and Zyskind (1966), Corsten (1976) and many others. We will review the work of these authors in Section 3 and relate it to Nelder's (1965) work.

Nelder (1965, 1968) has shown how a simple and unified approach may be adopted to the analysis of multistratum designed experiments satisfying (GB), including the estimation of stratum variances and the combination of information across strata. We summarise these facts in Section 4 and also prove a useful supplementary result: that (GB) is not only a sufficient but also a necessary condition (assuming known stratum variances) for the simple recovery of all information on every contrast from every stratum in which it is estimable. Our definition of (GB) is slightly different from Nelder's in that we accommodate unequal treatment replications, but it has all the same consequences, and the broad scope of the notion so defined is underlined by the fact that all block designs with equal block size are then generally balanced (assuming the standard dispersion model). It will be seen from our examples and the associated discussion that essentially all designs with orthogonal block structure which have ever been recommended for use satisfy (GB). It also provides a convenient basis for the classification of designs, one which is connected with the simple and directly interpretable analysis.

Section 5 below is devoted to examples, beginning with the balanced incomplete block design (BIBD) which is the prototype of all designs satisfying (GB). Instead of going on to prove directly that partially balanced incomplete block designs (PBIBDs) all satisfy (GB), we obtain the same conclusion for their natural generalisations to more general block structures. Following a brief discussion of some further examples, we close the paper with a row-column design *not* satisfying (GB).

2. Basic framework.

2.1. *Treatment structure.* Our data will be viewed as a random array $y = (y_i)_{i \in \mathbf{I}}$ indexed by a set $\mathbf{I}$ of $n = |\mathbf{I}|$ *unit labels* and taking values in the vector space $\mathscr{Y} = \mathbb{R}^{\mathbf{I}}$

Received May 1982, revised February 1983.

AMS 1980 *subject classifications.* Primary 62K05B.

Key words and phrases. Balance, designed experiment, recovery of information, simple combinability.

which has the inner product $\langle c \mid d \rangle = \sum_i c_i d_i$ and squared norm $\| c \|^2 = \langle c \mid c \rangle$. The models we consider for $\tau = \mathbb{E}y$, termed the *treatment structure*, will all be *linear*, i.e. of the form

$$\mathbb{E}y \in \mathscr{T} \tag{2.1}$$

where $\mathscr{T} \subseteq \mathscr{D}$ is a linear subspace of $\mathscr{D}$. In the theory of designed experiments this usually arises as follows: we have a set $\mathscr{X}$ of $v = |\mathscr{X}|$ *treatment labels*, a *design map* $x: \mathbf{I} \to \mathscr{X}$ which assigns a treatment to each unit, and a *design matrix* X satisfying $X(i, u) = 1$ if $x(i) = u$, $i \in I$, $u \in \mathscr{X}$, and $= 0$ otherwise. In this case $\mathscr{T} = \mathscr{R}(X)$, the range of X, and $\tau = X\alpha$ for some $\alpha \in \mathbb{R}^{\mathscr{X}}$. However none of the general discussion which follows assumes that $\mathscr{T}$ arises in this way. The (unweighted) orthogonal projection of $\mathscr{D}$ onto $\mathscr{T}$ will be denoted by T; if $\mathscr{T} = \mathscr{R}(X)$ then $T = X(X'X)^{-1}X'$.

A vector $c = (c_i) \in \mathscr{D}$ of constants satisfying $\sum c_i = 0$ is said to define (or be) a *contrast*; if $c \in \mathscr{T}$, then c defines (or is) a *treatment contrast*. This usage arises because least-squares estimation concentrates on the estimation of linear functions $\langle t \mid \tau \rangle$ of $\tau = \mathbb{E}y$ ($t \in \mathscr{T}$) based upon linear functions $\langle c \mid y \rangle$ of the data. Thus the term contrast refers in each case to the coefficients of these linear functions. In many analyses interest focuses on treatment contrasts $\langle t \mid \tau \rangle$ defined by elements t of specific *subspaces* of $\mathscr{T}$; for examples, we refer to Section 5 below. When $\mathscr{T} = \mathscr{R}(X)$ we say that *simple* treatment contrasts are those elements $t_{u,v} \in \mathscr{T}$ for which $\langle t_{u,v} \mid \tau \rangle$ is proportional to $\alpha_u - \alpha_v$, $u, v \in \mathscr{X}$, where $X\alpha = \tau$.

2.2. *Block structure.* Following Nelder (1965) we use the term *block structure* to mean the model for the dispersion matrix $V = \mathbb{D}y$, and all our models for V will have the form

$$\mathbb{D}y \in \mathscr{V} \tag{2.2}$$

where $\mathscr{V}$ is a suitably parameterized set of positive semi-definite (p.s.d.) matrices. We will say that we have *orthogonal block structure* (OBS) when $\mathscr{V}$ consists of all p.s.d. matrices $V(\xi) = \sum_\alpha \xi_\alpha S_\alpha$, where $\xi_\alpha \geq 0$ for all α, and the $\{S_\alpha\}$ are a family of known pairwise orthogonal projectors summing to the identity matrix, i.e. $S_\alpha = S_\alpha' = S_\alpha^2$, $S_\alpha S_\beta = S_\beta S_\alpha = 0$ if $\alpha \neq \beta$, and $\sum_\alpha S_\alpha = I$, the identity matrix. We call this representation of $V(\xi)$ its spectral form. In the theory of designed experiments such models usually arise in the following way: there is a system $\{A_a\}$ of *association matrices* defined over the set $\mathbf{I}$ of unit labels, and the dispersion matrix $V = \mathbb{D}y$ has the form $V = \sum_a \gamma_a A_a$ where $\{\gamma_a\}$ is a set of *covariances* varying freely subject only to the constraints ensuring that V is p.s.d. If the matrices $\{A_a\}$ satisfy the requirements of an *association scheme* then there always exist matrices $P = (p_{\alpha a})$ and $Q = (q_{a\alpha})$ of coefficients such that $S_\alpha = (1/n) \sum_a q_{a\alpha} A_a$ satisfies the properties listed above, and $\xi_\alpha = \sum_\alpha p_{\alpha a} \gamma_a$ constitutes an invertible linear reparametrization; see MacWilliams and Sloane (1978, Chapter 21, especially Section 2) for definitions and the results cited. Once more we remark that the general results which follow do not assume that our orthogonal block structure arose in this way although in practice the vast majority (block, row-column, split-plot designs etc.) do so. For example, any model $\mathscr{V}$ whose elements have the form $V = \sum_j \theta_j C_j$, where the $\{C_j\}$ are known symmetric idempotent matrices which *commute*, will be a submodel of a model of the form (OBS) above as the $\{C_j\}$ are simultaneously diagonalizable, but in general there will be more ξs than θs.

Summarising, we will be supposing that our data y is modeled by (2.1) and (2.2) where $\mathscr{T}$ is a linear subspace of $\mathscr{D}$ and $\mathscr{V}$ satisfies (OBS). The subspaces $\mathscr{S}_\alpha = \mathscr{R}(S_\alpha)$ are termed the *strata* of the dispersion model, the $\{S_\alpha\}$ are *strata projectors* and the $\{\xi_\alpha\}$ the *strata variances* (for it is easy to see that $\mathbb{D}S_\alpha y = \xi_\alpha S_\alpha$). *Multi-strata designs* are those with two or more strata variances in the dispersion model.

2.3. *Examples.*

EXAMPLE 1. The data y from an experiment consisting of v treatments applied across b blocks of k plots each are usually analysed under the *mixed model*

$$y = X\alpha + Z\gamma + \varepsilon, \tag{2.3}$$

where X and Z are the $n \times v$ and $n \times b$ treatment and block incidence matrices, respectively, α is a $v \times 1$ vector of treatment parameters, and γ is a $b \times 1$ vector of zero-mean block effects having dispersion matrix $\sigma_b^2 \mathrm{I}_b$ uncorrelated with the $n \times 1$ vector ε of errors which have dispersion matrix $\sigma^2 I_n$.

The dispersion matrix associated with such a model is $V = \sigma_b^2 ZZ' + \sigma^2 I_n$, and its spectral form is

$$V = \xi_0 G + \xi_1 (B - G) + \xi_2 (I_n - B) \tag{2.4}$$

where $G = n^{-1}11'$ is the grand mean averaging operator (1 is the $n \times 1$ vector of ones), $B = k^{-1}ZZ'$ is the block averaging operator, $\xi_0 = \xi_1 = k\sigma_b^2 + \sigma^2$ and $\xi_2 = \sigma^2$. Note that here we have the constraint $\xi_0 = \xi_1 \geq \xi_2 > 0$.

A *randomisation model* for y, see Nelder (1954), would generate a dispersion matrix of the form (2.4).

In order to include both types of model, we will assume when analysing data from block designs with equal block size (which are the only sort we consider) that $S_0 = G$, $S_1 = B - G$ and $S_2 = I - B$ defines our block structure satisfying (QBS). It will be simpler, and necessary for most results, to assume $\xi_0 > 0$, $\xi_1 > 0$ and $\xi_2 > 0$ as well. □

EXAMPLE 2. The data y from an experiment in which v treatments are allocated to the $n = rc$ plots of a row-column design consisting of r rows and c columns are usually analysed under the *mixed model*

$$y = X\alpha + Z_1\gamma_1 + Z_2\gamma_2 + \varepsilon \tag{2.5}$$

where X, Z_1 and Z_2 are the treatment, row and column incidence matrices, respectively, and γ_1, γ_2 and ε are uncorrelated zero-mean vectors having dispersion matrices $\sigma_r^2 I_r$, $\sigma_c^2 I_c$ and $\sigma^2 I_n$, respectively.

This time the dispersion matrix of y is $V = \sigma_r^2 Z_1 Z_1' + \sigma_c^2 Z_2 Z_2' + \sigma^2 I_n$ and its spectral form is

$$V = \xi_0 G + \xi_1 (R - G) + \xi_2 (C - G) + \xi_3 (1 - R - C + G) \tag{2.6}$$

where $G = (rc)^{-1}11'$, $R = c^{-1}Z_1 Z_1'$ and $C = r^{-1}Z_2 Z_2'$, $\xi_0 = c\sigma_r^2 + r\sigma_c^2 + \sigma^2$, $\xi_1 = c\sigma_r^2 + \sigma^2$, $\xi_2 = r\sigma_c^2 + \sigma^2$ and $\xi_3 = \sigma^2$. Again we have constraints: $\xi_1 \geq \xi_3 > 0$, $\xi_2 \geq \xi_3 > 0$ and $\xi_0 = \xi_1 + \xi_2 - \xi_3$.

A *randomisation model* for y would also generate a dispersion matrix of the form (2.6). Accordingly we will analyse row-column designs below with $S_0 = G$, $S_1 = R - G$, $S_2 = C - G$ and $S_3 = I - R - C + G$, a block structure satisfying (QBS). Again we will usually assume that $\xi_0 > 0$, $\xi_1 > 0$, $\xi_2 > 0$ and $\xi_3 > 0$. □

2.4. *Designed experiments.* The *design* of an experiment, i.e. the actual allocation of treatments to units, affects the least-squares analysis (under our model) of the data generated through the relationships it determines between the treatment subspace $\mathscr{T}$ and the strata subspaces $\{\mathscr{S}_\alpha\}$. For example, it is known that if T *commutes* with all the $\{S_\alpha\}$, then the analysis is easy; such designs are known as *orthogonal* designs, a class which includes completely randomised, randomised block, latin square and split-plot designs. For other designs, such as the balanced incomplete block designs (BIBDs), this commutativity fails, and a more elaborate analysis is required. Nelder's (1965) notion of general balance (GB) describes a relationship between T and the $\{S_\alpha\}$ which generalises, but in a sense is no more difficult than, that which arises with a BIBD, and as a consequence we find that essentially all designed experiments may be analysed in a manner almost identical to that of a BIBD. Note that the $\{C_i\}$ of Nelder (1965) correspond to our $\{S_\alpha\}$. Before giving any further details of these ideas, we devote the next section to reviewing the

antecedents of general balance and clarifying its connections with similar notions which have appeared since 1965. See also Bailey (1981) for a related discussion.

3. Eigenvectors, balance and simple combinability.

3.1. *Eigenvectors of information matrices.* It has long been known in linear regression analysis that contrasts which are eigenvectors of the information matrix have special properties which make inference concerning them particularly straightforward; the analogy with principal components analysis explains why this is so. However it appears that Morley Jones (1959) was the first person to examine these ideas in some detail in the context of block experiments, and because of their relevance to general balance we will summarise his results within the framework introduced in Example 1 of the previous section.

Morley Jones analysed the data y under the "fixed block effects" model: $\mathbb{E}y \in \mathscr{T} + \mathscr{B}$, $\mathbb{D}y \in \mathscr{V}$ where $\mathscr{B} = \mathscr{R}(B)$ and $\mathscr{V} = \{\sigma^2 I : \sigma^2 > 0\}$, and he concentrated upon the *intra-block analysis*, i.e. that using the reduced data $\bar{B}y$ ($\bar{B} = I - B$) consisting of the observations adjusted by their block means. Clearly $\mathbb{E}\bar{B}y \in \bar{B}\mathscr{T}$ and $\mathbb{D}\bar{B}y \in \bar{B}\mathscr{V}\bar{B}$, and the task of minimising $\| \bar{B}y - \bar{B}\tau \|^2$ over $\tau \in \mathscr{T}$ is equivalent to solving the reduced normal equations ("eliminating blocks"):

$$T\bar{B}T\tau = T\bar{B}y$$

for $\tau \in \mathscr{T}$. In this context the eigenvectors and eigenvalues of the information matrix $T\bar{B}T$ are likely to be of interest. (In fact Morley Jones studied a closely-related matrix with the same eigenvectors but eigenvalues one minus those of $T\bar{B}T$.) He made the following observations: (a) an element $t \in \mathscr{T}$ is an eigenvector of $T\bar{B}T$ iff there exists a constant k such that for all $u \in \mathscr{T}$, $\langle u \mid (B - G)t \rangle = k\langle u \mid \bar{B}t \rangle$; (b) if one of two orthogonal treatment contrasts t and u is an eigenvector of $T\bar{B}T$, then their inter-block components Bt, Bu (resp. intra-block components $\bar{B}t$, $\bar{B}u$) are also orthogonal; (c) the best linear unbiased estimators (BLUEs) of contrasts $\langle t \mid \tau \rangle$ defined by eigenvectors of $T\bar{B}T$ are easy to compute, as are their precisions, and these are related to the corresponding eigenvalue; (d) the eigenvalues of $T\bar{B}T$ are directly related to the Fisher *efficiency factors* describing the relative loss of information occurring by restricting attention only to the intrablock analysis; and (e) normalised contrasts defined by eigenvectors of $T\bar{B}T$ corresponding to the same eigenvalue are estimated with the same precision; in particular, all contrasts are estimated with the same precision in BIBDs.

Although not explicitly referring to eigenvector contrasts, similar ideas can be found in Kurkjian and Zelen (1963). Their "property A" is equivalent to the spectral decomposition $T\bar{B}T = \sum_\beta \lambda_\beta T_\beta$ where the $\{T_\beta\}$ are the orthogonal projections decomposing $\mathscr{T}$ into subspaces $\{\mathscr{T}_\beta\}$ corresponding to main effects and interactions in a factorial experiment laid out in blocks. Their conclusions included (c) above, with the BLUE of $\langle t_\beta \mid \tau \rangle$ based upon $\bar{B}y$ being $\lambda_\beta^{-1}\langle t_\beta \mid \bar{B}y \rangle$ for an arbitrary $t_\beta \in \mathscr{T}_\beta$, having variance $\sigma^2\lambda_\beta^{-1} \| t_\beta \|^2$, and they observed that BLUEs of contrasts defined by elements of the different subspaces $\{\mathscr{T}_\beta\}$ are uncorrelated (cf. (b) above). They also applied their results to other types of incomplete block designs including group divisible and direct product designs. A further paper, Zelen and Federer (1964) extended the same ideas to row-column designs, but still only in the context of the lowest stratum analysis, i.e. that based upon $(I - R - C + G)y$; cf. Example 2 above.

In Pearce, Caliński and Marshall (1974) the eigenvectors of $T\bar{B}T$ are called "basic contrasts", and these authors note that those with eigenvalue 1 can be estimated with full efficiency in the intra-block analysis, those with eigenvalue 0 are "totally confounded" with blocks, whilst the remainder are "partially confounded". They recommend that the spectral decomposition of $T\bar{B}T$ be used by experimenters to ensure that the design permits contrasts of particular interest to be estimated with maximum efficiency in the intra-block analysis.

Corsten's (1976) *canonical analysis* is also equivalent to the spectral analysis of $T\bar{B}T$. He calls the eigenvectors (with non-zero eigenvalues) "identifiable contrasts" and views the corresponding eigenvalues as the squared cosines of the *canonical angles* between the subspaces $\mathscr{T}$ and $\mathscr{B}^{\perp}$ the orthogonal complement of $\mathscr{B}$; the same geometric approach is used by James and Wilkinson (1971).

3.2. *Balance.* BIBDs were introduced by Yates (1935) as incomplete block designs with equal block sizes, equal replications, and having the combinatorial property that every pair of distinct treatments appeared together in a block the same number of times. It followed that simple treatment contrasts were all estimated with the same precision, and as a consequence, that normalised treatment contrasts were also estimated with the same precision. Thus combinatorial balance was related to the property of sets of contrasts being estimated with the same precison.

Generalised forms of these ideas appeared soon afterwards: PBIBDs were introduced by Bose and Nair (1939); designs with unequally replicated treatments having a restricted form of balance were studied by Nair and Rao (1942); designs with supplemented balance by Hoblyn, Pearce and Freeman (1954), and Pearce (1960, 1963). Morley Jones (1959) continued this line of development.

Balance in block designs was first linked to the spectral properties of the intra-block information matrix (or a closely related matrix) by V. R. Rao (1958) and Morley Jones (1959). The latter proved that a block design is balanced with respect to a set of treatment contrasts iff those contrasts span a subspace of an eigenspace of $T\bar{B}T$. The combinatorial aspects of balance are reviewed in Raghavarao (1971), although we will see that the approach through general balance is more relevant to the problem of analysing data from an experiment with a design exhibiting the given type of balance.

3.3. *Simple combinability.* The term *recovery of interblock information* has come to mean the double task of estimating the relevant strata variances and the calculation of weighted combinations of the inter- and intra-block estimates (where this is appropriate) of a given treatment contrast. Following earlier work with cubic lattice designs, Yates (1939), Yates (1940) showed that the overall (weighted least squares) BLUE of any treatment contrast in a BIBD was the linear combination of its BLUE calculated using the intra-block data $(I - B)y$ and that calculated using the inter-block data $(B - G)y$, each weighted inversely according to its variance. We shall call this result, which assumes that the strata variances are known, the property of *simple combinability*, which is valid for *all* contrasts in a BIBD. Yates also gave a method of estimating the usually unknown strata variances from the anova table.

Conditions on a design which ensure the simple combinability in PBIBDs of certain sets of treatment contrasts were described by Sprott (1956) in a paper which gave great insight into the relation between combinability and combinatorial balance. In particular Sprott showed that the property of simple combinability holds for all contrasts in a PBIBD only if the design is actually a BIBD. This and other results along the same lines are special cases of a general theorem proved in the next section.

A link between the spectral properties of $T\bar{B}T$ and simple combinability in an incomplete block design was established by Zyskind and Martin (1966), who showed that a treatment contrast is simply combinable iff it is an eigenvector of $T\bar{B}T$. Thus these three topics: the eigenspaces of $T\bar{B}T$, balance, in either the combinatorial sense or in the statistical sense of contrasts being estimable with the same precision, and simple combinability are all seen to be intimately related. With this introduction to general balance we now turn to its definition and study.

4. General balance. As we have explained in Section 2 above, our model for the data $y = (y_i)_{i \in \mathbf{I}}$ associated with our designed experiment is given by (2.1) $\mathbb{E}y \in \mathscr{T}$ and (2.2) $\mathbb{D}y \in \mathscr{V}$, where $\mathscr{T} \subseteq \mathscr{D}$ is a linear subspace and $\mathscr{V} = \{V(\xi) : V(\xi) = \sum_{\alpha} \xi_{\alpha} S_{\alpha},$

$\xi_\alpha > 0$ for all $\alpha\}$ is a dispersion model satisfying (OBS). General balance is a structural property relating $\mathscr{T}$ and the strata $\{\mathscr{S}_\alpha\}$.

4.1. *Definition of* (GB). We say that a design with (OBS) defined by $\{S_\alpha\}$ and treatment structure $\mathscr{T}$ is *generally balanced* with respect to the decomposition $\mathscr{T} = \oplus_\beta \mathscr{T}_\beta$ or just generally balanced if there exists a matrix $(\lambda_{\alpha\beta})$ of numbers such that for all α

$$TS_\alpha T = \sum_\beta \lambda_{\alpha\beta} T_\beta, \tag{GB}$$

where the $\{T_\beta\}$ are the orthogonal projectors onto the subspaces $\{\mathscr{T}_\beta\}$. It is clear that (GB) is equivalent to the requirement that the matrices $\{TS_\alpha T\}$ are simultaneously diagonalisible, with the $\{\mathscr{T}_\beta\}$ as their common eigenspaces. Another equivalent form is the following: there exists numbers $(\lambda_{\alpha\beta})$ such that for all α, β and β'

$$T_\beta S_\alpha T_{\beta'} = \begin{cases} \lambda_{\alpha\beta} T_\beta & \text{if } \beta = \beta', \\ 0 & \text{otherwise.} \end{cases}$$

Since the $\{S_\alpha\}$ and $\{T_\beta\}$ are all projectors, we must have $0 \le \lambda_{\alpha\beta} \le 1$ for all α and β, and it follows from $\sum_\alpha S_\alpha = I$ that for all β, $\sum_\alpha \lambda_{\alpha\beta} = 1$. A statistical interpretation of the $\lambda_{\alpha\beta}$ as *efficiency* factors will be explained in Section 4.3 below, and we refer to Fisher (1935) for the first use of such a two-way array. *Orthogonal designs* are just those for which each $\lambda_{\alpha\beta}$ is 0 or 1.

4.2. *Overall analysis assuming* (GB)*: known strata variances.* It is well known that the BLUE of $\tau = \mathbb{E}y$ based on y is given by the solution $\tau \in \mathscr{T}$ of the normal equation

$$TV^{-1}T\tau = TV^{-1}y; \tag{NE}$$

equivalently, that it is given by $\hat{\tau} = Uy$ where $U = P^V_{\mathscr{T}}$ is projection of $\mathscr{D}$ onto $\mathscr{T}$ orthogonal with respect to the weighted inner product $\langle c \mid d\rangle_V := \langle c \mid V^{-1}d\rangle$. Yet one further statement of this (Gauss's) result is the following: $\langle t \mid \hat{\tau}\rangle$ is the unique BLUE of $\langle t \mid \tau\rangle$ for every $t \in \mathscr{T}$.

Now $TV^{-1}T = \sum_\beta \nu_\beta T_\beta$ under (GB), where we write $\nu_\beta = \sum_\alpha \lambda_{\alpha\beta}\xi_\alpha^{-1}$, and so the unique matrix inverse of $TV^{-1}T$ on the subspace $\mathscr{T}$ is $\sum_\beta \nu_\beta^{-1} T_\beta$. Consequently the solution $\hat{\tau} = Uy$ of (NE) is given by

$$U = \sum_{\alpha,\beta} w_{\alpha\beta}\lambda_{\alpha\beta}^{-1} T_\beta S_\alpha \tag{4.1}$$

where we have written $w_{\alpha\beta} = \nu_\beta^{-1}\xi_\alpha^{-1}\lambda_{\alpha\beta}$. This expression is called the *weight* for the treatment term β within stratum α, a name which we will shortly justify. Here and later all summations involving $\lambda_{\alpha\beta}^{-1}$ will be restricted only to those α or β for which $\lambda_{\alpha\beta} > 0$.

As we have already observed, the unique BLUE of $\langle t \mid \tau\rangle$ for $t \in \mathscr{T}$ is $\langle t \mid \hat{\tau}\rangle$ and by (4.1) this is just

$$\langle t \mid \hat{\tau}\rangle = \sum_{\alpha,\beta} w_{\alpha\beta}\lambda_{\alpha\beta}^{-1}\langle t \mid T_\beta S_\alpha y\rangle \tag{4.2}$$

with *variance* $\sum_\beta \nu_\beta^{-1} \| T_\beta t \|^2$. If $t = t_\beta \in \mathscr{T}_\beta$, the BLUE simplifies to

$$\langle t_\beta \mid \hat{\tau}\rangle = \sum_\alpha w_{\alpha\beta}\lambda_{\alpha\beta}^{-1}\langle t_\beta \mid S_\alpha y\rangle \tag{4.3}$$

with variance $\nu_\beta^{-1} \| t_\beta \|^2$.

Finally, the *covariance* between two BLUEs $\langle t_1 \mid \hat{\tau}\rangle$ and $\langle t_2 \mid \hat{\tau}\rangle$ is just

$$\sum_\beta \nu_\beta^{-1}\langle T_\beta t_1 \mid T_\beta t_2\rangle,$$

and if $t_1 \in \mathscr{T}_\beta$, $t_2 \in \mathscr{T}_{\beta'}$, $\beta \neq \beta'$, this reduces to zero.

It is clear from the above that as long as the strata variances are known (up to a common scalar multiplier) and we can readily effect the projections $\{S_\alpha\}$ and $\{T_\beta\}$, the

weighted least squares analysis of data from a designed experiment with generally balanced block structure is particularly simple. We will deal with the problem of unknown strata variances in the next subsection and in Section 4.5 below. On the issue of the ease of calculation and computation of the projections we can say this: the $\{S_\alpha\}$ are commonly built up from simple averaging operators such as G and B in Example 1 or R, C and G in Example 2 above, and rarely give any difficulties. The common decompositions $\{T_\beta\}$ relative to which designed experiments satisfy (GB) are also of this form, although there are some that are quite different, and in general the problem is not: "how do we compute the projections $\{T_\beta\}$?" but: "how do we discover them?" This is essentially a combinatorial problem, which needs to be done for each new design or class of designs. The usual mathematical skills (trial and error, ingenuity, etc.) help, as does the occasional computer-aided spectral analysis, and it is only the broader classes of block designs for which general solutions are unavailable; see Section 5.4.

4.3. *Within strata analysis assuming* (GB). A reduction of the full data y to its strata projections $S_\alpha y$ permits analyses *within strata* without knowledge of the strata variances, for $\mathbb{E}S_\alpha y \in S_\alpha \mathscr{T}$ and $\mathbb{D}S_\alpha y = \xi_\alpha S_\alpha$; in particular, the dispersion matrix of $S_\alpha y$ is known up to a scalar, and this is adequate for the usual least-squares analyses.

The least-squares *fitted value* $\hat{y}_\alpha$ of y in stratum α is $\hat{y}_\alpha = P_{S_\alpha \mathscr{T}} y$, the unweighted projection of y onto $S_\alpha \mathscr{T}$, unweighted because the subspace $S_\alpha \mathscr{T}$ is invariant under $\mathbb{D}S_\alpha y$ whence unweighted and weighted projectors coincide. The normal equation within $\mathscr{S}_\alpha$ is

$$(\mathrm{NE}_\alpha) \qquad TS_\alpha T\tau = TS_\alpha y$$

and its solution $\hat{\tau}_\alpha = U_\alpha y$ is given by (cf. Nelder (1965) equation 3.3)

$$(4.4) \qquad U_\alpha y = \textstyle\sum_\beta \lambda_{\alpha\beta}^{-1} T_\beta S_\alpha y$$

where the sum is only over those β for which $\lambda_{\alpha\beta} > 0$. We can readily prove that $P_{S_\alpha \mathscr{T}} = S_\alpha U_\alpha$. It follows from (4.4) that the unique BLUE of a contrast $\langle t \mid \tau \rangle$ which is estimable in $\mathscr{S}_\alpha$ (i.e. for which there exists a BLUE based on $S_\alpha y$) is

$$(4.5) \qquad \langle t \mid \hat{\tau}_\alpha \rangle = \textstyle\sum_\beta \lambda_{\alpha\beta}^{-1} \langle T_\beta t \mid S_\alpha y \rangle$$

with variance $\xi_\alpha \sum_\beta \lambda_{\alpha\beta}^{-1} \| T_\beta t \|^2$. If $t = t_\beta \in \mathscr{T}_\beta$ the BLUE simplifies to

$$(4.6) \qquad \langle t_\beta \mid \hat{\tau}_\alpha \rangle = \lambda_{\alpha\beta}^{-1} \langle t_\beta \mid S_\alpha y \rangle \quad \text{(provided } \lambda_{\alpha\beta} > 0\text{)}$$

with variance $\lambda_{\alpha\beta}^{-1} \xi_\alpha \| t_\beta \|^2$, and if $\lambda_{\alpha\beta} = 0$ then *no* contrast $\langle t_\beta \mid \tau \rangle$ is estimable in $\mathscr{S}_\alpha$. Finally, we remark that the covariance between two BLUEs $\langle t_1 \mid \hat{\tau}_\alpha \rangle$ and $\langle t_2 \mid \hat{\tau}_\alpha \rangle$ is $\xi_\alpha \sum_\beta \lambda_{\alpha\beta}^{-1} \langle T_\beta t_1 \mid T_\beta t_2 \rangle$ and if $t_1 \in \mathscr{T}_\beta$, $t_2 \in \mathscr{T}_{\beta'}$, $\beta \neq \beta'$, this again reduces to zero.

There are a number of points in the formulae above and in the corresponding ones in the previous sub-section which are worth noting. First, it is clear from both (4.6) and (4.3) that estimation *is* especially simple for contrasts which are eigenvectors of *all* the information matrices $TS_\alpha T$, cf. Section 3.1 point (c). Secondly, BLUEs of contrasts from distinct (common) eigenspaces of the $TS_\alpha T$ *are* orthogonal, cf. Section 3.1 point (b), and so the BLUEs of contrasts $\langle t \mid \tau \rangle$ for arbitrary $t \in \mathscr{T}$ are sums of the uncorrelated BLUEs of $\langle T_\beta t \mid \tau \rangle$ which have the simple form. And finally, the overall BLUE (4.3) of $\langle t_\beta \mid \tau \rangle$ for $t_\beta \in \mathscr{T}_\beta$ is quite clearly the simple combination of its BLUEs (4.6) in each stratum in which it is estimable, each weighted inversely according to its variance: $\langle t_\beta \mid \hat{\tau} \rangle = \sum_\alpha w_{\alpha\beta} \langle t_\beta \mid \hat{\tau}_\alpha \rangle$. This justifies our use of the term *weight* for $w_{\alpha\beta}$ introduced following equation (4.1). Similarly we can compare the variance of $\langle t_\beta \mid \hat{\tau}_\alpha \rangle$ to that of $\langle t_\beta \mid \hat{\tau} \rangle$ when the ξ_α are assumed equal, and see why $\lambda_{\alpha\beta}$ is termed the *efficiency factor* for treatment term β in stratum α, cf. point (d) in Section 3.1.

In a sense there is no single analysis of variance table which summarises all aspects of the least-squares analysis of a designed experiment satisfying (GB), but rather one for each stratum and one overall. See Table 1, the anova table within stratum α. Examples of

TABLE 1
Anova table within stratum α

Source	d.f.	*Sum of squares*	$\mathbb{E}\{$*Mean square*$\}$
⋮	⋮	⋮	⋮
Treatment term $\mathscr{T}_\beta$ (assuming $\lambda_{\alpha\beta} > 0$)	$\dim \mathscr{T}_\beta$	$\lambda_{\alpha\beta}^{-1} \| T_\beta S_\alpha y \|^2$	$\xi_\alpha + \frac{\lambda_{\alpha\beta}}{\dim \mathscr{T}_\beta} \| T_\beta \tau \|^2$
⋮	⋮	⋮	⋮
Residual	d_α:By difference	By difference	ξ_α (if $d_\alpha > 0$)
	↑ may be zero	↑	
Total	$\dim \mathscr{S}_\alpha$	$\| S_\alpha y \|^2$	

designs with residual degrees of freedom $d_\alpha = 0$ in some strata are quite common, e.g. symmetric BIBDs, double, triple, $\cdots$ lattice designs, rectangular lattice designs all have zero residual d.f. in the inter-block stratum, and the best general way to estimate ξ_α is certainly not via the anova table for stratum α. For further comments on the estimation of ξ_α, see Section 4.5 below.

4.4. *Simple combinability: a converse to* (GB). We now prove a result asserting that under certain general circumstances, if a set of contrasts spanning $\mathscr{T}$ is simply combinable, then the design satisfies (GB). The following lemma has its straightforward proof omitted. Our framework is that of Section 2.4 without assuming (GB).

LEMMA. *If the treatment contrast $\langle t \mid \tau \rangle$ is estimable in stratum α, then there exists a unique $c_\alpha = c_\alpha(t) \in \mathscr{R}(S_\alpha T)$ such that $Tc_\alpha = t$. Furthermore, the unique* BLUE *of $\langle t \mid \tau \rangle$ based on $S_\alpha y$ is then $\langle c_\alpha \mid y \rangle$.* □

PROPOSITION 4.1. *Let $\langle t \mid \tau \rangle$ be a treatment contrast such that for each stratum $\mathscr{S}_\alpha$ it is either estimable in or orthogonal to $\mathscr{S}_\alpha$, and suppose that there is a set $\{w_\alpha\}$ of non-negative weights summing to unity such that*

$$\langle t \mid \hat{\tau} \rangle = \textstyle\sum_\alpha w_\alpha \langle c_\alpha \mid y \rangle, \quad (y \in \mathscr{D}) \tag{4.7}$$

where $\langle c_\alpha \mid y \rangle$ is the BLUE *of $\langle t \mid \tau \rangle$ based on $S_\alpha y$, if $\langle t \mid \tau \rangle$ is estimable in $\mathscr{S}_\alpha$, and $w_\alpha = 0$ if t is orthogonal to $\mathscr{S}_\alpha$. Then for all α, t is an eigenvector of $TS_\alpha T$ with eigenvalue $\lambda_\alpha = \xi_\alpha w_\alpha (\sum_\alpha \xi_\alpha w_\alpha)^{-1}$.*

PROOF. It is not hard to prove that the transpose U' of $U = P^V_{\mathscr{T}}$ coincides with $V^{-1}UV$. It follows from equation (4.7) that $V^{-1}UVt = \sum_\alpha w_\alpha c_\alpha$ and so

$$UVt = (\textstyle\sum_\alpha \xi_\alpha S_\alpha)(\sum_\alpha w_\alpha c_\alpha) = \sum_\alpha \xi_\alpha w_\alpha c_\alpha. \tag{4.8}$$

Now $TU = U$ and since $Tc_\alpha = t$ for all α, (4.8) implies

$$UVt = (\textstyle\sum_\alpha \xi_\alpha w_\alpha) t. \tag{4.9}$$

On the other hand, (4.8) also implies that $S_\alpha UVt = \xi_\alpha w_\alpha c_\alpha$, and so

$$TS_\alpha UVt = \xi_\alpha w_\alpha t. \tag{4.10}$$

The conclusion now follows from (4.9) and (4.10). □

Now let us suppose that the subspace $\mathscr{T}$ has a basis consisting of vectors t satisfying

the hypotheses of Proposition 4.1. Then for each such t there is a set $\{\lambda_{\alpha t}\}$ eigenvalues, and we can obtain a pairwise orthogonal system $\{\mathscr{T}_\beta\}$ of subspaces of $\mathscr{T}$ by grouping together all ts with a common set of eigenvalues, say $\{\lambda_{\alpha\beta}\}$ for each $t \in \mathscr{T}_\beta$. It is clear that the system $\{\mathscr{T}_\beta\}$ forms a complete set of eigenspaces common to all the matrices $\{TS_\alpha T\}$ and also that $\mathscr{T} = \oplus_\beta \mathscr{T}_\beta$. Thus we can obtain the following converse to (GB) implying equation (4.1).

PROPOSITION 4.2. *If there exists an orthogonal decomposition $\mathscr{T} = \oplus_\beta \mathscr{T}_\beta$ of $\mathscr{T}$ and a set $\{w^*_{\alpha\beta}\}$ of weights such that for all $V \in \mathscr{V}$ the projection U onto $\mathscr{T}$ orthogonal with respect to $\langle \cdot \mid \cdot \rangle_V$ is $U = \sum_{\alpha,\beta} w^*_{\alpha\beta} T_\beta S_\alpha$, where $w^*_{\alpha\beta}\xi_\alpha$ is independent of α, then the design satisfies* (GB) *with respect to* $\{\mathscr{T}_\beta\}$. □

The proof will be omitted; it can be found in Houtman (1980). A stronger result can be obtained when there are only two effective strata, i.e. $\mathscr{V}$ is spanned by $S_0 = \mathrm{G}$, S_1, S_2; for this case the hypothesis "for all $V \in \mathscr{V}$" in Proposition 4.2 is not required, as one suitable V leads to the same conclusion.

4.5. *The estimation of strata variances under* (GB). We remarked in Section 4.3 above that the residual operator $R_\alpha = S_\alpha - P_{S_\alpha \mathscr{T}}$ in stratum α may be zero, equivalently, that $d_\alpha = \mathrm{tr}\, R_\alpha = \dim \mathscr{S}_\alpha - \sum \{\dim \mathscr{T}_\beta : \lambda_{\alpha\beta} > 0\}$ may be zero. The reason for this is not hard to see: if $0 < \lambda_{\alpha\beta} < 1$, then treatment term $T_\beta\tau$ is being fitted and its full d.f. $\dim \mathscr{T}_\beta$ removed not only in stratum α, but also in one or more other strata in which it is estimable. In a sense we should only remove that fraction $w_{\alpha\beta}(\dim \mathscr{T}_\beta)$ of the d.f. corresponding to the amount of information on $\mathscr{T}_\beta$ in $\mathscr{S}_\alpha$ and the approach of Nelder (1968) amounts to just this.

More precisely, Nelder's approach is based upon equating the observed with expected mean square of the *actual* residual $S_\alpha(I - U)y = S_\alpha \bar{U} y$ in stratum α rather than doing so with the *apparent* residual $R_\alpha y$ as is done if only the anova table is consulted. To illustrate the difference between the two we cite the following without proof:

LEMMA (i) $\| S_\alpha \bar{U} y \|^2 = \| R_\alpha y \|^2 + \| (P_{S_\alpha \mathscr{T}} - S_\alpha U) y \|^2$.

(ii) $d'_\alpha = \mathrm{tr}(S_\alpha \bar{U}) = d_\alpha + \sum_\beta (1 - w_{\alpha\beta}) \dim \mathscr{T}_\beta$.

(iii) *When every treatment term is estimated in one of two strata, α and α' say, then*

$$\| (P_{S_\alpha \mathscr{T}} - S_\alpha U) y \|^2 = \sum_\beta w_{\alpha'\beta} \lambda_{\alpha\beta} \| \Delta_\beta y \|^2$$

where $\Delta_\beta y = \lambda^{-1}_{\alpha\beta} T_\beta S_\alpha y - \lambda^{-1}_{\alpha'\beta} T_\beta S_{\alpha'} y$ is the difference between the estimates of treatment term β in the two strata, and a similar equation holds with the roles of α and α' reversed. □

Now both U and d'_α involve the weights $\{w_{\alpha\beta}\}$ so if we are to make use of the identity $\mathbb{E} \| S_\alpha \bar{U} y \|^2 = d'_\alpha \xi_\alpha$ in estimating ξ_α, an *iterative* approach must be used. We proceed as follows:

(0) Begin with initial estimates $\{\xi^{(0)}_\alpha\}$ or $\{w^{(0)}_{\alpha\beta}\}$ of the strata variances or weights, possibly making use of the strata anova tables;

(1) Given a set $\{\xi_\alpha\}$ and $\{w_{\alpha\beta}\}$ of *working* estimates of the strata variances and weights, calculate U and d'_α and obtain *revised* estimates $\{\xi^*_\alpha\}$ by solving for $\{\xi_\alpha\}$ in

$$\| S_\alpha \bar{U} y \|^2 = \xi_\alpha d'_\alpha, \quad \alpha = 0, 1, \cdots. \tag{4.11}$$

It is interesting to note that equation (4.11) is in fact the likelihood equation for $\{\xi_\alpha\}$ based upon $\| (I - T) y \|^2$ under the assumption that y has a multivariate normal distribution, see Patterson and Thompson (1971) for details. The information matrix corre-

sponding to these *restricted* ML *estimates* $\{\hat{\xi}_\alpha\}$ under normality has elements

$$-2\mathbb{E}\left\{\frac{\partial^2 \log l}{\partial \xi_\alpha \xi_{\alpha'}}\right\} = \frac{1}{\xi_\alpha \xi_{\alpha'}} \times \begin{cases} [d_\alpha + \sum_\beta (1 - w_{\alpha\beta})^2 (\dim\ \mathscr{T}_\beta)] & \text{if } \alpha = \alpha' \\ [\sum_\beta w_{\alpha\beta} w_{\alpha'\beta} (\dim\ \mathscr{T}_\beta)] & \text{if } \alpha \neq \alpha' \end{cases}$$

where the sums are over all β for which $\lambda_{\alpha\beta}$ (or $\lambda_{\alpha'\beta}$) > 0.

4.6. *Inferential difficulties under* (GB). Even when a designed experiment with orthogonal block structure defined by the strata $\{\mathscr{S}_\alpha\}$ and treatment structure $\{\mathscr{T}_\beta\}$ satisfies (GB), there remain difficulties with estimation and testing the model.

Although the formula (4.1) gives a precise expression for $\hat{\tau}$ when the strata variances $\{\xi_\alpha\}$ are *known*, these considerations no longer apply when we use the estimates $\{\hat{\xi}_\alpha\}$ obtained as in Section 4.5. The general problem of combining information on a common mean when the weights require estimation has a large literature; see Brown and Cohen (1974) for a general discussion and further references. In some of these papers the problem of combining information on treatment contrasts in BIBDs is considered and it would be of interest to extend these conclusions to multi-strata designs with a number of treatment terms.

A second difficulty arises when the analyst wishes to test the hypothesis $T_\beta \tau = 0$ for some β, say under a normality assumption. This can be done by an F-test in every stratum α for which $\lambda_{\alpha\beta} > 0$ and the stratum residual d.f. $d_\alpha > 0$, and although such tests would be *independent*, there appears to be no accepted procedure for combining the tests into a single one. On the other hand, an overall test might be sought, fitting to $\mathscr{T}$ first and then to the orthogonal complement $\mathscr{T} \ominus \mathscr{T}_\beta$, of $\mathscr{T}_\beta$ in $\mathscr{T}$ which still satisfies (GB). The problem here is the fact that the likelihood ratio test for such hypotheses does not appear to have been studied when information concerning $\mathscr{T}_\beta$ resides in more than one stratum.

Both of these problems would seem to warrant further research. Until straightforward exact or approximate solutions are found, most analysts will follow Yates (1940) and others in substituting the estimated weights into (4.1), and testing hypotheses $T_\beta \tau = 0$ in the stratum α for which $\lambda_{\alpha\beta}$ is largest.

5. Examples.

5.1. BIBD*s*. The basic notation for block designs was introduced in Section 2.3: b blocks of k plots each, and the term *balanced* means that the $v \geq k$ different treatments are applied to the plots in such a way that each pair of distinct treatments appears together in a block the same number of times, λ say. The strata projections are G, $B - G$ and $I - B$, all derived from simple averaging operators, whilst the treatment decomposition $T = G + (T - G)$ is similarly straightforward. We readily find that

$$TGT = G, \quad T(B - G)T = \bar{e}(T - G), \quad T(I - B)T = e(T - G) \tag{5.1}$$

where $e = (1 - k^{-1})/(1 - v^{-1}) = 1 - \bar{e}$ is the *efficiency factor* of the design; Yates (1936). The computation which establishes the (GB) conditions most easily is the checking that $(T - G)B(T - G) = \bar{e}(T - G)$ by applying $(T - G)B$ to a simple contrast $t_{u,v}$; in this form it is nothing more than checking the balance condition.

The overall BLUE of a treatment contrast $\langle t \mid \tau \rangle$ is given by $\langle t \mid \hat{\tau} \rangle = \xi_1^{-1}(\bar{e}\xi_1^{-1} + e\xi_2^{-1})^{-1}\langle t \mid (B - G)y \rangle + \xi_2^{-1}(\bar{e}\xi_1^{-1} + e\xi_2^{-1})^{-1}\langle t \mid (I - B)y \rangle$, the correctly weighted linear combination of the inter- and intra-block BLUEs $\bar{e}^{-1}\langle t \mid (B - G)y \rangle$, and $e^{-1}\langle t \mid (I - B)y \rangle$, respectively.

When we turn to the estimation of ξ_1 and ξ_2, we note that the residual d.f. $d_1 = (b - 1) - (v - 1)$ in the inter-block stratum is usually small and is zero if $v = b$. Nelder's iterative method or its Fisher scoring variant can be used with initial values $\xi_1^{(0)} = \xi_2^{(0)} = d_2^{-1} \| R_2 y \|^2$ on $d_2 = b(k - 1) - (v - 1)$ d.f. from the intra-block stratum. The only quantities needed

for this calculation are the residual arrays

$$R_1 y = (B - G)y - \bar{e}^{-1}(B - G)T(B - G)y$$

$$R_2 y = \bar{B}y - e^{-1}\bar{B}T\bar{B}y$$

and the array of differences of effects estimated in the two strata:

$$\Delta_1 y = \bar{e}^{-1}T(B - G)y - e^{-1}T\bar{B}y.$$

The procedure generally converges quickly, and gives estimates which are close, although not identical, to those given by Yates' (1940) method based on anova tables, and the statistical properties of these estimates appear (by simulations) to be very similar to those of Yates' estimates.

5.2. *A natural generalisation of* PBIBDs. PBIBDs were introduced by Bose and Nair (1939) as generalisations of BIBDs and have been the subject of much study since then, mostly devoted to combinatorial aspects of the designs because the combinatorial objects now known as *association schemes* were first defined in this context, see MacWilliams and Sloane (1978) and Raghavarao (1971). The standard reference on the *analysis* of PBIBDs seems to be Clatworthy (1973). The idea behind PBIBDs is quite simple: where it is not possible for *every* pair of distinct treatment to be together in a block the same number λ of times, the pairs are partitioned into *association classes* forming an association scheme so that this can hold within classes, and the single number λ is replaced by a family λ_1, λ_2, $\cdots$ of numbers, one for each association class. Our generalisation carries this idea over to more general block structures than just blocks and plots such as nested BIBDs; Preece (1967).

Let us suppose that the orthogonal block structure of our design arises from a dispersion model based upon an association scheme $\{A_\alpha\}$ over the set $\mathbf{I}$ of unit labels as described in Section 2.2. That is, the strata projections $\{S_\alpha\}$ are given by $S_\alpha = (1/n)\sum_\alpha q_{\alpha\alpha}A_\alpha$ where $Q = (q_{\alpha\alpha})$ is a matrix of structure constants. The association matrices $\{A_\alpha\}$ are defined in terms of the strata projections by $A_\alpha = \sum_\alpha p_{\alpha\alpha}S_\alpha$ where $P = (p_{\alpha\alpha})$ is the "inverse" matrix of constants: $PQ = QP = nI$.

Similarly we suppose—as is customary with PBIBDs—that there is an association scheme $\{\hat{B}_b\}$ defined over the set $\mathscr{X}$ of treatment labels, see Section 2.2, with corresponding orthogonal projectors $\{\hat{T}_\beta\}$ given by $\hat{T}_\beta = (1/v)\sum_b \tilde{q}_{b\beta}\hat{B}_b$, where $\hat{Q} = (\tilde{q}_{b\beta})$ and $\hat{P} = (\tilde{p}_{\beta b})$ are the appropriate matrices of structure constants.

DEFINITION. A design map $x:\mathbf{I} \to \mathscr{X}$ is said to be $(\{A_a\}, \{\hat{B}_b\})$-*balanced* if for all association classes a over $\mathbf{I}$ and b over $\mathscr{X}$ and $u_1, u_2 \in \mathscr{X}$ with $\hat{B}_b(u_1, u_2) = 1$, the number $|\{(i, j) \in \mathbf{I} \times \mathbf{I} : A_a(i, j) = 1, x(i) = u_1, x(j) = u_2\}|$ depends only on b and not on the pair u_1, u_2 chosen. If we denote the number (of concurrences) in this definition by n_{ab} then, recalling the design matrix X introduced in Section 2.1 above, we see that an equivalent form of the definition is: there exists numbers n_{ab} such that for all a we have

$$X'A_aX = \sum_b n_{ab}B_b. \tag{5.2}$$

In particular if we consider A_e and $\hat{B}_e$ where e represents the identity association, we find that $n_{ee} = r$ defines the common replication number for the treatments of our design.

PROPOSITION 5.1. *An experiment with block structure arising from an association scheme $\{A_a\}$ over the set $\mathbf{I}$ of units, and having a design map which is $(\{A_a\}, \{\hat{B}_b\})$-balanced with respect to an association scheme $\{\hat{B}_b\}$ over the set $\mathscr{X}$ of treatments, satisfies* (GB). *In notation introduced above, the treatment decomposition is given by $\{T_\beta\}$ where $T_\beta =$*

$r^{-1}X\tilde{T}_\beta X'$, and the matrix $\Lambda = (\lambda_{\alpha\beta})$ of efficiency factors is given by

$$\lambda_{\alpha\beta} = (rn)^{-1} \sum_a \sum_b q_{a\alpha} n_{ab} \tilde{p}_{\beta b}$$

where $\mathbf{n} = (n_{ab})$ is the matrix of concurrences.

PROOF. We begin by noting that $T = r^{-1}XX'$. Then for all α

$$\begin{aligned}
TS_\alpha T &= n^{-1} \textstyle\sum_a q_{a\alpha} T A_a T && \text{(definition of } S_\alpha\text{)} \\
&= (r^2 n)^{-1} \textstyle\sum_a q_{a\alpha} X(X' A_a X) X' && \text{(definition of } T\text{)} \\
&= (r^2 n)^{-1} \textstyle\sum_a \sum_b q_{a\alpha} n_{ab} X \hat{B}_b X' && \text{(by (5.2))} \\
&= (r^2 n)^{-1} \textstyle\sum_a \sum_b \sum_\beta q_{a\alpha} n_{ab} \tilde{p}_{\beta b} X \hat{T}_\beta X' && \text{(definition of } \hat{T}_\beta\text{)} \\
&= \textstyle\sum_\beta \{(rn)^{-1} \sum_a \sum_b q_{a\alpha} n_{ab} \tilde{p}_{\beta b}\} T_\beta && \text{(definition of } T_\beta\text{)}
\end{aligned}$$

and the assertion is proved. □

EXAMPLE 1. It is not hard to see that a BIBD is built over an association scheme on its units with associations which can be labeled e (equality), 1 (same block but different unit) and 2 (different block), whilst its treatments have the trivial association scheme with associations e (equality) and 1 (inequality). We readily find that $(rn)^{-1}Q'\mathbf{n}\tilde{P}'$ takes the form

$$(rn)^{-1}\begin{bmatrix} 1 & 1 & 1 \\ b-1 & b-1 & -1 \\ b(k-1) & -b & 0 \end{bmatrix}\begin{bmatrix} r & 0 \\ 0 & \lambda \\ r(r-1) & r^2-\lambda \end{bmatrix}\begin{bmatrix} 1 & 1 \\ v-1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1-e \\ 0 & e \end{bmatrix}$$

making use of the relations $r(k-1) = \lambda(v-1)$ and $rv = bk = n$.

EXAMPLE 2. Kshirsagar (1957) gave the very interesting 6×6 row-column design with 9 treatments A, B, C, D, E, F, G, H, I shown in Table 2. Let us consider the association scheme defined on the treatments by imposing a row-column *pseudo-structure* on them as shown in Table 3. If we let e, 1, 2 and 3 denote the associations of equality, same row (but unequal), same column (but unequal) and different row and column for both schemes, then we have what is shown in Table 4, with a similar result holding for $X'A_3X$ by differencing, since $A_1 + A_2 + A_3 = J - I$, where J is the matrix of all 1s. These clearly satisfy our balance condition with matrix $\mathbf{n} = (n_{ab})$ of concurrences, shown in Table 5. With these preliminaries we can readily get $\tilde{P}$ and Q and calculate the matrix $\Lambda = (\lambda_{\alpha\beta})$ of efficiency factors; this turns out to be as given in Table 6.

For many further such designs see Preece (1968, 1976) and Cheng (1981a, b).

TABLE 2
Treatment allocation to 36 units with a 6×6 row-column block structure

B	D	H	G	F	C
C	E	G	B	D	I
E	F	C	A	G	H
D	I	A	H	C	E
F	G	I	E	A	B
A	H	B	D	I	F

TABLE 3
3×3 row-column pseudostructure on 9 treatments

A	B	C
D	E	F
G	H	I

TABLE 4

$X'A_1X =$

A	B	C	D	E	F	G	H	I	
0	2	2	2	3	3	2	3	3	A
	0	2	3	2	3	3	2	3	B
		0	3	3	2	3	3	2	C
			0	2	2	2	3	3	D
				0	2	3	2	3	E
by symmetry					0	3	3	2	F
						0	2	2	G
							0	2	H
								0	I

$X'A_2X =$

A	B	C	D	E	F	G	H	I	
0	3	3	3	2	2	3	2	2	A
	0	3	2	3	2	2	3	2	B
		0	2	2	3	2	2	3	C
			0	3	3	3	2	2	D
				0	3	2	3	2	E
by symmetry					0	2	2	3	F
						0	3	3	G
							0	3	H
								0	I

TABLE 5

$\mathbf{n} =$

e	1	2	3	
4	0	0	0	e
0	2	2	3	1
0	3	3	2	2
12	11	11	11	3

TABLE 6

Treatment pseudo-factor

$\Lambda =$

gm	r	c	r · c		
1	0	0	0	Grand mean	
0	1/8	1/8	0	Rows	*Block stratum*
0	0	0	1/8	Columns	
0	7/8	7/8	7/8	Rows · Columns	

5.3. *Supplemented balance and related notions.* Pearce (1960) described a class of block designs possessing what he termed *supplemented balance*, and later Pearce (1963) extended the notion to row-column and more general designs. A typical example is a BIBD consisting of b blocks of k plots each and a standard balanced allocation of v treatments, which is supplemented by the addition of an extra plot to each block to which a *control* is applied. The resulting block design has b blocks each of $k + 1$ plots and $v + 1$ "treatments", but is readily found to satisfy (GB) for the "treatment" decomposition

$$\mathcal{T} = \mathcal{G} \oplus \mathcal{T}_* \oplus \mathcal{T}_c \tag{5.3}$$

where $\mathcal{G} = \mathcal{R}(G)$, $\mathcal{T}_*$ is the $(v - 1)$-dimensional space of contrasts amongst the v original treatments, and $\mathcal{T}_c$ is the 1-dimensional subspace spanned by the contrast comparing the control to the average of the original treatments. This contrast is estimated with efficiency 1 in the intra-block stratum, whilst the contrasts in $\mathcal{T}_*$ are estimated intra-block with

efficiency e^* where $1 - e^* = k(k+1)^{-1}(1-e)$, e being the efficiency factor of the original BIBD.

A similar analysis holds for block designs which only satisfy (GB) with more complicated treatment decompositions, and also for row-column and other designs with supplemented balance: in these cases $\mathscr{T}_*$ is replaced by the direct sum of the terms relative to which the original (unsupplemented) design satisfied (GB).

Pearce's block designs with supplemented balance are a special case of a class of block designs introduced by Nair and Rao (1942), which are themselves a variant on those described in the previous sub-section. They are analogous to PBIBDs with group-divisible association schemes defined on the treatments, but do not necessarily have equal group sizes, in which case they do not define an association scheme. Despite this fact, even when the group sizes are unequal the line of argument used in Proposition 5.1 carries over. We illustrate the results with the case of two groups, as discussed in Nair and Rao (1942), supposing that there are v_1 "rare" treatments each replicated r_1 times, and v_2 "frequent" treatments each replicated r_2 times. Each pair of "rare" (resp. "frequent") treatments occurs together in the same block n_{11} (resp. n_{22}) times, whilst pairs of treatments one of which is "rare" and the other "frequent" occur together in a block $n_{12} = n_{21}$ times. It is easy to establish that such designs are balanced with respect to the treatment decomposition

$$\mathscr{T} = \mathscr{G} \oplus \mathscr{T}_1 \oplus \mathscr{T}_2 \oplus \mathscr{T}_c$$

where $\mathscr{T}_1$ (resp. $\mathscr{T}_2$) is the space of dimension $n_1 - 1$ (resp. $n_2 - 1$) spanned by contrasts between the "rare" (resp. "frequent") treatments, and $\mathscr{T}_c$ is spanned by the single d.f. contrast comparing the average of the "rare" treatments with the average of the "frequent" treatments. The array of efficiency factors is shown in Table 7.

5.4. *Designs satisfying* (GB). Nelder (1965) observed that most of the common designs in use satisfied his definition of general balance. With our extension (GB) to designs in which treatments are not necessarily equally replicated, we can go further and assert that *all* block designs (with equal block sizes, and the usual dispersion model) satisfy (GB), since it is quite obvious that TGT, $T(B-G)T$ and $T(I-B)T$ all commute. All row and column designs which we have seen in the literature satisfy (GB), see Kshirsagar (1957), Pearce (1963, 1975), Zelen and Federer (1964a) for examples, and so also do all designs known to us with orthogonal block structure having three or more strata.

Knowing that a block design must satisfy (GB) is one thing; having explicit expressions for the orthogonal projections $\{T_\beta\}$ is quite another matter. There are a very large number of types of PBIBDs, and although it is generally not difficult to describe the structure of their Bose-Mesner algebra, see MacWilliams and Sloane (1978, Chapter 21), and hence obtain the $\{T_\beta\}$, most writers in statistics have not taken this viewpoint. Corsten (1976) is an exception.

For classes of block designs which are not PBIBDs, other methods must be used; the details concerning rectangular lattice designs, linked block and a number of other classes

TABLE 7

Treatment term:	gm	1	2	c	*Stratum:*
$\Lambda =$	1	0	0	0	grand mean
	0	$\dfrac{r_1 - n_{11}}{kr_1}$	$\dfrac{r_2 - n_{22}}{kr_2}$	$\dfrac{r_1 r_2 - bn_{12}}{r_1 r_2}$	blocks
	0	$\dfrac{r_1(k-1) + n_{11}}{kr_1}$	$\dfrac{r_2(k-1) + n_{22}}{kr_2}$	$\dfrac{bn_{12}}{r_1 r_2}$	plots

are available on request. Recently the class of α-designs was introduced, Patterson and Williams (1976), these being obtained in a particularly simple way from a basic generating array. This class seems to be so large, including BIBDs, PBIBDs, square and rectangular lattice designs as well as many others, that it does not seem to be possible to give a general description of the subspaces $\{\mathscr{T}_\beta\}$ relative to which the designs satisfy (GB). However this should be regarded as a challenging unsolved problem.

5.5. *Designs not satisfying* (GB).

A black sheep. Although all block designs satisfy (GB) this is not necessarily the case for row-column designs as the following 4×4 example with four equally-replicated treatments is shown in Table 8. To see that (GB) fails, one simply notes that the contrast which compares treatment 1 with the average of treatments 2, 3 and 4 is an eigenvector of $T(C-G)T$ (notation as in Section 2 above) and *not* of $T(R-G)T$.

Other designs. Some designs in common use which may not satisfy (GB) are those in which repeated measures are taken on a number of units, when both time (e.g. periods) and subjects (say) are assumed to contribute to the dispersion model, i.e. are regarded as "random effects", and "residual" as well as "direct" treatment effects are included in the model, see Cochran and Cox (1957) for a general discussion. The problem here is that there are no residual effects applying to the first period. In general both time and subjects are regarded as "fixed", in which case no problems arise because the dispersion model is then trivial.

Another class of designs whose structure and accepted analysis does not satisfy (GB) is the class of so-called *two-phase experiments*, McIntyre (1955, 1956), Curnow (1959). The explanation here appears to be simply the amount of structure in the experiment.

5.6. *Concluding discussion.* Throughout this paper we have discussed the notion of balance and its generalisations from a purely theoretical point of view, focusing upon contrasts with particular mathematical properties. It has not been our concern whether these contrasts are natural, or of possible scientific interest, although this is clearly the case in many common examples.

The designer of an experiment has a quite different perspective. Amongst other things, he tries to ensure that contrasts of primary interest are estimated with as high a precision as possible, subject to the constraints imposed by the experimental material. It by no means follows that he should always design his experiment so that such contrasts are eigenvectors of all the $\{TS_\alpha T\}$ of Section 4.1; indeed in many cases this is impossible.

If a designed experiment with orthogonal block structure satisfies (GB), then the coarsest decomposition $\mathscr{T} = \oplus \mathscr{T}_\beta$ with respect to which it does so is uniquely defined by the design. Other decompositions of $\mathscr{T}$ which satisfy (GB) can only arise by further decomposition of the individual $\{\mathscr{T}_\beta\}$ in the coarsest one. When the designer is able to arrange that all of the subspaces $\{\mathscr{T}_\beta\}$ consist of contrasts of interest, the analysis of data from the experiment and the display of the results will be particularly straightforward; examples here include BIBDs and the designs of Section 5.3. In general, however, not all

TABLE 8
Design not satisfying (GB).

2	1	1	1
1	3	3	2
2	2	4	3
4	4	4	3

contrasts of interest will belong to one of the $\mathscr{T}_\beta$, and it will be necessary in the analysis to use the more complicated formula (4.2) involving the projections $\{T_\beta\}$; examples here include unbalanced lattice designs.

A final point concerning the subspaces $\{\mathscr{T}_\beta\}$ in (GB) is worth making. Even when they do not consist of contrasts of scientific interest, they are frequently recognisable as arising from a pseudo-structure on the treatments, i.e. an artificial view of the treatments relative to which the $\{\mathscr{T}_\beta\}$ are natural or interpretable. Examples here include many PBIBDs, most lattice designs and Example 2 of Section 5.2. The most general design satisfying (GB)—and we need go no further than block α-designs to find examples—involves a decomposition of $\mathscr{T}$ into subspaces $\{\mathscr{T}_\beta\}$ which have neither scientific interest nor any natural or interpretable structure, however we care to view the treatments. Our general theory applies to such designs, although it may be an affront to some to describe them as balanced in any sense. We hope that our readers will appreciate the value of tracing the path from balance in BIBDs through to the notion of general balance, and conclude that the unity of outlook achieved outweighs any terminological problems met along the way.

6. Acknowledgements. This work was begun in the Department of Statistics at Princeton University whilst the first author was a graduate student and the second a visitor to the department. Many thanks are due to Professor Geoffrey S. Watson for his joint role as adviser and host, and to others in the department for the congenial work atmosphere they helped to create. The paper was written whilst the authors were at the Polytechnic Institute of New York and CORE, Belgium (A.M.H.) and the University of Western Australia (T.P.S.) and thanks are due to these institutions for their support. The referees are also to be thanked for their many helpful comments.

REFERENCES

BAILEY, R. A. (1981). A unified approach to design of experiments. *J. Roy. Statist. Soc. Ser. A* **144** 214–223.

BOSE, R. C. and NAIR, K. R. (1939). Partially balanced incomplete block designs. *Sankhyā* **4** 337–372.

BROWN, L. D. and COHEN, ARTHUR (1974). Point and confidence estimation of a common mean and recovery of interblock information. *Ann. Statist.* **2** 963–976.

CHENG, C.-S. (1981a). Optimality and construction of pseudo Youden designs. *Ann. Statist.* **9** 201–205.

CHENG, C.-S. (1981b). A family of pseudo Youden designs with row size less than the number of symbols. *J. Comb. Theory Ser. A* **31** 219–221.

CLATWORTHY, W. H. (1973). *Tables of Two-Associate-Class Partially Balanced Designs.* Appl. Math. Ser. 63, National Bureau of Standards, Washington, D.C.

COCHRAN, WILLIAM G. and COX, GERTRUDE M. (1957). *Experimental Designs.* Second Edition. Wiley, New York.

CORSTEN, L. C. A. (1962). Balanced block designs with two different numbers of replicates. *Biometrics* **18** 499–519.

CORSTEN, L. C. A. (1976). Canonical correlation in incomplete blocks. *Essays in Probability and Statistics.* Shinko Tsusho Co., Ltd. Japan, 125–154.

CURNOW, R. N. (1959). The analysis of a two phase experiment. *Biometrics* **15** 60–73.

FISHER, R. A. (1935). Discussion following Yates' "Complex Experiments". *Collected Papers of R. A. Fisher* **3** 332–333.

HOBLYN, T. N., PEARCE, S. C. and FREEMAN, G. H. (1954). Some considerations in the design of successive experiments in fruit plantations. *Biometrics* **10** 503–515.

HOUTMAN, ANNE M. (1980). The analysis of designed experiments. Ph.D. thesis, Princeton University.

JAMES, A. T. and WILKINSON, G. N. (1971). Factorization of the residual operator and canonical decomposition of non-orthogonal factors in analysis of variance. *Biometrika* **58** 279–294.

KSHIRSAGAR, A. M. (1957). On balancing in designs in which heterogeneity is eliminated in two directions. *Calc. Statist. Assoc. Bull.* **7** 161–166.

KSHIRSAGAR, A. M. (1966). Balanced factorial designs. *J. Roy. Statist. Soc. Ser. B* **28** 559–567.

KURKJIAN, B. and ZELEN, M. (1963). Application of the calculus of factorial arrangements, I. Block and direct product designs. *Biometrika* **50** 63–73.

MACWILLIAMS, F. J. and SLOANE, N. J. A. (1978). *The Theory of Error Correcting Codes.* North Holland, Amsterdam.

BALANCE IN DESIGNED EXPERIMENTS

MARTIN, FRANK B. and ZYSKIND, GEORGE (1966). On combinability of information from uncorrelated linear models by simple weighting. *Ann. Math. Statist.* **37** 1338–1347.
MCINTYRE, G. A. (1955). Design and analysis of two-phase experiments. *Biometrics* **11** 324–334.
MCINTYRE, G. A. (1956). Query 123. *Biometrics* **12** 527–732.
MORLEY JONES, R. (1959). A property of incomplete blocks. *J. Roy. Statist. Soc. Ser. B* **21** 172–179.
NAIR, K. R. and RAO, C. R. (1942). Confounding in asymmetrtical factorial experiments. *J. Roy. Statist. Soc. Ser. B* **10** 109–131.
NELDER, J. A. (1954). The interpretation of negative components of variance. *Biometrika* **41** 544–548.
NELDER, J. A. (1965a). The analysis of randomized experiments with orthogonal block structure, I. Block structure and the null analysis of variance. *Proc. Roy. Soc. (London) Ser. A.* **273** 147–162.
NELDER, J. A. (1965b). The analysis of randomized experiments with orthogonal block structure, II. Treatment structure and the general analysis of variance. *Proc. Roy. Soc. (London) Ser. A* **273** 163–178.
NELDER, J. A. (1968). The combination of information in generally balanced designs. *J. Roy. Statist. Soc. Ser. B* **30** 303–311.
PATTERSON, H. D. and WILLIAMS, E. R. (1976). A new class of resolvable incomplete block designs. *Biometrika* **63** 83–92.
PATTERSON, H. D. and THOMPSON, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika* **58** 545–554.
PEARCE, S. C. (1960). Supplemented balance. *Biometrika* **47** 263–271.
PEARCE, S. C. (1963). The use and classification of non-orthogonal designs (with discussion). *J. Roy. Statist. Soc. Ser. A* **126** 353–377.
PEARCE, S. C. (1970). The efficiency of block designs in general. *Biometrika* **57** 339–346.
PEARCE, S. C. (1975). Row-and-column designs. *Applied Statistics* **24** 60–74.
PEARCE, S. C., CALINSKI, T. and MARSHALL, T. F. DE C. (1974). The basic contrasts of an experimental design with special referece to the analysis of data. *Biometrika* **61** 449–460.
PREECE, D. A. (1967). Nested balanced incomplete block designs. *Biometrika* **54** 479–486.
PREECE, D. A. (1968). Balanced 6 × 6 designs for 9 treatments. *Sankhyā Ser. B* **9** 201–205.
PREECE, D. A. (1976). A second domain of balanced 6 × 6 designs for 9 equally-replicated treatments. *Sankhyā Ser. B* **38** 192–194.
RAGHAVARAO, DAMARAJU (1971). *Construction and Combinatorial problems in Design of Experiments.* Wiley, New York.
RAO, V. R. (1958). A note on balanced designs. *Ann. Math. Statist.* **29** 290–294.
SPROTT, D. A. (1956). A note on combined interblock and intrablock estimation in incomplete block designs. *Ann. Math. Statist.* **27** 633–641.
YATES, F. (1936). Incomplete randomised blocks. *Ann. Eugenics London* **7** 121–140.
YATES, F. (1939). The recovery of inter-block information in variety trials arranged in three-dimensional lattices. *Ann. Eugenics* **9** 136–156.
YATES, F. (1940). The recovery of inter-block information in balanced incomplete block designs. *Ann. Eugenics* **10** 317–325.
ZELEN, M. and FEDERER, W. T. (1964a). Application of the calculus for factorial arrangements, II. Two-way elimination of heterogeneity. *Ann. Math. Statist.* **35** 658–672.
ZELEN, M. and FEDERER, W. T. (1964b). Application of the calculus for factorial arrangements, III. Analysis of factorials with unequal numbers of observations. *Sankhyā Ser. A* **25** 383–400.

56 AV. DES ARTS
BRUSSELS, BELGIUM

CSIRO
DIVISION OF MATHEMATICS AND STATISTICS
BOX 1965, GPO
CANBERRA, ACT 2601
AUSTRALIA

Austral. J. Statist., **26** (3), 1984, 227–246

THE ANALYSIS OF MULTISTRATA DESIGNED EXPERIMENTS WITH INCOMPLETE DATA[1]

A. HOUTMAN
A. C. Nielsen, Belgium

AND

T. P. SPEED
CSIRO Division of Mathematics & Statistics, Australia

Summary

Building upon early work of E. A. Cornish we show that G. N. Wilkinson's version of Yates' approach to the analysis of designed experiments with a single error stratum carries over completely to designs with an arbitrary non-singular covariance matrix, initially assumed known. We show that the equations, corrections, adjustments and algorithms all have their more general analogues and that these can be solved, computed or executed quite readily if the design has orthogonal block structure and satisfies Nelder's condition of general balance. The results are illustrated with a split-plot and a simple (square) lattice design.

1. Introduction

The problem of analysing designed experiments with incomplete data—for example, missing or mixed-up values—has received a lot of attention when the designs are analysed with only a single error line. The corresponding questions for designs analysed with more than one error line (which we term multistrata designs), such as split-plot designs or block designs in which inter-block information is recovered, have rarely been raised, and in our opinion the accepted answers in these areas are not completely satisfactory. The most frequently adopted approach is to change the model back to one with only a single error line, that corresponding to the lowest stratum, and to carry out the analysis appropriate for incomplete data under the model in which all other terms (including the other errors) are fixed. Such an approach has the merit of simplicity, but it has no theoretical basis

[1] Manuscript received May 30, 1983.

and in our experience can give replacement values which are undesirably discordant with the remaining data. This paper reports an attempt to give an analysis of multistrata designed experiments with incomplete data which is closer to the exact one under the model usually assumed for such data.

We will build upon the early work of Cornish (1943, 1944, 1956), showing that Wilkinson's (1958a,b) version of Yates' (1933) approach carries over completely to experimental designs with an arbitrary non-singular covariance matrix **V**, initially assumed known. More precisely, we show that the equations, corrections, adjustments and algorithms associated with the analysis of an experiment with incomplete data but only one error line all have their more general analogues; the main problem is their solution, evaluation or execution. To simplify the discussion, as well as to make contact with the common multistrata designs, we then specialise to designs which are generally balanced in the sense of Nelder (1965a,b, 1968). This means that we suppose **V** to have a very specific relationship to the treatment model under discussion, and we remark that all common designs—e.g. all those in Cochran & Cox (1957)—possess this property, see Nelder (1965b) and Houtman & Speed (1983). Indeed most of the common multistrata designs have only two effective strata, i.e. all of the information concerning treatment contrasts lies in only two strata, and for such designs our results are simplified substantially.

Our results are all exact as long as the covariance matrix **V** is known, and in the case of generally balanced designs a natural extension of Nelder's (1968) method for estimating an unknown **V** suggests itself. The discussion is then illustrated by giving our analysis of a split-plot and a simple (square) lattice design, each having a single missing value.

We have not attempted in this paper to describe what we regard as the best way to carry out the associated calculations. One reason for this is our desire to outline a general approach and avoid concentrating on particular designs, but the main reason is the absence of widely-used general algorithms which perform multistrata analyses and are capable of the few modifications necessary to do the calculations we require. The ANOVA algorithm and the associated Macro facilities which can be found in GENSTAT, see Alvey *et al.* (1977), provide the most convenient framework known to us for doing the job.

2. Previous Work on the Subject

Formulae for replacing a single missing value in randomized complete block designs and in Latin squares were given by Allan & Wishart (1930), but it was Yates (1933) who laid down the general principles for replacing missing values in designed experiments and for

correcting other aspects of the analysis of the completed data. Yates' method, suggested to him by Fisher, consists of using those replacement values which minimize the residual sum of squares when unknowns are substituted for the missing response values. When only one value is missing, this method leads to a simple direct formula for the replacement; when there are several missing values Yates suggests an iterative method for solving the equations. He notes that this method leads to the correct fitted values for the observed data, but with inflated treatment sums of squares, gives the correction for randomized blocks and for Latin squares and also computes the adjustment to the variance of a contrast for those two types of designs.

Yates' work was later generalized and expressed in a modern framework by Wilkinson (1958a,b, 1959) and a host of authors have made contributions to the formulation, interpretation, existence, uniqueness and solution of problems with incomplete data, see Hoyle (1971) for an extensive but incomplete bibliography. In particular we note the coordinate-free approach shown by Kruskal (1961) to include the estimation of mixed-up values and to provide an easy extension to the analysis of designs with extra observations.

A different approach to estimating missing values was introduced by Bartlett (1937). This method first assigns arbitrary response values to the missing plots and then adjusts the completed data by covariance upon pseudo-covariates, one being introduced for each missing plot and having value unity for that missing plot and zeroes elsewhere. It is easy to see that estimates so obtained are identical to those derived by Yates' method. Further contributions framed within the analysis of covariance approach with a single error can be found in Nair (1940), Truitt & Fairfield Smith (1956), Coons (1957), John & Prescott (1975), John & Lewis (1976) and P. L. Smith (1981).

Many authors have studied iterative methods to obtain estimates of missing values in single stratum experiments. The use of an iterative procedure was first recommended by Yates (1933). Later Healy & Westmacott (1956) gave a more general algorithm based on Yates' observation that the residuals after fitting the completed data must be zero in the cells corresponding to the missing values. Pearce (1965) improved the Healy–Westmacott algorithm by introducing an accelerating factor n/E where n is the total number of experimental units and E is the number of residual degrees of freedom for a complete experiment. This correction is also used in papers by Preece (1971) and Pearce & Jeffers (1971). More recently Rubin (1972), Haseman & Gaylor (1973), John & Prescott (1975) developed non-iterative methods involving $m+1$ uses of the same subroutine used for fitting the full model where m is the number of missing values. Jarrett (1978) describes the relationships between those various computing procedures.

The problem of mixed-up values was first considered by Nair (1940) using the analysis of covariance. Kruskal (1968) follows Yates' approach in a coordinate-free framework. Preece & Gower (1974) give an iterative procedure to deal with mixed-up values similar to the one advocated by Healy & Westmacott (1956) for missing values. John & Lewis (1976) give a direct procedure based on the appropriate analysis of variance.

Most of the literature on missing values concerns experiments with a single error stratum. The earliest efforts to adapt Yates' approach to designs with more than one error line, mainly lattices and BIB designs, are due to Cornish (1943, 1944, 1956) in three papers dealing with the recovery of interblock information. An influential early note of Anderson (1946) seems to end up recommending the lowest-stratum-only analysis for split-plot designs with missing data. Anderson's view has become accepted, see Cochran & Cox (1957), and is widely used to this day. We note in passing that little satisfaction can be gained by an appeal to the analysis of covariance, since, for multistrata designed experiments, this technique is not in much better state than the special cases which incomplete data pose. An exception is the unpublished report Cochran (1946) which discusses the analysis of covariance in split-plot designs and whose results may be modified to handle missing and mixed-up values. Recently Williams, Ratcliff & Speed (1981) showed how to get missing value estimates based on the information contained in the lowest two strata.

Finally we note that the EM algorithm described by Dempster, Laird & Rubin (1977) provides an iterative approach to the maximum likelihood estimation of parameters from incomplete data under quite general distributional assumptions. Under normality assumptions it can be shown that the recursion in the EM algorithm is the same as that satisfied by the estimates obtained at each step from Healy and Westmacott's algorithm, or from the extension we give of that algorithm for multistrata experiments. However the discussion below will only make the standard second-order assumptions usual in the analysis of designed experiments.

3. Derivation of the Basic Equations

We regard the observations as an array of numbers $\mathbf{y} = (y_i)_{i \in I}$ indexed by a set I of n unit labels and assume the following model for the expectation and dispersion of $\mathbf{y}$:

$$\mathbb{E}\mathbf{y} \in \mathcal{T} \tag{3.1}$$

$$\mathbb{D}\mathbf{y} = \mathbf{V}$$

where $\mathcal{T}$ is the subspace of arrays that are constant over treatments and $\mathbf{V}$ is a positive-definite matrix which is assumed to be known. The

n-dimensional vector space $\mathscr{D}$ of all possible arrays $\mathbf{y}$ is endowed with the natural inner product $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i \in I} x_i y_i$ and with the inner product $\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{V}} = \langle \mathbf{x}, \mathbf{V}^{-1}\mathbf{y} \rangle$ induced by $\mathbf{V}$; the associated norms are denoted respectively by $|.|$ and $|.|_{\mathbf{V}}$.

If the data are incomplete, the space $\mathscr{D}$ splits into the sum of two orthogonal sub-spaces

$$\mathscr{D} = \mathscr{D}_1 \oplus \mathscr{D}_2$$

reflecting the decomposition

$$\mathbf{y} = \mathbf{y}_1 + \mathbf{y}_2$$

of the data into the observed part $\mathbf{y}_1$ and the "missing" part $\mathbf{y}_2$. This notation was shown by Kruskal (1968) to include both the case of missing values and the case of mixed-up values: in the first case, $\mathbf{y}_1$ has zeroes in all units corresponding to missing observations and in the second case, $\mathbf{y}_1$ has a quantity $z = m^{-1}S$ in all m units corresponding to the observations whose sum S only was observed.

For the observed part $\mathbf{y}_1$ of the data, the model (3.1) now becomes

$$\begin{aligned} \mathbb{E}\mathbf{y}_1 &\in \mathscr{T}_1 \\ \mathbb{D}\mathbf{y}_1 &= \mathbf{V}_1 \end{aligned} \tag{3.2}$$

where $\mathscr{T}_1 = \mathbf{D}_1\mathscr{T}$, the orthogonal projection of $\mathscr{T}$ onto $\mathscr{D}_1$ and $\mathbf{V}_1 = \mathbf{D}_1\mathbf{V}\mathbf{D}_1$. (In this paper we will always use script letters to denote linear spaces and the corresponding capital letters to denote the orthogonal projections onto those spaces, with a superscript $\mathbf{V}$ if the projection is orthogonal with respect to $\langle .\,, .\rangle_{\mathbf{V}}$ rather than $\langle .\,, .\rangle$). Although the data could be fitted by estimating $\mathbb{E}\mathbf{y}_1$ by its BLUE $\hat{\boldsymbol{\tau}}_1 = \mathbf{T}_1^{\mathbf{V}_1}\mathbf{y}_1$, it is usually not straightforward to do so, since any special relationship that existed between $\mathbf{V}$ and $\mathscr{T}$ (e.g. orthogonality or general balance) would not usually continue to hold between $\mathbf{V}_1$ and $\mathscr{T}_1$. Accordingly, following Cornish (1956), we minimize

$$|\mathbf{y}_1 + \mathbf{y}_2 - \boldsymbol{\tau}|^2_{\mathbf{V}},$$

over $\boldsymbol{\tau} \in \mathscr{T}$ and $\mathbf{y}_2 \in \mathscr{D}_2$, and assume that $(\tilde{\boldsymbol{\tau}}, \tilde{\mathbf{y}}_2)$ is a pair at which the minimum is achieved. Arguing as Yates (1933) did in the single stratum case, we may minimize first over $\mathscr{T}$ and then over $\mathscr{D}_2$ to get

$$\tilde{\mathbf{y}}_2 = \mathbf{D}_2^{\mathbf{V}}(\tilde{\boldsymbol{\tau}} - \mathbf{y}_1),$$

and in the reverse order, obtaining

$$\tilde{\boldsymbol{\tau}} = \mathbf{T}^{\mathbf{V}}(\mathbf{y}_1 + \tilde{\mathbf{y}}_2).$$

Each of the two relations can be substituted into the other, leading to

$$\mathbf{D}_2^{\mathbf{V}}\bar{\mathbf{T}}^{\mathbf{V}}\tilde{\mathbf{y}}_2 = -\mathbf{D}_2^{\mathbf{V}}\bar{\mathbf{T}}^{\mathbf{V}}\mathbf{y}_1 \tag{3.3}$$

and

$$\mathbf{T}^{\mathbf{V}}\bar{\mathbf{D}}_2^{\mathbf{V}}\tilde{\boldsymbol{\tau}} = \mathbf{T}^{\mathbf{V}}\bar{\mathbf{D}}_2^{\mathbf{V}}\mathbf{y}_1 \tag{3.4}$$

where $\bar{\mathbf{T}}^{\mathbf{V}} = \mathbf{I} - \mathbf{T}^{\mathbf{V}}$ and similarly $\bar{\mathbf{D}}_2^{\mathbf{V}} = \mathbf{I} - \mathbf{D}_2^{\mathbf{V}}$.

In the next section section we will show how to solve these equations. For the moment we simply state the result which justifies their solution as follows: the restriction $\mathbf{D}_1\mathbf{T}^{\mathbf{V}}(\mathbf{y}_1 + \tilde{\mathbf{y}}_2)$ to the subspace corresponding to the observed data of the fitted values $\mathbf{T}^{\mathbf{V}}(\mathbf{y}_1 + \tilde{\mathbf{y}}_2)$ obtained by analysing the observed data $\mathbf{y}_1$ completed with any solution $\tilde{\mathbf{y}}_2$ of (3.3) coincides with the fitted values $\mathbf{T}_1^{\mathbf{V}_1}\mathbf{y}_1$ of the observed data $\mathbf{y}_1$ to the appropriate submodel, i.e.

$$\hat{\boldsymbol{\tau}}_1 = \mathbf{D}_1\tilde{\boldsymbol{\tau}}. \tag{3.5}$$

A proof of this result is given in the appendix.

From (3.3) or (3.4) it is easy to verify that the vector of residuals $\tilde{\mathbf{r}} = \bar{\mathbf{T}}^{\mathbf{V}}(\mathbf{y}_1 + \tilde{\mathbf{y}}_2)$ satisfies the equation

$$\mathbf{D}_2^{\mathbf{V}}\tilde{\mathbf{r}} = \mathbf{0}, \tag{3.6}$$

which is similar but not equivalent to the property noticed by Yates in the single stratum case, that the residuals after fitting the completed data to $\mathcal{T}$ must be zeroes in the units corresponding to missing values.

4. Solutions of the Equations

In the simple case in which there is a single missing value (or only two mixed-up values) a direct formula may be obtained. Letting $\boldsymbol{\varepsilon}$ denote a dummy vector with unity in place of the missing value (or $+1$ and -1 in place of the two mixed-up values) and zeroes everywhere else, the unobserved vector has the form $\mathbf{y}_2 = a\boldsymbol{\varepsilon}$ where a is to be estimated, and $\mathbf{D}_2^{\mathbf{V}}\mathbf{z} = |\boldsymbol{\varepsilon}|_{\mathbf{V}}^{-2}\langle\boldsymbol{\varepsilon}, \mathbf{z}\rangle_{\mathbf{V}}\boldsymbol{\varepsilon}$. It follows then immediately from (3.3) that

$$\tilde{\mathbf{y}}_2 = -\frac{\langle\boldsymbol{\varepsilon}, \bar{\mathbf{T}}^{\mathbf{V}}\mathbf{y}_1\rangle_{\mathbf{V}}}{|\bar{\mathbf{T}}^{\mathbf{V}}\boldsymbol{\varepsilon}|_{\mathbf{V}}^2}\boldsymbol{\varepsilon}. \tag{4.1}$$

A more manageable form of $\tilde{\mathbf{y}}_2$ will be obtained in §6 for generally balanced designs and it will be illustrated with examples in §8. We will now suppose that $\dim \mathcal{D} > 1$, i.e. that there is more than one missing value or there are more than two mixed-up values, and study iterative methods for computing the solutions of (3.3) and (3.4) The following recursion formulae suggest themselves for $\tilde{\mathbf{y}}_2$, $\tilde{\mathbf{y}} = \mathbf{y}_1 + \tilde{\mathbf{y}}_2$, $\tilde{\boldsymbol{\tau}}$ and $\tilde{\mathbf{r}} = \bar{\mathbf{T}}^{\mathbf{V}}\tilde{\mathbf{y}}$:

(i) $\mathbf{y}_2^{(0)} = \mathbf{0}$; $\quad \tilde{\mathbf{y}}_2^{(m+1)} = \mathbf{D}_2^{\mathbf{V}}[\mathbf{T}^{\mathbf{V}}(\mathbf{y}_1 + \tilde{\mathbf{y}}_2^{(m)}) - \mathbf{y}_1]$, $\quad m \geq 0$.

(ii) $\tilde{\mathbf{y}}^{(0)} = \mathbf{y}_1$; $\quad \tilde{\mathbf{y}}^{(m+1)} = (\mathbf{I} - \mathbf{D}_2^{\mathbf{V}}\bar{\mathbf{T}}^{\mathbf{V}})\tilde{\mathbf{y}}^{(m)}$, $\quad m \geq 0$.

(iii) $\tilde{\boldsymbol{\tau}}^{(0)} = \mathbf{T}^{\mathbf{V}}\mathbf{y}_1$; $\quad \tilde{\boldsymbol{\tau}}^{(m+1)} = \mathbf{T}^{\mathbf{V}}(\bar{\mathbf{D}}_2^{\mathbf{V}}\mathbf{y}_1 + \mathbf{D}_2^{\mathbf{V}}\tilde{\boldsymbol{\tau}}^{(m)})$, $\quad m \geq 0$.

(iv) $\tilde{\mathbf{r}}^{(0)} = \bar{\mathbf{T}}^{\mathbf{V}}\mathbf{y}_1$; $\quad \tilde{\mathbf{r}}^{(m+1)} = \bar{\mathbf{T}}^{\mathbf{V}}\bar{\mathbf{D}}_2^{\mathbf{V}}\tilde{\mathbf{r}}^{(m)}$, $\quad m \geq 0$.

All of these recursions are essentially the same, each being obtainable from the others by simple algebraic manipulations. Recursion (i) is a generalization of Healy and Westmacott's algorithm, and, under normality assumptions, (iii) can be shown to be equivalent to the EM algorithm of Dempster *et al.* (1977). It is the last recursion which most clearly indicates why convergence must take place, since $\tilde{\mathbf{T}}^{\mathbf{V}}$ and $\tilde{\mathbf{D}}_2^{\mathbf{V}}$ are projection operators, and so $|\tilde{\mathbf{r}}^{(m+1)}|_{\mathbf{V}} \leq |\mathbf{r}^{(m)}|_{\mathbf{V}}$ for all $m = 0, 1, 2, \ldots$, with equality if and only if $\tilde{\mathbf{r}}^{(m+1)} = \tilde{\mathbf{r}}^{(m)}$, in which case the algorithm stops and $\tilde{\mathbf{r}}^{(m)}$ is the solution $\tilde{\mathbf{r}}$ by virtue of (3.6). An alternative proof of convergence uses a theorem of Von Neumann (1950, p. 55) showing that (iv) converges to the projection of $\mathbf{y}_1$ onto $\mathbf{V}\mathcal{T}^{\perp} \cap \mathbf{V}\mathcal{D}_1$, orthogonal with respect to $\langle .\,, . \rangle_{\mathbf{V}}$. We also notice that each algorithm is equivalent to a Taylor expansion.

The speed of convergence may be improved by the introduction of an appropriate acceleration factor ω. With the same initial values as before, the algorithms are modified as follows for $m \geq 0$:

(i)′ $\tilde{\mathbf{y}}_2^{(m+1)} = \tilde{\mathbf{y}}_2^{(m)} - \omega^{-1}\mathbf{D}_2^{\mathbf{V}}\tilde{\mathbf{T}}^{\mathbf{V}}(\mathbf{y}_1 + \tilde{\mathbf{y}}_2^{(m)})$.

(ii)′ $\tilde{\mathbf{y}}^{(m+1)} = \tilde{\mathbf{y}}^{(m)} - \omega^{-1}\mathbf{D}_2^{\mathbf{V}}\tilde{\mathbf{T}}^{\mathbf{V}}\tilde{\mathbf{y}}^{(m)}$.

(iii)′ $\tilde{\boldsymbol{\tau}}^{(m+1)} = \tilde{\boldsymbol{\tau}}^{(m)} - \omega^{-1}\mathbf{T}^{\mathbf{V}}\tilde{\mathbf{D}}_2^{\mathbf{V}}(\tilde{\boldsymbol{\tau}}^{(m)} - \mathbf{y}_1)$.

(iv)′ $\tilde{\mathbf{r}}^{(m+1)} = \tilde{\mathbf{r}}^{(m)} - \omega^{-1}\tilde{\mathbf{T}}^{\mathbf{V}}\mathbf{D}_2^{\mathbf{V}}\tilde{\mathbf{r}}^{(m)}$.

As all four algorithms are equivalent we study the convergence of (i)′. Using (3.3) one obtains

$$\tilde{\mathbf{y}}_2 - \tilde{\mathbf{y}}_2^{(m+1)} = (\mathbf{I} - \omega^{-1}\mathbf{D}_2^{\mathbf{V}}\tilde{\mathbf{T}}^{\mathbf{V}}\mathbf{D}_2^{\mathbf{V}})(\tilde{\mathbf{y}}_2 - \tilde{\mathbf{y}}_2^{(m)}).$$

It follows that the algorithm converges to a solution of (3.3) for all $\mathbf{y}_2^{(0)}$ in $\mathcal{D}_2$ if and only if the spectral radius ρ of $I - \omega^{-1}\mathbf{D}_2^{\mathbf{V}}\tilde{\mathbf{T}}^{\mathbf{V}}\mathbf{D}_2^{\mathbf{V}}$ is strictly smaller than unity. The solution is unique if we assume that $\mathcal{D}_2 \cap \mathcal{T} = \{0\}$, excluding, in particular, situations where all the observations on a treatment combination are missing. Under this assumption, the algorithm converges to the unique solution $\tilde{\mathbf{y}}_2$ if and only if

$$\omega > \tfrac{1}{2}\lambda_{\max}(\mathbf{D}_2^{\mathbf{V}}\tilde{\mathbf{T}}^{\mathbf{V}}\mathbf{D}_2^{\mathbf{V}})$$

where $\lambda_{\max}(\mathbf{A})$ is the largest eigenvalue of $\mathbf{A}$. The fastest convergence is obtained for the value ω_{opt} which minimizes ρ and hence

$$\omega_{\text{opt}} = \tfrac{1}{2}[\lambda_{\min}(\mathbf{D}_2^{\mathbf{V}}\tilde{\mathbf{T}}^{\mathbf{V}}\mathbf{D}_2^{\mathbf{V}}) + \lambda_{\max}(\mathbf{D}_2^{\mathbf{V}}\tilde{\mathbf{T}}^{\mathbf{V}}\mathbf{D}_2^{\mathbf{V}})],$$

where $\lambda_{\min}(\mathbf{A})$ is the smallest nonzero eigenvalue of $\mathbf{A}$.

5. Modifications to the Subsequent Analysis

The analysis performed on the data completed with a solution of (3.3) will produce the correct residual sum of squares (although the number of degrees of freedom must be reduced by the dimension of

$\mathcal{D}_2$) but treatment sums of squares (differences between residual sums of squares for a pair of nested treatment models) and variances of contrasts will need adjustment. Yates had already pointed this out in 1933 and he gave the corrections for the designs he studied. We now give the appropriate adjustments for multistrata experiments.

For a submodel $\mathbb{E}\mathbf{y} \in \mathcal{U} \subset \mathcal{T}$ of our original model, the treatment sum of squares $|\bar{\mathbf{U}}^{\mathbf{V}}\mathbf{y}|^2_{\mathbf{V}} - |\bar{\mathbf{T}}^{\mathbf{V}}\mathbf{y}|^2_{\mathbf{V}}$ will be inflated by the quantity

$$|\bar{\mathbf{U}}^{\mathbf{V}}\mathbf{y}(\mathcal{T})|^2_{\mathbf{V}} - |\bar{\mathbf{U}}^{\mathbf{V}}\mathbf{y}(\mathcal{U})|^2_{\mathbf{V}} = |\bar{\mathbf{U}}^{\mathbf{V}}[\mathbf{y}_2(\mathcal{T}) - \mathbf{y}_2(\mathcal{U})]|^2_{\mathbf{V}} \tag{5.1}$$

where $\mathbf{y}(\mathcal{T}) = \mathbf{y}_1 + \mathbf{y}_2(\mathcal{T})$ and $\mathbf{y}(\mathcal{U}) = \mathbf{y}_1 + \mathbf{y}_2(\mathcal{U})$ denote the completed data obtained by solving the missing values equations (3.3) respectively for the models $\mathbb{E}\mathbf{y} \in \mathcal{T}$ and $\mathbb{E}\mathbf{y} \in \mathcal{U}$. Equation (5.1) is simple the difference between the *apparent* sum of squares

$$|\bar{\mathbf{U}}^{\mathbf{V}}\mathbf{y}(\mathcal{T})|^2_{\mathbf{V}} - |\bar{\mathbf{T}}^{\mathbf{V}}\mathbf{y}(\mathcal{T})|^2_{\mathbf{V}}$$

and the *correct* sum of squares

$$|\bar{\mathbf{U}}^{\mathbf{V}}\mathbf{y}(\mathcal{U})|^2_{\mathbf{V}} - |\bar{\mathbf{T}}^{\mathbf{V}}\mathbf{y}(\mathcal{T})|^2_{\mathbf{V}},$$

the latter being smaller than the former since $\mathbf{y}(\mathcal{U})$ minimizes $|\bar{\mathbf{U}}^{\mathbf{V}}\mathbf{y}|^2_{\mathbf{V}}$ over $\mathcal{D}_2$. The algebra leading to the RHS of (5.1) is given in the appendix.

On the other hand, the variance of a contrast $\langle \mathbf{t}, \tilde{\boldsymbol{\tau}} \rangle$ where $\tilde{\boldsymbol{\tau}}$ satisfies (3.4) can be decomposed into the sum of the variance of that contrast when the data are complete and an adjustment due to the loss of precision encountered when estimating missing data. We have the identity

$$\operatorname{cov}(\langle \mathbf{t}, \tilde{\boldsymbol{\tau}} \rangle, \langle \mathbf{u}, \tilde{\boldsymbol{\tau}} \rangle) = \langle \mathbf{t}, \mathbf{T}^{\mathbf{V}}\mathbf{V}\mathbf{u} \rangle + \langle \mathbf{t}, \mathbf{T}^{\mathbf{V}}(\mathbf{D}_2^{\mathbf{V}}\bar{\mathbf{T}}^{\mathbf{V}}\mathbf{D}_2^{\mathbf{V}})^{-}\mathbf{D}_2^{\mathbf{V}}\mathbf{T}^{\mathbf{V}}\mathbf{V}\mathbf{u} \rangle. \tag{5.2}$$

If $\mathcal{D}_2$ is of dimension one only, this expression simplifies to

$$\operatorname{cov}(\langle \mathbf{t}, \tilde{\boldsymbol{\tau}} \rangle, \langle \mathbf{u}, \tilde{\boldsymbol{\tau}} \rangle) = \langle \mathbf{t}, \mathbf{T}^{\mathbf{V}}\mathbf{V}\mathbf{u} \rangle + \frac{\langle \mathbf{t}, \mathbf{T}^{\mathbf{V}}\boldsymbol{\varepsilon} \rangle \langle \mathbf{u}, \mathbf{T}^{\mathbf{V}}\boldsymbol{\varepsilon} \rangle}{|\bar{\mathbf{T}}^{\mathbf{V}}\boldsymbol{\varepsilon}|^2_{\mathbf{V}}} \tag{5.3}$$

where $\boldsymbol{\varepsilon}$ is the dummy vector introduced in §4. Again we leave the algebra to the appendix and illustrations to §8.

6. Generally Balanced Experiments

Following Nelder (1965a) we now assume that the design has an orthogonal block structure which determines the eigenstructure of the covariance operator

$$\mathbf{V} = \sum_{\alpha} \xi_{\alpha} \mathbf{S}_{\alpha} \tag{6.1}$$

where the $\{\xi_{\alpha}\}$ are (usually unknown) positive eigenvalues and the $\{\mathbf{S}_{\alpha}\}$ are known symmetric and idempotent projectors such that $\sum_{\alpha} \mathbf{S}_{\alpha} = \mathbf{I}$.

Designs with such dispersion models are said to be *generally balanced* for an orthogonal decomposition $\mathcal{T} = \bigoplus_\beta \mathcal{T}_\beta$ of the treatment space if for all α

$$\mathbf{T}\mathbf{S}_\alpha\mathbf{T} = \sum_\beta \lambda_{\alpha\beta}\mathbf{T}_\beta \tag{6.2}$$

for a set of eigenvalues $\{\lambda_{\alpha\beta}\}$ such that $0 \le \lambda_{\alpha\beta} \le 1$ and $\sum_\alpha \lambda_{\alpha\beta} = 1$. This condition will be assumed to hold for the rest of this section. The *effect* corresponding to treatment β in stratum α is then calculated by $\mathbf{Q}_{\alpha\beta} = \lambda_{\alpha\beta}^{-1}\mathbf{T}_\beta\mathbf{S}_\alpha$ (unless there is no information $\mathcal{T}_\beta$ in $\mathcal{S}_\alpha$, and $\lambda_{\alpha\beta} = 0$) and, assuming that the $\{\xi_\alpha\}$ are known, the overall effect of treatment term β is $\mathbf{Q}_\beta = \sum_\alpha w_{\alpha\beta}\mathbf{Q}_{\alpha\beta}$, a linear combination of the within strata effects with weights

$$w_{\alpha\beta} = \xi_\alpha^{-1}\lambda_{\alpha\beta}\Big(\sum_{\alpha'} \xi_{\alpha'}^{-1}\lambda_{\alpha'\beta}\Big)^{-1}$$

(the sum being over all α' such that $\lambda_{\alpha'\beta} \neq 0$). Further, the $\{\mathbf{Q}_\beta\}$ are mutually orthogonal and $\mathbf{T}^{\mathbf{V}} = \sum_\beta \mathbf{Q}_\beta$. We refer to Houtman & Speed (1983) for a fuller discussion.

For a single missing value the solution (4.1) may be written using (6.1) as

$$\tilde{\mathbf{y}}_2 = -\left[\frac{\sum_\alpha \xi_\alpha^{-1}\langle \mathbf{S}_\alpha\bar{\mathbf{T}}^{\mathbf{V}}\boldsymbol{\varepsilon}, \mathbf{S}_\alpha\bar{\mathbf{T}}^{\mathbf{V}}\mathbf{y}_1\rangle}{\sum_\alpha \xi_\alpha^{-1}|\mathbf{S}_\alpha\bar{\mathbf{T}}^{\mathbf{V}}\boldsymbol{\varepsilon}|^2}\right]\boldsymbol{\varepsilon}. \tag{6.3}$$

If the design is generally balanced and there are only two effective strata, the lowest, say α', and another, say α'', with eigenvalues $\lambda_{\alpha'\beta}$ and $\lambda_{\alpha''\beta}$ for treatment $\mathcal{T}_\beta$, and $\mathbf{Q}_{\alpha'\beta}$ and $\mathbf{Q}_{\alpha''\beta}$ as effects in those strata, we will write $\boldsymbol{\Delta}_\beta = \mathbf{Q}_{\alpha'\beta} - \mathbf{Q}_{\alpha''\beta}$ for their difference. Then

$$\tilde{\mathbf{y}}_2 = -\left[\frac{\sum_\alpha \xi_\alpha^{-1}\langle \mathbf{R}_\alpha\boldsymbol{\varepsilon}, \mathbf{R}_\alpha\mathbf{y}_1\rangle + \sum_\beta \mu_\beta\langle \boldsymbol{\Delta}_\beta\boldsymbol{\varepsilon}, \boldsymbol{\Delta}_\beta\mathbf{y}_1\rangle}{\sum_\alpha \xi_\alpha^{-1}|\mathbf{R}_\alpha\boldsymbol{\varepsilon}|^2 + \sum_\beta \mu_\beta|\boldsymbol{\Delta}_\beta\boldsymbol{\varepsilon}|^2}\right]\boldsymbol{\varepsilon}, \tag{6.4}$$

where $\mu_\beta = \xi_{\alpha'}^{-1}\xi_{\alpha''}^{-1}\lambda_{\alpha'\beta}\lambda_{\alpha''\beta}(\xi_{\alpha'}^{-1}\lambda_{\alpha'\beta} + \xi_{\alpha''}^{-1}\lambda_{\alpha''\beta})^{-1}$, and $\mathbf{R}_\alpha$ is the residual operator after fitting to $\mathcal{T}$ in stratum α.

If no treatment term is estimated in more than one stratum, then (6.4) simplifies to

$$\tilde{\mathbf{y}}_2 = -\left[\frac{\sum_\alpha \xi_\alpha^{-1}\langle \mathbf{R}_\alpha\boldsymbol{\varepsilon}, \mathbf{R}_\alpha\mathbf{y}_1\rangle}{\sum_\alpha \xi_\alpha^{-1}|\mathbf{R}_\alpha\boldsymbol{\varepsilon}|^2}\right]\boldsymbol{\varepsilon}. \tag{6.5}$$

This is the case for all orthogonal designs i.e. designs for which $\lambda_{\alpha\beta} = 0$ or 1. For example, the covariance operator of a complete randomized block design has spectral decomposition $\mathbf{V} = \xi_0\mathbf{G} + \xi_b(\mathbf{B} - \mathbf{G}) + \xi_w(\mathbf{I} - \mathbf{B})$ where $\mathbf{G}$ is the overall averaging operator (replacing all the components of $\mathbf{y}$ by the grand mean) and $\mathbf{B}$ is the block averaging operator (replacing all the components of $\mathbf{y}$ by the average of those components belonging to the same block). If there are b blocks and t treatments, a

single missing value is estimated using (6.5) by

$$\frac{\xi_w^{-1}(t\sum_T + b\sum_B - \sum_G) + \xi_b^{-1}(\sum_G - b\sum_B)}{\xi_w^{-1}(t-1)(b-1) + \xi_b^{-1}(b-1)}$$

where $\sum_G$, $\sum_B$ and $\sum_T$ denote respectively the sum of all the observations, the sum of the observations in the block containing the missing observation and the sum of the observations that received the same treatment as the missing unit.

In a $t \times t$ Latin square with one missing value, the replacement using (6.5) is given by

$$\frac{\xi_p^{-1}[t(\sum_R + \sum_C + \sum_T) - 2\sum_G] + \xi_r^{-1}(\sum_G - t\sum_R) + \xi_c^{-1}(\sum_G - t\sum_C)}{\xi_p^{-1}(t-1)(t-2) + \xi_r^{-1}(t-1) + \xi_c^{-1}(t-1)}.$$

Here $\mathbf{V} = \xi_0\mathbf{G} + \xi_r(\mathbf{R}-\mathbf{G}) + \xi_c(\mathbf{C}-\mathbf{G}) + \xi_p(\mathbf{I}-\mathbf{C}-\mathbf{R}+\mathbf{G})$ ($\mathbf{R}$ and $\mathbf{C}$ are the row and column averaging operators and $\mathbf{G}$ is as before), $\sum_R$ and $\sum_C$ are respectively the sum of the observations in the same row and same column as the missing observation and $\sum_G$ and $\sum_T$ are as before. The case of a split plot design will be discussed in §8.

In practice the operator $\mathbf{V}$ is only partially known: the projectors $\{\mathbf{S}_\alpha\}$ are determined by the structure of the design whilst the strata variances $\{\xi_\alpha\}$ need to be estimated, and we outline a method for doing so in the next paragraph. The $\{\xi_\alpha\}$ are not needed however if we assume that most of the information on treatments is concentrated in the lowest stratum in which they are estimable. If attention is restricted to that stratum only, we let $\xi_\alpha^{-1} \approx 0$ in all the other strata and write $\mathbf{Ry}$ for the vector of residuals in that stratum, then

$$\tilde{\mathbf{y}}_2 = -\left[\frac{\langle \mathbf{Ry}_1, \mathbf{R}\boldsymbol{\varepsilon}\rangle}{|\mathbf{R}\boldsymbol{\varepsilon}|^2}\right]\boldsymbol{\varepsilon}.$$

This gives all the usual missing value estimates, see, for example, Cochran & Cox (1957).

7. Estimation of Strata Variances

When information on some or all treatment terms is available from more than one stratum, we saw in the previous paragraph that the missing value estimators involve the strata variances $\{\xi_\alpha\}$. We now outline a method for estimating the $\{\xi_\alpha\}$ in a generally balanced design. The method we propose is an extension of Nelder's way of handling the problem for a complete design. The main steps are as follows:

(i) complete the data with initial estimates $\mathbf{y}_2^{(0)}$ computed using lowest stratum information only;

(ii) calculate estimates $\{\xi_\alpha^{(0)}\}$ using the data completed with $\mathbf{y}_2^{(0)}$;

(iii) calculate $\mathbf{y}_2^{(1)}$ using one of the methods indicated in §§4 or 6;

(iv) calculate new estimates $\{\xi_\alpha^{(1)}\}$ and then go back to (iii), continuing as often as seems necessary.

Estimates of the $\{\xi_\alpha\}$ in (ii) and (iv) may be obtained in two ways. If, after estimating treatments, there are enough degrees of freedom left in stratum α, equating the error mean square in that stratum to its expectation easily provides an estimator of ξ_α. Indeed we have under (6.2)

$$\begin{aligned} \mathbb{E}\,|\mathbf{R}_\alpha \mathbf{y}|^2 &= \xi_\alpha \text{ trace } \mathbf{R}_\alpha \\ &= \xi_\alpha\Big(\text{trace } \mathbf{S}_\alpha - \sum_\beta \text{trace } \mathbf{T}_\beta\Big) \\ &= \xi_\alpha d_\alpha \end{aligned} \tag{7.1}$$

where the sum is over all β such that $\lambda_{\alpha\beta} > 0$. But d_α is often small and may even be zero, and so we would rather use the *actual* residual in stratum α given by $\mathbf{S}_\alpha \bar{\mathbf{T}}^{\mathbf{V}}\mathbf{y}$. If (6.2) holds we have

$$\begin{aligned} \mathbb{E}\,|\mathbf{S}_\alpha \bar{\mathbf{T}}^{\mathbf{V}}\mathbf{y}|^2 &= \xi_\alpha \text{ trace } \mathbf{S}_\alpha \bar{\mathbf{T}}^{\mathbf{V}} \\ &= \xi_\alpha\Big\{d_\alpha + \sum_\beta (1 - w_{\alpha\beta}) \text{ trace } \mathbf{T}_\beta\Big\} \\ &= \xi_\alpha d'_\alpha \end{aligned} \tag{7.2}$$

where the sum is again over all βs such that $\lambda_{\alpha\beta} > 0$. The "degrees of freedom" d'_α is larger than d_α in (7.1), but, like $|\mathbf{S}_\alpha \bar{\mathbf{T}}^{\mathbf{V}}\mathbf{y}|^2$, it involves the unknown $\{\xi_\alpha\}$ through the $w_{\alpha\beta}$. Nelder suggested an iterative method, choosing initial values $\{\xi_\alpha^{(0)}\}$, for example from (7.1) in the lowest stratum, obtaining $|\mathbf{S}_\alpha \bar{\mathbf{T}}^{\mathbf{V}(0)}\mathbf{y}|^2$ and $d_\alpha'^{(0)}$, and then revised estimators

$$\xi_\alpha^{(1)} = \frac{|\mathbf{S}_\alpha \bar{\mathbf{T}}^{\mathbf{V}(0)}\mathbf{y}|^2}{d_\alpha'^{(0)}}.$$

Again this needs to be continued as often as seems necessary.

In our experience the estimates of $\{\xi_\alpha\}$ do not change very much, and unless there is a lot of missing data, one would not expect them to. No result guaranteeing convergence is available even with complete data.

8. Examples

(i) *Split-plot:* Let us consider a general split-plot experiment with r replications of a plots (levels of A) each of b subplots (levels of B). The block structure determines the spectral form of $\mathbb{D}\mathbf{y} = \mathbf{V}$ as

$$\mathbf{V} = \xi_0\mathbf{G} + \xi_r(\mathbf{R} - \mathbf{G}) + \xi_p(\mathbf{P} - \mathbf{R}) + \xi_s(\mathbf{I} - \mathbf{P}) \tag{8.1}$$

where $\mathbf{G}, \mathbf{R}$ and $\mathbf{P}$ are respectively the overall averaging operator, the operator averaging over replications, and the operator averaging over plots. The factorial treatment structure determines the decomposition of the treatment space

$$\mathcal{T} = \mathcal{G} \oplus (\mathcal{A} \ominus \mathcal{G}) \oplus (\mathcal{B} \ominus \mathcal{G}) \oplus (\mathcal{T} \ominus (\mathcal{A} + \mathcal{B})) \tag{8.2}$$

corresponding to a decomposition of the vector of means

$$\mathbb{E}\mathbf{y} = \tau = \mathbf{G}\tau + (\mathbf{A} - \mathbf{G})\tau + (\mathbf{B} - \mathbf{G})\tau + \mathbf{T}_{\mathbf{A.B}}\tau$$

for all $\tau \in \mathcal{T}$. The operator $\mathbf{G}$ is the same as above, $\mathbf{A}$ and $\mathbf{B}$ average respectively over the levels of A and over the levels of B, and $\mathbf{T}_{\mathbf{A.B}} = \mathbf{T} - \mathbf{A} - \mathbf{B} + \mathbf{G}$ where $\mathbf{T}$ is the treatment averaging operator. We note that $\mathbf{G}$, $\mathbf{A}$, $\mathbf{B}$, $\mathbf{T}$ and $\mathbf{T}_{\mathbf{A.B}}$ are orthogonal projectors with respective ranges $\mathcal{G}$, $\mathcal{A}$, $\mathcal{B}$, $\mathcal{T}$ and $\mathcal{T} \ominus (\mathcal{A} + \mathcal{B})$.

The experiment is generally balanced with respect to the treatment structure (8.2), with a set of eigenvalues all equal to zero or one, this always being the case for orthogonal designs. All the information on contrasts comparing levels of A (contrasts in $\mathcal{A} \ominus \mathcal{G}$) is contained in the main plot means adjusted by their replicate means $(\mathbf{P} - \mathbf{R})\mathbf{y}$, whilst all the information on contrasts comparing levels of B (contrasts in $\mathcal{B} \ominus \mathcal{G}$) and on those describing interaction between A and B (contrasts in $\mathcal{T} \ominus (\mathcal{A} + \mathcal{B})$) is contained in the stratum of subplot comparisons $(\mathbf{I} - \mathbf{P})\mathbf{y}$. And so a single missing value can be estimated using (6.5) by x where x is

$$\frac{\xi_s^{-1}(ra\sum_P + ab\sum_{AB} - a\sum_A) + \xi_p^{-1}(r\sum_R + a\sum_A - ra\sum_P - \sum_G) + \xi_r^{-1}(\sum_G - r\sum_R)}{\xi_s^{-1}[a(r-1)(b-1)] + \xi_p^{-1}[(r-1)(a-1)] + \xi_r^{-1}(r-1)}. \tag{8.3}$$

Here $\sum_P$ is the total of the observations in the plot containing the missing observation,

$\sum_R$ is the sum of all the observations in the replicate containing the missing observation,

$\sum_{AB}$ is the total of all the subplots that received the same treatment combination as the missing unit,

$\sum_A$ is the total of all the subplots that received the same level of treatment A as the missing one,

$\sum_G$ is the sum of all the observations.

This formula may be compared with the estimate obtained by Anderson (1946) by minimizing the subplot error only:

$$x' = \frac{r\sum_P + b\sum_{AB} - \sum_A}{(r-1)(b-1)}. \tag{8.4}$$

If ξ_p and ξ_r are both very large in comparison with ξ_s, then (8.3) reduces to (8.4). In the example treated in his paper, Anderson obtains a replacement of 763 whereas (8.3) gives 726 (using (7.1) to estimate

the strata variances); the latter value is in better harmony with the rest of the data whose mean was 492.

The correction to the variance var $(\langle \mathbf{t}, \bar{\tau} \rangle)$ where $\bar{\tau}$ satisfies (3.4), will be zero whenever the missing value is at another treatment level than the levels compared in the contrast t. When this correction is not zero we will follow Cochran & Cox (1957, p. 303) and consider four simple kinds of contrasts. Let us write

$\mathbf{t}_A$ (resp. $\mathbf{t}_B$) = a difference between two A means (resp. B means),

$\mathbf{t}_{Ba}$ = a difference between two B means at the same level of A,

$\mathbf{t}_{Ab}$ = a difference between two A means at the same level of B, or at different levels of B.

With the denominator of the correction given in (5.3) equal to $|\bar{\mathbf{T}}^{\mathbf{V}}\boldsymbol{\varepsilon}|_{\mathbf{V}}^2 = (r-1)(rab)^{-1}[\xi_r^{-1} + \xi_p^{-1}(a-1) + \xi_s^{-1}(b-1)a] = d$, we have

$$\begin{aligned}
\text{var}\,(\langle \mathbf{t}_A, \bar{\tau} \rangle) &= 2\xi_p(rb)^{-1} + (r^2b^2d)^{-1},\\
\text{var}\,(\langle \mathbf{t}_B, \bar{\tau} \rangle) &= 2\xi_s(ra)^{-1} + (r^2a^2d)^{-1},\\
\text{var}\,(\langle \mathbf{t}_{Ba}, \bar{\tau} \rangle) &= 2\xi_s r^{-1} + (r^2d)^{-1},\\
\text{var}\,(\langle \mathbf{t}_{Ab}, \bar{\tau} \rangle) &= [2\xi_p(rb)^{-1} + 2\xi_{ss}(b-1)(rb)^{-1}] + (r^2d)^{-1}.
\end{aligned}$$

Our corrections (second terms) reduce to the ones obtained by Cochran & Cox (1957, p. 303) and based on a lowest stratum estimate of the missing value by setting ξ_r^{-1} and ξ_p^{-1} equal to zero.

If we now consider the submodel having no AB-interaction term, namely $\mathbb{E}\mathbf{y} \in \mathscr{A} + \mathscr{B}$, the apparent sum of squares due to the interaction, $|\mathbf{T}_{\mathbf{A}.\mathbf{B}}^{\mathbf{V}}\mathbf{y}(\mathscr{T})|_{\mathbf{V}}^2 = \xi_s^{-1}\,|(\mathbf{T} - \mathbf{A} - \mathbf{B} + \mathbf{G})\mathbf{y}(\mathscr{T})|^2$ must be adjusted. This is done by subtracting from it the correction term

$$|\bar{\mathbf{P}}_{\mathscr{A}+\mathscr{B}}^{\mathbf{V}}(\mathbf{y}(\mathscr{T}) - \mathbf{y}(\mathscr{A} + \mathscr{B}))|_{\mathbf{V}}^2 = (d-f)^2\,|\bar{\mathbf{P}}_{\mathscr{A}+\mathscr{B}}^{\mathbf{V}}\boldsymbol{\varepsilon}|_{\mathbf{V}}^2$$

where d is the replacement (8.3) under the full model, and f is the replacement under the submodel given by f where f is

$$\frac{\xi_s^{-1}(ra\sum_P + b\sum_B - \sum_G) + \xi_P^{-1}(r\sum_R + a\sum_A - ar\sum_P - \sum_G) + \xi_r^{-1}(\sum_G - r\sum_R)}{\xi_s^{-1}(ra-1)(b-1) + \xi_p^{-1}(a-1)(r-1) + \xi_r^{-1}(r-1)}. \tag{8.5}$$

The notation is as in (8.3) with $\sum_B$ denoting the sum of the observations that received the same level of treatment B as the missing one, and finally

$$|\bar{\mathbf{P}}_{\mathscr{A}+\mathscr{B}}^{\mathbf{V}}\boldsymbol{\varepsilon}|_{\mathbf{V}}^2 = (abr)^{-1}[\xi_s^{-1}(ra-1)(b-1) + \xi_p^{-1}(a-1)(r-1) + \xi_r^{-1}(r-1)].$$

(ii) *Simple (square) lattice:* We consider an experiment performed to compare k^2 treatments in two replicates of k blocks of k plots each. As in a split-plot experiment, the block structure here is doubly nested and so there are three strata (other than the grand

mean): between replicates, within replicates between blocks, and within blocks. This defines the spectral decomposition of the dispersion matrix $\mathbb{D}\mathbf{y} = \mathbf{V}$ as

$$\mathbf{V} = \xi_0\mathbf{G} + \xi_r(\mathbf{R}-\mathbf{G}) + \xi_b(\mathbf{B}-\mathbf{G}) + \xi_p(\mathbf{I}-\mathbf{B}) \tag{8.6}$$

where $\mathbf{G}$, $\mathbf{B}$ and $\mathbf{R}$ are respectively the overall, block and replicate averaging operators. This design does not satisfy the conditions (6.2) of general balance with respect to the natural treatment decomposition

$$\mathbf{T} = \mathbf{G} + (\mathbf{T}-\mathbf{G})$$

corresponding to "no structure" on treatments but general balance is obtained by introducing a factorial "pseudo-structure". This is determined by the following scheme: the treatments are arranged in a $k \times k$ square and treatments belonging to the same row (resp. column) of the square are allocated to the same block in the first (resp. second) replicate. Let us use M and N for the pseudo-factors corresponding to the rows and columns of the treatment array. The experiment is generally balanced with respect to the treatment decomposition

$$\tau = \mathbf{G}\tau + (\mathbf{M}-\mathbf{G})\tau + (\mathbf{N}-\mathbf{G})\tau + \mathbf{T}_{M.N}\tau \tag{8.7}$$

where $\tau = \mathbb{E}\mathbf{y} \in \mathcal{T}$. The notation here is as in (i) with $\mathbf{M}$ and $\mathbf{N}$ instead of $\mathbf{A}$ and $\mathbf{B}$. The relationships (6.2) are

$$\begin{aligned}
\mathbf{T}(\mathbf{R}-\mathbf{G})\mathbf{T} &= 0,\\
\mathbf{T}(\mathbf{B}-\mathbf{R})\mathbf{T} &= \tfrac{1}{2}(\mathbf{M}-\mathbf{G}) + \tfrac{1}{2}(\mathbf{N}-\mathbf{G}),\\
\mathbf{T}(\mathbf{I}-\mathbf{B})\mathbf{T} &= \tfrac{1}{2}(\mathbf{M}-\mathbf{G}) + \tfrac{1}{2}(\mathbf{N}-\mathbf{G}) + \mathbf{T}_{M.N};
\end{aligned}$$

and so the effects are

$$\begin{aligned}
\mathbf{Q}_{bM} &= (\tfrac{1}{2})^{-1}\mathbf{M}(\mathbf{B}-\mathbf{R}), & \mathbf{Q}_{pM} &= (\tfrac{1}{2})^{-1}\mathbf{M}(\mathbf{I}-\mathbf{B}),\\
\mathbf{Q}_{bN} &= (\tfrac{1}{2})^{-1}\mathbf{N}(\mathbf{B}-\mathbf{R}), & \mathbf{Q}_{pN} &= (\tfrac{1}{2})^{-1}\mathbf{N}(\mathbf{I}-\mathbf{B}),\\
\mathbf{Q}_{bM.N} &= \mathbf{0}, & \mathbf{Q}_{pM.N} &= \mathbf{T}_{M.N}(\mathbf{I}-\mathbf{B}) = \mathbf{T}_{M.N},
\end{aligned}$$

and the weights are

$$\begin{aligned}
w_{bM} = w_{bN} = w &= \tfrac{1}{2}\xi_b^{-1}(\tfrac{1}{2}\xi_b^{-1} + \tfrac{1}{2}\xi_p^{-1})^{-1} = \xi_b^{-1}(\xi_b^{-1} + \xi_p^{-1})^{-1},\\
w_{rM} = w_{rN} &= 1 - w = \xi_p^{-1}(\xi_b^{-1} + \xi_p^{-1})^{-1}.\\
w_{bM.N} = 0, \quad w_{pM.M} &= 1.
\end{aligned}$$

Assuming known strata variances $\{\xi_\alpha\}$, a single missing value may be estimated using (6.4) by x where x is

$$\frac{\xi_p^{-1}(k^2\Sigma_T - 2\Sigma_R + \Sigma_G - 2kC - kC') - 2(\xi_b + \xi_p)^{-1}(2\Sigma_G - 4\Sigma_R - 2kC + kC')}{\xi_p^{-1}(k-1)^2 + 4(\xi_b + \xi_p)^{-1}(k-1)} \tag{8.8}$$

where Σ_G, Σ_R and Σ_T denote respectively the sum of all the observa-

tions, the total of the observations in the replicate containing the missing observation and the total of the observations that received the same treatment as the missing unit. We have borrowed from Cochran & Cox (1957) the notation

C = [total (over replicates) of all treatments in the block to which the missing unit belongs] $-2\sum_B$,

where $\sum_B$ is the sum of all the observations in the block containing the missing unit, and

C' = sum of the C values for all blocks containing the treatment that was allocated to the missing unit.

Since there is no natural submodel of the treatment model assumed, corrections to sums of squares due to treatments will not usually be needed. And so we only compute the adjustments to be added to the variances of elementary contrasts (contrasts between pairs of treatments). For a complete experiment, the usual formulae for those variances are

$$\frac{2}{k}\left\{\frac{1}{\xi_b^{-1}+\xi_p^{-1}}+\frac{k-1}{2\xi_p^{-1}}\right\} \tag{8.9}$$

if the two treatments belong to the same block (in either replicate) and

$$\frac{2}{k}\left\{\frac{2}{\xi_b^{-1}+\xi_p^{-1}}+\frac{k-2}{2\xi_p^{-1}}\right\} \tag{8.10}$$

if the two treatments never appear in the same block. The correction to those variances, due to the estimation of a single missing value is given in (5.3) by $(\langle \mathbf{t}, \mathbf{T}^{\mathbf{V}}\boldsymbol{\varepsilon}\rangle)^2(|\bar{\mathbf{T}}^{\mathbf{V}}\boldsymbol{\varepsilon}|_{\mathbf{V}}^2)^{-1}$ with

$$\begin{aligned} D = |\bar{\mathbf{T}}^{\mathbf{V}}\boldsymbol{\varepsilon}|_{\mathbf{V}}^2 = \frac{1}{2k^4}[&4\xi_0^{-1}(k-1)^2+\xi_r^{-1}k^2 \\ &+\xi_b^{-1}\{4k^2(k-1)w^2+(k-1)(k-2)^2\} \\ &+\xi_p^{-1}\{4k^2(k-1)\bar{w}^2(k-1)k^3\}] \end{aligned} \tag{8.11}$$

where $\bar{w} = 1 - w$. Using D for the denominator, the correction to (8.9) is

$$\frac{1}{4k^2D}(k-2\bar{w})^2$$

for a comparison between the treatment allocated to the missing unit, say τ_m, and a treatment in the same block as the missing unit, and

$$\frac{1}{4k^2D}(k-2w)^2$$

for a comparison between τ_m and a treatment appearing in the same

block as τ_m in the replicate that does not contain the missing unit. The correction to (8.10) is

$$\frac{1}{4D}$$

for a comparison between τ_m and a treatment never appearing in the same block as τ_m; it is

$$\frac{1}{k^2 D}(1-2\bar{w})^2$$

for a comparison between a treatment appearing in the same block as τ_m in the same replicate as the missing unit and such a treatment in the other replicate; finally, it is

$$\frac{1}{D}\left(\frac{\bar{w}}{k}\right)^2 \left(\text{resp. } \frac{1}{D}\left(\frac{w}{k}\right)^2\right)$$

for a comparison between a treatment appearing in the same block as τ_m in the same replicate as the missing unit (resp. in the replicate that does not contain the missing unit) and a treatment never appearing in the same block as τ_m. The corrections in all other cases are zero.

9. Appendix

(*i*) *Proof of* (3.5)

Step 1. We define the vector of residuals

$$\tilde{\mathbf{r}} = \bar{\mathbf{T}}^{\mathbf{V}}(\mathbf{y}_1 + \tilde{\mathbf{y}}_2) \tag{A1}$$

where $\tilde{\mathbf{y}}_2$ satisfies the missing values equations (3.3), and first prove that

$$\tilde{\mathbf{r}} = (\mathbf{I} - \mathbf{P}^{\mathbf{V}}_{\mathscr{D}_2+\mathscr{T}})\bar{\mathbf{T}}^{\mathbf{V}}\mathbf{y}_1 \tag{A2}$$

where $\mathbf{P}^{\mathbf{V}}_{\mathscr{D}_2+\mathscr{T}}$ is the projection onto the space $\mathscr{D}_2+\mathscr{T} = \{\mathbf{y}_2+\mathbf{t} \mid \mathbf{y}_2 \in \mathscr{D}_2, \mathbf{t} \in \mathscr{T}\}$, orthogonal w.r.t. $\langle .\,, .\rangle_{\mathbf{V}}$.
Using recursion (iv) of Section 4, we have

$$\begin{aligned}\tilde{\mathbf{r}} &= \lim_{n\to\infty} (\bar{\mathbf{T}}^{\mathbf{V}}\bar{\mathbf{D}}_2^{\mathbf{V}})^n \bar{\mathbf{T}}^{\mathbf{V}}\mathbf{y}_1 \\ &= \mathbf{P}^{\mathbf{V}}_{\mathbf{V}\mathscr{T}\text{-}\cap \mathbf{V}\mathscr{D}_1}\bar{\mathbf{T}}^{\mathbf{V}}\mathbf{y}_1 \qquad \text{Von Neumann (1950, p. 55)} \\ &= (\mathbf{I} - \mathbf{P}^{\mathbf{V}}_{\mathscr{T}+\mathscr{D}_2})\bar{\mathbf{T}}^{\mathbf{V}}\mathbf{y}_1.\end{aligned}$$

This last relationship and the previous one both use the fact that the orthogonal complements w.r.t. $\langle .\,, .\rangle_{\mathbf{V}}$ of $\mathscr{T}$ and $\mathscr{D}_2$ are respectively $\mathbf{V}\mathscr{T}^{\perp}$ and $\mathbf{V}\mathscr{D}_1$.

Step 2.

$$\begin{aligned}
\mathbf{D}_1\tilde{\boldsymbol{\tau}} &= \mathbf{D}_1\mathbf{T}^{\mathbf{V}}(\mathbf{y}_1+\tilde{\mathbf{y}}_2) && \text{where } \tilde{\mathbf{y}}_2 \text{ satisfies (3.3)}\\
&= \mathbf{D}_1(\mathbf{I}-\bar{\mathbf{T}}^{\mathbf{V}})(\mathbf{y}_1+\tilde{\mathbf{y}}_2)\\
&= \mathbf{y}_1-\mathbf{D}_1\tilde{\mathbf{r}} && \text{using (A1)}\\
&= \mathbf{y}_1-\mathbf{D}_1(\mathbf{I}-\mathbf{P}^{\mathbf{V}}_{\mathscr{T}+\mathscr{D}_2})\bar{\mathbf{T}}^{\mathbf{V}}\mathbf{y}_1 && \text{using (A2)}\\
&= \mathbf{y}_1-\mathbf{D}_1\mathbf{y}_1+\mathbf{D}_1\mathbf{T}^{\mathbf{V}}\mathbf{y}_1-\mathbf{D}_1\mathbf{P}^{\mathbf{V}}_{\mathscr{T}+\mathscr{D}_2}\mathbf{T}^{\mathbf{V}}\mathbf{y}_1+\mathbf{D}_1\mathbf{P}^{\mathbf{V}}_{\mathscr{D}_2+\mathscr{T}}\mathbf{y}_1\\
&= \mathbf{D}_1\mathbf{P}^{\mathbf{V}}_{\mathscr{D}_2+\mathscr{T}}\mathbf{y}_1. && \text{since } \mathbf{P}^{\mathbf{V}}_{\mathscr{T}+\mathscr{D}_2}\mathbf{T}^{\mathbf{V}}=\mathbf{T}^{\mathbf{V}}.
\end{aligned}$$

Step 3. We first observe that

$$\mathscr{D}_2+\mathscr{T}=\mathscr{D}_2+\mathscr{T}_1=\mathscr{D}_2\oplus\bar{\mathbf{D}}_2^{\mathbf{V}}\mathscr{T}_1$$

where the last sum is orthogonal with respect to $\langle .\,,.\rangle_{\mathbf{V}}$. Thus the weighted projection onto $\mathscr{D}_2+\mathscr{T}$ decomposes into the sum of the weighted projections onto $\mathscr{D}_2$ and $\mathscr{A}=\bar{\mathbf{D}}_2^{\mathbf{V}}\mathscr{T}_1$ respectively. We then have

$$\begin{aligned}
\mathbf{D}_1\tilde{\boldsymbol{\tau}} &= \mathbf{D}_1\mathbf{D}_2^{\mathbf{V}}\mathbf{y}_1+\mathbf{D}_1\mathbf{A}^{\mathbf{V}}\mathbf{y}_1\\
&= \mathbf{D}_1\mathbf{A}^{\mathbf{V}}\mathbf{y}_1.
\end{aligned}$$

Now $\mathscr{R}(\mathbf{A}^{\mathbf{V}})=\mathscr{R}(\mathbf{V}(\mathbf{D}_1\mathbf{V}\mathbf{D}_1)^{-}\mathbf{T}_1)$, and so

$$\begin{aligned}
\mathbf{D}_1\tilde{\boldsymbol{\tau}} &= \mathbf{D}_1\mathbf{V}(\mathbf{D}_1\mathbf{V}\mathbf{D}_1)^{-}\mathbf{T}_1[\mathbf{T}_1(\mathbf{D}_1\mathbf{V}\mathbf{D}_1)^{-}\mathbf{V}\mathbf{V}^{-1}\\
&\quad\times\mathbf{V}(\mathbf{D}_1\mathbf{V}\mathbf{D}_1)^{-}\mathbf{T}_1]^{-}\mathbf{T}_1(\mathbf{D}_1\mathbf{V}\mathbf{D}_1)^{-}\mathbf{V}\mathbf{V}^{-1}\mathbf{y}_1\\
&= \mathbf{T}_1[\mathbf{T}_1(\mathbf{D}_1\mathbf{V}\mathbf{D}_1)^{-}\mathbf{T}_1]^{-}\mathbf{T}_1(\mathbf{D}_1\mathbf{V}\mathbf{D}_1)^{-}\mathbf{y}_1\\
&= \mathbf{T}_1^{\mathbf{V}_1}\mathbf{y}_1 \qquad (\text{where } \mathbf{V}_1=\mathbf{D}_1\mathbf{V}\mathbf{D}_1)\\
&= \hat{\boldsymbol{\tau}}_1
\end{aligned}$$

which proves (3.5).

(*ii*) *Verification of* (5.1)

$$\begin{aligned}
&|\bar{\mathbf{U}}^{\mathbf{V}}\mathbf{y}(\mathscr{T})|^2_{\mathbf{V}}-|\bar{\mathbf{U}}^{\mathbf{V}}\mathbf{y}(\mathscr{U})|^2_{\mathbf{V}}\\
&= \langle\bar{\mathbf{U}}^{\mathbf{V}}\mathbf{y}(\mathscr{T}),\bar{\mathbf{U}}^{\mathbf{V}}\mathbf{y}(\mathscr{T})\rangle_{\mathbf{V}}-\langle\bar{\mathbf{U}}^{\mathbf{V}}\mathbf{y}(\mathscr{U}),\bar{\mathbf{U}}^{\mathbf{V}}\mathbf{y}(\mathscr{U})\rangle_{\mathbf{V}}\\
&= \langle\mathbf{y}(\mathscr{T}),\bar{\mathbf{U}}^{\mathbf{V}}\mathbf{y}(\mathscr{T})\rangle_{\mathbf{V}}-\langle\mathbf{y}(\mathscr{U}),\bar{\mathbf{U}}^{\mathbf{V}}\mathbf{y}(\mathscr{U})\rangle_{\mathbf{V}}\\
&\quad+\langle\mathbf{y}(\mathscr{T}),\bar{\mathbf{U}}^{\mathbf{V}}\mathbf{y}(\mathscr{U})\rangle_{\mathbf{V}}-\langle\mathbf{y}(\mathscr{T}),\bar{\mathbf{U}}^{\mathbf{V}}\mathbf{y}(\mathscr{U})\rangle_{\mathbf{V}}\\
&= \langle\mathbf{y}(\mathscr{T}),\bar{\mathbf{U}}^{\mathbf{V}}[\mathbf{y}(\mathscr{T})-\mathbf{y}(\mathscr{U})]\rangle_{\mathbf{V}}+\langle\mathbf{y}(\mathscr{T})-\mathbf{y}(\mathscr{U}),\bar{\mathbf{U}}^{\mathbf{V}}\mathbf{y}(\mathscr{U})\rangle_{\mathbf{V}}.
\end{aligned}$$

Now, $\langle\mathbf{y}(\mathscr{T})-\mathbf{y}(\mathscr{U}),\bar{\mathbf{U}}^{\mathbf{V}}\mathbf{y}(\mathscr{U})\rangle_{\mathbf{V}}=\langle\mathbf{y}_2(\mathscr{T})-\mathbf{y}_2(\mathscr{U}),\bar{\mathbf{U}}^{\mathbf{V}}\mathbf{y}(\mathscr{U})\rangle_{\mathbf{V}}$

$$\begin{aligned}
&= \langle\mathbf{D}_2^{\mathbf{V}}[\mathbf{y}_2(\mathscr{T})-\mathbf{y}_2(\mathscr{U})],\bar{\mathbf{U}}^{\mathbf{V}}\mathbf{y}(\mathscr{U})\rangle_{\mathbf{V}}\\
&= \langle\mathbf{y}_2(\mathscr{T})-\mathbf{y}_2(\mathscr{U}),\mathbf{D}_2^{\mathbf{V}}\bar{\mathbf{U}}^{\mathbf{V}}\mathbf{y}(\mathscr{U})\rangle_{\mathbf{V}},
\end{aligned}$$

and, using (3.6) for the model $\mathscr{U}$, we have $\mathbf{D}_2^{\mathbf{V}}\bar{\mathbf{U}}^{\mathbf{V}}\mathbf{y}(\mathscr{U})=0$. Thus we

may write

$$
\begin{aligned}
&|\bar{\mathbf{U}}^{\mathbf{V}}(\mathcal{T})|^2_{\mathbf{V}} - |\bar{\mathbf{U}}^{\mathbf{V}}\mathbf{y}(\mathcal{U})|^2_{\mathbf{V}} \\
&= \langle \mathbf{y}(\mathcal{T}), \bar{\mathbf{U}}^{\mathbf{V}}[\mathbf{y}(\mathcal{T}) - \mathbf{y}(\mathcal{U})]\rangle_{\mathbf{V}} - \langle \mathbf{y}(\mathcal{T}) - \mathbf{y}(\mathcal{U}), \bar{\mathbf{U}}^{\mathbf{V}}\mathbf{y}(\mathcal{U})\rangle_{\mathbf{V}} \\
&= \langle \mathbf{y}(\mathcal{T}), \bar{\mathbf{U}}^{\mathbf{V}}[\mathbf{y}_2(\mathcal{T}) - \mathbf{y}_2(\mathcal{U})]\rangle_{\mathbf{V}} - \langle \mathbf{y}(\mathcal{U}), \bar{\mathbf{U}}^{\mathbf{V}}[\mathbf{y}_2(\mathcal{T}) - \mathbf{y}_2(\mathcal{U})]\rangle_{\mathbf{V}} \\
&= \langle \mathbf{y}_2(\mathcal{T}) - \mathbf{y}_2(\mathcal{U}), \bar{\mathbf{U}}^{\mathbf{V}}[\mathbf{y}_2(\mathcal{T}) - \mathbf{y}_2(\mathcal{U})]\rangle_{\mathbf{V}} \\
&= |\bar{\mathbf{U}}^{\mathbf{V}}[\mathbf{y}_2(\mathcal{T}) - \mathbf{y}_2(\mathcal{U})]|^2_{\mathbf{V}}.
\end{aligned}
$$

(*iii*) *Verifications of* (5.2) *and* (5.3)

We start from equation (3.4) giving the vector of fitted values $\tilde{\boldsymbol{\tau}}$ when the data are incomplete, writing the equation in the form

$$(\mathbf{I} - \mathbf{T}^{\mathbf{V}}\mathbf{D}_2^{\mathbf{V}}\mathbf{T}^{\mathbf{V}})\tilde{\boldsymbol{\tau}} = \mathbf{T}^{\mathbf{V}}\bar{\mathbf{D}}_2^{\mathbf{V}}\mathbf{y}_1. \tag{A3}$$

We will only consider the case where $\mathcal{T} \cap \mathcal{D}_2 = \{0\}$ so that all the eigenvalues of $\mathbf{T}^{\mathbf{V}}\mathbf{D}_2^{\mathbf{V}}\mathbf{T}^{\mathbf{V}}$ are strictly smaller than one and hence all those of $\mathbf{I} - \mathbf{T}^{\mathbf{V}}\mathbf{D}_2^{\mathbf{V}}\mathbf{T}^{\mathbf{V}} = \mathbf{A}$ are strictly positive and the operator $\mathbf{A}$ is invertible. The unique solution of (A3) is thus

$$\tilde{\boldsymbol{\tau}} = \mathbf{A}^{-1}\mathbf{T}^{\mathbf{V}}\bar{\mathbf{D}}_2^{\mathbf{V}}\mathbf{y}_1$$

and so we have

$$
\begin{aligned}
\operatorname{var}\tilde{\boldsymbol{\tau}} &= \mathbf{A}^{-1}\mathbf{T}^{\mathbf{V}}\bar{\mathbf{D}}_2^{\mathbf{V}}\mathbf{D}_1\mathbf{V}\mathbf{D}_1(\bar{\mathbf{D}}_2^{\mathbf{V}})^*(\mathbf{T}^{\mathbf{V}})^*[\mathbf{A}^{-1}]^* \\
&= \mathbf{A}^{-1}\mathbf{T}^{\mathbf{V}}\bar{\mathbf{D}}_2^{\mathbf{V}}\mathbf{T}^{\mathbf{V}}\mathbf{A}^{-1}\mathbf{V} \\
&= \mathbf{A}^{-1}\mathbf{A}\mathbf{T}^{\mathbf{V}}\mathbf{A}^{-1}\mathbf{V} \\
&= \mathbf{T}^{\mathbf{V}}(\mathbf{I} - \mathbf{T}^{\mathbf{V}}\mathbf{D}_2^{\mathbf{V}}\mathbf{T}^{\mathbf{V}})^{-1}\mathbf{V}.
\end{aligned}
$$

We may re-express this as follows

$$
\begin{aligned}
\operatorname{var}\tilde{\boldsymbol{\tau}} &= \mathbf{T}^{\mathbf{V}}\left[\sum_{i=0}^{\infty}(\mathbf{T}^{\mathbf{V}}\mathbf{D}_2^{\mathbf{V}}\mathbf{T}^{\mathbf{V}})^i\right]\mathbf{V} \\
&= \mathbf{T}^{\mathbf{V}}\mathbf{V} + \sum_{i=1}^{\infty}(\mathbf{T}^{\mathbf{V}}\mathbf{D}_2^{\mathbf{V}}\mathbf{T}^{\mathbf{V}})^i\mathbf{V} \\
&= \mathbf{T}^{\mathbf{V}}\mathbf{V} + \mathbf{T}^{\mathbf{V}}\mathbf{D}_2^{\mathbf{V}}\mathbf{T}^{\mathbf{V}}\left[\sum_{i=0}^{\infty}(\mathbf{T}^{\mathbf{V}}\mathbf{D}_2^{\mathbf{V}}\mathbf{T}^{\mathbf{V}})^i\right]\mathbf{V} \\
&= \mathbf{T}^{\mathbf{V}}\mathbf{V} + \mathbf{T}^{\mathbf{V}}\left[\sum_{i=0}^{\infty}(\mathbf{D}_2^{\mathbf{V}}\mathbf{T}^{\mathbf{V}}\mathbf{D}_2^{\mathbf{V}})^i\right]\mathbf{D}_2^{\mathbf{V}}\mathbf{T}^{\mathbf{V}}\mathbf{V} \\
&= \mathbf{T}^{\mathbf{V}}\mathbf{V} + \mathbf{T}^{\mathbf{V}}(\mathbf{D}_2^{\mathbf{V}}\bar{\mathbf{T}}^{\mathbf{V}}\mathbf{D}_2^{\mathbf{V}})^{-}\mathbf{D}_2^{\mathbf{V}}\mathbf{T}^{\mathbf{V}}\mathbf{V}\mathbf{u}.
\end{aligned}
$$

From this we have for all $\mathbf{t}, \mathbf{u} \in \mathcal{T}$

$$\operatorname{cov}(\langle \mathbf{t}, \tilde{\boldsymbol{\tau}}\rangle, \langle \mathbf{u}, \tilde{\boldsymbol{\tau}}\rangle) = \langle \mathbf{t}, \mathbf{T}^{\mathbf{V}}\mathbf{V}\mathbf{u}\rangle + \langle \mathbf{t}, \mathbf{T}^{\mathbf{V}}(\mathbf{D}_2^{\mathbf{V}}\bar{\mathbf{T}}^{\mathbf{V}}\mathbf{D}_2^{\mathbf{V}})^{-}\mathbf{D}_2^{\mathbf{V}}\mathbf{T}^{\mathbf{V}}\mathbf{V}\mathbf{u}\rangle$$

which is (5.2).

Now consider the case where $\dim \mathcal{D}_2 = 1$ so that $\mathcal{D}_2$ is spanned

by the vector $\boldsymbol{\varepsilon}$ and $\mathbf{D}_2^{\mathbf{V}}\mathbf{z} = |\boldsymbol{\varepsilon}|_{\mathbf{V}}^{-2}\langle \boldsymbol{\varepsilon}, \mathbf{z}\rangle_{\mathbf{V}}\boldsymbol{\varepsilon}$. Let $c \in R$ be such that

$$(\mathbf{D}_2^{\mathbf{V}}\bar{\mathbf{T}}^{\mathbf{V}}\mathbf{D}_2^{\mathbf{V}})^{-}\mathbf{D}_2^{\mathbf{V}}\mathbf{T}^{\mathbf{V}}\mathbf{V}\mathbf{u} = c\boldsymbol{\varepsilon}. \tag{A4}$$

Since $\mathbf{D}_2^{\mathbf{V}}\mathbf{T}^{\mathbf{V}}\mathbf{V}\mathbf{u}$ is a vector in $\mathcal{D}_2$, premultiplying both sides of (A4) by $\mathbf{D}_2^{\mathbf{V}}\bar{\mathbf{T}}^{\mathbf{V}}\mathbf{D}_2^{\mathbf{V}}$ gives

$$\mathbf{D}_2^{\mathbf{V}}\mathbf{T}^{\mathbf{V}}\mathbf{V}\mathbf{u} = c\mathbf{D}_2^{\mathbf{V}}\bar{\mathbf{T}}^{\mathbf{V}}\mathbf{D}_2^{\mathbf{V}}\boldsymbol{\varepsilon} = c\mathbf{D}_2^{\mathbf{V}}\bar{\mathbf{T}}^{\mathbf{V}}\boldsymbol{\varepsilon},$$

and using the above expression for the projector $\mathbf{D}_2^{\mathbf{V}}$ we get

$$|\boldsymbol{\varepsilon}|_{\mathbf{V}}^{-2}\langle \boldsymbol{\varepsilon}, \mathbf{T}^{\mathbf{V}}\mathbf{V}\mathbf{u}\rangle_{\mathbf{V}}\boldsymbol{\varepsilon} = c|\boldsymbol{\varepsilon}|_{\mathbf{V}}^{-2}\langle \boldsymbol{\varepsilon}, \bar{\mathbf{T}}^{\mathbf{V}}\boldsymbol{\varepsilon}\rangle_{\mathbf{V}}.$$

Hence

$$\begin{aligned} c &= \langle \boldsymbol{\varepsilon}, \mathbf{T}^{\mathbf{V}}\mathbf{V}\mathbf{u}\rangle_{\mathbf{V}}[\langle \boldsymbol{\varepsilon}, \bar{\mathbf{T}}^{\mathbf{V}}\boldsymbol{\varepsilon}\rangle_{\mathbf{V}}]^{-1} \\ &= \langle \mathbf{T}^{\mathbf{V}}\boldsymbol{\varepsilon}, \mathbf{u}\rangle[|\bar{\mathbf{T}}^{\mathbf{V}}\boldsymbol{\varepsilon}|_{\mathbf{V}}^{2}]^{-1}. \end{aligned}$$

Using this value of c in (A4) and substituting (A4) into (5.2), we see that the latter expression simplifies in this particular case to

$$\operatorname{cov}(\langle \mathbf{t}, \tilde{\boldsymbol{\tau}}\rangle, \langle \mathbf{u}, \tilde{\boldsymbol{\tau}}\rangle) = \langle \mathbf{t}, \mathbf{T}^{\mathbf{V}}\mathbf{u}\rangle + \frac{\langle \mathbf{t}, \mathbf{T}^{\mathbf{V}}\boldsymbol{\varepsilon}\rangle\langle \mathbf{u}, \mathbf{T}^{\mathbf{V}}\boldsymbol{\varepsilon}\rangle}{|\bar{\mathbf{T}}^{\mathbf{V}}\boldsymbol{\varepsilon}|_{\mathbf{V}}^{2}}.$$

This completes the verification.

References

ALLAN, F. E. & WISHART, J. (1930). A method of estimating the yield of a missing plot in field experimental work. *Jour. Agric. Sci.* **20,** 399–406.

ALVEY, N. G. *et al.* (1977) GENSTAT. *A general statistical programme.* Rothamsted Experimental Station.

ANDERSON, R. L. (1946). Missing-plot techniques. *Biometrics* **2,** 41–47.

BARTLETT, M. S. (1937). Some examples of statistical methods of research in agriculture and applied biology. *J. Roy. Statist. Soc., Suppl.* **4,** 137–183.

COCHRAN, W. G. (1946). Analysis of covariance. *Institute of Statistics Mimeo Series No. 6,* University of North Carolina.

COCHRAN, W. G. & COX, G. M. (1957). *Experimental designs.* (2nd Edition) New York: Wiley.

COONS, I. (1957). The analysis of covariance as a missing-plot technique. *Biometrics* **13,** 387–405.

CORNISH, E. A. (1943). The recovery of inter-block information in quasi-factorial designs with incomplete data. I. Square, triple and cubic lattices. *C.S.I.R. Bull., No. 158.*

CORNISH, E. A. (1944). The recovery of inter-block information in quasi-factorial designs with incomplete data. II. Lattice squares. *C.S.I.R. Bull., No. 175.*

CORNISH, E. A. (1956). The recovery of inter-block information in quasi-factorial designs with incomplete data. III. Balanced incomplete blocks. *C.S.I.R.O. Divn. Math. Stats. Tech. Paper, No.* 4.

DEMPSTER, A. P., LAIRD, N. M. & RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. B* **39,** 1–38.

HASEMAN, J. K. & GAYLOR, D. W. (1973). An algorithm for non-iterative estimation of multiple missing values for crossed classifications. *Technometrics* **15,** 631–636.

HEALY, M. J. R. & WESTMACOTT, M. H. (1956). Missing values in experiments analysed on automatic computers. *Appl. Statist.* **5,** 203–206.

HOUTMAN, A. M. (1980). *The analysis of designed experiments.* Ph.D. thesis, Princeton University.

HOUTMAN, A. M. & SPEED, T. P. (1983). Balance in designed experiments with orthogonal block structure. *Ann. Statist.,* **11,** 1069–1085.

HOYLE, M. H. (1971). Spoilt data—an introduction and bibliography. *J. Roy. Statist. Soc. A* **134,** 429–439.

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.

JARRETT, R. G. (1978). The analysis of designed experiments with missing observations. *Appl. Statist.* **27,** 38–46.

JOHN, J. A. & LEWIS, S. M. (1976). Mixed-up values in experiments. *Appl. Statist.* **25,** 61–63.

JOHN, J. A. & PRESCOTT, P. (1975). Estimating missing values in experiments. *Appl. Statist.* **24,** 190–192.

KRUSKAL, W. H. (1961). The coordinate-free approach to Gauss-Markov estimation, and its application to missing and extra observations. *Fourth Berkeley Symp. Math. Statist. Prob.* **1,** 435–451.

NAIR, K. R. (1940). The application of the technique of analysis of covariance to field experiments with several missing or mixed-up plots. *Sankhyā* **4,** 581–588.

NELDER, J. A. (1965a). The analysis of randomised experiments with orthogonal block structure. I. Block structure and the null analysis of variance. *Proc. Roy. Soc. (London) Ser. A.* **273,** 147–162.

NELDER, J. A. (1965b). The analysis of randomised experiments with orthogonal block structure II. Treatment structure and the general analysis of variance. *Ibid.* 163–178.

NELDER, J. A. (1968). The combination of information in generally balanced designs. *J. Roy. Statist. Soc. B* **30,** 303–311.

PEARCE, S. C. (1965). *Biological statistics: An Introduction.* New York: McGraw Hill.

PEARCE, S. C. & JEFFERS, J. R. N. (1971). Block designs and missing data. *J. Roy. Statist. Soc. B* **33,** 131–136.

PREECE, D. A. (1971). Iterative procedures for missing values in experiments. *Technometrics* **13,** 743–753.

PREECE, D. A. & GOWER, J. C. (1974). An iterative procedure for mixed-up values in experiments. *Appl. Statist.* **23,** 73–74.

RUBIN, D. B. (1972). A non-iterative method for L.S. estimation of missing values in any analysis of variance design. *Appl. Statist.* **21,** 136–141.

RUBIN, D. B. (1976). Non-iterative least squares estimates, standard errors and *F*-tests for analysis of variance with missing data. *J. Roy. Statist. Soc. B* **38,** 270–274.

RUBIN, D. B. (1976). Inference and missing data (with discussion). *Biometrika* **63,** 581–592.

SMITH, P. L. (1981). The use of Analysis of Covariance to analyse data from designed experiments with missing or mixed-up values. *Appl. Statist.* **30,** 1–8.

TRUITT, J. T. & FAIRFIELD SMITH, H. (1956). Adjustment by covariance and consequent tests of significance in split-plot experiments. *Biometrics* **12,** 23–39.

VON NEUMANN, J. (1950). *Functional operators, vol. II: The geometry of orthogonal spaces.* Princeton University Press.

WILKINSON, G. N. (1958a). Estimation of missing values for the analysis of incomplete data. *Biometrics* **14,** 257–286.

WILKINSON, G. N. (1958b). The analysis of variance and derivation of standard errors for incomplete data. *Biometrics* **14,** 360–384.

WILKINSON, G. N. (1960). Comparison of missing value procedures. *Austr. J. Statist.* **2,** 53–65.

WILLIAMS, E. R., RATCLIFF, D. & SPEED, T. P. (1981). Estimating missing values in multi-stratum experiments. *Appl. Statist.* **30,** 71–72.

YATES, F. (1933). The analysis of replicated experiments when the field results are incomplete. *Emp. Jour. Exp. Agric.* **1,** 129–142.

The Annals of Statistics
1986, Vol. 14, No. 3, 874–895

RECTANGULAR LATTICE DESIGNS: EFFICIENCY FACTORS AND ANALYSIS

BY R. A. BAILEY AND T. P. SPEED

Rothamsted Experimental Station and C.S.I.R.O.

Rectangular lattice designs are shown to be generally balanced with respect to a particular decomposition of the treatment space. Efficiency factors are calculated, and the analysis, including recovery of interblock information, is outlined. The ideas are extended to rectangular lattice designs with an extra blocking factor.

1. Introduction. The class of incomplete block designs known as *rectangular lattice* designs was introduced by Harshbarger (1946), with further details and extensions being given in a subsequent series of papers by Harshbarger (1947, 1949, 1951) and Harshbarger and Davis (1952). Apart from a contribution by Grundy (1950) concerning the efficient estimation of the stratum variances and the papers by Nair (1951, 1952, 1953) relating rectangular lattice designs to partially balanced designs, little further theoretical discussion of this class of designs seems to have occurred. Expositions of the basic results about rectangular lattice designs in two and three replicates, as well as tables of designs, can be found in Robinson and Watson (1949) and Cochran and Cox (1957). Discussions exist in other standard texts on the design and analysis of experiments, for example Kempthorne (1952), but, apart from recent contributions by Williams (1977) and Williams and Ratcliff (1980), the literature seems to end in the early 1950's. [In his recent note, Thompson (1983) uses the results in the present paper, as he acknowledges.] A possible explanation of this fact may be the observations of Nair (1951, 1953) that every 2-replicate rectangular lattice design is a partially balanced incomplete block design with four associate classes, whilst the obvious extension of the argument to r-replicate rectangular lattice designs for $r \geq 3$ fails in general, although the classes of rectangular lattice designs for $n(n-1)$ treatments in $n-1$ or n replicates again turn out to be partially balanced. Perhaps it was felt that, in not being partially balanced, rectangular lattice designs were rather too complicated.

In his fundamental papers on designed experiments with *simple orthogonal block structure* Nelder (1965a, b) introduced the notion of *general balance*, this being a relationship between the treatment structure and the block structure of the design. It is immediate from his definition that all block experiments (in the usual sense of the term) are generally balanced for some treatment structure [see Houtman and Speed (1983)], although here we might more properly use the term *treatment pseudo-structure*, and when this structure is elucidated for a given class of designs they can be regarded as understood and readily analysed. In a

Received June 1985; revised September 1985.

AMS 1980 *subject classifications.* Primary 62K10; secondary 62J10, 05B15.

Key words and phrases. Analysis of variance, block structure, combination of information, efficiency factor, general balance, Latin square, rectangular lattice, resolvable design, stratum, treatment decomposition.

later paper, Nelder (1968) showed the importance of general balance in permitting the straightforward estimation of stratum variances, introducing a method equivalent to that which has come to be known as *restricted maximum likelihood estimation* of variances [see Patterson and Thompson (1971) and Harville (1977)]. The definition of general balance in block designs is intimately connected with the eigenspaces of a certain linear transformation, denoted by $\mathbf{L}_B$ in this paper, and in this form a number of other authors have recently emphasised the same concept [see, for example, Pearce, Caliński, and Marshall (1974), who called the eigenvectors of $\mathbf{L}_B$ *basic contrasts*, and Corsten (1976)].

In Sections 3 and 4 of this paper we obtain an orthogonal decomposition of the space of all treatment contrasts associated with a general r-replicate rectangular lattice design. In Section 5 we use this decomposition to identify all the eigenspaces of the linear transformation $\mathbf{L}_B$. An equivalent description of our results is that we determine the treatment pseudo-structure relative to which the designs are generally balanced; equivalently again, we describe the basic contrasts of the design. Using these results, a full analysis, modelled on Nelder's (1965b, 1968) general approach, of rectangular lattice designs is given in Section 6, involving the derivation of a fully orthogonal analysis of variance and estimates of the stratum variances, and the calculations of estimates of treatment contrasts, together with their standard errors. A recursive analysis along the lines of Wilkinson (1970) is most satisfactory, as the eigenspaces are orthogonal complements of subspaces each of which has a simple formula for its orthogonal projection in terms of averaging operators, and so these subspaces can be swept out successively in a quite straightforward manner. Our general approach to the analysis of designed experiments is framed in vector space terms, similar to that used by James and Wilkinson (1971) and Bailey (1981), but in the multistratum framework of Nelder's papers.

Finally, we use the foregoing ideas to sketch the design and analysis of an experiment in which an extra blocking factor was imposed on a rectangular lattice design. Two examples are used throughout the paper to illustrate the theory.

EXAMPLE 1. This is a rectangular lattice for 20 treatments in three replicates of five blocks of four plots. Although this is an entirely abstract example, there being no associated experiment, it illustrates the general theory well because it has no special features: the design is *not* partially balanced, and its construction does *not* use a complete set of mutually orthogonal Latin squares. Tables 1, 3–5, 7, and 12–15 refer to Example 1.

EXAMPLE 2. In an experiment into the digestibility of stubble, 12 feed treatments were applied to sheep. There were 12 sheep, in three rooms of four animals each. There were three test periods of four weeks each, separated by two-week recovery periods. Each sheep was fed three treatments, one in each test period. During the recovery periods all animals received their usual feed, so that they would return to normal conditions before being subjected to a new treatment.

TABLE 1
Transversal of a 5 × 5 Latin square

①	2	3	4	5
2	1	④	5	3
3	⑤	1	2	4
4	3	5	1	②
5	4	2	③	1

It was desired that each treatment should be fed once in each room and once in each period. If periods are ignored, a suitable design is a rectangular lattice design in which sheep are blocks and rooms are replicates. We shall ignore the periods until Section 7, where we show how to deal with this extra blocking factor. Tables 9–11 and 18–19 refer to Example 2.

2. Construction. In this section we review the construction of rectangular lattice designs, partly in order to establish our terminology and notation.

A rectangular lattice design is a resolvable incomplete block design for t treatments in r replicates of n blocks of size $n - 1$, where $t = n(n - 1)$ and $2 \leq r \leq n$, for some integer n. We write b for rn, the total number of blocks, and N for $b(n - 1)$, the total number of plots. The design has the property that any pair of treatments occur together in at most one block. The design is constructed from a set of $r - 2$ mutually orthogonal $n \times n$ Latin squares $\Lambda_1, \ldots, \Lambda_{r-2}$.

A *transversal* of such a set of Latin squares is defined [see Dénes and Keedwell (1974), pages 28 and 331] to be a set of n cells with one cell in each row and one in each column, which between them have all the letters of all the squares $\Lambda_1, \ldots, \Lambda_{r-2}$. In Table 1 a transversal of a single 5×5 Latin square is indicated with circles. Transversals do not always exist: Table 2 shows a 4×4 Latin square with no transversal. A sufficient condition for the existence of a transversal is the existence of a Latin square Λ_{r-1} orthogonal to each of $\Lambda_1, \ldots, \Lambda_{r-2}$, for then each letter of Λ_{r-1} corresponds to a transversal. Such a set of mutually orthogonal $n \times n$ Latin squares $\Lambda_1, \ldots, \Lambda_{r-1}$ exists whenever n is a prime or prime power and r is less than or equal to n [see Dénes and Keedwell (1974), page 165]. However, this condition is not necessary, because the square in Table 1 has no orthogonal mate.

It is convenient (although not essential) to permute the rows and columns of $\Lambda_1, \ldots, \Lambda_{r-2}$ simultaneously so that the transversal lies down the main diagonal.

TABLE 2
A 4 × 4 Latin square with no transversal

1	2	3	4
4	1	2	3
3	4	1	2
2	3	4	1

RECTANGULAR LATTICE DESIGNS

TABLE 3a
Table 1 *with rows permuted*

1	2	3	4	5
3	5	1	2	4
2	1	4	5	3
5	4	2	3	1
4	3	5	1	2

TABLE 3b
Table 3a *with letters permuted*

1	5	4	3	2
4	2	1	5	3
5	1	3	2	4
2	3	5	4	1
3	4	2	1	5

This is achieved by moving the ith row to the jth row if the unique transversal cell in row i is in column j. It is also convenient to rename the "letters" of each square independently so that the letters on the main diagonal are in natural order. Tables 3a and 3b show the results of applying these processes to the square in Table 1.

An $n \times n$ square array is drawn. The diagonal cells are left blank, and the t treatments are allocated to the remaining cells, as in Table 4. In this example we have labelled the treatments $A, B, \ldots, T$, but we shall usually use ω to denote a general treatment, to avoid confusion with other symbols. We denote the n diagonal cells by $i, j, \ldots$ and the r *classifications* (that is, rows, columns, letters of $\Lambda_1, \ldots,$ letters of Λ_{r-2}) by $a, b, \ldots$.

We define subsets of the treatments called *spokes* and *fans*. A *1-spoke* is the set of $n - 1$ treatments in any row; a *2-spoke* is the set of $n - 1$ treatments in any column. For $a = 3, \ldots, r$, an *a-spoke* is the set of $n - 1$ treatments in the positions of any one letter of square Λ_{a-2}. For $a = 1, \ldots, r$ and $i = 1, \ldots, n$ we denote by $\mathscr{S}_{ai}$ the unique a-spoke which would naturally go through the ith diagonal cell if the diagonal cells were not excluded. For each fixed i, the *fan* $\mathscr{F}_i$ through the ith diagonal cell is defined to be the union of all spokes through that

TABLE 4
Treatment array for Example 1

*	*A*	*B*	*C*	*D*
E	*	*F*	*G*	*H*
I	*J*	*	*K*	*L*
M	*N*	*O*	*	*P*
Q	*R*	*S*	*T*	*

TABLE 5
Rectangular lattice block design (Example 1)
(blocks are columns)

replicate 1					replicate 2					replicate 3				
A	*E*	*I*	*M*	*Q*	*E*	*A*	*B*	*C*	*D*	*F*	*D*	*C*	*B*	*A*
B	*F*	*J*	*N*	*R*	*I*	*J*	*F*	*G*	*H*	*J*	*K*	*H*	*E*	*G*
C	*G*	*K*	*O*	*S*	*M*	*N*	*O*	*K*	*L*	*P*	*M*	*N*	*L*	*I*
D	*H*	*L*	*P*	*T*	*Q*	*R*	*S*	*T*	*P*	*T*	*S*	*Q*	*R*	*O*

diagonal cell; that is,

$$\mathscr{F}_i = \mathscr{S}_{1i} \cup \mathscr{S}_{2i} \cup \cdots \cup \mathscr{S}_{ri}.$$

The terminology is suggested by the fact that all spokes in a fan have the corresponding diagonal cell in common, while no two spokes in the same fan have any further cells in common. In the example given by Tables 3b and 4, we have

$$\mathscr{S}_{11} = \{A, B, C, D\},$$
$$\mathscr{S}_{24} = \{C, G, K, T\},$$
$$\mathscr{S}_{32} = \{D, K, M, S\},$$
$$\mathscr{F}_5 = \{Q, R, S, T, D, H, L, P, A, G, I, O\}.$$

The design is now constructed very easily. For $a = 1, \ldots, r$, the blocks of the ath replicate are just the a-spokes. Table 5 shows the (unrandomized) design which emerges in this way from Tables 3b and 4. Thus spokes have a genuine statistical meaning, as each spoke gives a block of the design. Fans have no direct statistical meaning, but they are a combinatorial consequence of the spokes which prove useful for the analysis of the design.

Orthogonal cyclic Latin squares may be constructed by the automorphism method of Mann (1942), which is described in Section 7.2 of Dénes and Keedwell (1974). If p is the smallest prime divisor of n then $p - 1$ orthogonal squares are obtained, and hence rectangular lattice designs may be constructed for $r \leq p$ (reserving one of the squares for the transversal). The same designs may also be constructed as α-designs [Patterson and Williams (1976)]. Let $q_1, q_2, \ldots, q_{r-1}$ be any integers such that no two are congruent modulo p and none is divisible by p. Without loss of generality we may take $q_1 = 1$. The generating α-array is in Table 6, in the format used by Patterson and Williams (1976), whose series I, II, and IV are all examples of the array shown here.

3. Decomposition of the treatment space. Let $\mathbb{R}^t$ be the real vector space of vectors indexed by the t treatments. We need to find an orthogonal decomposition of $\mathbb{R}^t$ that will enable us to analyse data from experiments with the rectangular lattice design. To this end, we define certain special vectors in and subspaces of $\mathbb{R}^t$.

Let $\mathbf{u}$ be the vector $(1, 1, \ldots, 1)$. For $a = 1, \ldots, r$ and $i = 1, \ldots, n$ let $\mathbf{v}_{ai}$ be the characteristic vector of the spoke $\mathscr{S}_{ai}$; that is, the ω-entry $(\mathbf{v}_{ai})_\omega$ of $\mathbf{v}_{ai}$ is 1 if

RECTANGULAR LATTICE DESIGNS

TABLE 6
Generators for α-designs which are also rectangular lattice designs (entries in the array should be reduced modulo n)

0	0	0	⋯	0
0	1	q_2	⋯	q_{r-1}
0	2	$2q_2$	⋯	$2q_{r-1}$
⋮	⋮	⋮		⋮
0	$n-2$	$(n-2)q_2$	⋯	$(n-2)q_{r-1}$
0	$n-1$	$(n-1)q_2$	⋯	$(n-1)q_{r-1}$

$\omega \in \mathscr{S}_{ai}$ and 0 otherwise. Similarly, for $i = 1, \ldots, n$, let $\mathbf{w}_i$ be the characteristic vector of the fan $\mathscr{F}_i$, so that

$$\mathbf{w}_i = \mathbf{v}_{1i} + \mathbf{v}_{2i} + \cdots + \mathbf{v}_{ri}.$$

Let U_μ be the subspace spanned by $\mathbf{u}$; let U_f be the subspace spanned by the fan vectors $\mathbf{w}_i$; let U_s be the subspace spanned by the spoke vectors $\mathbf{v}_{ai}$; and let U_ε be the whole space $\mathbb{R}^t$. [Our conventions for labelling the first and last of these spaces agree with those used by Throckmorton (1961) and Kempthorne (1982).] Then

$$U_\mu \subseteq U_f \subseteq U_s \subseteq U_\varepsilon.$$

For Example 1 we display each vector in $\mathbb{R}^{20}$ in a two-dimensional array corresponding to Table 4. Tables 7a and 7b give examples of vectors in $U_s \setminus U_f$ and in U_f respectively.

The dimension of U_μ is 1. The space $\mathbb{R}^t$ has an inner product $\langle\ ,\ \rangle$ on it defined by

$$\langle \mathbf{z}, \mathbf{z}' \rangle = \sum_{\omega=1}^{t} z_\omega z'_\omega.$$

TABLE 7a
The vector $\mathbf{v}_{11} - 2\mathbf{v}_{24} + 5\mathbf{v}_{32}$

*	1	1	−1	6
0	*	0	−2	0
0	0	*	3	0
5	0	0	*	0
0	0	5	−2	*

TABLE 7b
The vector $\mathbf{w}_1 + 3\mathbf{w}_5$

*	4	1	1	4
1	*	1	3	3
4	1	*	0	3
1	0	3	*	4
4	3	3	4	*

We use this to find the dimensions of the spaces U_f and U_s. Note that

$$\begin{aligned}\langle \mathbf{v}_{ai}, \mathbf{v}_{bj} \rangle &= |\mathscr{S}_{ai} \cap \mathscr{S}_{bj}| \\ &= \begin{cases} n-1 & \text{if } a = b \quad \text{and} \quad i = j, \\ 0 & \text{if } a = b \quad \text{and} \quad i \neq j, \\ 0 & \text{if } a \neq b \quad \text{and} \quad i = j, \\ 1 & \text{if } a \neq b \quad \text{and} \quad i \neq j, \end{cases}\end{aligned} \tag{3.1}$$

so that

$$\begin{aligned}\langle \mathbf{w}_i, \mathbf{w}_j \rangle &= |\mathscr{F}_i \cap \mathscr{F}_j| \\ &= \begin{cases} r(n-1) & \text{if } i = j, \\ r(r-1) & \text{if } i \neq j. \end{cases}\end{aligned} \tag{3.2}$$

Moreover, $\Sigma_i \mathbf{w}_i = r\mathbf{u}$. Suppose that $\Sigma_i \lambda_i \mathbf{w}_i = \mathbf{0}$ for some real numbers λ_i. If $r \neq n$, taking inner products with individual $\mathbf{w}_i$ shows that $\lambda_1 = \cdots = \lambda_n$, and hence that $\lambda_1 = \cdots = \lambda_n = 0$: thus the fan vectors are linearly independent and so U_f has dimension n. On the other hand, if $r = n$ then $\mathbf{w}_i = \mathbf{u}$ for $i, \ldots, n$: thus $U_f = U_\mu$. Now suppose that $\Sigma_a \Sigma_i \lambda_{ai} \mathbf{v}_{ai} = \mathbf{0}$ for some real numbers λ_{ai}. Taking inner products with individual $\mathbf{v}_{ai}$ shows that there are real numbers θ_a and ϕ_i such that $\lambda_{ai} = \theta_a + \phi_i$ for all a and i. Since

$$\mathbf{v}_{a1} + \mathbf{v}_{a2} + \cdots + \mathbf{v}_{an} = \mathbf{u}$$

for $a = 1, \ldots, r$, this implies that $(\Sigma_a \theta_a)\mathbf{u} + \Sigma_i \phi_i \mathbf{w}_i = \mathbf{0}$. Hence U_s has dimension $nr - (r-1)$ if $r \neq n$, and $nr - (r-1) - (n-1)$ if $r = n$.

For Example 1, Equations (3.1) and (3.2) are demonstrated in Tables 7a and 7b, respectively. For example, the six entries equal to 4 in Table 7b correspond to the elements of $\mathscr{F}_1 \cap \mathscr{F}_5$. In this case the five fan vectors form a basis for U_f; while a basis of U_s consists of $\mathbf{u}$ and all but three spoke vectors, one being omitted for each classification.

We can form the orthogonal complements of the U-subspaces, and thus obtain the subspaces that really interest us. Specifically, we put

$$\begin{aligned} V_\mu &= U_\mu, \\ V_f &= \text{the orthogonal complement of } U_\mu \text{ in } U_f, \\ V_s &= \text{the orthogonal complement of } U_f \text{ in } U_s, \\ V_e &= \text{the orthogonal complement of } U_s \text{ in } U_e. \end{aligned}$$

Then V_f is spanned by vectors of the form $\mathbf{w}_i - \mathbf{w}_j$; while V_s is spanned by vectors of the form $\mathbf{v}_{ai} - \mathbf{v}_{bi}$. Now $\mathbb{R}^t$ is the orthogonal direct sum

$$\mathbb{R}^t = V_\mu \oplus V_f \oplus V_s \oplus V_e.$$

We record the important facts about this decomposition in Table 8.

In two special cases this decomposition can be described in simpler terms. If $r = n$ then the set $\{\Lambda_1, \ldots, \Lambda_{r-2}\}$ is only one square short of a complete set of mutually orthogonal Latin squares. Thus there exists a (unique) Latin square

RECTANGULAR LATTICE DESIGNS

TABLE 8
Decomposition of the treatment subspace

subspace description	V_μ mean	V_f contrasts between fans	V_s contrasts between spokes within fans	V_ε orthogonal to spokes
dimension ($r < n$)	1	$n - 1$	$(n - 1)(r - 1)$	$(n - r)(n - 1) - 1$
dimension ($r = n$)	1	0	$(n - 1)^2$	$n - 2$

Λ_{n-1} orthogonal to all the others, by Theorem 1.6.1 of Rhagavarao (1971). One letter of Λ_{n-1} must correspond to the transversal. Each other letter of Λ_{n-1} occurs just once in each a-spoke, for each classification a. Hence the contrasts between these $n - 1$ other letters are orthogonal to spokes, and so they form the whole space V_ε. Since V_f is null in this case, V_s must consist of all treatment contrasts which are orthogonal to the letters of Λ_{n-1}. Thus the treatments have the simple nested structure $(n - 1) \to n$ [in the notation of Nelder (1965a)], and the treatment space decomposition is the familiar one into mean, between letters of Λ_{n-1} and within letters.

If $r = n - 1$ and $n \neq 4$, the results of Shrikhande (1961) and Bruck (1963) show that there is a unique complete orthogonal set $\{\Lambda_1, \ldots, \Lambda_{n-1}\}$ containing the original set $\{\Lambda_1, \ldots, \Lambda_{n-3}\}$ and that the original transversal corresponds to a letter of one of the two extra squares, say Λ_{n-2}. The same result is true even when $n = 4$, because the existence of the original transversal prevents Λ_1 from being isotopic to the square in Table 2, which is the only 4×4 Latin square (up to isotopy) which is not uniquely embeddable in a complete set of mutually orthogonal Latin squares [isotopy classes are also called transformation sets (see Fisher and Yates (1934))]. The treatments now have the simple crossed factorial structure $Q_1 \times Q_2$, where the levels of Q_1 are the $n - 1$ other letters of Λ_{n-2} and the levels of Q_2 are the n letters of Λ_{n-1}. Now V_ε is the main effect of Q_1; while V_f is the main effect of Q_2 and V_s is the Q_1Q_2 interaction.

Example 2 has $r = n - 1 = 3$. The rectangular lattice design is constructed from the set of mutually orthogonal 4×4 Latin squares in Table 9 : the rows, columns, and letters of Λ_1 are the three classifications; letter 1 of Λ_2 gives the transversal; the remaining letters of Λ_2 and Λ_3 give the 3×4 factorial treatment structure described above and shown in Table 10. The design is that shown in Table 11, ignoring periods.

In both these special cases the factorial treatment decomposition has no direct statistical meaning, but is merely an aid to the analysis. The factors Q_1 and Q_2 are entirely analogous to the pseudo-factors used in the construction and analysis of square lattice designs [Yates (1936)].

4. Treatment projection. Let $\mathbf{z}$ be a vector in $\mathbb{R}^t$. In order to use the spaces V_μ, V_f, V_s, and V_ε in the analysis of an experiment we need to know how to calculate the projections of $\mathbf{z}$ onto these spaces. This is done in terms of the

TABLE 9a
Three mutually orthogonal 4 × 4 *Latin squares*

Λ_1 (gives 3rd replicate)				Λ_2 ("1" gives transversal; other letters are levels of Q_1)				Λ_3 (letters are levels of Q_2)			
1	4	2	3	1	2	3	4	1	2	3	4
3	2	4	1	2	1	4	3	3	4	1	2
4	1	3	2	3	4	1	2	4	3	2	1
2	3	1	4	4	3	2	1	2	1	4	3

TABLE 9b
Array of twelve treatments for Example 2

*	*A*	*B*	*C*
D	*	*E*	*F*
G	*H*	*	*I*
J	*K*	*L*	*

TABLE 10
3 × 4 *factorial structure for Example* 2

treatment	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	*L*
level of Q_1	2	3	4	2	4	3	3	4	2	4	3	2
level of Q_2	2	3	4	3	1	2	4	3	1	2	1	4

TABLE 11
Design which is not generally balanced

room		1				2				3			
sheep		1	2	3	4	5	6	7	8	9	10	11	12
time period	**1**	*B*	*D*	*I*	*L*	*K*	*E*	*F*	*G*	*A*	*J*	*C*	*H*
	2	*C*	*E*	*H*	*K*	*A*	*L*	*I*	*J*	*G*	*B*	*D*	*F*
	3	*A*	*F*	*G*	*J*	*H*	*B*	*C*	*D*	*E*	*I*	*K*	*L*

following totals:

$$\text{grand total } G(\mathbf{z}) = \sum_{\omega} \mathbf{z}_{\omega},$$

$$\text{spoke total } S_{ai}(\mathbf{z}) = \sum \{\mathbf{z}_{\omega} \colon \omega \in \mathscr{S}_{ai}\} = \langle \mathbf{z}, \mathbf{v}_{ai} \rangle,$$

$$\text{fan total } F_i(\mathbf{z}) = \sum \{\mathbf{z}_{\omega} \colon \omega \in \mathscr{F}_i\} = \langle \mathbf{z}, \mathbf{w}_i \rangle.$$

RECTANGULAR LATTICE DESIGNS

TABLE 12
A particular vector $\mathbf{z}$ *in* $\mathbb{R}^{20}$

*	7	3	2	3
6	*	5	9	4
5	2	*	6	7
4	5	8	*	1
2	4	2	5	*

It is immediate that

$$\sum_i S_{ai}(\mathbf{z}) = G(\mathbf{z}), \tag{4.1}$$

$$\sum_a S_{ai}(\mathbf{z}) = F_i(\mathbf{z}), \tag{4.2}$$

$$\sum_i F_i(\mathbf{z}) = rG(\mathbf{z}). \tag{4.3}$$

Define the *fan totals vector* $\mathbf{f}(\mathbf{z})$ and the *spoke totals vector* $\mathbf{s}(\mathbf{z})$ by

$$\mathbf{f}(\mathbf{z}) = \sum_i F_i(\mathbf{z})\mathbf{w}_i,$$

$$\mathbf{s}(\mathbf{z}) = \sum_a \sum_i S_{ai}(\mathbf{z})\mathbf{v}_{ai}.$$

We also need the *grand totals vector* $\mathbf{g}(\mathbf{z})$, all of whose entries are equal to $G(\mathbf{z})$.

Continuing our Example 1, a vector $\mathbf{z}$ is shown in Table 12. Its spoke totals are in Table 13: the column margins are the fan totals, and the row totals are all the grand total. The vectors $\mathbf{f}(\mathbf{z})$ and $\mathbf{s}(\mathbf{z})$ are shown in Table 14.

We aim to give the projections of $\mathbf{z}$ onto the spaces V_μ, V_f, V_s, and V_ε in terms of $\mathbf{f}(\mathbf{z})$, $\mathbf{s}(\mathbf{z})$, and $\mathbf{g}(\mathbf{z})$. The necessary calculations are contained in the following two lemmas.

LEMMA 1.

(i) $$\langle \mathbf{s}(\mathbf{z}), \mathbf{v}_{ai} \rangle = nS_{ai}(\mathbf{z}) + (r-1)G(\mathbf{z}) - F_i(\mathbf{z}),$$

(ii) $$\langle \mathbf{f}(\mathbf{z}), \mathbf{v}_{ai} \rangle = (n-r)F_i(\mathbf{z}) + r(r-1)G(\mathbf{z}),$$

(iii) $$\langle \mathbf{f}(\mathbf{z}), \mathbf{w}_i \rangle = r(n-r)F_i(\mathbf{z}) + r^2(r-1)G(\mathbf{z}).$$

TABLE 13
Spoke totals of $\mathbf{z}$

i	1	2	3	4	5	total
row (a = 1)	15	24	20	18	13	90
column (a = 2)	17	18	18	22	15	90
letter (a = 3)	13	15	13	20	29	90
fan totals	45	57	51	60	57	270

TABLE 14

fan totals vector f(z)					spoke totals vector s(z)				
*	159	156	156	159	*	62	53	50	45
162	*	153	174	165	61	*	55	75	52
153	153	*	168	168	66	51	*	57	55
162	168	168	*	162	50	49	65	*	46
153	174	165	162	*	43	51	46	48	*

PROOF. To simplify the expressions, we omit "(z)", the vector z being understood.

(i)
$$\begin{aligned}\langle \mathbf{s}, \mathbf{v}_{ai}\rangle &= \sum_b \sum_j S_{bj}\langle \mathbf{v}_{bj}, \mathbf{v}_{ai}\rangle \\ &= (n-1)S_{ai} + \sum_{b\neq a}\sum_{i\neq j} S_{bj} \quad \text{(by (3.1))} \\ &= (n-1)S_{ai} + \sum_{b\neq a}(G - S_{bi}) \quad \text{(by (4.1))} \\ &= nS_{ai} + (r-1)G - \sum_b S_{bi} \\ &= nS_{ai} + (r-1)G - F_i \quad \text{(by (4.2))}.\end{aligned}$$

(ii)
$$\begin{aligned}\langle \mathbf{f}, \mathbf{v}_{ai}\rangle &= \sum_j F_j\langle \mathbf{w}_j, \mathbf{v}_{ai}\rangle \\ &= (n-1)F_i + (r-1)\sum_{j\neq i} F_j \quad \text{(by (3.1))} \\ &= (n-r)F_i + (r-1)\sum_j F_j \\ &= (n-r)F_i + r(r-1)G \quad \text{(by (4.3))}.\end{aligned}$$

(iii) Summing the equation in (ii) over all the spokes in $\mathcal{F}_i$ gives

$$\langle \mathbf{f}, \mathbf{w}_i\rangle = r(n-r)F_i + r^2(r-1)G. \qquad \square$$

LEMMA 2. *The orthogonal projections of* z *onto* $U_\mu, U_f, U_s, U_\varepsilon$, *respectively, are*

$$\frac{\mathbf{g}(\mathbf{z})}{n(n-1)}, \qquad \frac{\mathbf{f}(\mathbf{z})}{r(n-r)} - \frac{(r-1)\mathbf{g}(\mathbf{z})}{(n-1)(n-r)},$$

$$\frac{\mathbf{s}(\mathbf{z})}{n} + \frac{\mathbf{f}(\mathbf{z})}{n(n-r)} - \frac{(r-1)\mathbf{g}(\mathbf{z})}{(n-1)(n-r)}, \qquad \mathbf{z}$$

when $r \neq n$. *When* $r = n$ *then* $U_f = U_\mu$ *and the orthogonal projection of* z *onto* U_s *is*

$$\frac{\mathbf{s}(\mathbf{z})}{n} - \frac{(n-2)\mathbf{g}(\mathbf{z})}{n(n-1)}.$$

RECTANGULAR LATTICE DESIGNS

PROOF. Put $\mathbf{x} = [r(n-r)]^{-1}\mathbf{f} - (r-1)[(n-1)(n-r)]^{-1}\mathbf{g}$ when $r \neq n$. Since $\mathbf{f}$ and $\mathbf{g}$ are both sums of fan vectors, $\mathbf{x} \in U_f$. Thus it suffices to show that $\mathbf{z} - \mathbf{x}$ is orthogonal to U_f. This is so if $\langle \mathbf{z} - \mathbf{x}, \mathbf{w}_i \rangle = 0$ for each fan $\mathscr{F}_i$. By Lemma 1(iii) and (3.2),

$$\langle \mathbf{x}, \mathbf{w}_i \rangle = \frac{r(n-r)F_i + r^2(r-1)G}{r(n-r)} - \frac{r(n-1)(r-1)G}{(n-1)(n-r)} = F_i = \langle \mathbf{z}, \mathbf{w}_i \rangle.$$

Similarly, put $\mathbf{y} = n^{-1}\mathbf{s} + [n(n-r)]^{-1}\mathbf{f} - (r-1)[(n-1)(n-r)]^{-1}\mathbf{g}$. Then $\mathbf{y} \in U_s$, because $\mathbf{s}$, $\mathbf{f}$, and $\mathbf{g}$ are all sums of spoke vectors, so it suffices to show that $\langle \mathbf{z} - \mathbf{y}, \mathbf{v}_{ai} \rangle = 0$ for all spokes $\mathscr{S}_{ai}$. Lemmas 1(i) and (ii) show that $\langle \mathbf{y}, \mathbf{v}_{ai} \rangle$ is equal to

$$\frac{nS_{ai} + (r-1)G - F_i}{n} + \frac{(n-r)F_i + r(r-1)G}{n(n-r)} - \frac{(n-1)(r-1)G}{(n-1)(n-r)},$$

which is S_{ai}, which is $\langle \mathbf{z}, \mathbf{v}_{ai} \rangle$.

Now let $r = n$ and put $\mathbf{y} = n^{-1}\mathbf{s} - (n-2)[n(n-1)]^{-1}\mathbf{g}$. Then

$$\langle \mathbf{y}, \mathbf{v}_{ai} \rangle = \frac{nS_{ai} + (n-2)G}{n} - \frac{(n-2)(n-1)G}{n(n-1)} = S_{ai} = \langle \mathbf{z}, \mathbf{v}_{ai} \rangle$$

so that $\mathbf{y} \in U_s$ and $\mathbf{z} - \mathbf{y}$ is orthogonal to U_s. □

Now subtraction gives the orthogonal projection of $\mathbf{z}$ onto V_μ, V_f, V_s, V_ε.

THEOREM 1. *Let $\mathbf{T}_\mu, \mathbf{T}_f, \mathbf{T}_s, \mathbf{T}_\varepsilon$ be the operators of orthogonal projection from $\mathbb{R}^t$ onto $V_\mu, V_f, V_s, V_\varepsilon$, respectively. Then, for all $\mathbf{z}$ in $\mathbb{R}^t$,*

$$\mathbf{T}_\mu \mathbf{z} = \frac{\mathbf{g}(\mathbf{z})}{n(n-1)},$$

$$\mathbf{T}_f \mathbf{z} = \frac{\mathbf{f}(\mathbf{z})}{r(n-r)} - \frac{r\mathbf{g}(\mathbf{z})}{n(n-r)} \quad \textit{when } r \neq n \textit{ and zero otherwise},$$

$$\mathbf{T}_s \mathbf{z} = \frac{\mathbf{s}(\mathbf{z})}{n} - \frac{\mathbf{f}(\mathbf{z})}{rn},$$

$$\mathbf{T}_\varepsilon \mathbf{z} = \mathbf{z} - (\mathbf{T}_\mu \mathbf{z} + \mathbf{T}_f \mathbf{z} + \mathbf{T}_s \mathbf{z}).$$

In Example 1 we have $n = 5$ and $r = 3$, so $\mathbf{T}_\mu \mathbf{z} = \mathbf{g}(\mathbf{z})/20$; $\mathbf{T}_f(\mathbf{z}) = \mathbf{f}(\mathbf{z})/6 - 3\mathbf{g}(\mathbf{z})/10$; $\mathbf{T}_s \mathbf{z} = \mathbf{s}(\mathbf{z})/5 - \mathbf{f}(\mathbf{z})/15$, and $\mathbf{T}_\varepsilon \mathbf{z}$ is best obtained by subtraction. For the particular vector $\mathbf{z}$ shown in Table 12, these four components of $\mathbf{z}$ are shown in Table 15. The orthogonality of the decomposition may be verified by noting that

$$\|\mathbf{T}_\mu \mathbf{z}\|^2 + \|\mathbf{T}_f \mathbf{z}\|^2 + \|\mathbf{T}_s \mathbf{z}\|^2 + \|\mathbf{T}_\varepsilon \mathbf{z}\|^2$$

$$= 405 + 24 + 47.2 + 21.8 = 498 = \|\mathbf{z}\|^2.$$

TABLE 15

$\mathbf{T}_\mu\mathbf{z}$					$\mathbf{T}_f\mathbf{z}$				
*	4.5	4.5	4.5	4.5	*	−0.5	−1.0	−1.0	−0.5
4.5	*	4.5	4.5	4.5	0.0	*	−1.5	2.0	0.5
4.5	4.5	*	4.5	4.5	−1.5	−1.5	*	1.0	1.0
4.5	4.5	4.5	*	4.5	0.0	1.0	1.0	*	0.0
4.5	4.5	4.5	4.5	*	−1.5	2.0	0.5	0.0	*
$\mathbf{T}_s\mathbf{z}$					$\mathbf{T}_\varepsilon\mathbf{z}$				
*	1.8	0.2	−0.4	−1.6	*	1.2	−0.7	−1.1	0.6
1.4	*	0.8	3.4	−0.6	0.1	*	1.2	−0.9	−0.4
3.0	0.0	*	0.2	−0.2	−1.0	−1.0	*	0.3	1.7
−0.8	−1.4	1.8	*	−1.6	0.3	0.9	0.7	*	−1.9
−1.6	−1.4	−1.8	−1.2	*	0.6	−1.1	−1.2	1.7	*

5. General balance. The block structure of a rectangular lattice design is the double nested classification of plots within blocks within replicates. This is one of the *simple orthogonal block structures* defined by Nelder (1965a). In what follows we retain the notation of Nelder (1965a, b, 1968) and Bailey (1981) as far as possible.

Let $\mathbb{R}^N$ be the real vector space associated with the N plots. Each grouping of the plots according to the block structure defines an averaging operation $\mathbf{P}$ on $\mathbb{R}^N$. In our case there are four averaging operators: the grand mean averaging operator $\mathbf{P}_\mu = \mathbf{J}/N$, where $\mathbf{J}$ is the all-1's matrix; the replicates averaging operator $\mathbf{P}_R$; the blocks averaging operator $\mathbf{P}_B$; and the identity $\mathbf{P}_\varepsilon = \mathbf{I}$. Nelder (1965a) showed that there is an orthogonal direct sum decomposition $\oplus_\alpha W_\alpha$ of $\mathbb{R}^N$ such that each W_α is an eigenspace of every $\mathbf{P}$. Let $\mathbf{C}_\alpha$ be the operator of orthogonal projection from $\mathbb{R}^N$ onto W_α. Nelder (1965a) showed that each $\mathbf{C}_\alpha$ is a linear combination of the $\mathbf{P}$'s with integer coefficients: Speed and Bailey (1982) gave explicit formulae for these coefficients. In our case we have

$$\mathbf{C}_\mu = \mathbf{P}_\mu, \qquad \mathbf{C}_R = \mathbf{P}_R - \mathbf{P}_\mu,$$

$$\mathbf{C}_B = \mathbf{P}_B - \mathbf{P}_R, \qquad \mathbf{C}_\varepsilon = \mathbf{P}_\varepsilon - \mathbf{P}_B.$$

The spaces W_α are called *strata*: they play an important role in analysis of variance [see Nelder (1965b) and Bailey (1981)]. Our covariance model for the data vector $\mathbf{y}$ is

$$\text{Cov}(\mathbf{y}) = \xi_\mu\mathbf{C}_\mu + \xi_R\mathbf{C}_R + \xi_B\mathbf{C}_B + \xi_\varepsilon\mathbf{C}_\varepsilon \tag{5.1}$$

for unknown scalars ξ_μ, ξ_R, ξ_B, and ξ_ε.

Denote by $\mathbf{X}$ the $N \times t$ design matrix; that is, $\mathbf{X}_{p\omega}$ is 1 if plot p receives treatment ω and 0 otherwise. For each stratum W_α, the matrix $\mathbf{L}_\alpha$ defined by $\mathbf{L}_\alpha = \mathbf{X}'\mathbf{C}_\alpha\mathbf{X}$ is called the *information matrix* for that stratum. For designs with equal replication r, we have $\mathbf{L}_\mu = r\mathbf{T}_\mu$. If $\mathbf{L}_\alpha = \mathbf{0}$ there is no information about

RECTANGULAR LATTICE DESIGNS

treatments in stratum W_α. Strata, other than W_μ, for which $\mathbf{L}_\alpha \neq \mathbf{0}$, are called *effective strata.*

Suppose that $\oplus_\theta V_\theta$ is an orthogonal direct sum decomposition of $\mathbb{R}^t$. Nelder (1965b) defined an equally replicated design to be *generally balanced* with respect to this treatment decomposition if each V_θ is an eigenspace of every information matrix; that is, there are numbers $\lambda_{\alpha\theta}$ such that $\mathbf{L}_\alpha = \sum_\theta \lambda_{\alpha\theta}\mathbf{T}_\theta$, where $\mathbf{T}_\theta$ denotes orthogonal projection onto V_θ. We have $0 \leq \lambda_{\alpha\theta} \leq r$ for all α and θ; and $\sum_\alpha \lambda_{\alpha\theta} = r$ for all θ. The quantity $\lambda_{\alpha\theta}/r$ is the *efficiency factor* for treatment term V_θ in the stratum W_α. In a simple block design with blocks stratum W_B, examination of the trace of $\mathbf{L}_B$ shows that $\sum_\theta \lambda_{B\theta}\dim(V_\theta)/r = b/r - 1$, the so-called *loss of information due to blocks.*

Houtman and Speed (1983) have shown that in any design with only two effective strata there must be *some* decomposition $\oplus V_\theta$ of $\mathbb{R}^t$ with respect to which the design is generally balanced. However, the decomposition may not be easy to find, use or interpret. Our claim is that a rectangular lattice design is generally balanced with respect to the treatment decomposition given in Section 3.

LEMMA 3. *For $a = 1, \ldots, r$ and $i = 1, \ldots, n$,*

$$\mathbf{X}'\mathbf{P}_B\mathbf{X}\mathbf{v}_{ai} = (n\mathbf{v}_{ai} - \mathbf{w}_i + (r-1)\mathbf{u})/(n-1).$$

PROOF. If $\mathscr{B}$ is any block and $\mathbf{v}$ is any vector in $\mathbb{R}^t$ then the entries of $\mathbf{P}_B\mathbf{X}\mathbf{v}$ for the plots in $\mathscr{B}$ are all equal to the average of the entries of $\mathbf{v}$ for those treatments which occur in $\mathscr{B}$. If $\mathbf{v} = \mathbf{v}_{ai}$ and $\mathscr{B}$ consists of $\mathscr{S}_{bj}$ then this average is equal to $\langle \mathbf{v}_{ai}, \mathbf{v}_{bj}\rangle/(n-1)$. Denote the characteristic vector of this block by $\mathbf{x}_{bj}$. Then

$$(n-1)\mathbf{P}_B\mathbf{X}\mathbf{v}_{ai} = \sum_b \sum_j \langle \mathbf{v}_{ai}, \mathbf{v}_{bj}\rangle \mathbf{x}_{bj}.$$

Since $\mathbf{X}'\mathbf{x}_{bj} = \mathbf{v}_{bj}$ we have

$$\begin{aligned}(n-1)\mathbf{X}'\mathbf{P}_B\mathbf{X}\mathbf{v}_{ai} &= \sum_b \sum_j \langle \mathbf{v}_{ai}, \mathbf{v}_{bj}\rangle \mathbf{v}_{bj} \\ &= (n-1)\mathbf{v}_{ai} + \sum_{b \neq a} (\mathbf{u} - \mathbf{v}_{bi}) \quad \text{(by (3.1))} \\ &= n\mathbf{v}_{ai} + (r-1)\mathbf{u} - \mathbf{w}_i. \qquad \square\end{aligned}$$

THEOREM 2. *Rectangular lattice designs are generally balanced with respect to the treatment decomposition given in Section* 3.

PROOF. We always have $\mathbf{L}_\mu\mathbf{u} = r\mathbf{u}$, and $\mathbf{L}_\mu\mathbf{z} = \mathbf{0}$ whenever $\mathbf{z}$ is orthogonal to $\mathbf{u}$. By definition of replicate, $\mathbf{X}'\mathbf{P}_R\mathbf{X}\mathbf{z} = r\mathbf{g}(\mathbf{z})/n(n-1) = \mathbf{X}'\mathbf{P}_\mu\mathbf{X}\mathbf{z}$, so $\mathbf{L}_R = \mathbf{0}$. Moreover, $\mathbf{L}_B = \mathbf{X}'\mathbf{P}_B\mathbf{X} - \mathbf{X}'\mathbf{P}_R\mathbf{X}$, and so

$$\mathbf{L}_B(\mathbf{v}_{ai} - \mathbf{v}_{bi}) = n(n-1)^{-1}(\mathbf{v}_{ai} - \mathbf{v}_{bi})$$

by Lemma 3. Since V_s is spanned by vectors of the form $\mathbf{v}_{ai} - \mathbf{v}_{bi}$, this shows that

TABLE 16
Efficiency factors of a rectangular lattice design

	treatment subspace			
	V_μ	V_f	V_s	V_ε
stratum				
mean W_μ	1	0	0	0
replicates W_R	0	0	0	0
blocks W_B	0	$\frac{n-r}{r(n-1)}$	$\frac{n}{r(n-1)}$	0
plots W_ε	0	$\frac{n(r-1)}{r(n-1)}$	$\frac{rn-r-n}{r(n-1)}$	1

V_s is an eigenspace of $\mathbf{L}_B$ with eigenvalue $\lambda_{Bs} = n/(n-1)$. Similarly, Lemma 3 shows that

$$\mathbf{L}_B(\mathbf{w}_i - \mathbf{w}_j) = (n-r)(n-1)^{-1}(\mathbf{w}_i - \mathbf{w}_j),$$

so V_f is an eigenspace of $\mathbf{L}_B$ with eigenvalue $\lambda_{Bf} = (n-r)/(n-1)$. Whether or not $r = n$, Table 8 now shows that $\lambda_{Bs}\dim(V_s) + \lambda_{Bf}\dim(V_f) = b - r$, so there can be no further nonzero eigenvalues in the blocks stratum. Thus V_ε must be an eigenspace of $\mathbf{L}_B$ with $\lambda_{B\varepsilon} = 0$.

By the result of Houtman and Speed (1983), the spaces V_f, V_s, V_ε are also eigenspaces of $\mathbf{L}_\varepsilon$. □

The eigenvalues in stratum W_ε are calculated by subtraction. Division by r gives the efficiency factors, which are shown in Table 16, which is laid out like the table in Section 4.2 of Nelder (1968).

Block designs are often classified by a single measure of efficiency: the harmonic mean of the efficiency factors (taking account of multiplicity) in stratum W_ε. It follows from Tables 8 and 16, that, whether $r = n$ or $r < n$, the harmonic mean efficiency factor for a rectangular lattice design is

$$\frac{n(r-1)(rn-r-n)(n^2-n-1)}{(r-1)^2n^2(n^2-n-1) - r^2(n-1)^2 + rn(r-1)}.$$

This efficiency factor is proportional to the reciprocal of the average variance of the intrablock estimates of simple treatment differences, and so may also be obtained from this average variance, which is given by Williams (1977, page 413).

6. Analysis. Since rectangular lattice designs are generally balanced, their analysis follows the pattern described by Nelder (1965b, 1968), Wilkinson (1970), and James and Wilkinson (1971). In this section we specialize their results to rectangular lattice designs, retaining most of Nelder's notation. We outline the procedure for fitting the model, deriving a complete analysis of variance, estimating the stratum variances ξ_R, ξ_B, and ξ_ε, and obtaining minimum variance

unbiased linear estimates (with estimated weights) of arbitrary treatment contrasts, together with their estimated variances.

Let $\mathbf{t}$ be the $t \times 1$ vector of individual treatment effects and let $\mathbf{y}$ be the $N \times 1$ vector of observations. If $\lambda_{\alpha\theta} \neq 0$, the treatment *effect* $\mathbf{T}_\theta\mathbf{t}$ is estimated in stratum W_α by $\mathbf{h}_{\alpha\theta}$, where $\mathbf{h}_{\alpha\theta} = \mathbf{T}_\theta\mathbf{X}'\mathbf{C}_\alpha\mathbf{y}/\lambda_{\alpha\theta}$. The contribution of treatment term V_θ to the *fitted value* in stratum W_α is $\mathbf{C}_\alpha\mathbf{X}\mathbf{h}_{\alpha\theta}$, with the sum of squares $\lambda_{\alpha\theta}\|\mathbf{h}_{\alpha\theta}\|^2$. Thus the overall fitted value in stratum W_α is $\Sigma'_\theta\mathbf{C}_\alpha\mathbf{X}\mathbf{h}_{\alpha\theta}$, where Σ'_θ denotes summation over those θ for which $\lambda_{\alpha\theta} \neq 0$. The residual sum of squares, RSS_α, in stratum W_α, and its number of degrees of freedom, d_α, are obtained by subtraction:

$$\mathrm{RSS}_\alpha = \mathbf{y}'\mathbf{C}_\alpha\mathbf{y} - \sum_\theta{}'\lambda_{\alpha\theta}\|\mathbf{h}_{\alpha\theta}\|^2, \tag{6.1}$$

$$d_\alpha = \dim(W_\alpha) - \sum_\theta{}'\dim(V_\theta). \tag{6.2}$$

Thus we obtain the analysis of variance shown in Tables 17a ($r < n$) and 17b ($r = n$).

If the stratum variances ξ_α are known, we put $w_\theta = \Sigma_\alpha\lambda_{\alpha\theta}/\xi_\alpha$ and define *weights* $w_{\alpha\theta}$ by $w_{\alpha\theta} = \lambda_{\alpha\theta}/\xi_\alpha w_\theta$. The *weighted effect* corresponding to treatment term V_θ is $\Sigma_\alpha w_{\alpha\theta}\mathbf{h}_{\alpha\theta}$, and the overall *weighted fitted value* $\hat{\mathbf{t}}$ is $\Sigma_\theta\Sigma_\alpha w_{\alpha\theta}\mathbf{h}_{\alpha\theta}$. If $\mathbf{x}$ is any treatment contrast (that is, $\mathbf{x} \in \mathbb{R}^t$ and $\langle \mathbf{x}, \mathbf{u} \rangle = 0$) then the minimum variance unbiased linear estimate of $\langle \mathbf{x}, \mathbf{t} \rangle$ is $\langle \mathbf{x}, \hat{\mathbf{t}} \rangle$, with variance $\Sigma_\theta\|\mathbf{T}_\theta\mathbf{x}\|^2/w_\theta$.

TABLE 17a
Analysis of variance when $r < n$

stratum	source of variation	df	SS	EMS
mean		1	$\mathbf{y}'\mathbf{C}_\mu\mathbf{y}$	$r\|\mathbf{T}_\mu\mathbf{t}\|^2 + \xi_\mu$
replicates		$r - 1$	$\mathbf{y}'\mathbf{C}_R\mathbf{y}$	ξ_R
blocks	V_f	$n - 1$	$\lambda_{Bf}\|\mathbf{h}_{Bf}\|^2$	$\dfrac{\lambda_{Bf}\|\mathbf{T}_f\mathbf{t}\|^2}{n-1} + \xi_B$
	V_s	$(n-1)(r-1)$	$\lambda_{Bs}\|\mathbf{h}_{Bs}\|^2$	$\dfrac{\lambda_{Bs}\|\mathbf{T}_s\mathbf{t}\|^2}{(n-1)(r-1)} + \xi_B$
	total	$r(n-1)$	$\mathbf{y}'\mathbf{C}_B\mathbf{y}$	
plots	V_f	$n - 1$	$\lambda_{\varepsilon f}\|\mathbf{h}_{\varepsilon f}\|^2$	$\dfrac{\lambda_{\varepsilon f}\|\mathbf{T}_f\mathbf{t}\|^2}{n-1} + \xi_\varepsilon$
	V_s	$(n-1)(r-1)$	$\lambda_{\varepsilon s}\|\mathbf{h}_{\varepsilon s}\|^2$	$\dfrac{\lambda_{\varepsilon s}\|\mathbf{T}_s\mathbf{t}\|^2}{(n-1)(r-1)} + \xi_\varepsilon$
	V_ε	$(n-r)(n-1)-1$	$\lambda_{\varepsilon\varepsilon}\|\mathbf{h}_{\varepsilon\varepsilon}\|^2$	$\dfrac{\lambda_{\varepsilon\varepsilon}\|\mathbf{T}_\varepsilon\mathbf{t}\|^2}{(n-r)(n-1)-1} + \xi_\varepsilon$
	error	$n(rn-2r-n+1)+1$	RSS_ε	ξ_ε
	total	$rn(n-2)$	$\mathbf{y}'\mathbf{C}_\varepsilon\mathbf{y}$	

TABLE 17b
Analysis of variance when $r = n$

stratum	source of variation	df	SS	EMS
mean		1	$\mathbf{y}'\mathbf{C}_\mu\mathbf{y}$	$r\|\|\mathbf{T}_\mu\mathbf{t}\|\|^2 + \xi_\mu$
replicates		$r - 1$	$\mathbf{y}'\mathbf{C}_R\mathbf{y}$	ξ_R
blocks	V_s	$(n-1)^2$	$\lambda_{Bs}\|\|\mathbf{h}_{Bs}\|\|^2$	$\dfrac{\lambda_{Bs}\|\|\mathbf{T}_s\mathbf{t}\|\|^2}{(n-1)^2} + \xi_B$
	error	$n - 1$	RSS_B	ξ_B
	total	$n(n-1)$	$\mathbf{y}'\mathbf{C}_B\mathbf{y}$	
plots	V_s	$(n-1)^2$	$\lambda_{\varepsilon s}\|\|\mathbf{h}_{\varepsilon s}\|\|^2$	$\dfrac{\lambda_{\varepsilon s}\|\|\mathbf{T}_s\mathbf{t}\|\|^2}{(n-1)^2} + \xi_\varepsilon$
	V_ε	$n - 2$	$\lambda_{\varepsilon\varepsilon}\|\|\mathbf{h}_{\varepsilon\varepsilon}\|\|^2$	$\dfrac{\lambda_{\varepsilon\varepsilon}\|\|\mathbf{T}_\varepsilon\mathbf{t}\|\|^2}{n-2} + \xi_\varepsilon$
	error	$(n-1)(n^2-2n-1)$	RSS_ε	ξ_ε
	total	$n^2(n-2)$	$\mathbf{y}'\mathbf{C}_\varepsilon\mathbf{y}$	

Usually the stratum variances ξ_α are not known. If $d_\alpha \neq 0$ then $\mathrm{RSS}_\alpha/d_\alpha$ provides an unbiased estimate of ξ_α, but in general such estimates are based on too few degrees of freedom, because one or more treatment terms have been fitted and removed in more than one stratum. For a rectangular lattice design with $r < n$ there is no such estimate of ξ_B, because $d_B = 0$.

The solution to this difficulty is to estimate the stratum variances and the weights simultaneously. With the weighted fitted value $\hat{\mathbf{t}}$ given above, the sum of squares, R_α, for the residual in stratum W_α is given by

$$R_\alpha = \mathrm{RSS}_\alpha + \sum_\theta \lambda_{\alpha\theta} \sum_\beta \sum_\gamma w_{\beta\theta} w_{\gamma\theta} \langle \mathbf{h}_{\alpha\theta} - \mathbf{h}_{\beta\theta}, \mathbf{h}_{\alpha\theta} - \mathbf{h}_{\gamma\theta} \rangle, \tag{6.3}$$

with expected value $d'_\alpha \xi_\alpha$, where

$$d'_\alpha = \dim(W_\alpha) - \sum_\theta{}' w_{\alpha\theta} \dim(V_\theta). \tag{6.4}$$

Equating observed and expected values of the R_α gives a set of equations in the ξ_α. As Nelder (1968) observed, (6.3) simplifies considerably when there are only two effective strata. Thus for rectangular lattice designs we obtain the following equations for ξ_B and ξ_ε:

$$\mathrm{RSS}_B + \sum_\theta \lambda_{B\theta} w_{\varepsilon\theta}^2 \|\mathbf{h}_{B\theta} - \mathbf{h}_{\varepsilon\theta}\|^2 = \xi_B \left[r(n-1) - \sum_\theta{}' w_{B\theta} \dim(V_\theta) \right],$$

$$\mathrm{RSS}_\varepsilon + \sum_\theta \lambda_{\varepsilon\theta} w_{B\theta}^2 \|\mathbf{h}_{\varepsilon\theta} - \mathbf{h}_{B\theta}\|^2 = \xi_\varepsilon \left[rn(n-2) - \sum_\theta{}' w_{\varepsilon\theta} \dim(V_\theta) \right].$$

Note that RSS_B is zero when $r < n$, and that the weights $w_{\alpha\theta}$ also involve the unknown ξ_α. However, these equations may be solved, iteratively if necessary, to

give us estimates ξ_B and ξ_ε, which, under normality, correspond to the so-called *restricted maximum likelihood estimates*, and these may be used to give the best available estimates of linear combinations $\langle \mathbf{x}, \mathbf{t} \rangle$ and the estimated variances of those estimates.

It is clear that the analysis depends on the availability of the projection operators $\mathbf{C}_\alpha$ and $\mathbf{T}_\theta$. The former are quite standard, and correspond to fitting and removing the grand mean, replicate means, and block means. The latter are given by the fan and spoke totals, and so are straightforward to calculate, even by hand. If the statistical programming language GENSTAT is used, spoke totals are automatically calculated if r treatment pseudo-factors are declared, one for each classification: the levels of the ath pseudo-factor are the a-spokes. An alternative strategy is to input r copies of the data and use just two treatment pseudo-factors, FAN and SPOKE. In the ath copy, treatments in spoke $\mathscr{S}_{ai}$ are declared to have level i of FAN and level a of SPOKE. The treatment declaration FAN/SPOKE ensures that all the correct major calculations are done, using the *sweeps* of Wilkinson (1970), although minor adjustments have to be made to the output to allow for the multiple copies. Thompson (1983) explains this method, and its difficulties, in more detail, using the general methods of Thompson (1984), and shows that this type of pseudo-factorial structure is also useful for diallel experiments.

Thus, apart from the use of estimated weights because the stratum variances are in general not known, a completely satisfactory analysis of any rectangular lattice design can be made once the operators $\mathbf{T}_\theta$ are available. Given these, the analysis is analogous to that of a balanced incomplete block design with recovery of interblock information.

Williams and Ratcliff (1980) gave a procedure for the analysis of rectangular lattice designs which differs from ours in two respects. In the first place, their covariance model is of the form

$$\mathrm{Cov}[(\mathbf{I} - \mathbf{P}_R)\mathbf{y}] = \gamma_B \mathbf{P}_B + \gamma_\varepsilon \mathbf{I},$$

which differs from our equation (5.1). Secondly, our iterative analysis ensures that the final estimates of ξ_B, ξ_ε and the treatment effects are consistent with each other, while the Williams–Ratcliff procedure, which is based on that given by Yates (1940) and Cochran and Cox (1957, Section 1.3), is, roughly speaking, only the first cycle of the restricted maximum likelihood analysis of Patterson and Thompson (1971). The differences between these methods, which apply not only to rectangular lattice designs, will be discussed in more detail elsewhere.

7. Rectangular lattices with cross-blocking. The foregoing ideas may be extended to a more complicated block structure.

In Example 2 we have so far ignored the periods. However, it was desirable that each treatment should be fed once in each period. The experimenter concerned found that, for the rectangular lattice design constructed at the end of Section 3, the treatments could be permuted within sheep so that each treatment occurred once in each period: his proposed design is shown in Table 11.

Unfortunately, this design takes no account of the grouping of the 36 experimental units into nine room-periods: each room-period consists of the four observations made in the same test period in the same room. In the notation of Nelder (1965a), the block structure is

$$3 \text{ periods} \times (3 \text{ rooms} \rightarrow 4 \text{ sheep}).$$

The stratum projection matrices are given by

$$\begin{aligned}
\mathbf{C}_\mu &= \mathbf{P}_\mu, \\
\mathbf{C}_R &= \mathbf{P}_R - \mathbf{P}_\mu, \\
\mathbf{C}_P &= \mathbf{P}_P - \mathbf{P}_\mu, \\
\mathbf{C}_{RP} &= \mathbf{P}_{RP} - \mathbf{P}_P - \mathbf{P}_R + \mathbf{P}_\mu, \\
\mathbf{C}_S &= \mathbf{P}_S - \mathbf{P}_R, \\
\mathbf{C}_\varepsilon &= \mathbf{P}_\varepsilon - \mathbf{P}_S - \mathbf{P}_{RP} + \mathbf{P}_R,
\end{aligned}$$

where, for example, $\mathbf{P}_{RP}$ is the averaging matrix for room-periods. Although V_μ, V_f, V_s, and V_ε are eigenspaces of $\mathbf{C}_\mu$, $\mathbf{C}_R$, $\mathbf{C}_P$, and $\mathbf{C}_S$, they are *not* eigenspaces of $\mathbf{C}_{RP}$ and $\mathbf{C}_\varepsilon$, because the block design given by the room-periods alone is not in any sense balanced with respect to the treatment decomposition $V_\mu \oplus V_f \oplus V_s \oplus V_\varepsilon$. Thus the design is not generally balanced.

However, it is possible to permute the treatments given to each sheep so that each treatment occurs once in each period and the design is generally balanced. This may be done for $n(n-1)$ treatments in the simple orthogonal block structure

$$(n-1) \text{ periods} \times [(n-1) \text{ rooms} \rightarrow n \text{ sheep}]$$

as follows. Ignoring periods, the design is constructed from a set of mutually orthogonal Latin squares $\Lambda_1, \ldots, \Lambda_{n-1}$, as in Section 2. A supplementary $(n-1) \times (n-1)$ Latin square Δ is needed, whose letters are the remaining letters of Λ_{n-2}. Let δ_{ap} be the letter in row a and column p of Δ. Then the treatment in the pth period and the ith animal of the ath room is the unique treatment which is in spoke $\mathscr{S}_{ai}$ and in letter δ_{ap} of Λ_{n-2}. In our particular example we may take the supplementary square Δ shown in Table 18: the resulting design is in Table 19.

In the notation of Section 3, V_ε is the main effect of Q_1, where the levels of Q_1 are the remaining letters of Λ_{n-2}. By our construction, Q_1 is completely confounded with room-periods, while all treatment vectors which are orthogonal to Q_1 are also orthogonal to room-periods. Hence the efficiency factors for this extension of the rectangular lattice design are those shown in Table 20.

TABLE 18
Supplementary Latin square

2	3	4
3	4	2
4	2	3

RECTANGULAR LATTICE DESIGNS

TABLE 19
Generally balanced design for [periods × (rooms → sheep)]

room		1				2				3			
sheep		1	2	3	4	5	6	7	8	9	10	11	12
time period	1	*A*	*D*	*I*	*L*	*G*	*K*	*B*	*F*	*H*	*J*	*C*	*E*
	2	*B*	*F*	*G*	*K*	*J*	*H*	*E*	*C*	*L*	*I*	*D*	*A*
	3	*C*	*E*	*H*	*J*	*D*	*A*	*L*	*I*	*F*	*B*	*K*	*G*

TABLE 20
Efficiency factors of an extended rectangular lattice design

	treatment subspace			
	V_μ	$V_f = Q_2$	$V_s = Q_1Q_2$	$V_e = Q_1$
stratum				
mean W_μ	1	0	0	0
rooms W_R	0	0	0	0
periods W_P	0	0	0	0
room-periods W_{PR}	0	0	0	1
sheep W_s	0	$\dfrac{1}{(n-1)^2}$	$\dfrac{n}{(n-1)^2}$	0
units W_e	0	$\dfrac{n(n-2)}{(n-1)^2}$	$\dfrac{n^2-3n+1}{(n-1)^2}$	0

Acknowledgments. The contribution of T.P.S. to the work reported here was carried out whilst he was a visitor at the Indian Statistical Institute in Calcutta, and he would like to thank the then Acting Director, Dr. A. Maitra, and other staff and scholars, for providing such an enjoyable environment for study. Part of R.A.B.'s contribution was made while she was visiting the Mathematics Department of the University of Western Australia, to whose staff she would like to extend similar thanks. We are grateful to A. Grassia of C.S.I.R.O., Perth, for drawing our attention to the problem described in Example 2.

REFERENCES

BAILEY, R. A. (1981). A unified approach to design of experiments. *J. Roy. Statist. Soc. Ser. A* **144** 214–223.

BRUCK, R. H. (1963). Finite nets. II. Uniqueness and embedding. *Pacific J. Math.* **13** 421–457.

COCHRAN, W. G. and COX, G. M. (1957). *Experimental Designs*, 2nd ed. Wiley, New York.

CORSTEN, L. C. A. (1976). Canonical correlation in incomplete blocks. In *Essays in Probability and Statistics. A Volume in Honour of Prof. Junjiro Ogawa* (S. Ikeda et al., eds.) 125–154. Shinko Tsusho Co. Ltd., Tokyo.

DÉNES, J. and KEEDWELL, A. D. (1974). *Latin Squares and Their Applications*. English Universities Press Limited, London.

FISHER, R. A. and YATES, F. (1934). The 6 × 6 Latin squares. *Proc. Cambridge Philos. Soc.* **30** 492–507.

GRUNDY, P. M. (1950). The estimation of error in rectangular lattices. *Biometrics* **6** 25–33.

HARSHBARGER, B. (1946). Preliminary report on the rectangular lattices. *Biometrics* **2** 115–119.

HARSHBARGER, B. (1947). *Rectangular Lattices.* Virginia Agricultural Experiment Station Memoir 1.

HARSHBARGER, B. (1949). Triple rectangular lattices. *Biometrics* **5** 1–13.

HARSHBARGER, B. (1951). Near balance rectangular lattices. *Virginia J. Sci.* **2** 13–27.

HARSHBARGER, B. and DAVIS, L. L. (1952). Latinized rectangular lattices. *Biometrics* **8** 73–84.

HARVILLE, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems (with discussion). *J. Amer. Statist. Assoc.* **72** 320–340.

HOUTMAN, A. M. and SPEED, T. P. (1983). Balance in designed experiments with orthogonal block structure. *Ann. Statist.* **11** 1069–1085.

JAMES, A. T. and WILKINSON, G. H. (1971). Factorization of the residual operator and canonical decomposition of nonorthogonal factors in the analysis of variance. *Biometrika* **58** 279–294.

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

KEMPTHORNE, O. (1982). Classificatory data structures and associated linear models. In *Statistics and Probability: Essays in Honor of C. R. Rao* (G. Kallianpur, P. R. Krishnaiah and J. K. Ghosh, eds.) 397–410. North-Holland, Amsterdam.

MANN, H. B. (1942). The construction of orthogonal Latin squares. *Ann. Math. Statist.* **13** 418–423.

NAIR, K. R. (1951). Rectangular lattices and partially balanced incomplete block designs. *Biometrics* **7** 145–154.

NAIR, K. R. (1952). Analysis of partially balanced incomplete block designs illustrated on the simple square and rectangular lattices. *Biometrics* **8** 122–155.

NAIR, K. R. (1953). A note on rectangular lattices. *Biometrics* **9** 101–106.

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

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

NELDER, J. A. (1968). The combination of information in generally balanced designs. *J. Roy. Statist. Soc. Ser. B* **30** 303–311.

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

PATTERSON, H. D. and WILLIAMS, E. R. (1976). A new class of resolvable incomplete block designs. *Biometrika* **63** 83–92.

PEARCE, S. C., CALIŃSKI, T. and MARSHALL, T. F. DE C. (1974). The basic contrasts of an experimental design with special reference to the analysis of data. *Biometrika* **61** 449–460.

RAGHAVARAO, D. (1971). *Constructions and Combinatorial Problems in Design of Experiments.* Wiley, New York.

ROBINSON, H. F. and WATSON, G. S. (1949). Analysis of simple and triple rectangular lattice designs. North Carolina Agricultural Experimental Station Tech. Bull. 88.

SHRIKHANDE, S. S. (1961). A note on mutually orthogonal latin squares. *Sankhyā Ser. A* **23** 115–116.

SPEED, T. P. and BAILEY, R. A. (1982). On a class of association schemes derived from lattices of equivalence relations. In *Algebraic Structures and Applications* (P. Schultz, C. E. Praeger and R. P. Sullivan, eds.) 55–74. Marcel Dekker, New York.

THOMPSON, R. (1983). Diallel crosses, partially balanced incomplete block designs with triangular association schemes and rectangular lattices. *Genstat Newsletter* **10** 16–32.

THOMPSON, R. (1984). The use of multiple copies of data in forming and interpreting analysis of variance. In *Experimental Design, Statistical Models, and Genetic Statistics. Essays in Honor of Oscar Kempthorne* (K. Hinkelmann, ed.) 155–173. Marcel Dekker, New York.

THROCKMORTON, T. N. (1961). Structures of classificatory data. Ph.D. thesis, Iowa State Univ.

WILKINSON, G. N. (1970). A general recursive algorithm for analysis of variance. *Biometrika* **57** 19–46.

WILLIAMS, E. R. (1977). A note on rectangular lattice designs. *Biometrics* **33** 410–414.

WILLIAMS, E. R. and RATCLIFF, D. (1980). A note on the analysis of lattice designs with repeats. *Biometrika* **67** 706–708.

YATES, F. (1936). A new method of arranging variety trials involving a large number of varieties. *J. Agric. Sci.* **26** 424–455.

YATES, F. (1940). The recovery of inter-block information in balanced incomplete block designs. *Ann. Eugenics* **10** 317–325.

STATISTICS DEPARTMENT
ROTHAMSTED EXPERIMENTAL STATION
HARPENDEN
HERTFORDSHIRE AL5 2JQ
ENGLAND

C.S.I.R.O.
DIVISION OF MATHEMATICS AND STATISTICS
G.P.O. BOX 1965
CANBERRA ACT 2601
AUSTRALIA

The Annals of Statistics
1987, Vol. 15, No. 3, 885-910

SPECIAL INVITED PAPER

WHAT IS AN ANALYSIS OF VARIANCE?

BY T. P. SPEED

CSIRO, Canberra, Australia

The analysis of variance is usually regarded as being concerned with sums of squares of numbers and independent quadratic forms of random variables. In this paper, an alternative interpretation is discussed. For certain classes of dispersion models for finite or infinite arrays of random variables, a form of generalized spectral analysis is described and its intuitive meaning explained. The analysis gives a spectral decomposition of each dispersion in the class, incorporating an analysis of the common variance, and an associated orthogonal decomposition of each of the random variables. One by-product of this approach is a clear understanding of the similarity between the spectral decomposition for second-order stationary processes and the familiar linear models with random effects.

"... the analysis of variance, which may perhaps be called a statistical method, because the term is a very ambiguous one—is not a mathematical theorem, but rather a convenient method of arranging the arithmetic."

R. A. Fisher (1934)

1. Introduction. To most of us the expression analysis of variance or anova conjures up a subset of the following: multiindexed arrays of numbers, sums of squares, anova tables with lines; perhaps, somewhat more mathematically, independent quadratic forms of random variables, chi-squared distributions, and F-tests. We would also think of linear models and the associated notions of main effects and interactions of various orders; indeed the standard text on the subject, Scheffé (1959, page 5) essentially defines the analysis of variance to be regression analysis where the regressor variables (x_{ij}) take only the values 0 or 1, although he mentions in a footnote that -1 and 2 have also arisen. What is anova? Is there a variance being analysed? Is there a mathematical theorem, contrary to Fisher's assertion? Or is it just a body of techniques, a statistical method,..., a convenient method of arranging the arithmetic?

Signs that there might be an underlying mathematical structure began to appear in the late 1950s and early 1960s. James (1957) emphasised the role of the algebra of projectors in the analysis of experimental designs, Tukey (1961) outlined the connection between anova and spectrum analysis [something which was made more explicit by Hannan (1961, 1965), who focussed on the decomposition of permutation representations of groups], whilst Graybill and Hultquist

Received February 1986; revised August 1986.
AMS 1980 *subject classification.* 62J10.
Key words and phrases. Analysis of variance, dispersion model, sums of squares, association scheme, group, symmetry, manova.

(1961) gave a definition of anova (assuming joint normality of all random variables concerned) which incorporated many of the same ideas as the others mentioned: the commuting of projectors and the spectral decomposition of a covariance matrix.

Of course, anova is just a word (or three) and people can give it any meaning they wish, so there is no sense in which the definition I offer in the following text has any greater claim to be the correct one than any other. What I do believe is that it is a mathematically fruitful definition, that it covers most if not all situations which statisticians would regard as being instances of anova and that its generality and simplicity are both pedagogically and scientifically helpful. And yes, I believe there are relevant mathematical theorems, although as we will see it is perhaps unreasonable to expect a single theorem to cover all existing cases.

2. Two simple examples. Let us begin with an array $y = (y_{ij})$ of mn random variables where $i = 1, \ldots, m$ and $j = 1, \ldots, n$ is nested within i, i.e., j only has meaning within the values of i. The following decomposition of the sum of squares is familiar to all who have met anova:

$$\sum_k \sum_l y_{kl}^2 = mny_{..}^2 + n\sum_h (y_{h.} - y_{..})^2 + \sum_i \sum_j (y_{ij} - y_{i.})^2, \tag{2.1}$$

and we denote the three terms on the right by SS_0, SS_1 and SS_2. Here $y_{i.} = n^{-1}\sum_j y_{ij}$, $y_{..} = m^{-1}\sum_i y_{i.}$, etc. It is not hard to derive (2.1) by the standard juggling which many believe characterises anova. Of what interest or use is this decomposition? To answer this question, we must make some assumptions about the y_{ij}, and one set—the ones Fisher (1934) probably had in mind when he made the remark quoted—is the following: $Ey_{ij} = \mu_i$, where (μ_i) is a set of m unknown parameters, the (y_{ij}) are pairwise uncorrelated and they have a common variance σ^2; i.e., the dispersion matrix Dy of y is just $\sigma^2 I$. Under these assumptions we can prove (see the following text) that $E\{SS_0\} = mn\mu_.^2 + \sigma^2$, $E\{SS_1\} = (m-1)\sigma^2 + n\sum_i(\mu_i - \mu_.)^2$ and $E\{SS_2\} = m(n-1)\sigma^2$. It is here that we can see the point of Fisher's remark about "the arithmetic," for when the (y_{ij}) are jointly normal, SS_0/σ^2, SS_1/σ^2 and SS_2/σ^2 are mutually independent with chi-squared distributions on 1, $m-1$ and $m(n-1)$ degrees of freedom, respectively, and the ratio $F = m(n-1)SS_1/(m-1)SS_2$ permits a test of the hypothesis H: $\mu_1 = \mu_2 = \cdots = \mu_m$, having a central F-distribution with $(m-1, m(n-1))$ degrees of freedom when H is true. The F-test of this hypothesis has many desirable properties [Hsu (1941, 1945), Wald (1942), Wolfowitz (1949), Herbach (1959) and Gautschi (1959)] and the decomposition (2.1) is indeed a convenient method of arranging the arithmetic.

But all of this is just sums of squares—quadratic forms in normal variates if you wish; the only variance in sight is the common σ^2 and that does not appear to be undergoing any analysis. However, let us look closely at the proof of some of the foregoing assertions. How do we see that the quadratic forms SS_0, SS_1 and SS_2 are independent under the assumption $Dy = \sigma^2 I$ and joint normality? One approach, owing to Tang (1938), uses the fact that their (unsquared and un-

summed) components $y..$, $y_{h\cdot} - y..$ and $y_{ij} - y_{i\cdot}$ are uncorrelated, and hence, by the joint normality, independent, and this property is retained when the components are squared and summed.

How do we see that these components are uncorrelated? Each is a linear combination of elements in the array y with easily calculated coefficients and, with the assumption that $Dy = \sigma^2 I$, their covariances are simply σ^2 times the sums of the products of these coefficients. For example, the coefficient of y_{kl} in $y_{h\cdot} - y..$ is $-1/mn$ if $k \neq h$ and $1/n - 1/mn$ if $k = h$, whilst that of y_{kl} in $y_{ij} - y_{i\cdot}$ is 0 if $k \neq i$, $-1/n$ if $k = i$ and $l \neq j$ and $1 - 1/n$ if $k = i$ and $l = j$. Thus if $h = i$,

$$\operatorname{cov}(y_{h\cdot} - y.., y_{ij} - y_{i\cdot})$$

$$= \sigma^2\left[-\frac{1}{mn}0(m-1)n + \left(\frac{1}{n} - \frac{1}{mn}\right)\left(-\frac{1}{n}\right)(n-1) + \left(\frac{1}{n} - \frac{1}{mn}\right)\left(1 - \frac{1}{n}\right)1\right],$$

which is zero as stated; the case $h \neq i$ is dealt with similarly. Similar calculations prove that $\operatorname{cov}(y.., y_{h\cdot} - y..) = \operatorname{cov}(y.., y_{ij} - y..) = 0$ and, further, that $E\{y_{\cdot\cdot}^2\} = \mu_{\cdot}^2 + (1/mn)\sigma^2$, $E\{(y_{h\cdot} - y..)^2\} = ((m-1)/mn)\sigma^2 + (\mu_{h\cdot} - \mu_{\cdot})^2$ and $E\{(y_{ij} - y_{i\cdot})^2\} = (m(n-1)/mn)\sigma^2$.

It has just been proved that the three components in the sum

$$y_{ij} = y.. \quad + \quad y_{i\cdot} - y.. \quad + \quad y_{ij} - y_{i\cdot} \tag{2.2}$$

are uncorrelated; their variances thus add and we may write this as

$$\sigma^2 = \frac{1}{mn}\sigma^2 + \frac{m-1}{mn}\sigma^2 + \frac{m(n-1)}{mn}\sigma^2. \tag{2.3}$$

Here at last is a variance being analysed! But before we examine this any further let us see with a minimum of further algebra how the sums of squares of the components in (2.2) must add up and give (2.1). Denoting the coefficients of y_{kl} in $y..$, $y_{i\cdot} - y..$ and $y_{ij} - y_{i\cdot}$ by $S_0(ij, kl)$, $S_1(ij, kl)$ and $S_2(ij, kl)$, respectively, we can easily check that the $mn \times mn$ matrices S_0, S_1 and S_2 so defined are symmetric, idempotent, pairwise orthogonal and sum to the $mn \times mn$ identity matrix I. Symmetry is quickly apparent from their definition; orthogonality is implicit in the calculation which proved the components in (2.2) uncorrelated, whilst idempotence is proved by a similar calculation; and clearly they sum to the identity. Thus we can write $y = S_0 y + S_1 y + S_2 y$ as

$$(y_{ij}) = (y..) + (y_{i\cdot} - y..) + (y_{ij} - y_{i\cdot}), \tag{2.4}$$

where the S_α act on arrays $u = (u_{ij})$ of real numbers as follows $(S_\alpha u)_{ij} = \Sigma_k \Sigma_l S_\alpha(ij, kl) u_{kl}$, $\alpha = 0, 1, 2$. But then (2.4) is a decomposition of the array into component arrays which are orthogonal with respect to the inner product $\langle u, v\rangle = \Sigma_i \Sigma_j u_{ij} v_{ij}$, whilst (2.1) is simply the Pythagorean relationship

$$|y|^2 = |S_0 y|^2 + |S_1 y|^2 + |S_2 y|^2,$$

where $|y|^2 = \langle y, y\rangle$ is the associated squared norm.

An unexpected bonus. Without any further calculations we may assert that (2.2) remains an orthogonal decomposition of y_{ij} when the dispersion matrix $Dy = \Gamma$ has the form

$$\Gamma = \xi_0 S_0 + \xi_1 S_1 + \xi_2 S_2, \tag{2.5}$$

where the eigenvalues ξ_0, ξ_1 and ξ_2 are positive real numbers. A modified version of (2.3) also holds, namely

$$\operatorname{var}(y_{ij}) = \frac{1}{mn}\xi_0 + \frac{m-1}{mn}\xi_1 + \frac{m(n-1)}{mn}\xi_2. \tag{2.6}$$

These assertions are readily checked. For example,

$$\operatorname{cov}(y_{i\cdot} - y_{\cdot\cdot}, y_{ij} - y_{i\cdot}) = (S_1\Gamma S_2)(ij, ij) = 0,$$

and

$$\operatorname{var}(y_{ij} - y_{i\cdot}) = (S_2\Gamma S_2)(ij, ij) = \xi_2 S_2(ij, ij) = \frac{m(n-1)}{mn}\xi_2.$$

The question this observation now raises is: How wide is the class of matrices of the form (2.5)? Perhaps unexpectedly, it coincides with a class which arises frequently, namely the set of all matrices Γ having the form

$$\Gamma = \gamma_2 A_2 + \gamma_1 A_1 + \gamma_0 A_0, \tag{2.7}$$

where $A_2 = I$ is the identity matrix, $A_1(ij, kl) = 1$ if $i = k$, $j \neq l$ and 0 otherwise, $A_0(ij, kl) = 1$ if $i \neq k$ and 0 otherwise, and γ_2, γ_1 and γ_0 are a variance and two covariances constrained only to ensure that Γ is positive definite. The easiest way to see that Γ's of the form (2.5) and (2.7) coincide is to list the index ij lexicographically and write the matrices in tensor product form. We find that $A_2 = I_m \otimes I_n$, $A_1 = I_m \otimes (J_n - I_n)$ and $A_0 = (J_m - I_m) \otimes J_n$, whilst $S_0 = (1/m)J_m \otimes (1/n)J_n$, $S_1 = (I_m - (1/m)J_m) \otimes (1/n)J_n$ and $S_2 = I_m \otimes (I_n - (1/n)J_n)$, where I_m and J_m are the $m \times m$ identity and matrix of 1's, respectively. The eigenvalues ξ and the entries γ correspond in the following way:

$$\begin{bmatrix} \xi_0 \\ \xi_1 \\ \xi_2 \end{bmatrix} = \begin{bmatrix} 1 & n-1 & n(m-1) \\ 1 & n-1 & -n \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} \gamma_2 \\ \gamma_1 \\ \gamma_0 \end{bmatrix}, \tag{2.8a}$$

$$\begin{bmatrix} \gamma_2 \\ \gamma_1 \\ \gamma_0 \end{bmatrix} = \frac{1}{mn} \begin{bmatrix} 1 & m-1 & m(n-1) \\ 1 & m-1 & -m \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} \xi_0 \\ \xi_1 \\ \xi_2 \end{bmatrix}. \tag{2.8b}$$

Where have we gotten to? We have exhibited a set of covariance matrices (2.7) for a random array $y = (y_{ij})$ which are simultaneously diagonalisable, cf. (2.5); their eigenvalues are invertible linear combinations (2.8) of their entries; their common eigenspace projectors decompose the elements of the array into statistically orthogonal (i.e., uncorrelated) components (2.2) whilst also decomposing the arrays themselves into geometrically orthogonal arrays (2.4). Pythagoras' theorem applied to the decomposition of array elements gives an analysis of variance

qua variance (2.6), whilst it gives the sum of squares decomposition (2.1) of an anova table when applied to the decomposition of arrays. We might also add that these decompositions all make "statistical sense."

How special is this example? Before answering this question let us look at a second example, which is not normally regarded as being an instance of anova. This time our array has a circular nature: A sequence $y = (y_t: t = 0, 1, \ldots, n-1)$ of $n = 2m + 1$ random variables with $\operatorname{cov}(y_s, y_t) = \gamma_{|t-s|}$, $0 \le s, t < n$, i.e., $\Gamma = Dy$ is a symmetric circulant with first row $(\gamma_0\gamma_1 \cdots \gamma_m\gamma_m \cdots \gamma_1)$. To emphasize the similarity with (2.7) we write it as

$$\Gamma = \sum_0^m \gamma_a A_a, \tag{2.9}$$

where A_a is the symmetric circulant having first row $(0 \cdots 010 \cdots 010 \cdots 0)$ with 1's in the ath and $(n-a)$th position, $1 \le a \le m$, and $A_0 = I$, the $n \times n$ identity matrix. It is well known that the class of all such matrices is simultaneously diagonalisable with common projectors $S_0 = (1/n)J_n$ and $S_\alpha(s, t) = (2/n)\cos(2\pi(s-t)\alpha/n)$, $0 < \alpha \le m$, $0 \le s, t < n$, whilst their eigenvalues are linear combinations of their entries

$$\xi_\alpha = \gamma_0 + 2\sum_1^m \gamma_a \cos\left(\frac{2\pi}{n} a\alpha\right), \qquad \alpha = 0, \ldots, m, \tag{2.10a}$$

with inverses

$$\gamma_a = \frac{1}{n}\xi_0 + \frac{2}{n}\sum_1^m \xi_\alpha \cos\left(\frac{2\pi}{n} a\alpha\right), \qquad a = 0, \ldots, m. \tag{2.10b}$$

Further, we have an orthogonal decomposition of the random variables similar to (2.2):

$$y_t = y_\cdot + \sum_1^m S_\alpha y_t, \tag{2.11}$$

where $S_\alpha y_t = (2/n)\sum_0^{n-1} y_s \cos(2\pi(s-t)\alpha/n)$, $1 \le \alpha \le m$, cf. Hannan (1960, I.2), and the variances of each component add, corresponding to $a = 0$ in (2.10b).

Finally, we remark that a decomposition of the n-dimensional vector space analogous to (2.4) and its associated sum of squares decomposition may also be derived; it is just the (real form of the) discrete Fourier transform. The analogy with the view of the classical anova we have just presented is complete.

3. Sums of squares. Let $y = (y_t: t \in T)$ be a finite array of random variables with mean $Ey = 0$ and dispersion matrix $Dy = \Gamma \in \mathbf{V}$, where $\mathbf{V}$ is a family of positive definite matrices over T. The formal definition of anova given by Graybill and Hultquist (1961) refers to a decomposition of $|y|^2$ into a sum of quadratic forms under an assumption of joint normality of y. It had two aspects which we will recall shortly: one which in essence refers to properties of the individual matrices $\Gamma \in \mathbf{V}$, and one which was clearly a property of the model as

a whole. Later writers on the same topic include Albert (1976), Brown (1984) and Harville (1984), and in all of these papers the role of anova as a property of a model $\mathbf{V}$ has tended to get emphasised less than the consequences of the definition for arrays y with $Dy \in \mathbf{V}$. In what follows we modify the Graybill and Hultquist (1961) definition slightly, removing some details without, we hope, losing its essence. We also express the definition solely in terms of the class $\mathbf{V}$ of dispersion matrices, removing the joint normality assumption. Finally we argue that the definition is most fruitful when applied to a particular parametrization of $\mathbf{V}$, one which is not usual in this context, although as we will see it coincides with that used in developing the spectral theory of second-order stationary processes over index sets of various kinds.

Initially we will suppose that $\mathbf{V}$ is a class of positive definite matrices having the form

$$\Gamma(\theta) = \sum_{a=1}^{s} \theta_a A_a, \tag{3.1}$$

where the $\{A_a\}$ are known symmetric matrices and $\theta = (\theta_a)$ is an s-dimensional real parameter belonging to $\Theta \subset R^s$. It will be convenient to suppose that the $\{A_a\}$ are linearly independent matrices over T and that $\mathbf{V}$ contains s linearly independent elements. Dispersion models of this form have been studied by a number of authors over the years including Anderson (1969, 1970, 1973) and Jensen (1975), but our emphasis is quite different from theirs. Essentially following Graybill and Hultquist (1961) we say that an anova exists for $\mathbf{V}$ if there exists a family $\{S_\alpha\}$ of s known pairwise orthogonal symmetric idempotent matrices summing to the identity matrix I over T such that

(a) for every $\theta \in \Theta$ and α there exists $\xi_\alpha(\theta)$ such that

$$\Gamma(\theta)S_\alpha = \xi_\alpha(\theta)S_\alpha; \tag{3.2}$$

(b) the map $\theta = (\theta_a) \to \xi(\theta) = (\xi_\alpha(\theta))$ is linear and invertible.

Condition (a) replaces the condition that for each $\theta \in \Theta$ the s quadratic forms $\{|S_\alpha y|^2\}$ are mutually independent scale multiples of chi-squares under the assumption $y \sim N(0, \Gamma(\theta))$ [see Albert (1976, Theorem 1(a))], whilst condition (b) asserts that the multipliers $\xi_\alpha(\theta) = E\{d_\alpha^{-1}|S_\alpha y|^2\}$, where $d_\alpha = \operatorname{rank} S_\alpha$, are independent linear functions of the $\{\theta_a\}$.

It is clear from (a) that the matrices $\{S_\alpha\}$ simultaneously reduce all $\Gamma \in \mathbf{V}$, i.e., that $\Gamma = \sum_\alpha \xi_\alpha S_\alpha$, where we omit the dependence on θ if no confusion can result, and thus every element of $\mathbf{V}$ commutes with every other. As long as $\mathbf{V}$ contains s linearly independent elements, these conclusions extend to *all* matrices of the form $\sum_a \theta_a A_a$ with $\theta \in R^s$ and in particular we deduce that the $\{A_a\}$ commute. It also follows from (b) that, in general, $\Gamma(\theta)$ has s distinct eigenvalues.

Conversely, if the $\{A_a\}$ all commute, a well known theorem in linear algebra tells us that there is a family $\{S_\alpha\}$ of t (say) pairwise orthogonal symmetric idempotent matrices summing to I such that $A_a S_\alpha = p_{\alpha a} S_\alpha$ for constants $p_{\alpha a}$, $\alpha = 1, \ldots, t$, $a = 1, \ldots, s$. It follows that an element $\Gamma \in \mathbf{V}$ will have spectral

decomposition $\Gamma = \Sigma_\alpha \xi_\alpha S_\alpha$, where $\xi_\alpha = \Sigma_a p_{\alpha a}\theta_a$, and if, in general, such a Γ has s distinct eigenvalues, then we deduce that $t = s$ and that $P = (p_{\alpha a})$ is an invertible $s \times s$ matrix.

Where have we gotten to? Without giving full details we have seen the reason why the preceding (a) and (b) are jointly equivalent to the two conditions

(c) the matrices $\{A_a\}$ commute,
(d) in general, $\Gamma(\theta)$ has s distinct eigenvalues.

This is in essence the content of Graybill and Hultquist (1961, Theorem 6). Note that under (c) and (d) we can write $A_a = \Sigma_\alpha p_{\alpha a} S_\alpha$ and $S_\alpha = (1/n)\Sigma_a q_{a\alpha} A_a$, where we have inserted a scale factor $n = |T|$ for later convenience, and where $\Sigma_a p_{\alpha a} q_{a\beta} = n\delta_\beta^\alpha$ and $\Sigma_\alpha q_{a\alpha} p_{\alpha b} = n\delta_b^a$, δ here being Kronecker's delta. These equations combine to give

$$A_a A_b = A_b A_a = \sum_c \left\{(1/n)\sum_\alpha p_{\alpha a} p_{\alpha b} q_{c\alpha}\right\} A_c, \tag{3.3}$$

implying that $\mathbf{V}$ may be extended to the linear algebra generated by the $\{A_a\}$ without invalidating anything we have said to date.

If the $\{A_a\}$ all have the property that all their row (column) sums are the same, i.e., if for each a there exists k_a such that $\Sigma_s A_a(s, t) = \Sigma_t A_a(s, t) = k_a$, then the matrix $S_0 = (1/n)J$, where J is the matrix of 1's over T, is always one of the $\{S_\alpha\}$.

Let us leave the matrices $\Gamma \in \mathbf{V}$ for a moment and turn to the elements y_t of random array $y = (y_t\colon t \in T)$ with $Dy = \Gamma \in \mathbf{V}$, still assuming that $\mathbf{V}$ satisfies (c) and (d). The prescription $S_\alpha y_t = \Sigma_s S_\alpha(s, t) y_s$ defines a family of random variables such that

$$y_t = \sum_\alpha S_\alpha y_t. \tag{3.4}$$

Now $\operatorname{cov}(S_\alpha y_t, S_\beta y_u) = (S_\alpha \Gamma S_\beta)(t, u) = \xi_\alpha S_\alpha(t, u)\delta_\beta^\alpha = 0$ if $\alpha \neq \beta$ and so the different terms on the R.H.S. of (3.4) are uncorrelated. Further $\operatorname{var}(S_\alpha y_t) = \xi_\alpha S_\alpha(t, t)$. Next suppose that $\operatorname{var}(y_t) = \sigma^2$ is the same for all $t \in T$, i.e., that the matrices $\{A_a\}$ are all constant down their diagonals. Then $S_\alpha(t, t) = n^{-1}d_\alpha$, where $d_\alpha = \operatorname{rank}(S_\alpha) = \operatorname{trace}(S_\alpha)$, and we can sum the variances in (3.4) obtaining

$$\sigma^2 = \sum_\alpha \phi_\alpha, \tag{3.5}$$

where $\phi_\alpha = n^{-1}d_\alpha \xi_\alpha = \operatorname{var}(S_\alpha y_t)$, independent of $t \in T$. Clearly this is an analysis of variance. The connection between it and the sum of squares decomposition

$$|y|^2 = \sum_\alpha |S_\alpha y|^2 \tag{3.6}$$

resulting from the geometric orthogonality of the terms in

$$y = \sum_\alpha S_\alpha y \tag{3.7}$$

is clear: The eigenvalues ξ_α are the expected mean squares:

$$\xi_\alpha = E\{d_\alpha^{-1}|S_\alpha y|^2\}. \tag{3.8}$$

Is this the correct anova? Does it have all the properties one might hope for? I would like to suggest that the answer to these questions is no, and that although the definition is basically correct, it is really only appropriate for a particular class $\{A_a\}$ of basis matrices and parameters $\{\theta_a\}$, namely, when the entries of the basis matrices are either 0 or 1 and the parameters are covariances. With this class we will find that we have a notion that extends fruitfully far beyond sums of squares.

4. Anova: Finite arrays. In this section we will sketch the most natural framework within which the special properties of our examples hold generally. The restriction to finite arrays is vital because there are many sorts of infinities and, perhaps surprisingly, no single mathematical framework is yet available which covers all the cases.

As before we begin with an array $y = (y_t: t \in T)$ of random variables indexed by a finite set T with $Ey = 0$ and we will consider a very special sort of parametrization of its dispersion matrix $\Gamma = Dy$, namely that defined by equality constraints among the elements of Γ. More fully, we will suppose that

$$\Gamma = \sum_a \gamma_a A_a, \tag{4.1}$$

where $\{A_a: a \in X\}$ is a class of matrices over T whose elements are 0 and 1 only satisfying (i) each matrix A_a is symmetric; (ii) $\Sigma_a A_a = J$, the matrix of 1's over T; (iii) one of these matrices, A_e say, is the identity matrix I over T; and (iv) there exist integers (n_{abc}), $a, b, c \in X$ such that $A_a A_b = \Sigma_c n_{abc} A_c$. Finally, $\{\gamma_a: a \in X\}$ is a set of covariances which are such that Γ given by (4.1) is positive definite.

Such matrices $\{A_a\}$ are the adjacency matrices of the association scheme over T defined by saying that s and t are a-associates, $a(s, t) = a$, say, if $A_a(s, t) = 1$, $s, t \in T$, $a \in X$; see MacWilliams and Sloane (1977, Chapter 21) for fuller background and the theory which follows.

We proceed to analyse the class of all Γ of the form (4.1). From (i) all such Γ are symmetric; from (ii) the $\{A_a\}$ are linearly independent and hence the dimension of the vector space $\mathbf{A}$ of all such Γ (forgetting positive definiteness for the moment) is $s = |X|$; from (iii) $\mathbf{A}$ contains the identity and from (iv) we deduce that $\mathbf{A}$ is a commutative algebra. The theorem in linear algebra already cited tells us that there exists a unique basis of $\mathbf{A}$ of primitive idempotents $\{S_\alpha: \alpha \in Z\}$, where $S_\alpha = S_\alpha^2 = S_\alpha'$, $S_\alpha S_\beta = S_\beta S_\alpha = 0$, $\alpha \neq \beta$, $\Sigma_\alpha S_\alpha = I$, containing $(1/n)J = S_0$, say. Further the transformation from this basis to the original one consisting of the $\{A_a\}$ is linear and invertible:

$$S_\alpha = \frac{1}{n} \sum_a q_{a\alpha} A_a, \tag{4.2a}$$

$$A_a = \sum_\alpha p_{\alpha a} S_\alpha, \tag{4.2b}$$

WHAT IS AN ANALYSIS OF VARIANCE?

where $P = (p_{\alpha a})$ and $Q = (q_{a\alpha})$ are matrices of coefficients satisfying $PQ = QP = nI$, $n = |T|$ and I here is the identity matrix of order $s = |X| = |Z|$. Since the eigenvalues of A_a are $(p_{\alpha a})$ from (4.2b), those of $\Gamma = \Sigma_a \gamma_a A_a = \Sigma_\alpha \xi_\alpha S_\alpha$ are

$$\xi_\alpha = \sum_a p_{\alpha a}\gamma_a \tag{4.3a}$$

whilst the entries γ_a of Γ in (4.1) are recoverable from the eigenvalues via

$$\gamma_a = (1/n)\sum_\alpha q_{a\alpha}\xi_\alpha. \tag{4.3b}$$

Writing $k_a = |\{t \in T\colon A_a(s, t) = 1\}|$, independent of $s \in T$, and $d_\alpha = \operatorname{rank}(S_\alpha)$, we summarise some basic facts concerning these numbers and the matrices P and Q. Here δ denotes the Kronecker delta.

THEOREM (cf. MacWilliams and Sloane 1977, Chapter 21, Section 2).

(i) $p_{\alpha e} = q_{a0} = 1$; $p_{0a} = k_a$; $q_{e\alpha} = d_\alpha$; $d_\alpha p_{\alpha a} = k_a q_{a\alpha}$.
(ii) $\Sigma_\alpha d_\alpha p_{\alpha a} p_{\alpha b} = n k_a \delta_b^a$, $\Sigma_a k_a q_{a\alpha} q_{a\beta} = n d_\alpha \delta_\beta^\alpha$.
(iii) $p_{\alpha a} p_{\alpha b} = \Sigma_c n_{abc} p_{\alpha c}$.

All of these facts give us great insight into the structure of matrices of the form (4.1) and many examples can be found in the literature; see MacWilliams and Sloane (1977) and references therein. Speed and Bailey (1982) show that all standard ("balanced complete," "orthogonal") anova models arise from such schemes where X is a modular lattice of equivalence relations on T, and the Möbius function on X (together with the number of levels of each index) determines the matrices P and Q. These results are summarized in Section 6. For most but not all classical anova models, results equivalent to the preceding were given by Nelder (1954, 1965) when Γ is induced by randomisation; see Speed (1985) for more details concerning the connexions. Early forms of (3.4) and (3.6) can be found in Kempthorne (1952, Chapter 8), again with a randomisation distribution defining Γ.

Let us turn now to the elements y_t of the array y. As in Section 3 we write $S_\alpha y_t = \Sigma_u S_\alpha(t, u) y_u$, and find that (3.4) is a decomposition of y_t into uncorrelated components which in this context satisfies

$$E\{(S_\alpha y_t)(S_\beta y_u)\} = n^{-1}\xi_\alpha q_{a(t,u)\alpha}\delta_\beta^\alpha, \tag{4.4}$$

and in particular this equals $n^{-1}d_\alpha\xi_\alpha = \phi_\alpha$ say, if $t = u$ and $\alpha = \beta$. Here $a(t, u)$ is the unique $a \in X$ such that $A_a(t, u) = 1$. With this notation we may write (4.3b) in the form

$$\gamma_a = \sum_\alpha (d_\alpha^{-1} q_{a\alpha})\phi_\alpha, \tag{4.5}$$

noting that the special case $a = e$ (the identity association) gives us the analysis of variance (3.5) corresponding to the decomposition (3.4). The index α labels the "lines" of the anova table—we call them strata—and the projectors S_α will be termed stratum projectors.

Summarising, we have seen that if $\Gamma = Dy$ has the form (4.1) where the $\{A_a\}$ satisfy conditions (i), (ii), (iii) and (iv) following (4.1), then, from Section 3, our variants (a) and (b) [equivalently, (c) and (d)] of Graybill and Hultquist's (1961) definition are certainly satisfied. Do we get anything extra which might justify our belief that it is only with these sorts of basis matrices and corresponding parameters that the term anova is appropriate? I believe we do, and make the following supporting observations:

(i) the present framework has a common variance (that to be analysed) as part of its formulation;

(ii) the $\{A_a\}$ matrices already have the property that their row (column) sums are the same, which implies that $S_0 = (1/n)J$ is one of the $\{S_\alpha\}$;

(iii) the $\{A_a\}$ matrices are all constant down their diagonal, a property which combines with (i) to give the analysis of the common variance;

(iv) we have the compact and extremely useful formula (4.4).

In the more general discussion of Section 3 each of the preceding (i), (ii) and (iii) had to be assumed in order to obtain the desired consequences, whilst (iv) shows the great simplification which results from covariance parametrization: With it, we need only know $\{A_a\}$, $\{d_\alpha\}$, $\{k_a\}$ and the function $s_\alpha(a) = k_a^{-1}p_{\alpha a} = d_\alpha^{-1}q_{a\alpha}$; without it (cf. Section 3) we need the entries of the $\{A_a\}$, the $\{S_\alpha\}$ and the change-of-basis matrices $(p_{\alpha a})$ and $(q_{a\alpha})$.

In a sense the reasons just given for selecting this formulation as the one deserving the title anova are mere details; the real reason is the fact that almost all examples and the natural generalisations and variants all derive from the present and no other approach. This will become more apparent in the next section, but first we give an example.

EXAMPLE. Suppose that $T = \prod_1^r\{1,\dots,n_j\}$ and that the indices are nested in a hierarchical structure t_1 nesting t_2 which nests t_3, etc. If we write $t = (t_1,\dots,t_r)$ then there is an obvious way to define a set of matrices $\{A_a: a = 0,\dots,r\}$ satisfying (i), (ii), (iii) and (iv), namely, $A_a(s,t) = 1$ if $s_h = t_h$, $h = 1,\dots,a$, $s_{a+1} \neq t_{a+1}$, $A_a(s,t) = 0$ otherwise, $0 \le a < r$; $A_r = I$ $(= A_e)$. When working with this example it is helpful to introduce the equivalence matrices $\{R_a: a = 0,\dots,r\}$ defined by $R_a(s,t) = 1$ if $s_h = t_h$, $h = 1,\dots,a$, $R_a(s,t) = 0$ otherwise; clearly $R_a = A_a + \cdots + A_r$, $0 \le a \le r$, while $A_a = R_a - R_{a+1}$, $0 \le a < r$, and $A_r = R_r = I$. This is because the primitive idempotents $\{S_\alpha\}$ are now readily defined by

$$S_0 = (n_1 \cdots n_r)^{-1}R_0,$$

$$S_\alpha = (n_{\alpha+1} \cdots n_r)^{-1}R_\alpha - (n_\alpha \cdots n_r)^{-1}R_{\alpha-1}, \qquad 1 \le \alpha < r,$$

$$S_r = I - n_r^{-1}R_{r-1}.$$

It is easy to calculate that $k_r = 1 = d_0$, $k_a = (n_{a+1} - 1)n_{a+2} \cdots n_r$, $0 \le a < r$,

$d_\alpha = n_1 \cdots (n_\alpha - 1)$, $0 < \alpha \le r$, and

$$d_\alpha^{-1} q_{a\alpha} = \begin{cases} 0, & a = 0, \dots, \alpha - 2, \\ -(n_\alpha - 1)^{-1}, & a = \alpha - 1, \\ 1, & a = \alpha, \dots, r. \end{cases}$$

The decomposition of $y_t = y_{t_1 \cdots t_r}$ is totally straightforward:

$$y_{t_1 t_2 \cdots t_{r-1} t_r} = y_{\cdot\cdots\cdot} + (y_{t_1 \cdots\cdots} - y_{\cdot\cdots\cdot}) + \cdots + (y_{t_1 t_2 \cdots t_{r-1} t_r} - y_{t_1 t_2 \cdots t_{r-1}\cdot})$$

and the other results follow immediately. This is one of the examples where X (and hence Z) have a lattice structure, namely the $(r+1)$-chain $\{\phi, \{1\}, \{1,2\}, \dots, \{1,2,\dots,r\}\}$; see Section 6.

5. Anova for infinite arrays. From the viewpoint presented in this paper one of the earliest instances of anova in statistics was the spectral representation of weakly stationary time series $y = (y_t\colon t \in \mathbb{Z})$, essentially put in its modern form by Cramér (1940) following earlier work by Khinchin (1934). Here the covariance matrix $\Gamma(s,t) = \operatorname{cov}(y_s, y_t)$ satisfies $\Gamma(s,t) = \Gamma(u,v)$ whenever $t - s = v - u$ and so may, formally at least, be written

$$\Gamma = \sum_0^\infty \gamma_a A_a, \tag{5.1}$$

where $A_0 = I$ is the doubly infinite identity matrix and A_a is the doubly infinite symmetric circulant having zeroth row $(\cdots 010 \cdots 0 \cdots 010 \cdots)$ with a 1 in the ath and $-a$th position, $a = 1, 2, \dots$. Because Γ is positive definite, a theorem of Herglotz tells us that for such a matrix there exists a uniquely defined positive measure on $[-\pi, \pi)$ whose Fourier coefficients are the $\{\gamma_a\}$. Since $\gamma_{-a} = \gamma_a$, this measure must be symmetric about 0 and so we can obtain the real spectral representation

$$\gamma_a = \int_{[0,\pi)} \cos(a\alpha)\phi(d\alpha), \qquad a \in \mathbb{Z}, \tag{5.2}$$

a formula which can readily be compared with (2.10b). The corresponding (real) representation of y_t with $E\{y_t\} \equiv 0$ takes the form

$$y_t = y_{\cdot} + 2\int_{(0,\pi)} [\cos(t\alpha)u(d\alpha) + \sin(t\alpha)v(d\alpha)], \tag{5.3}$$

where u and v are additive and mean-square continuous random set functions defined on the Borel subsets of $(0, \pi)$, spanning the Hilbert space generated by $y = (y_t\colon t \in \mathbb{Z})$ having zero means and satisfying

$$\begin{aligned} E\{u(A)u(B)\} &= E\{v(A)v(B)\} = \phi(A \cap B), \\ E\{u(A)v(B)\} &= 0, \end{aligned} \tag{5.4}$$

for A, B Borel subsets of $(0, \pi)$. Finally $y_\cdot$ is the mean-square limit of $T^{-1}\sum_1^T y_t$ as

$T \to \infty$, which is easily shown to exist. To compare (5.3) and (2.11) one simply expands the $\cos(2\pi(s-t)\alpha/n)$ and separates out random variables from non-random coefficients.

This is one kind of "infinite anova"; there are many similar ones in the literature of stochastic processes; see Hannan (1970, Chapter 1) and references therein.

At this point we do not stop to consider the method of proof of (5.3); in essence it reduces to the spectral decomposition of a unitary operator in Hilbert space and this will be covered by the discussion in Section 6. Rather we turn to another kind of infinite array.

Our original example $y = (y_{ij}: i = 1, \ldots, m;\ j = 1, \ldots, n)$ with j nested within i and having $\Gamma = Dy$ of the form (2.7) makes perfect sense if m or n (or both) is (are) countably infinite. Indeed one such example is the "random effects model"

$$y_{ij} = \varepsilon_0 + \varepsilon_i + \varepsilon_{ij}, \tag{5.5}$$

where (ε_i) and (ε_{ij}) are uncorrelated infinite sequences of uncorrelated random variables with zero means and variances σ_1^2 and σ_2^2, respectively, and ε_0 is a zero mean random variable uncorrelated with the ε_i and the ε_{ij} with variance σ_0^2. In this case the parameters γ_2, γ_1 and γ_0 of (2.7) are

$$\gamma_2 = \sigma_0^2 + \sigma_1^2 + \sigma_2^2, \qquad \gamma_1 = \sigma_0^2 + \sigma_1^2, \qquad \gamma_0 = \sigma_0^2. \tag{5.6}$$

What is the analogue of (2.4), (2.5) and (2.6) for an array $y = (y_{ij})$ with $\Gamma = Dy$ satisfying (2.7) for $m = n = \infty$? Clearly we can truncate i and j (within i) to the ranges $1, \ldots, m$ and $1, \ldots, n$, respectively, and see what results as $m, n \to \infty$, and doing this leads to some simple and interesting conclusions. Denoting the parameters and other objects associated with the truncated array by a superscript (m, n), we can prove directly that $\phi_\alpha^{(m,n)} = (mn)^{-1} d_\alpha^{(m,n)} \xi_\alpha^{(m,n)}$ and $[d_\alpha^{(m,n)}]^{-1} q_{a\alpha}^{(m,n)}$ both converge as m and $n \to \infty$ to ϕ_α and $s_\alpha(a)$ say, $\alpha = 0, 1, 2$ and $a = 0, 1, 2$. It follows that the terms $\xi_\alpha S_\alpha$ in the spectral representation (2.5) also converge as m and $n \to \infty$, since $\xi_\alpha^{(m,n)} S_\alpha^{(m,n)}(ij, kl) = \xi_\alpha^{(m,n)} (mn)^{-1} q_{a(ij,kl)\alpha}^{(m,n)}$, and we find that the limiting form of (2.5) is

$$\Gamma = \phi_0 J \otimes J + \phi_1 I \otimes J + \phi_2 I \otimes I, \tag{5.7}$$

where I and J are the infinite identity matrix and matrix of all 1's, respectively. Although (5.7) is not a spectral representation in any obvious sense, it can be proved that the most general positive definite matrix of the form

$$\Gamma = \gamma_2 I \otimes I + \gamma_1 I \otimes (J - I) + \gamma_0 (J - I) \otimes J \tag{5.8}$$

has a unique representation in the form (5.7) with ϕ_0, ϕ_1 and ϕ_2 all positive. The relations between ϕ's and γ's are simple enough:

$$\gamma_2 = \phi_0 + \phi_1 + \phi_2, \qquad \gamma_1 = \phi_0 + \phi_1, \qquad \gamma_0 = \phi_0 \tag{5.9a}$$

with inverse

$$\phi_2 = \gamma_2 - \gamma_1, \qquad \phi_1 = \gamma_1 - \gamma_0, \qquad \phi_0 = \gamma_0. \tag{5.9b}$$

In an obvious notation we can also prove that for $m' \geq m$ and $n' \geq n$,

$$\text{(5.10a)} \qquad \left\| y_{\cdot\cdot}^{(m,n)} - y_{\cdot\cdot}^{(m',n')} \right\|^2 = \phi_2\left[\frac{1}{mn} - \frac{1}{m'n'}\right] + \phi_1\left[\frac{1}{m} - \frac{1}{m'}\right],$$

$$\text{(5.10b)} \qquad \begin{aligned} &\left\| \left(y_{i\cdot}^{(m,n)} - y_{\cdot\cdot}^{(m,n)} \right) - \left(y_{i\cdot}^{(m',n')} - y_{\cdot\cdot}^{(m',n')} \right) \right\|^2 \\ &= \phi_2\left[\left(\frac{1}{n} - \frac{1}{n'}\right)\left(1 - \frac{1}{m}\right) + \frac{1}{n'}\left(\frac{1}{m} - \frac{1}{m'}\right)\right] + \phi_1\left[\frac{1}{m} - \frac{1}{m'}\right], \end{aligned}$$

$$\text{(5.10c)} \qquad \left\| \left(y_{ij}^{(m,n)} - y_{i\cdot}^{(m,n)} \right) - \left(y_{ij}^{(m',n')} - y_{i\cdot}^{(m',n')} \right) \right\|^2 = \phi_2\left[\frac{1}{n} - \frac{1}{n'}\right],$$

from which it follows that $y_{\cdot\cdot}^{(m,n)}$, $y_{i\cdot}^{(m,n)} - y_{\cdot\cdot}^{(m,n)}$ and $y_{ij}^{(m,n)} - y_{i\cdot}^{(m,n)}$—the components in (2.2)—all converge in mean square as $m, n \to \infty$. Denoting their limits by ε_0, ε_i and ε_{ij}, respectively, it can also be proved that not only are ε_0, ε_h and ε_{ij} pairwise orthogonal—they come from different strata in the limiting form of (2.2)—but also ε_h and ε_i are orthogonal if $h \neq i$, and similarly ε_{ij} and ε_{kl} are orthogonal if $i \neq k$ or $i = k$ and $j \neq l$. But all this has proved that (5.5) is (up to second order) the most general form for an array $y = (y_{ij})$ with $Dy = \Gamma$ satisfying (5.8), and that (5.7) is the most general form for such Γ. In this sense the standard random effects models arise naturally as the spectral decompositions of infinite arrays of multiindexed random variables with the appropriate dispersion models. For further details including a proof of this general result we refer to Speed (1986).

For our final illustration of an anova for an infinite array we return to the Example at the end of Section 4 and suppose that the repeated nesting goes on ad infinitum, i.e., that $T = \prod_1^\infty \{1, \ldots, n_j\}$ with each index of $t = (t_1, t_2, \ldots) \in T$ nesting all subsequent ones. As with the finite version, we can define association matrices $\{A_a\colon a = 0, 1, \ldots\}$ to which we must add $A_\infty = I\ (= A_e$ in our general notation). The relationship matrices $\{R_a\colon a = 0, 1, \ldots, \infty\}$ are defined in the same way as we did earlier and the passage from A-matrices to R-matrices is as before. We now look for a spectral representation for the positive definite matrices of the form

$$\text{(5.11)} \qquad \Gamma = \sum_{a=0}^{a=\infty} \gamma_a A_a.$$

As with our previous discussion, it is instructive to look at a truncated version of T, and the obvious candidate here is $T^{(r)} = \{t \in T\colon t_{r+1} = t_{r+2} = \cdots = 1\}$.

Denoting parameters and other expressions associated with the subarray $y^{(r)} = (y_t\colon t \in T^{(r)})$ with a superscript (r), we note that $s_\alpha(a) = [d_\alpha^{(r)}]^{-1} q_{a\alpha}^{(r)}$ does not depend upon r as long as $0 \leq a,\ a \leq r$. Furthermore, a straightforward calculation proves that $\phi_\alpha^{(r)} = (n_1 \cdots n_r)^{-1} d_\alpha^{(r)} \xi_\alpha^{(r)}$ satisfies

$$\text{(5.12)} \qquad \phi_\alpha^{(r)} - \phi_\alpha^{(r+1)} = \left(1 - n_\alpha^{-1}\right) n_{\alpha+1}^{-1} \cdots n_r^{-1} \left(1 - n_{r+1}^{-1}\right)(\gamma_e - \gamma_r),$$

which is nonnegative since $\gamma_a \leq \gamma_e$ for all a. Since $0 \leq \phi_\alpha^{(r)} \leq \gamma_e$ for all $r \geq 1$ and $\alpha \leq r$, we deduce that $\phi_\alpha^{(r)}$ converges, to ϕ_α say, as $r \to \infty$. Thus the elements of

$\xi_\alpha^{(r)} S_\alpha^{(r)}$ also converge as $r \to \infty$ and so we conjecture a unique representation for Γ in (5.11) taking the form of an ordinary infinite series

$$\Gamma = \sum_\alpha \phi_\alpha S_\alpha, \tag{5.13}$$

where the ϕ_α are positive (summing to γ_e—the anova) and the S_α satisfy $S_\alpha(s, t) = s_\alpha(a(s, t))$, i.e.,

$$S_\alpha = \sum_{a=0}^{a=\infty} s_\alpha(a) A_a. \tag{5.14}$$

These facts are readily proved and are perhaps most easily seen by using formal infinite tensor products. In an obvious notation $S_0 = J = J_{n_1} \otimes J_{n_2} \otimes \cdots$, whilst for $\alpha > 0$ we can use the expression for $s_\alpha(a)$ to get

$$\begin{aligned} S_\alpha &= \sum_{a \geq \alpha} A_a - (n_\alpha - 1)^{-1} A_{\alpha - 1} \\ &= \frac{n_\alpha}{n_\alpha - 1} I_{n_1} \otimes \cdots \otimes I_{n_{\alpha-1}} \otimes \left(I_{n_\alpha} - \frac{1}{n_\alpha} J_{n_\alpha} \right) \otimes J_{n_{\alpha+1}} \otimes J_{n_{\alpha+2}} \otimes \cdots . \end{aligned}$$

This completes our discussion of the spectral decomposition of Dy and we turn to that of y_t, $t \in T$. As with our previous example, its components are defined as mean-square limits, and in this case it is perhaps no surprise to see that these exist for

$$S_\alpha^{(r)} y_t = y_{t_1 \cdots t_\alpha \cdot \cdots \cdot t_{r+1} t_{r+2} \cdots} - y_{t_1 \cdots t_{\alpha-1} \cdot \cdots \cdot t_{r+1} t_{r+2} \cdots}$$

as $r \to \infty$. Indeed $\|S_\alpha^{(r)} y_t - S_\alpha^{(r')} y_t\|^2 = \phi_\alpha^{(r)} - \phi_\alpha^{(r')}$ for $1 \leq r \leq r'$, and by (5.12) this converges to zero as $r, r' \to \infty$ (assuming $n_r \geq 2$ for all r). Of course the mean-square limit $S_\alpha y_t$, say, of $S_\alpha^{(r)} y_t$, satisfies $\|S_\alpha y_t\|^2 = \phi_\alpha$, and so the spectral representation of y_t is the infinite sum, defined as a mean-square limit

$$y_t = \sum_\alpha S_\alpha y_t, \tag{5.15}$$

with associated anova $\gamma_e = \Sigma_\alpha \phi_\alpha$. Note that (5.15) is not the same as the expression

$$y_{t_1 t_2 t_3 \cdots} = \varepsilon_0 + \varepsilon_{t_1} + \varepsilon_{t_1 t_2} + \varepsilon_{t_1 t_2 t_3} + \cdots,$$

where $\{\varepsilon_0\}, \{\varepsilon_{t_1}\}, \{\varepsilon_{t_1 t_2}\}, \{\varepsilon_{t_1 t_2 t_3}\}, \ldots$ are uncorrelated sets of uncorrelated effects having variances $\phi_0, \phi_1, \phi_2, \phi_3, \ldots$; to get such a representation we would also need to let $n_1 \to \infty, n_2 \to \infty, n_3 \to \infty, \ldots$ in the preceding discussion.

These three examples of anovas for infinite arrays give a good idea of the range of possibilities. With the finite cyclic structure going over to the infinite one, we obtain a "continuous infinity" of strata; with the classical anova models illustrated by our second example, we simply recover standard random effects models, the number of strata remaining constant; whilst our final example shows how limits can be taken along infinite chains in the partially-ordered subset defining the nesting relationships on the set of indices, with the number of strata going to a countable infinity.

In none of these infinite examples does there appear to be a full analogue of the geometrically orthogonal decomposition of arrays y of real numbers, nor any associated sum of squares decompositions. Given that we never observe an infinite array of real numbers, this is no real limitation of the theory, and for many examples—most importantly the standard anova models in statistics—these decompositions for finite subarrays give useful information concerning aspects of the full array. Some details are sketched in Speed (1985) in a discussion relating the anova of a subarray, where it exists, to the anova of a full array.

The conclusion we come to after this discussion is that there is more to anova than sums of squares. Our view, already stated in the previous section, is that anova is a feature of certain models $\mathbf{V}$ which impose *equality constraints* on the covariances between pairs of elements of arrays of random variables.

6. Classical anova: Factorial dispersion models. The historically important anovas with multiply indexed arrays are the random effects models, dating back beyond Fisher (1925) to the last century, the randomization or permutation models following those discussed by Neyman, Iwaskiewicz and Kolodziejczyk (1935) and the more recent generalisations of de Finetti's exchangeability, studied by Aldous (1981) and others. Because of the importance of these ideas in statistics, I will sketch their common second-order theory.

We begin with a set F of factors $f_1, f_2, \ldots$, and a partial order $\leq$ on F where $f_1 \leq f_2$ means that the factor f_1 is nested within the factor f_2; cf. Nelder (1965). A subset $a \subseteq F$ is said to be a filter if $f_1 \in a$ and $f_1 \leq f_2$ implies that $f_2 \in a$, the need for such subsets arising because it is frequently necessary, when referring to the levels of a given factor f, to refer at the same time to all factors within which f is nested. The set of all filters of the partially ordered set $(F; \leq)$ forms a distributive lattice $L(F)$ under the operations of set union and intersection [see Aigner (1979, page 33)] and we refer to this book for all other order-theoretic terminology and results used in what follows. We remark in passing that our use of partially ordered sets in this context is closely related to, but does not coincide with, that of Throckmorton (1961), adopted by Kempthorne and Folks (1971, Section 16.11).

Next we suppose that the set of levels of factor f is T_f, $f \in F$, and we write $T = \prod_f T_f$ for the set of all combinations of levels of factors in F, denoting a typical element by $t = (t_f\colon f \in F)$. For any pair $s, t \in T$ we write $a(s, t)$ for the largest filter $a \in L(F)$ such that $s_f = t_f$ for all $f \in a$; e.g., if $s = ijk$ and $t = i'j'k'$, where we have three factors whose levels are denoted by the usual ijk rather than (s_1, s_2, s_3), and the second factor j is nested within the first i, then $a(s, t) = \{1, 2\}$ if $i = i'$, $j = j'$ and $k \neq k'$, whereas $a(s, t) = \{3\}$ if $i \neq i'$, $j = j'$ and $k = k'$, for $\{2, 3\}$ is not a filter of the partially ordered set of factors.

With these preliminaries we turn to the definition of factorial dispersion models. These are for arrays $y = (y_t\colon t \in T)$ of real random variables indexed by the set T of all combinations of levels of a set F of factors whose nesting relationships are defined by the partially ordered set $(F; \leq)$. The factorial dispersion model $\mathbf{V} = \mathbf{V}(F, T)$ is the class of all covariance matrices $\Gamma = Dy$ over

T which satisfy

$$\text{cov}(y_s, y_t) = \text{cov}(y_u, y_v) \tag{6.1}$$

whenever $a(s, t) = a(u, v)$, $s, t, u, v \in T$. Such classes are slightly more general than ones introduced by Nelder (1965), and we note that it has not yet been necessary to state whether or not the sets T_f are finite. For our summary of the structure of these models, we consider the two cases $|T_f| < \infty$ for all $f \in F$, and $|T_f| = \infty$ for all $f \in F$.

Finite factorial dispersion models. If $|T_f| = n_f < \infty$ for all $f \in F$, and we write $n = \prod_f n_f$, then $\mathbf{V}(F, T)$ is a class of $n \times n$ matrices whose structure is readily exhibited; see Speed and Bailey (1982) for full details. First we define the family $\{A_a: a \in L(F)\}$ of matrices over T by writing $A_a(s, t) = 1$ if $a(s, t) = a$ and $A_a(s, t) = 0$ otherwise, $s, t \in T$, $a \in L(F)$. Each element $\Gamma \in \mathbf{V}(F, T)$ satisfying (6.1) may then be represented uniquely in the form $\Gamma = \Sigma_a \gamma_a A_a$, the sum being over $L(F)$, with the parameters $\{\gamma_a: a \in L(F)\}$ being covariances.

It can be shown that the $\{A_a\}$ so defined form an association scheme, i.e., that (i), (ii), (iii) and (iv) of Section 4 and hence the consequences of these conditions hold, but here we can construct the structure constants $\{k_a\}, \{d_a\}$ and the functions $\{s_a(a)\}$ directly. To do this we introduce a second representation of $\mathbf{V}(F, T)$ involving relationship matrices $\{R_b: b \in L(F)\}$, where $R_b(s, t) = 1$ if $s_f = t_f$ for all $f \in b$ and $R_b(s, t) = 0$ otherwise, $s, t \in T$ and $b \in L(F)$. Clearly $R_b = \Sigma_{a \supseteq b} A_a$ and the representation we refer to is

$$\Gamma = \sum_b f_b R_b, \tag{6.2}$$

where the parameters $\{f_b: b \in L(F)\}$ have been called canonical components of variance by Fairfield-Smith (1955), Σ-quantities by Wilk and Kempthorne (1956), and f-quantities by Nelder (1965), although he later called them components of excess variance [Nelder (1977)]. Unfortunately it would take us too far afield to explain fully the frameworks of these other writers and the correspondence of the different parameters.

Relating the $\{f_b\}$ to the $\{\gamma_a\}$ requires the zeta function of the lattice $L(F)$, defined by $\zeta(a, b) = 1$ if $a \subseteq b$, $\zeta(a, b) = 0$ otherwise, and the associated Möbius function μ defined by $\Sigma\zeta(a, b)\mu(b, c) = \Sigma\mu(a, b)\zeta(b, c) = \delta(a, c) = 1$ if $a = c$ and 0 otherwise; here a, b and $c \in L(F)$ and the sums are over all $b \in L(F)$; see Aigner (1979, page 141) for further details. In this notation

$$f_b = \sum_a \mu(a, b)\gamma_a \tag{6.3a}$$

and

$$\gamma_a = \sum_a \zeta(b, a) f_b = \sum_{b \subseteq a} f_b. \tag{6.3b}$$

It can be shown that for all lattices of the form $L(F)$ the Möbius function μ takes only the values 1, -1 or 0; indeed the following concise formula for μ can

be proved:

$$\mu(a, b) = \begin{cases} (-1)^{|b \setminus a|}, & \text{if } b \supset a \text{ and } b \setminus a \subseteq b_m, \\ 0, & \text{otherwise,} \end{cases} \tag{6.4}$$

where b_m denotes the set of minimal elements of $b \subseteq F$.

The final representation of elements of $\mathbf{V}(F, T)$ we present is an explicit form of their common spectral decomposition. If we write $\bar{n}_a = \prod\{n_f\colon f \notin a\}$ for an element $a \in L(F)$, then the formula

$$S_\alpha = \sum_a \mu(a, \alpha)\bar{n}_a^{-1} R_a, \qquad \alpha \in L(F) \tag{6.5}$$

defines a set of pairwise orthogonal symmetric idempotent matrices summing to the identity matrix I over T. Further the formula

$$\xi_\alpha = \sum_b \zeta(a, b)\bar{n}_b f_b \tag{6.6}$$

gives the eigenvalues of $\Gamma = \Sigma_b f_b R_b$ and its spectral decomposition is then $\Gamma = \Sigma_\alpha \xi_\alpha S_\alpha$. Thus the eigenvalues $\{\xi_\alpha\colon \alpha \in L(F)\}$ constitute a third set of parameters whose positivity succinctly defines the parameter space, and there are two related sets of parameters which also have been used: the specific components of variance $\{\sigma_\alpha^2\colon \alpha \in L(F)\}$ of Cornfield and Tukey (1956), given by $\sigma_\alpha^2 = \bar{n}_\alpha^{-1}\xi_\alpha$, and the spectral components of variance $\{\phi_\alpha\colon \alpha \in L(F)\}$, cf. Daniels (1939), given by $\phi_\alpha = n^{-1}d_\alpha \xi_\alpha$, where $d_\alpha = \operatorname{rank}(S_\alpha)$.

If we combine the relationships between the $\{\gamma_a\}$ and the $\{f_b\}$ with those connecting the $\{f_b\}$ and the $\{\xi_\alpha\}$ we can obtain (4.3a) and (4.3b) where a and $\alpha \in L(F)$ and the sums are over $L(F)$, and of course (4.2a) and (4.2b) also hold with the same coefficients $(p_{a\alpha})$ and $(q_{a\alpha})$. The following formulas give expressions for the key quantities:

$$d_\alpha = \prod_{f \in \alpha \setminus \alpha_m} n_f \times \prod_{f \in \alpha_m} (n_f - 1), \tag{6.7}$$

where α_m denotes the set of minimal elements of α,

$$k_a = \prod_{f \in \bar{a} \setminus \bar{a}^m} n_f \prod_{f \in \bar{a}^m} (n_f - 1), \tag{6.8}$$

where $\bar{a}^m$ denotes the set of maximal elements of $\bar{a} = F \setminus a$, and the common value $s_\alpha(a)$ of $d_\alpha^{-1} q_{a\alpha} = k_a^{-1} p_{a\alpha}$ is

$$s_\alpha(a) = \begin{cases} \prod_{f \in \alpha_m \setminus a} \{-1/(n_f - 1)\}, & \text{if } \alpha \setminus \alpha_m \subseteq a, \\ 0, & \text{otherwise,} \end{cases} \tag{6.9}$$

where an empty product is defined to be unity.

The foregoing discussion enables a fairly complete analysis of finite factorial dispersion models to be given and we now indicate the changes necessary when $|T_f| = n_f = \infty$ for all $f \in F$. The main conclusion is the fact that the first two representations, $\Gamma = \Sigma_a \gamma_a A_a$ and $\Gamma = \Sigma_b f_b R_b$, continue to apply because we never need to multiply these matrices. After a suitable normalization and

limiting argument, the third representation turns out to coincide with the second. In particular the limiting forms of the two parametrizations, which are essentially normalized eigenvalues $\{\sigma_\alpha^2\}$ and $\{\phi_\alpha\}$, coincide with the corresponding $\{f_\alpha\}$. Finally, the limiting form of the function $s_\alpha(a)$ is just the zeta function $\zeta(\alpha, a) = 1$ if $\alpha \subseteq a$ and 0 otherwise.

We turn now to the spectral decompositions (3.4) and (3.6) in our classical anova context. It is easy to see that for finite arrays the matrices $\{\bar{n}_a^{-1}R_a: a \in L(F)\}$ act on $y_t(t \in T)$ by simply averaging out all indices t_f with $f \notin a$, and so by (6.4) the expression (6.5) for S_α reduces to an alternating sum of averaging operators starting with $\bar{n}_\alpha^{-1}R_\alpha$. For infinite arrays it all carries through using mean-square limits; cf. Section 5. In the finite case this is just the familiar anova decomposition of multi-indexed arrays into admissible main effects and interactions termed the population identity by Kempthorne (1952, Chapter 8) (his arrays having permutation or sampling distributions) and called the yield identity by Nelder (1965). For infinite arrays we recover the standard random effects linear models appropriate to the nesting structure on the indices: the components $S_\alpha y_t$ are not only uncorrelated across strata but (when $n_f \equiv \infty$) also, when distinct, within strata. Again we refer to Speed (1986) for more details.

7. Anova and groups. In all the particular examples we have given so far, and in the vast majority of those which occur in practice, there is an underlying group G acting transitively on the index set T, denoted $(g, t) \to t^g$, in such a way that the class of covariance matrices $\Gamma = Dy$ of $y = (y_t: t \in T)$ which we consider for our anovas coincides with the class of positive definite functions Γ on $T \times T$ which are G-invariant in the sense that

$$\Gamma(s, t) = \Gamma(s^g, t^g), \qquad (s, t) \in T \times T, g \in G. \tag{7.1}$$

It will follow from a few simple manipulations that the mathematical parts of our anovas, getting the spectral representation of the matrices Γ and the corresponding orthogonal decompositions of the array elements y_t $(t \in T)$, are only a slightly disguised form of a standard problem in harmonic analysis. This should hardly come as a surprise given the earlier discussion of finite and infinite circular arrays $(y_t: t = 0, 1, \ldots, n-1)$ and $(y_t: t \in \mathbb{Z})$.

We will only sketch the connexion here; the interested reader is referred to Hannan (1965, Section 5) and Dieudonné (1978) for further details. Choosing and fixing an arbitrary $t_0 \in T$, we define the subgroup $K = \{g \in G: t_0^g = t_0\}$ of G and observe that the homogeneous space G/K of cosets of G modulo K corresponds naturally with T, gK corresponding to t iff $t^g = t_0$. Now a function Φ on T is said to be spherically symmetric (relative to K) if $\Phi(t) = \Phi(t^k)$, $t \in T$, $k \in K$; similarly a function Ψ on G is said to be bi-invariant (relative to K) if $\Psi(kgk') = \Psi(g)$, $g \in G$, $k, k' \in K$, whilst we have called a function Γ on $T \times T$ G-invariant if it satisfied (7.1). The simple manipulations previously referred to show that these three classes of functions are essentially the same one, e.g., if Γ is G-invariant on $T \times T$, then $\Psi(g) = \Gamma(t_0^g, t_0)$ is bi-invariant on G whilst $\Phi(t) = \Gamma(t, t_0)$ is spherically symmetric on T. Conversely, if Ψ is bi-invariant on G and g_s, g_t are elements g and $h \in G$ for which $s^g = t_0$, $t^h = t_0$,

respectively, then $\Gamma(s, t) = \Psi(g_s^{-1} g_t)$ is G-invariant on $T \times T$. Finally, we let Y denote the space of all orbits of G over $T \times T$; clearly functions γ over Y correspond in an obvious way to G-invariant functions Γ on $T \times T$ and hence to the other classes previously mentioned. With this background our initial anova problems take the form: Describe the class of all functions γ on Y, in particular those for which $\Gamma(s, t) = \gamma_{b(s,t)}$ is positive definite over T, where $b(s, t)$ is the unique element of Y containing $(s, t) \in T \times T$.

Solutions to the problem just posed exist for many group actions, the most elegant case apparently being when (G, K) is a Gel'fand pair [Dieudonné (1978, page 55)] usually discussed when G is a unimodular separable metrizable locally compact group and K a compact subgroup. When (G, K) is a Gel'fand pair there is a class Z of functions called zonal spherical functions which plays a prominent role and in our terms these are the functions on Y defined by $s_\alpha(a) = d_\alpha^{-1} q_{a\alpha}$, $a \in Y$, $\alpha \in Z$. We note in passing that this class includes all characters of locally compact abelian groups, so our anova decomposition of the matrix Γ is a form of generalised Bochner–Godement theorem.

In his expositions Letac (1981, 1982) presents a wide range of applications of the theory of Gel'fand pairs in probability theory and we can clearly add anova to his list. The example in Letac (1982) which he calls the infinite symmetric tree is just the third example we discussed in the previous section—the infinitely nested hierarchical anova model—and so we have given an alternative approach to its harmonic analysis. It is also of interest to note that the theory of discrete Gel'fand pairs which Letac summarises in his paper is included within the theory of association schemes: All of his formulas can be found in the theorem we cited in Section 4, e.g., $\pi(a) = k_a$ is the measure on X induced by the uniform measure on T, the spherical functions are $s_\alpha(a) = d_\alpha^{-1} q_{a\alpha}$ as has already been noted and the Plancherel measure on Z is $\nu(\alpha) = n^{-1} d_\alpha$.

What of the spectral decompositions for the elements y_t $(t \in T)$ of the arrays? These arise from the decomposition of the permutation representation $g \to U_g$ of G into its irreducible constituents, where U_g is defined on the Hilbert space H spanned by the $(y_t\colon t \in T)$ [using the inner product $\langle y_s, y_t \rangle = \Gamma(s, t)$] by extending the assignment $U_g y_t = y_{ts}$, $t \in T$, $g \in G$ to the whole of H. In seeking to derive the decomposition in any particular case there are issues concerning the compactness of K, separability and local compactness of G, the nature of the representation $\{U_g\}$ and so on, which must be verified before general theory can be applied; we refer to Dieudonné (1978, 1980) for details. Perhaps surprisingly, none of the simple (infinite) classical anova models gives rise to pairs (G, K) for which these conditions hold, and so the ad hoc approach adopted in Speed (1986) still seems to be necessary. Even defining the groups for these classical anova models is a formidable task; see Bailey, Praeger, Rowley and Speed (1983) for details of the finite cases and Speed (1986) for some remarks on their infinite analogues.

8. Manova. The multivariate analysis of variance or manova does for arrays of random vectors what anova does for arrays of (real-valued) random variables, that is, gives suitable spectral decompositions of their dispersion matrices,

orthogonal decompositions of both the elements of the arrays and the arrays themselves; associated with these are analysis of the variances and covariances and decompositions of the sums of squares and products. There are some twists, however, which require us to generalise slightly our earlier formulation involving association matrices. For example, suppose that $w = (w_t\colon t = 0, \ldots, n-1)$ is a circular array of zero mean random vectors $w_t = (x_t, y_t)'$ with dispersion matrix

$$\Gamma = D\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \Gamma^{xx} & \Gamma^{xy} \\ \Gamma^{yx} & \Gamma^{yy} \end{bmatrix}.$$

We assume that $\Gamma^{xx} = Dx$ and $\Gamma^{yy} = Dy$ both have the form (2.9) whilst $\Gamma^{xy} = \operatorname{cov}(x, y)$ satisfies $\Gamma^{xy}(s, t) = \Gamma^{xy}(u, v)$ if $t - s = v - u$, i.e., Γ^{xy} is a circulant, although not necessarily a symmetric one. Indeed $\operatorname{cov}(x_s, y_t)$ and $\operatorname{cov}(y_s, x_t)$ are in general different. What is the decomposition of Γ^{xy} analogous to the diagonalisation of Γ^{xx} and Γ^{yy}?

The solution in this case is easy enough because the structure of arbitrary circulants is as transparent as that of symmetric circulants: Write $\Gamma^{xy} = \sum_0^{n-1}\gamma_b^{xy}B_b$, where B_b is the $n \times n$ circulant having a single 1 in the bth position and 0's elsewhere in its first row. Assuming that $n = 2m + 1$ as before—the case $n = 2m$ is just as readily dealt with—we recover our earlier association matrices by noting that $A_0 = B_0$, whilst $A_a = B_a + B_a'$, $a = 1, \ldots, m$. The $(m+1) \times (m+1)$ structural matrices $P = (p_{a\alpha})$ and $Q = (q_{a\alpha})$ are best described by the equations

$$k_a^{-1}p_{a\alpha} = d_a^{-1}q_{a\alpha} = \cos\left(\frac{2\pi}{n}a\alpha\right), \tag{8.1}$$

where $k_0 = d_0 = 1$, $k_a = d_a = 2$, $1 \le a$, $\alpha \le m$. We now need to introduce another inverse pair of $m \times m$ matrices of structural constants, namely $T = (t_{b\alpha})$ and $L = (l_{\alpha b})$:

$$t_{b\alpha} = l_{\alpha b} = 2\sin\left(\frac{2\pi}{n}b\alpha\right), \qquad 1 \le \alpha, b \le m. \tag{8.2}$$

It is not hard to prove that $TL = LT = nI_m$. With these constants defined, we supplement the $\{S_\alpha\}$ defined following (2.9) with T_α defined by $T_\alpha(s, t) = (1/n)t_{b(s,t)\alpha}$ where $b(s, t) = (t - s) \pmod n$. This is equivalent to

$$T_\alpha = (1/n)\sum_1^m t_{b\alpha}(B_b - B_b'), \qquad \alpha = 1, \ldots, m. \tag{8.3}$$

In these terms we have

$$B_b = S_0 + \tfrac{1}{2}\sum_1^m (p_{\alpha b}S_\alpha + l_{\alpha b}T_\alpha), \qquad b = 1, \ldots, m, \tag{8.4}$$

which, incidentally, agrees with our earlier notation since

$$A_a = B_a + B_a' = 2S_0 + \sum_1^m p_{a\alpha}S_\alpha = \sum_0^m p_{a\alpha}S_\alpha, \qquad a = 1, \ldots, m.$$

Also we see that $B_b - B_b' = \sum_1^m l_{b\alpha}T_\alpha$, a consequence of the relation $LT = TL =$

nI_m. It is not hard to check that $T_\alpha' = -T_\alpha$, $T_\alpha^2 = -S_\alpha$, $\alpha = 1, \ldots m$, and with all these preliminaries we can write the real form of the spectral decomposition of Γ^{xy} as

$$\Gamma^{xy} = c_0^{xy} S_0 + \sum_1^m (c_\alpha^{xy} S_\alpha + q_\alpha^{xy} T_\alpha), \tag{8.5}$$

where c_α^{xy} and q_α^{xy} are given by

$$c_\alpha^{xy} = \gamma_0^{xy} + \sum_{a=1}^m \cos\left(\frac{2\pi}{n} a\alpha\right)\left[\gamma_a^{xy} + \gamma_{n-a}^{xy}\right], \tag{8.6a}$$

$$q_\alpha^{xy} = \sum_{a=1}^m \sin\left(\frac{2\pi}{n} a\alpha\right)\left[\gamma_{n-a}^{xy} - \gamma_a^{xy}\right], \tag{8.6b}$$

with inverse

$$\gamma_b^{xy} = \frac{1}{n} c_0^{xy} + \frac{2}{n} \sum_{\alpha=1}^m \left[c_\alpha^{xy} \cos\left(\frac{2\pi}{n} b\alpha\right) + q_\alpha^{xy} \sin\left(\frac{2\pi}{n} b\alpha\right)\right]. \tag{8.6c}$$

In fact $c_\alpha^{xy} = \mathrm{Re}(\xi_\alpha^{xy})$ and $q_\alpha^{xy} = -\mathrm{Im}(\xi_\alpha^{xy})$, $\alpha = 0, 1, \ldots, m$, where ξ_α^{xy}, $\alpha = 0, \ldots, n$, are the eigenvalues of Γ^{xy}, in general complex, although they do satisfy the reality constraint $\bar{\xi}_\alpha^{xy} = \xi_{n-\alpha}^{xy}$.

The element γ_b^{xy} can be viewed as the bth entry in Γ^{xy} or as the xy entry in Γ_b, the lag b cross covariance matrix of the two sequences (x_t) and (y_t):

$$\Gamma_b = \begin{bmatrix} \gamma_b^{xx} & \gamma_b^{xy} \\ \gamma_b^{yx} & \gamma_b^{yy} \end{bmatrix}.$$

Grouping the c_α and q_α into matrices we may combine (8.6c) with the corresponding results for γ_b^{xx} and γ_b^{yy} to get

$$\Gamma_b = \frac{1}{n} C_0 + \frac{2}{n} \sum_\alpha \left[C_\alpha \cos\left(\frac{2\pi}{n} b\alpha\right) + Q_\alpha \sin\left(\frac{2\pi}{n} b\alpha\right)\right]. \tag{8.7}$$

This is the real spectral representation of Γ_b with $\{C_\alpha\}$ and $\{Q_\alpha\}$ being termed the cospectral and quadrature spectral matrices, respectively. The former are positive definite and the latter antisymmetric, as we will see in due course. Either (8.5) (together with the corresponding result for Γ^{xx} or Γ^{yy}) or (8.7) leads to the real spectral representation of a Γ having the form

$$\Gamma = A_e \otimes \Gamma_e + \sum_1^m [B_b \otimes \Gamma_b + B_{n-b} \otimes \Gamma_{n-b}], \tag{8.8}$$

which is

$$\Gamma = S_0 \otimes C_0 + \sum_1^m [S_\alpha \otimes C_\alpha + T_\alpha \otimes Q_\alpha]. \tag{8.9}$$

Now that we have the equivalent of the relations (4.2a) and (4.3a) for this class of covariance matrices, we can consider the corresponding decomposition of the elements w_t and the arrays w. The orthogonal decomposition of elements is

just what one would expect, namely

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \sum_\alpha \begin{bmatrix} S_\alpha x_t \\ S_\alpha y_t \end{bmatrix}, \tag{8.10}$$

where $S_\alpha x_t = \Sigma_u S_\alpha(t, u)x_u$ are similar for $S_\alpha y_t$; cf. (2.11). The terms are of course orthogonal across strata and obey the following rules within strata:

$$\text{cov}(S_\alpha x_t, S_\alpha y_t) = n^{-1}d_\alpha c_\alpha^{xy}, \qquad \text{cov}(T_\alpha x_t, S_\alpha y_t) = n^{-1}d_\alpha q_\alpha^{xy}. \tag{8.11}$$

We can combine (8.11) with the corresponding results for x_t and y_t alone and obtain the formulas

$$D\begin{bmatrix} S_\alpha x_t \\ S_\alpha y_t \end{bmatrix} = n^{-1}d_\alpha C_\alpha, \qquad D\left[\begin{bmatrix} T_\alpha x_t \\ T_\alpha y_t \end{bmatrix}, \begin{bmatrix} S_\alpha x_t \\ S_\alpha y_t \end{bmatrix}\right] = n^{-1}d_\alpha Q_\alpha, \tag{8.12}$$

from which it is clear that C_α is positive definite; since $T_\alpha' = -T_\alpha$, $T_\alpha x_t$ is orthogonal to $S_\alpha x_t$ and so Q_α is antisymmetric.

The preceding discussion gives a good illustration of the extra difficulties encountered when nonsymmetric elements B_b appear in the class of basis matrices describing the cross covariances between different components of a vector element of a random array. How general can the class of $\{B_b\}$ of matrices be and still permit a satisfactory manova? Condition (i) of symmetry on our family of adjacency matrices can be modified—the matrices would then be described as the adjacency matrices of a homogeneous coherent configuration [Higman (1975, 1976)], but more is needed to give a reasonable theory. The appropriate conditions on a class $\{B_b\colon b \in Y\}$ of matrices over a set T with entries 0 and 1 only are the following:

(i) the transpose B_b' belongs to the class $\{B_b\}$, i.e., there exists $b^\vee$ such that $B_b' = B_{b^\vee}$;
(ii) $\Sigma_b B_b = J$, the matrix of 1's over T;
(iii) one of the matrices, B_e say, is the identity matrix over T;
(iv) $B_b B_c = \Sigma_d n_{bcd} B_d$ for suitable integers (n_{bcd});
(v) the symmetric elements of the algebra $\mathbf{B}$ of all linear combinations of the $\{B_b\}$ commute, i.e., $(B_b + B_b')(B_c + B_c') = (B_c + B_c')(B_b + B_b')$.

The last condition was introduced in a similar context by McLaren (1963).

Some of the B_b may already be symmetric: Let us list them first and write them as A_a; the remaining A-matrices are the symmetrized B-matrices $A_a = B_a + B_a'$, and we can list the remaining B-matrices in transpose pairs.

A dispersion model for an array $w = (w_t\colon t \in T)$ of random vectors which has the form

$$\Gamma = \sum_a A_a \otimes \Gamma_a + \sum_b [B_b \otimes \Gamma_b + B_{b^\vee} \otimes \Gamma_{b^\vee}], \tag{8.13}$$

where the first sum is over the symmetric relations and the second over the appropriate half of the nonsymmetric relations will have a manova decomposition provided that (v) is satisfied as well as (i), (ii), (iii) and (iv). The general

spectral decomposition of such a Γ then takes the form

$$\begin{aligned}\Gamma = &\sum_{\alpha}^{(1)} S_\alpha \otimes C_\alpha + \sum_{\alpha}^{(2)} [S_\alpha \otimes C_\alpha + T_\alpha \otimes Q_\alpha] \\ &+ \sum_{\alpha}^{(3)} [S_\alpha \otimes C_\alpha + T_\alpha \otimes Q_\alpha + U_\alpha \otimes D_\alpha + V_\alpha \otimes E_\alpha],\end{aligned} \tag{8.14}$$

where the sums $\Sigma^{(1)}$, $\Sigma^{(2)}$ and $\Sigma^{(3)}$ are over what we term the real, complex and quaternionic types of strata, respectively; $T_\alpha' = -T_\alpha$, $U_\alpha' = -U_\alpha$, $V_\alpha' = -V_\alpha$, $T_\alpha^2 = U_\alpha^2 = V_\alpha^2 = -S_\alpha$, $T_\alpha U_\alpha = V_\alpha$, $U_\alpha V_\alpha = T_\alpha$ and $V_\alpha T_\alpha = U_\alpha$. In the representation (8.14) the parameter matrices $\{C_\alpha\}$ are positive definite whilst $\{Q_\alpha\}$, $\{D_\alpha\}$ and $\{E_\alpha\}$ are all antisymmetric; cf. (8.12). There are further sets of structure matrices beyond $P = (p_{\alpha a})$ and $Q = (q_{a\alpha})$ which continue to relate the $\{S_\alpha\}$ and the $\{A_a\}$; where complex strata occur we need matrices $T = (t_{ba})$ and $L = (l_{\alpha b})$ to pass from the $\{B_b\}$ to the $\{T_\alpha\}$ as we did in the cyclic example; and where quaternionic strata arise we also need two further pairs of mutually inverse structure matrices to permit the passage between the $\{B_b\}$ and the $\{U_\alpha\}$ and $\{V_\alpha\}$. The details are straightforward but lengthy and will not be given here; they will appear in Chapter 11 of Bailey, Praeger, Speed and Taylor (1987).

When the structure of the vector space $\mathbf{B}$ spanned by the $\{B_b\}$ is fully exhibited, the decompositions of w_t and (w_t) follow as before. We have the familiar expression

$$w_t = \sum_{\alpha} S_\alpha w_t, \tag{8.15}$$

where, as usual, $S_\alpha w_t = \Sigma_u S_\alpha(t, u) w_u$ (i.e., S_α effectively acts componentwise) and the terms in (8.15) are orthogonal across strata and satisfy relations similar to (8.12) within complex or quaternionic strata. For example, if α is quaternionic we have

$$D(S_\alpha w_t) = \frac{d_\alpha}{n} C_\alpha, \qquad D(T_\alpha w_t, S_\alpha w_t) = \frac{d_\alpha}{n} Q_\alpha,$$

$$D(U_\alpha w_t, S_\alpha w_t) = \frac{d_\alpha}{n} D_\alpha, \qquad D(V_\alpha w_t, S_\alpha w_t) = \frac{d_\alpha}{n} E_\alpha,$$

whereas $D(U_\alpha w_t, V_\alpha w_t)$ must be worked out from (8.14) using the formulas given after it. The anova in this context is simply

$$\Gamma_e = \sum_{\alpha} \Phi_\alpha, \tag{8.16}$$

where $\Phi_\alpha = n^{-1} d_\alpha C_\alpha$ is the (matrix) spectral component of variance of stratum $\alpha \in X$.

9. What is an anova? It must be abundantly clear by now that we regard anova as a property of certain special classes of dispersion models for arrays of random variables, or vectors, namely, for certain models defined by equality

constraints amongst (co)variances. There should be an appropriate (real) spectral decomposition for all the dispersion matrices in the model, and a corresponding orthogonal decomposition for elements of the array. The components in these decompositions have interpretations which range from the notions of (random) main effects and interactions, in the classical anovas, through to harmonics at different wavelengths, wave numbers, etc., in the more classical harmonic analyses. For finite arrays there are also decompositions of sums of squares.

All of this is in marked contrast to the current use of the term in regression analysis and variance component analysis, where analysis of variance decompositions is more-or-less arbitrary orthogonal decomposition of sums of squares relating to "fixed" or "random" effects in assumed linear models. At this point it is worth explaining why our theory concerns only those structures described as "balanced" or "orthogonal." The reason is simple: Arrays with an anova as we use the term—one might add unique and complete—all have a high degree of symmetry, and in a sense the underlying index set is "complete." By comparison, the so-called "unbalanced" or "nonorthogonal" (random effects) anova models are in general rather messy subarrays of arrays with anova, and do not have an anova in their own right. For some further discussion of these points, see Speed (1985).

Although the vast majority of anova decompositions—of the matrices (or functions) and the random variables—are associated with a group action, and hence could be viewed as a part of a theory of generalised harmonic analysis, this line of thinking is by no means the best or the most general approach. For many arrays of random variables, including the standard multi-indexed ones of classical anova, the permutation groups are extremely complicated, whilst a direct combinatorial approach by-passing all representation theory is quite efficient; see Speed and Bailey (1982). Also in the reference just cited, an example of an association scheme which is *not* induced by a group action is given which shows that there are cases without an underlying group action.

Is there a single general theorem? It is hard to believe that one theorem will ever be formulated which covers all the examples mentioned so far. It would have to include all homogeneous coherent configurations satisfying condition (v) of Section 8, all limits of finite association schemes such as those illustrated in Section 5, the theory of Gel'fand pairs mentioned in Section 7, and much more. For example James (1982) has discussed the classical diallel cross in genetics from essentially our viewpoint; the triallel, double cross and other genetic structures give further interesting examples.

In closing we state what must be quite obvious to the reader: This paper has concentrated on the question, "What is an anova?" We have not discussed any of the many questions, which are both mathematically and statistically interesting, which arise when the array of random variables has an anova.

Acknowledgments. I am very pleased to be able to acknowledge the assistance given to me over the years by my collaborators Rosemary Bailey, Cheryl Praeger and Don Taylor. Much of the material outlined in this paper is or has arisen out of joint work with one or more of them, and without their skills

and enthusiasm this paper could not have been written. I would also like to thank the Associate Editor for his very careful reading of the manuscript.

REFERENCES

AIGNER, M. (1979). *Combinatorial Theory*. Springer, New York.

ALBERT, A. (1976). When is a sum of squares an analysis of variance? *Ann. Statist.* **4** 775–778.

ALDOUS, D. J. (1981). Representations for partially exchangeable arrays of random variables. *J. Multivariate Anal.* **11** 581–598.

ANDERSON, T. W. (1969). Statistical inference for covariance matrices with linear structure. In *Proc. Second Internat. Symp. on Multivariate Anal.* (P. R. Krishnaiah, ed.) 55–66. Academic, New York.

ANDERSON, T. W. (1970). Estimation of covariance matrices which are linear combinations or whose inverses are linear combinations of given matrices. In *Essays in Probability and Statistics* (R. C. Bose et al., eds.) 1–24. Univ. of North Carolina Press, Chapel Hill, N.C.

ANDERSON, T. W. (1973). Asymptotically efficient estimation of covariance matrices with linear structure. *Ann. Statist.* **1** 135–141.

BAILEY, R. A., PRAEGER, C. E., ROWLEY, C. A. and SPEED, T. P. (1983). Generalised wreath products of permutation groups. *Proc. London Math. Soc.* (*3*) **47** 69–82.

BAILEY, R. A., PRAEGER, C. E., SPEED, T. P. and TAYLOR D. E. (1987). *Analysis of Variance*. To appear.

BROWN, K. G. (1984). On analysis of variance in the mixed model. *Ann. Statist.* **12** 1488–1499.

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

CRAMÉR, H. (1940). On the theory of stationary random processes. *Ann. of Math.* **41** 215–230.

DANIELS, H. (1939). The estimation of components of variance. *J. Roy. Statist. Soc.* (*Suppl.*) **6** 186–197.

DIEUDONNÉ, J. (1978). *Treatise on Analysis* **6**. Academic, New York.

DIEUDONNÉ, J. (1980). *Special Functions and Linear Representations of Lie Groups* Amer. Math. Soc., Providence, R.I.

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

FISHER, R. A. (1925). *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh.

FISHER, R. A. (1934). Discussion in Wishart, J. Statistics in agricultural research. *J. Roy. Statist. Soc. Suppl.* **1** 26–61 (see especially page 52).

GAUTSCHI, W. (1959). Some remarks on Herbach's paper "Optimum nature of the F-test for model II in the balanced case." *Ann. Math. Statist.* **30** 960–963.

GRAYBILL, F. A. and HULTQUIST, R. A. (1961). Theorems concerning Eisenhart's model II. *Ann. Math. Statist.* **32** 261–269.

HANNAN, E. J. (1960). *Time Series Analysis*. Methuen, London.

HANNAN, E. J. (1961). Application of the representation theory of groups and algebras to some statistical problems, Research Reports, Part 1, Summer Research Institute, Australian Math. Soc.

HANNAN, E. J. (1965). Group representations and applied probability. *J. Appl. Probab.* **2** 1–68.

HANNAN, E. J. (1970). *Multiple Time Series*. Wiley, New York.

HARVILLE, D. A. (1984). A generalised version of Albert's theorem, with applications to the mixed linear model. In *Experimental Design Statistical Models and Genetic Statistics, Essays in Honor of Oscar Kempthorne* (K. Hinkelmann, ed.) 231–238. Dekker, New York.

HERBACH, L. H. (1959). Properties of model II—type analysis of variance tests, A: Optimum nature of the F-test for model II in the balanced case. *Ann. Math. Statist.* **30** 939–959.

HIGMAN, D. G. (1975). Coherent configuration I. *Geom. Dedicata* **4** 1–32.

HIGMAN, D. G. (1976). Coherent configurations II. *Geom. Dedicata* **5** 413–424.

HSU, P. L. (1941). Analysis of variance from the power function standpoint. *Biometrika* **32** 62–69.

HSU, P. L. (1945). On the power function of the E^2 test and the T^2 test. *Ann. Math. Statist.* **16** 278–286.

JAMES, A. T. (1957). The relationship algebra of an experimental design. *Ann. Math. Statist.* **28** 993–1002.

JAMES, A. T. (1982). Analysis of variance determined by symmetry and combinatorial properties of zonal polynomials. In *Statistics and Probability: Essays in Honor of C. R. Rao* (G. Kallianpur, P. R. Krishnaiah and J. K. Ghosh, eds.) 329–341. North-Holland, Amsterdam.

JENSEN, S. T. (1975). Covariance hypotheses which are linear in both the covariance and the inverse covariance. Preprint No. 1, Institute of Mathematical Statistics, Univ. Copenhagen.

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

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

KHINCHIN, A. YA. (1934). Korrelationstheorie der Stationären stochastischen Prozesse. *Math. Ann.* **109** 604–615.

LETAC, G. (1981). Problèmes classiques de probabilité sur un couple de Gel'fand. *Analytical Methods in Probability Theory. Lecture Notes in Math.* **861** 93–120. Springer, Berlin.

LETAC, G. (1982). Les fonctions sphériques d'un couple de Gel'fand symmétrique et les chains de Markov. *Adv. in Appl. Probab.* **14** 272–294.

MACWILLIAMS, F. J. and SLOANE, N. J. A. (1077). *The Theory of Error-Correcting Codes*. North-Holland, Amsterdam.

MCLAREN, A. D. (1963). On group representations and invariant stochastic processes. *Proc. Camb. Philos. Soc.* **59** 431–450.

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

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

NELDER, J. A. (1977). A reformulation of linear models (with discussion). *J. Roy. Statist. Soc. Ser. A* **140** 48–76.

NEYMAN, J., IWASKIEWICZ, K. and KOLODZIEJCZYK, ST. (1935). Statistical problems in agricultural experimentation (with discussion). *J. Roy. Statist. Soc. Suppl.* **2** 107–189.

SCHEFFÉ, H. (1959). *The Analysis of Variance*. Wiley, New York.

SPEED, T. P. (1985). Dispersion models for factorial experiments. *Bull. Internat. Statist. Inst.* **51** 24.1-1–24.1-16.

SPEED, T. P. (1986). Anova models with random effects: An approach via symmetry. In *Essays in Time Series and Allied Processes: Papers in Honor of E. J. Hannan* (J. Gani and M. B. Priestley, eds.) 355–368. Applied Probability Trust, Sheffield.

SPEED, T. P. and BAILEY, R. A. (1982). On a class of association schemes derived from lattices of equivalence relations. In *Algebraic Structures and Applications* (P. Schultz, C. E. Praeger and R. P. Sullivan, eds.) 55–74. Dekker, New York.

TANG, P. C. (1938). The power function of the analysis of variance tests with tables and illustrations of their use. *Statist. Res. Mem.* **2** 126–149.

THROCKMORTON, T. N. (1961). Structures of classification data. Ph.D. thesis, Iowa State Univ.

TUKEY, J. W. (1961). Discussion, emphasising the connection between analysis of variance and spectrum analysis. *Technometrics* **3** 191–219. (Follows papers by G. M. Jenkins and E. Parzen on spectrum analysis.)

WALD, A. (1942). On the power function of the analysis of variance test. *Ann. Math. Statist.* **13** 434–439.

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

WOLFOWITZ, J. (1949). The power of the classical tests associated with the normal distribution. *Ann. Math. Statist.* **20** 540–551.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720

International Statistical Review (1987), **55**, 3, pp. 261–277. Printed in Great Britain

Factorial Dispersion Models

T.P. Speed[1] and R.A. Bailey[2]

[1]CSIRO Division of Mathematics and Statistics, Canberra, Australia ACT 2601. [2]Statistics Department, Rothamsted Experimental Station, Harpenden, Herts., AL5 2JQ, UK.

Summary

A class of dispersion models for multi-indexed arrays of random variables is introduced and discussed. These models generalize the second-order properties of variance component, randomization and exchangeability models, and lead naturally to general techniques for calculating the orthogonal decompositions, expected mean squares and other aspects of the analysis of variance of such arrays.

Key words: Analysis of variance; Association scheme; Canonical component of variance; Exchangeability; Linear model; Permutation model; Randomization; Sample; Symmetry; Variance components.

1 Introduction

The analysis of variance of multi-indexed arrays, i.e. data from factorial experiments, interpreting this expression in the widest possible sense, had its origins in the quantitative genetic research of R.A. Fisher. By the time of the publication of Fisher (1925) these ideas had also been applied to comparative experiments in agriculture and in the following 15 years the range of applications was broadened to include sampling (Yates & Zacopanay, 1935; Youden & Mehlich 1937; Cochran, 1939) and industrial statistics (Daniels, 1938, 1939). Over the same period the models and assumptions underlying the analysis of variance were closely scrutinized: see especially Eden & Yates (1933), who examined the z-test using nonnormal data, and the later work of Pitman (1938) and Welch (1937) on the same topic, and the critical study by Neyman et al. (1935) of the use of Fisher's methods in agricultural experiments. Somewhat different problems were being tackled within a similar framework in psychometrics (Spearman, 1910; Brown, 1913) and animal breeding (Lush, 1931; Lush et al., 1933). In both of these fields there were measurements with two components of error; in modern terms they were concerned with the estimation in the presence of random effects, a topic whose origins can be found in nineteenth century astronomy: see Scheffé (1956) for further details.

Many modern writers on what has come to be called *variance component* analysis take as their starting point a *linear model* for their data array built up from independent sets of independent random effects, with one set of effects for each appropriate index or set of indices: some of these effects are termed *main effects,* the remainder *interactions.* Such effects induce a variety of covariances between elements of the array, although it is not common to regard the estimation of these covariances as an issue of particular statistical interest. This linear model approach is not appropriate if the underlying distribution of the array is a permutation distribution, a viewpoint adopted by a number of writers from Fisher onwards, including Kempthorne (1952) and Nelder (1965), who have chosen to emphasize the *randomization* aspects of analysis of variance. Nor is it appropriate if the effects or indeed the whole array are to be regarded as *randomly sampled* without replacement from one or more finite populations, an approach also adopted by Fairfield

Smith (1955), Cornfield & Tukey (1956) and Hooke (1956a,b). Finally, we mention that a statistician may wish to assume nothing more than a certain amount of symmetry, invariance or stationarity, such as the generalization to multi-indexed arrays of de Finetti's exchangeability: see, for example, Dawid (1977) and Aldous (1981). Do these symmetry assumptions still permit us to carry out the usual analysis of variance calculations in a meaningful way?

All of the different views or models mentioned above lead to the same structure for the covariance matrix of the data, although the form and interpretation of the parameters and the problems attacked naturally differ between models. Our approach is therefore based upon this common *dispersion model.* We shall see that all of the second-order calculations associated with analysis of variance can be derived straightforwardly from the relevant aspects of the dispersion model. Here we are taking the point of view of Cox (1960) and Speed (1987) that the term 'analysis of variance' means the decomposition of the common variance of several random variables into variance components which are of intrinsic interest, rather than the calculations required to analyse data from a so-called 'fixed effects model' with a single variance which is a nuisance parameter. Of course, many of the calculations are the same for both cases, because they are merely the mathematical decompositions of quadratic forms or geometric projections, as Bryant (1984) and Saville & Wood (1986) have recently pointed out. However, the underlying philosophy is quite different.

The work reported in this paper had as its starting point the paper by Nelder (1965), which concerns the second-order properties of the class of multi-indexed arrays which can be built up by successively nesting and crossing simpler ones, starting from a single unstructured factor. Although ostensibly set within a randomization framework, Nelder's results have a broader applicability, and § 2 below refines them somewhat and extends them to random arrays with more general (not necessarily permutation) distributions, to a wider class of index sets, and to the case where the number of values, or levels, of the indices, or factors, may be countably infinite. Nelder's (1965) work was primarily motivated by the need to systematize analysis of variance techniques so that a general computer program could be written to replace large collections of subroutines, each appropriate for a particular 'design'. An independent stream of work, initiated by Kempthorne (1952) and continued throughout the 1950's and early 1960's, see for example Wilk (1955), Wilk & Kempthorne (1956a,b; 1957), was concerned with the objective development and interpretation of linear models for randomized experiments. This body of research, from what we shall call the Iowa (State University) school, includes a number of valuable techniques for calculating the averages of certain quadratic forms over random sampling and randomly selecting designs. At around the same time Cornfield & Tukey (1956) reported on work done by them some years earlier addressing essentially the same problem: the calculation of expected mean squares in analysis of variance tables, or, as they term it, 'average values of mean squares in factorials'. One of our aims was to derive the main results of these authors within the modified Nelder framework outlined above.

Much of the paper is devoted to the broad problem of relating the characteristics of a subset of a multi-indexed array, which we can think of as a *sample,* to those of the full array, thought of as the *population.* The results of Cornfield & Tukey (1956) and the Iowa school do come out naturally, as do some less familiar ones concerning the prediction of unobserved from observed random variables, a topic usually referred to in this context as the estimation of random effects; see, for example, Harville (1976).

We hope that our methods, which attempt to treat finite and infinite populations, the different models or approaches noted above, as well as various kinds of samples, in a

uniform manner, will lead to:

(i) an understanding of the different parameterizations of factorial dispersion models, together with their interpretations;
(ii) the various orthogonal decompositions of random variables, of arrays of numbers, and of sums of squares, including the associated numbers of degrees of freedom;
(iii) techniques for calculating expected values of mean squares under a range of assumptions including linear models, over randomization, random sampling, and symmetry, obtaining answers in terms of the desired parameters;
(iv) procedures for getting 'best' estimates of all parameters;
(v) formulae for obtaining 'best' linear predictors of key unobserved random variables.

Throughout the paper we illustrate our results with a triply indexed array $y = (y_{ijk})$ and its associated dispersion matrix $\Gamma = Dy$, assuming that the second index, j, is nested within the first, i, and that these two are crossed with the third, k. As well as being complicated enough to exhibit most of the possibilities, this example allows the reader to specialize the results to a simple nesting, by suppressing k, and to a simple crossing, by suppressing j.

A good deal of the new work reported here is closely related to joint work with C.E. Praeger and D.E. Taylor which we hope will appear soon in a monograph entitled *Analysis of Variance*. We should like to thank them both for their collaboration over the years.

2 Factorial dispersion models

2.1 Preliminaries

We will suppose given a set Π of factors $p, q, \ldots$ and a partial order $\leqslant$ on Π, where $q \leqslant p$ means that the factor q is *nested* in the factor p, in the sense of Nelder (1965). It is helpful to draw the partially ordered set $(\Pi; \leqslant)$, which we term the *nesting poset*, with p above q if $q \leqslant p$ and connected to q if there is no r distinct from p and q for which $q \leqslant r$ and $r \leqslant p$; this is the so-called Hasse diagram; see Fig. 1. We refer to Aigner (1979) for terminology and further details concerning ordered sets. A subset $a \subseteq \Pi$ is said to be a *filter* if $p \in a$ whenever both $q \leqslant p$ and $q \in a$; such subsets have also been termed *admissible* by the Iowa school, but we shall adhere to (one) standard order-theoretic terminology. The need for such subsets arises because, in referring to the levels of a factor, it is frequently necessary to refer at the same time to all factors within which that factor is nested. The class $L(\Pi)$ of all filters of a poset $(\Pi; \leqslant)$ is readily found to be a *distributive lattice* under the operations of set union and set intersection, containing the empty set $\varnothing$ and the whole set Π (Aigner, 1979, p. 33). The lattice $L(\Pi)$ is also a poset under set-inclusion and so we can draw its Hasse diagram as well. It is convenient for our purposes to draw the subset lattice diagram 'upside down', using the reverse ordering from set inclusion. Figures 1 and 2 depict an *example* of a simple poset of three factors and its associated distributive lattice of filters; note that $2 < 1$ (meaning $2 \leqslant 1$ and $2 \neq 1$) is the sole nontrivial nesting relationship. This example will be used to illustrate much of what follows. Note that our use of Hasse diagrams in this context is quite different from that of Throckmorton (1961).

In a completely general factorial model, if factor q is nested in factor p then there is no need for q to have the same number of levels within each level of p. However, models

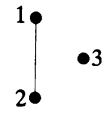

Figure 1. *The poset* Π *in the example.*

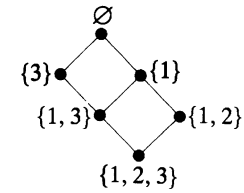

Figure 2. *The lattice* $L(\Pi)$ *in the example.*

based on assumptions of exchangeability or randomization do imply that q has the same number of levels with each level of p, and § 4 makes it clear that models based on random sampling also imply this balance condition if all the random variables are to have the same variance. Of all the viewpoints discussed in § 1, only the linear model approach permits q to have different numbers of levels within different levels of p. Since this paper is concerned with the theory that is common to all the approaches in § 1, we may assume that q has the same number of levels within each level of p; this paper has nothing to say about so-called 'unbalanced' data.

Next we suppose that the set of *levels* of factor p is T_p for p in Π. Of course, if q is nested in p then the levels of q at different levels of p bear no relation to one another at all even if the number of levels is constant. Nevertheless, it is extremely common to use the same set T_q to denote the levels of q within each level of p; see, for example, John (1971) or Kempthorne (1952). Although this might appear somewhat confusing, there are two good reasons for this convention: it facilitates both the algorithm for analysis of variance calculations (Nelder, 1965) and some of the formal mathematics (Bailey et al.,1983). We write $T = \prod_p T_p$ for the set of all *combinations* of levels of factors in Π, denoting a typical element of T by $t = (t_p : p \in \Pi)$. For any pair t, u in T we write $a(t, u)$ for the largest filter a in $L(\Pi)$ such that $t_p = u_p$ for all p in a.

Example. With Π as in Fig. 1, let $t = ijk$ and $u = i'j'k'$. Then $a(t, u) = \{1, 2\}$ if $i = i'$, $j = j'$ and $k \neq k'$; whilst $a(t, u) = \{3\}$ if $i \neq i'$, $j = j'$ and $k = k'$, since $\{2, 3\}$ is not a filter of Π in this case. Here and below, when discussing our example, it is convenient to write (t_1, t_2, t_3) as (i, j, k) and abbreviate this to ijk.

With these preliminaries we can now define the *dispersion models* of our title. They are for arrays $y = (y_t : t \in T)$ of real random variables indexed by the set T of combinations of levels of a set Π of factors whose nesting relationships are described by the partially ordered set $(\Pi; \leqslant)$. The covariance matrix Dy is defined over T and is said to be *factorial* if $\operatorname{cov}(y_t, y_u) = \operatorname{cov}(y_v, y_w)$ whenever $a(t, u) = a(v, w)$ for t, u, v, w in T, and the class of all such covariance matrices is denoted by $V(\Pi, T)$; briefly, a covariance matrix is factorial if the covariance between any two elements y_t and y_u depends only on the (largest) subset (filter) of the factors corresponding to the components on which the indices t and u agree. This class is more general than that introduced by Nelder (1965), and we note that it has not yet been necessary to state whether or not the sets T_p are *finite* for p in Π. See Tjur (1984) for a discussion of an even wider class of models, and Bailey (1984) for a discussion of the relationship between Tjur's work and the present paper: factorial dispersion models correspond to Bailey's *poset block structures*.

2.2 *Finite index sets: Algebraic theory*

If $|T_p| = n_p < \infty$ for p in Π, and we write $n = \prod_p n_p$, then $V(\Pi, T)$ is a class of $n \times n$ positive-definite matrices whose structure is readily exhibited: see Speed & Bailey (1982)

Factorial Dispersion Models

for fuller details and proofs. Firstly, we can define the *association matrices* $\{A_a : a \in L(\Pi)\}$ over T by writing $A_a(t, u) = 1$ if $a(t, u) = a$, and $A_a(t, u) = 0$ otherwise. A general element Γ of $V(\Pi, T)$ thus has the form $\Gamma = \sum_a \gamma_a A_a$, the sum being over $L(\Pi)$, with the parameters $\{\gamma_a : a \in L(\Pi)\}$ being *covariances*.

For reference in § 4, we note that the set of matrices $\{A_a : a \in L(\Pi)\}$ forms an *association scheme*. This means that the following conditions are satisfied:

(i) for all a in $L(\Pi)$, every entry in A_a is equal to 0 or 1, but A_a is not the zero matrix;
(ii) for all a in $L(\Pi)$, the matrix A_a is symmetric;
(iii) the sum $\sum_{a \in L(\Pi)} A_a$ is the matrix J with every entry equal to 1;
(iv) one of the matrices (in this case A_Π) is equal to the identity matrix I;
(v) there are integers n_{abc} for a, b, c in $L(\Pi)$ such that, for all a, b in $L(\Pi)$,

$$A_a A_b = \sum_{c \in L(\Pi)} n_{abc} A_c.$$

See, for example, MacWilliams & Sloane (1977, Ch. 21) for a good discussion of the general theory of association schemes.

A second, useful representation of elements of $V(\Pi, T)$ involves the *relationship matrices* $\{R_b : b \in L(\Pi)\}$, where $R_b(t, u) = 1$ if $t_p = u_p$ for all p in b, and $R_b(t, u) = 0$ otherwise. Clearly

$$R_b = \sum_{a \supseteq b} A_a$$

and the representation we refer to is $\Gamma = \sum_b f_b R_b$, where the parameters $\{f_b : b \in L(\Pi)\}$ are called *canonical components of variance* by Fairfield Smith (1955), *Σ-quantities* by Wilk & Kempthorne (1956a); and simply *f-quantities* by Nelder (1965), although later he called them *components of excess variance* (Nelder, 1977). Relating the f's to the γ's requires the *zeta* function of the lattice $L(\Pi)$ given by $\zeta(a, b) = 1$ if $a \subseteq b$ and $\zeta(a, b) = 0$ otherwise, and the associated *Möbius function* μ defined by

$$\sum_b \zeta(a, b)\mu(b, c) = \sum_b \mu(a, b)\zeta(b, c) = \delta(a, c),$$

where $\delta(a, c) = 1$ if $a = c$, and $\delta(a, c) = 0$ otherwise; the sums are over all b in $L(\Pi)$: see Aigner (1979, p. 141) for further details. In this notation $f_b = \sum_a \mu(a, b)\gamma_a$.

Because it plays such a prominent role in the discussion which follows, we explain briefly how the Möbius function μ of a lattice L is calculated from a Hasse diagram. An easy reformulation of the definition is the following: $\mu(a, c) = 0$ unless $a \subseteq c$; $\mu(a, a) = 1$; and for $a \subset c$:

$$\mu(a, c) = -\sum_{a \subseteq b \subset c} \mu(a, b);$$

equivalently,

$$\mu(a, c) = -\sum_{a \subset b \subseteq c} \mu(b, c).$$

Example (cont.). Let us calculate some values of μ. From $\mu(\varnothing, \varnothing) = \mu(\{1\}, \{1\}) = 1$ and either of the above, we deduce that $\mu(\varnothing, \{1\}) = -1$. A similar argument applies to any pair connected by an edge in the Hasse diagram. Turning to $\mu(\varnothing, \{1, 3\})$ we can use the first of the above formulae to find that

$$\mu(\varnothing, \{1, 3\}) = -[\mu(\varnothing, \varnothing) + \mu(\varnothing, \{1\}) + \mu(\varnothing, \{3\})] = +1,$$

Table 1

The matrices of the zeta and Möbius functions for the example.

(a) Z

	∅	{1}	{1, 2}	{3}	{1, 3}	{1, 2, 3}
∅	1	1	1	1	1	1
{1}	0	1	1	0	1	1
{1, 2}	0	0	1	0	0	1
{3}	0	0	0	1	1	1
{1, 3}	0	0	0	0	1	1
{1, 2, 3}	0	0	0	0	0	1

(b) Z^{-1}

	∅	{1}	{1, 2}	{3}	{1, 3}	{1, 2, 3}
∅	1	−1	0	−1	1	0
{1}	0	1	−1	0	−1	1
{1, 2}	0	0	1	0	0	−1
{3}	0	0	0	1	−1	0
{1, 3}	0	0	0	0	1	−1
{1, 2, 3}	0	0	0	0	0	1

whilst similar reasoning shows that

$$\mu(\varnothing, \{1, 2\}) = -[\mu(\varnothing, \varnothing) + \mu(\varnothing, \{1\})] = 0.$$

Alternatively, the Möbius function may be calculated by matrix inversion. Let Z be the $L(\Pi) \times L(\Pi)$ matrix with entries $\zeta(a, b)$: Table 1(a) shows Z for our example. Since Z is upper triangular (if the elements of $L(\Pi)$ are written in a suitable order), it is easily inverted. The values of μ are simply the entries of Z^{-1}. For example, Table 1(b) shows that $\mu(\{1\}, \{1, 2, 3\}) = 1$.

It can be shown that, for all lattices of the form $L(\Pi)$ that we are considering, μ takes only the values 0, +1 or −1. A concise formula for the values of μ can be given but we do not need it here.

The final representation of elements of $V(\Pi, T)$ is the explicit form of their common spectral decomposition. If we write $\bar{n}_a = \Pi(n_p : p \notin a)$ for a in $L(\Pi)$, then the formula

$$S_\alpha = \sum_a \mu(a, \alpha)\bar{n}_a^{-1} R_a, \quad \alpha \in L(\Pi),$$

gives a set of pairwise orthogonal, that is $S_\alpha S_\beta = 0 = S_\beta S_\alpha$ if $\alpha \neq \beta$, idempotent ($S_\alpha^2 = S_\alpha$) symmetric matrices which sum to the identity ($I = \sum_\alpha S_\alpha$). Further, the formula

$$\xi_\alpha = \sum_b \zeta(\alpha, b)\bar{n}_b f_b, \quad \alpha \in L(\Pi),$$

gives the eigenvalues of Γ, whose spectral decomposition is then $\Gamma = \sum_\alpha \xi_\alpha S_\alpha$. Thus the *eigenvalues* $\{\xi_\alpha : \alpha \in L(\Pi)\}$ constitute a third set of parameters for $V(\Pi, T)$, and there are two related sets of parameters which have also been used: the *specific components of variance* $\{\sigma_\alpha^2 : \alpha \in L(\Pi)\}$ of Cornfield & Tukey (1956), where $\sigma_\alpha^2 = \bar{n}_\alpha^{-1}\xi_\alpha$, and the *spectral components of variance* $\{\phi_\alpha : \alpha \in L(\Pi)\}$, of Daniels (1939), where $\phi_\alpha = n^{-1} d_\alpha \xi_\alpha$ and $d_\alpha = \text{rank}\,(S_\alpha)$. Table 2 summarizes the main representations of a factorial covariance matrix.

The nonnegativity of the eigenvalues $\{\xi_\alpha : \alpha \in L(\Pi)\}$ or, equivalently, $\{\sigma_\alpha^2 : \alpha \in L(\Pi)\}$ or $\{\phi_\alpha : \alpha \in L(\Pi)\}$, succintly defines the parameter space for $V(\Pi, T)$. There is no such simple characterization in terms of the covariances $\{\gamma_a : a \in L(\Pi)\}$, nor, in general, of the $\{f_b : b \in L(\Pi)\}$. In the linear model approach the latter parameters are the variances of

Table 2

Representations of a factorial covariance matrix

	$\sum_{a \in L(\Pi)} \gamma_a A_a$	$\sum_{b \in L(\Pi)} f_b R_b$	$\sum_{\alpha \in L(\Pi)} \xi_\alpha S_\alpha$
Using	Association matrices	Relationship matrices	Orthogonal projectors

Factorial Dispersion Models

independent sets of random variables, and so are constrained to be nonnegative. This constraint is stronger than the nonnegative-definiteness of Γ, and so the linear model approach is, in general, a proper subset of $V(\Pi, T)$. We shall show in § 2.4 that the classes coincide when T_p is infinite for all p in Π.

It is convenient to combine the relationships between γ's and f's and f's and ξ's to give

$$\xi_\alpha = \sum_a p_{\alpha a}\gamma_a, \quad \gamma_a = n^{-1}\sum_\alpha q_{a\alpha}\xi_\alpha, \tag{1}$$

where a, $\alpha \in L(\Pi)$ and the sums are over $L(\Pi)$. It is also true that

$$S_\alpha = n^{-1}\sum_a q_{a\alpha}A_a, \quad A_a = \sum_\alpha p_{\alpha a}S_\alpha \tag{2}$$

and the matrices $P = (p_{\alpha a})$ and $Q = (q_{a\alpha})$ thus hold the key to the solution of many later problems.

The $\varnothing$-row of the matrix P consists of the elements $p_{\varnothing a} = k_a = |\{u : a(t, u) = a\}|$, independent of t, whilst the Π-row of Q consists of $q_{\Pi\alpha} = d_\alpha = \text{rank}\,(S_\alpha)$, the number of so-called *degrees of freedom* in the *stratum* α: see below for an explanation of this terminology.

It can be shown that $k_a q_{a\alpha} = d_\alpha p_{\alpha a}$ and the simplest way to describe the entries of P and Q is via formulae for d_α, k_a and the common value $s_\alpha(a)$ of $d_\alpha^{-1}q_{a\alpha} = k_a^{-1}p_{\alpha a}$. These are as follows:

$$d_\alpha = \prod_{p\in\alpha\backslash\alpha_m} n_p \times \prod_{p\in\alpha_m}(n_p - 1),$$

where α_m denotes the set of *minimal* elements of α;

$$k_a = \prod_{p\in\bar{a}\backslash\bar{a}^m} n_p \times \prod_{p\in\bar{a}^m}(n_p - 1),$$

where $\bar{a} = \Pi\backslash a$ and $\bar{a}^m$ denotes the set of *maximal* elements of $\bar{a}$; and

$$s_\alpha(a) = \begin{cases} \displaystyle\prod_{p\in\alpha_m\backslash a}\{-1/(n_p - 1)\} & \text{if } \alpha\backslash\alpha_m \subseteq a, \\ 0 & \text{otherwise,} \end{cases}$$

where an empty product is defined to be unity.

For our *example* given in Fig. 1 and 2 the values of d_a, k_a and $s_\alpha(a)$ are shown in Table 3.

2.3 Finite index sets: Decomposition of arrays

The preceding approach permits a full discussion of the structure of matrices in the class $V(\Pi, T)$. We now turn to the random array $y = (y_t : t \in T)$ with dispersion matrix Γ in $V(\Pi, T)$. The matrices $\{S_\alpha : \alpha \in L(\Pi)\}$ are pairwise orthogonal projectors summing to the identity and so define an orthogonal decomposition of the n-dimensional space $\mathbb{R}^T$ of T-indexed arrays of real numbers, and hence also of the space of random arrays taking values in $\mathbb{R}^T$. Thus the decomposition

$$y = \sum_{\alpha\in L(\Pi)} S_\alpha y \tag{3}$$

of the array y into component arrays $S_\alpha y$ is orthogonal with respect to the standard inner product $\langle x, y\rangle = \sum_t x_t y_t$. Therefore we have the sum of squares decomposition $|y|^2 = \sum_\alpha |S_\alpha y|^2$, where $|y|^2 = \langle y, y\rangle = \sum_t y_t^2$.

Table 3

(*a*) *Values of d and k in the Example*

a	d_a	k_a
$\varnothing$	1	$(n_1-1)n_2(n_3-1)$
$\{1\}$	n_1-1	$(n_2-1)(n_3-1)$
$\{1,2\}$	$n_1(n_2-1)$	(n_3-1)
$\{3\}$	n_3-1	$(n_1-1)n_2$
$\{1,3\}$	$(n_1-1)(n_3-1)$	n_2-1
$\{1,2,3\}$	$n_1(n_2-1)(n_3-1)$	1

(*b*) *Values of* $s_\alpha(a)$ *in the Example*

α	$a=\varnothing$	$a=\{1\}$	$a=\{1,2\}$	$a=\{3\}$	$a=\{1,3\}$	$a=\{1,2,3\}$
$\varnothing$	1	1	1	1	1	1
$\{1\}$	$\frac{-1}{n_1-1}$	1	1	$\frac{-1}{n_1-1}$	1	1
$\{1,2\}$	0	$\frac{-1}{n_2-1}$	1	0	$\frac{-1}{n_2-1}$	1
$\{3\}$	$\frac{-1}{n_3-1}$	$\frac{-1}{n_3-1}$	$\frac{-1}{n_3-1}$	1	1	1
$\{1,3\}$	$\frac{1}{(n_1-1)(n_3-1)}$	$\frac{-1}{n_3-1}$	$\frac{-1}{n_3-1}$	$\frac{-1}{n_1-1}$	1	1
$\{1,2,3\}$	0	$\frac{1}{(n_2-1)(n_3-1)}$	$\frac{-1}{n_3-1}$	0	$\frac{-1}{n_2-1}$	1

By taking components of equation (3) and writing $S_\alpha y_t$ for $(S_\alpha y)_t$ we obtain the decomposition

$$y_t = \sum_{\alpha \in L(\Pi)} S_\alpha y_t \tag{4}$$

of the element y_t, for t in T, into components $S_\alpha y_t$ which depend only on indices in α. This is the *population identity* of Kempthorne (1952, Ch. 8), his arrays having permutation or sampling distributions, which we discuss below; it is also called the *yield identity* by Nelder (1965).

Example (cont.). This decomposition is the familiar one:

$$y_{ijk} = y_{...} + (y_{i..} - y_{...}) + (y_{ij.} - y_{i..}) + (y_{..k} - y_{...}) + (y_{i.k} - y_{i..} - y_{..k} + y_{...}) + (y_{ijk} - y_{ij.} - y_{i.k} + y_{i..}), \tag{4a}$$

where the terms correspond to $\alpha = \varnothing$, $\{1\}$, $\{1,2\}$, $\{3\}$, $\{1,3\}$ and $\{1,2,3\}$ respectively, and we denote the *averaging* over an index by a dot in that position.

Factorial Dispersion Models

For simplicity, suppose that $Ey = 0$. Since $\Gamma = \sum_\alpha \xi_\alpha S_\alpha$, the components of (3) are uncorrelated, as are the components of (4); thus the $S_\alpha y_t$ are the *principal components* of y. We find that $E(S_\alpha y_t)^2 = \phi_\alpha$ for α in $L(\Pi)$ and so $\text{var}(y_t) = \sum_\alpha \phi_\alpha$. Moreover, since $\phi_\alpha = n^{-1} d_\alpha \xi_\alpha$, we have

$$E\{d_\alpha^{-1} |S_\alpha y|^2\} = \xi_\alpha. \tag{5}$$

Example (cont.) As an illustration we consider $\alpha = \{1, 2, 3\}$. From Table 3 we have $d_{\{1,2,3\}} = n_1(n_2 - 1)(n_3 - 1)$ whilst (4a) gives $S_{\{1,2,3\}} y = (y_{ijk} - y_{ij.} - y_{i.k} + y_{i..})$; hence for any array $y = (y_{ijk})$ with zero mean and dispersion matrix in $V(\Pi, T)$ we have

$$E\left\{\frac{1}{n_1(n_2-1)(n_3-1)} \sum_i \sum_j \sum_k (y_{ijk} - y_{ij.} - y_{i.k} + y_{i..})^2\right\} = \xi_{\{1,2,3\}}.$$

From (1) and the values of k_a and $s_\alpha(a)$ in Table 3 we find that

$$\xi_{\{1,2,3\}} = \gamma_{\{1,2,3\}} - \gamma_{\{1,3\}} - \gamma_{\{1,2\}} + \gamma_{\{1\}}.$$

It should now be apparent from our Example, if not the general discussion, that we are providing an alternative interpretation of the *analysis of variance* of a multi-indexed array $y = (y_t)$. We have indicated how the components $S_\alpha y_t$ of (4) are the principal components of the random array y, provided that $Dy \in V(\Pi, T)$, and seen that their variances ϕ_α are, when suitably normalized by their multiplicities d_α and the array size n, the eigenvalues ξ_α of Dy. Because of the double role of the projectors S_α these components are also the terms which, when squared and summed, give the *sums of squares* decompositions that are such a familiar feature of analysis of variance *tables*. The *lines* of the analysis of variance table, termed *strata* by analogy with stratified sampling, are labelled by the filters of Π, that is by the elements α of the lattice $L(\Pi)$, and the number of *degrees of freedom* for the line labelled α is d_α, coinciding with the multiplicity of the corresponding eigenvalue ξ_α. And, finally, the *expected mean square* in line α given by (5) is the link between the principal components and the sum of squares decompositions. All this has been done by assuming only that $Dy \in V(\Pi, T)$; we have not *assumed* any linear model for the array y, although (4) is in a sense an *implicit* linear model. Note that we have assumed throughout that $Ey = 0$ and so our discussion is truly an analysis of *variance* qua variance; the introduction of structured *mean values* is an additional complication which we do not discuss here. All of this seems very similar to the discussion by Hannan (1965, § 5.2) and indeed the connection with spectral analysis can be made complete.

2.4 Infinite index sets

Most of the foregoing extends to the situation where some or all of the factors have countably infinitely many levels, and for simplicity we discuss the case that T_p is countably infinite for *all* p in Π. The representations $\Gamma = \sum_a \gamma_a A_a = \sum_b f_b R_b$ continue to hold (as these matrices are never multiplied), and we find that the spectral representation $\Gamma^{(\mathbf{n})} = \sum_\alpha \xi_\alpha^{(\mathbf{n})} S_\alpha^{(\mathbf{n})}$ of the truncation $\Gamma^{(\mathbf{n})}$ of Γ, converges, after a suitable normalization, to the representation $\sum_\alpha f_\alpha R_\alpha$. Here we use the superscript $(\mathbf{n})$ to denote the truncation to $t_p \leqslant n_p$ for all p in Π and our limits are all as $n_p \to \infty$ for all p in Π. Indeed $\phi_\alpha^{(\mathbf{n})} = n^{-1} d_\alpha^{(\mathbf{n})} \xi_\alpha^{(\mathbf{n})}$ converges to f_α, as does $\bar{n}_\alpha^{-1} \xi_\alpha^{(\mathbf{n})} = (\sigma_\alpha^{(\mathbf{n})})^2$, and we find that the components $S_\alpha^{(\mathbf{n})} y_t$ of y_t, where $t_p \leqslant n_p$ for all p in Π, also converge in mean square. The decomposition (4) of y_t continues to hold in the limit with the additional property that $E\{(S_\alpha y_t)(S_\alpha y_u)\} = 0$ if $(t_p : p \in \alpha) \neq (u_p : p \in \alpha)$, and so in this case (4) is (to second-order) just the often *assumed linear model* with random effects used in variance component analysis with the set Π of factors exhibiting the nesting structure characterized by $(\Pi; \leqslant)$.

T.P. Speed and R.A. Bailey

Example (cont.). The above implies that the linear model

$$y_{ijk} = \mu + \alpha_i + \beta_{ij} + \gamma_k + \delta_{ik} + \varepsilon_{ijk},$$

where μ, $\{\alpha_i\}$, $\{\beta_{ij}\}$, $\{\gamma_k\}$, $\{\delta_{ik}\}$ and $\{\varepsilon_{ijk}\}$ are uncorrelated sets of uncorrelated effects with zero means and variances $\sigma^2_{\varnothing}$, $\sigma^2_{\{1\}}$, $\sigma^2_{\{1,2\}}$, $\sigma^2_{\{3\}}$, $\sigma^2_{\{1,3\}}$ and $\sigma^2_{\{1,2,3\}}$ respectively is, to second order, the most general array y with $Ey = 0$, $Dy \in V(\Pi, T)$ with Π as in Fig. 1 and $T_1 = T_2 = T_3 = \{1, 2, \ldots\}$.

3 Permutation distributions

Suppose that $\eta = (\eta_t : t \in T)$ is a finite array of real numbers indexed by T as in § 2 above and that we define an array of random variables $y = (y_t : t \in T)$ by the rule $y_t = \eta_{\pi(t)}$ for t in T, where π is a *random permutation* of the index set T which respects the nesting relationships; see Bailey et al. (1983) for full details of the group G of all such permutations. Following Nelder (1954, 1965) we ask: What are the covariances induced by this randomization? It is not hard to see that to answer this question we do not need to know anything about the group G other than the following facts:

$$P(y_t = \eta_v) = n^{-1}, \quad P(y_u = \eta_w \mid y_t = \eta_v) = k^{-1}_{a(t,u)} \delta(a(t, u), a(v, w)),$$

where t, u, v and w are in T. With this information it is clear that $Ey_t = n^{-1} \sum_v \eta_v$, and, if $a(t, u) = a$,

$$Ey_t y_u = (nk_a)^{-1} \sum_v \sum_w A_a(v, w) \eta_v \eta_w.$$

By using the relations between the matrices $\{A_a\}$ and $\{S_\alpha\}$ given in (2) above, we find that $\Gamma = Dy = \sum_{\alpha \neq \varnothing} \xi_\alpha S_\alpha$, where $\xi_\alpha = d_\alpha^{-1} |S_\alpha \eta|^2$. Since $\sum_t y_t$ has the constant value $\sum_t \eta_t$ and $S_\varnothing y = y$, the eigenvalue $\xi_\varnothing$ is equal to zero. Using (1) we can obtain the covariances $\{\gamma_a\}$ of y in terms of the $\{\xi_\alpha\}$.

Example (cont.). We might ask for the covariance $\gamma_{\{3\}}$, which is $\operatorname{cov}(y_{ijk}, y_{i'j'k})$ with $i \neq i'$. Since $\gamma_{\{3\}} = n^{-1} \sum_\alpha q_{\{3\},\alpha} \xi_\alpha$, Table 3 shows that

$$n_1 n_2 n_3 \gamma_{\{3\}} = \xi_\varnothing - \xi_{\{1\}} + (n_3 - 1)(\xi_{\{3\}} - \xi_{\{1,3\}}).$$

We therefore need expressions for $\xi_{\{1\}}$, $\xi_{\{3\}}$ and $\xi_{\{1,3\}}$, as $\xi_\varnothing = 0$. From (4a) and (5) we see that $(n_1 - 1)\xi_{\{1\}} = n_2 n_3 \sum_i (\eta_{i..} - \eta_{...})^2$, and the corresponding expressions for $\xi_{\{3\}}$ and $\xi_{\{1,3\}}$ are as readily obtained.

4 Restricting to subsets

In many situations, including all of those for which T is infinite, we can observe only a finite part $y^{(U)} = (y_u : u \in U)$ of our random array $y = (y_t : t \in T)$, where U is some subset of T. What can we learn from $y^{(U)}$ about the various sets of parameters $\{\gamma_a\}$, $\{f_b\}$ and $\{\xi_\alpha\}$ of Dy? It is evident from the simplest examples that the restriction of a factorial dispersion model $V(\Pi, T)$ to a subset U of T does not necessarily result in a factorial dispersion model over U, so we are led into some broader aspects of analysis of variance which it is beyond our scope to cover fully here. For a in $L(\Pi)$, let $A_a^{(U)}$ be the restriction $A_a \,|_{U \times U}$ of the association matrix A_a over T to the subset U. The restriction $A_a^{(U)}$ may be zero for some values of a: however, if $A_a^{(U)}$ and $A_b^{(U)}$ are both nonzero for distinct filters a,

Factorial Dispersion Models

b in $L(\Pi)$, then $A_a^{(U)} \neq A_b^{(U)}$. Let $M = \{a \in L(\Pi) : A_a^{(U)} \neq 0\}$. It is clear that $\{A_a^{(U)} : a \in M\}$ satisfies conditions (i)–(iv) of § 2.2. It may happen that condition (v) is also satisfied, so that $\{A_a^{(U)} : a \in M\}$ forms an association scheme over U: in this case we shall call U a *tractable* subset of T. We note that, even if U is a tractable subset of T, the association scheme over U is not necessarily of the same kind as that over T: indeed, it may not even be a factorial association scheme.

Example (cont.). Let U be the subset $\{(i, j, j) : 1 \leqslant i \leqslant r,\ 1 \leqslant j \leqslant v\}$, where r and v are integers with $r \leqslant n_1$, $v \leqslant n_2$ and $v \leqslant n_3$. Then $A_{\{1,2\}}^{(U)} = A_{\{1,3\}}^{(U)} = 0$ and $M = \{\varnothing, \{1\}, \{3\}, \{1, 2, 3\}\}$. It may be checked that $\{A_a^{(U)} : a \in M\}$ satisfies condition (v), and so U is tractable. In fact, the association scheme over U is the factorial one corresponding to the simple crossed structure obtained from the poset Π in Fig. 1 omitting the factor 2. In § 5 we shall give an example where the association schemes on U and T are given by the *same* poset Π.

For simplicity let us suppose T to be finite; the extension to infinite T is quite straightforward. The standard theory of association schemes (MacWilliams & Sloane, 1977, p. 655) shows that, if U is a tractable subset of T, the set $\{A_a^{(U)} : a \in M\}$ of association matrices can be simultaneously diagonalized, with $|M|$ distinct common eigenspaces. Thus there are orthogonal projectors $\{S_\lambda^{(U)} : \lambda \in \Lambda\}$, where $|\Lambda| = |M|$, and, as in § 2.2, a $\Lambda \times M$ matrix $P^{(U)}$ and $M \times \Lambda$ matrix $Q^{(U)}$ such that

$$S_\lambda^{(U)} = |U|^{-1} \sum_{a \in M} q_{a\lambda}^{(U)} A_a^{(U)}, \quad A_a^{(U)} = \sum_{\lambda \in \Lambda} p_{\lambda a}^{(U)} S_\lambda^{(U)}.$$

We shall comment later on how the coefficients $q_{a\lambda}^{(U)}$ and $p_{\lambda a}^{(U)}$ may be found explicitly.

Just as in § 2.3, the projectors $\{S_\lambda^{(U)} : \lambda \in \Lambda\}$ define orthogonal decompositions of random arrays $y^{(U)}$ and of elements y_u of $y^{(U)}$ for u in U, and sum of squares decompositions with known expressions for expected mean squares. In short, when U is tractable then the dispersion model $\{\Gamma|_{U \times U} : \Gamma \in V(\Pi, T)\}$ exhibits the main features of an analysis of variance.

Write $\xi_\lambda^{(U)}$ for the eigenvalue of $Dy^{(U)}$ corresponding to the projector $S_\lambda^{(U)}$, where $\lambda \in \Lambda$. Then $\xi_\lambda^{(U)}$ is not, in general, equal to an eigenvalue of Dy. However, since $Dy^{(U)} = Dy\,|_{U \times U}$, we know that every covariance $\gamma_a^{(U)}$ in $Dy^{(U)}$ must appear, as γ_a in Dy. Thus we can combine the expression $\xi_\lambda^{(U)} = \sum_{a \in M} p_{\lambda a}^{(U)} \gamma_a^{(U)}$ for $\xi_\lambda^{(U)}$ with the formula (1) for γ_a to relate the expected mean square parameters $\{\xi_\lambda^{(U)} : \lambda \in \Lambda\}$ in the analysis of variance of $y^{(U)}$ to the analogous parameters $\{\xi_\alpha : \alpha \in L(\Pi)\}$ of y. This relationship may be expressed in matrix form as

$$\xi^{(U)} = P^{(U)} I^{(U,T)} |T|^{-1} Q \xi, \tag{6}$$

where $I^{(U,T)}$ is an $M \times L(\Pi)$ matrix with $I^{(U,T)}(a, b) = \delta_{ab}$.

A sufficient condition for tractability is that the restricted relationship matrices $\{R_a^{(U)} : a \in L(\Pi)\}$, where $R_a^{(U)} = R_a|_{U \times U}$, arise from a lattice of commuting uniform equivalence relations on U; see Speed & Bailey (1982) for definitions and fuller details. We note that in such cases, if at least one of $A_a^{(U)}$, $A_b^{(U)}$ is zero, then $R_a^{(U)}$ may be equal to $R_b^{(U)}$ even if $a \neq b$. For convenience let L^* be any subset of $L(\Pi)$ containing M such that $\{R_a^{(U)} : a \in L^*\} = \{R_a^{(U)} : a \in L(\Pi)\}$ but that $R_a^{(U)} \neq R_b^{(U)}$ whenever a and b are distinct elements of L^*. Then, for a in $L(\Pi)$, let a^* be the unique element of L^* such that $R_a^{(U)} = R_{a^*}^{(U)}$. In many cases L^* is a (not necessarily distributive) lattice of the type considered by Speed & Bailey (1982, § 2), who proved that analogues of all the earlier formulae and results hold with μ and ζ replaced by the Möbius and zeta functions of the lattice L^*. In particular, we have the following simple relationship between the

f-parameters of the two systems:

$$f_a^{(U)} = \sum_{b^* = a} f_b \quad (a \in L^*). \tag{7}$$

In most examples $M = L^*$ and the theory of § 2 applies to $y^{(U)}$ with no difficulty: we shall give an example of such a subset U in § 5. However, there are minor technical difficulties when $M \subset L^*$. One example where this happens is a 3×3 Graeco–Latin square viewed as a subset of a full $3 \times 3 \times 3 \times 3$ array. The difficulties are caused by the facts that $A_a^{(U)} = 0$ for $a \in L^* \backslash M$, and, correspondingly, some of the projectors defined in terms of the $R_a^{(U)}$ and the Möbius function of L^* are zero.

It should be clear that the foregoing discussion covers classical variance component estimates, at least in principle, although we do not discuss any general ways of disentangling estimates of population (that is, T) parameters from, say, quadratic forms in observed subsets. Equations (6) and (7) are most useful in the so-called 'balanced' or 'orthogonal' cases, where the observed subarray has a high degree of symmetry closely related to that of the full array. In most recent literature the full array is taken to be infinite, arising from an assumed linear model rather than an assumed covariance structure, but the results quoted in § 2.4 show that these two sets of assumptions are equivalent for our present purposes; see also Speed (1986).

5 Random sampling from structured populations

As in § 3, suppose that $\eta = (\eta_t : t \in T)$ is a finite population of real numbers indexed by T, and let U be an arbitrary subset of T. Let us consider *sampling* a random copy of U within T; that is, obtaining a random subset $\psi(U)$ of T which is labelled by U via the random injection ψ. Note that different choices of ψ may give rise to the same *set* $\psi(U)$ and yet must be considered different *samples*, because the labelling by the elements of U is a crucial feature of the sample. Thus random sampling amounts to random choice of ψ from some set Ψ of injections from U to T. We can now define a random array $y^{(U)} = (y_u : u \in U)$ indexed by U by putting $y_u = \eta_{\psi(u)}$ for u in U. The randomness underlying the distribution of $y^{(U)}$ is provided by the random sampling of ψ.

If we take Ψ to be (the restriction to U of) the group G mentioned in § 3, then $y^{(U)}$ is identical to the array obtained by restricting to U the random array $y = (y_t : t \in T)$ having the permutation distribution defined in § 3. Since all our results depend only on second-order properties, it follows from § 3 that all we require of Ψ is that it satisfy the following condition: for all u, v in U and all t, w in T,

$$P(\psi(u) = t) = n^{-1},$$
$$P(\psi(u) = t \mid \psi(v) = w) = k_{a(t,w)}^{-1} \delta(a(t, w), a(u, v)). \tag{*}$$

Thus *-random sampling an array of numbers is (to second order) the same as restricting to a subset of an array of random variables having an appropriate permutation distribution. Indeed the array η could well consist itself of random variables; provided that $D\eta \in V(\Pi, T)$, a sampling procedure satisfying (*) has no effect on the form of the dispersion matrix of the random variables selected, although the values of the individual covariances will change if $E\eta$ is not constant. If η has zero mean and dispersion matrix in $V(\Pi, T)$ then we may restrict at the outset to the subarray $\eta^{(U)} = (\eta_u : u \in U)$ of the desired form.

With this background we can carry out calculations concerning the sampled array using the structure on U derived from that on T. If U is a tractable subset of T then we may use

Factorial Dispersion Models

(6) and (7) to derive many expressions for average values of mean squares over sampling distributions: compare this with the work of Cornfield & Tukey (1956), Wilk (1955) and Wilk & Kempthorne (1956a,b; 1957).

Example. A triply-indexed population of numbers,

$$\eta = (\eta_{IJK} : I = 1, \ldots, N_1, J = 1, \ldots, N_2, K = 1, \ldots, N_3),$$

where the three factors 'rows', 'subrows' and 'columns' have the nesting relationship shown in Fig. 1, may be sampled as follows. Using simple random sampling without replacement, obtain n_1 rows $\psi_1(1), \ldots, \psi_1(n_1)$, and, independently within each of these, n_2 subrows $\psi_2(1, 1), \ldots, \psi_2(1, n_2), \ldots, \psi_2(n_1, 1), \ldots, \psi_2(n_1, n_2)$, and, independently of all the foregoing, n_3 columns $\psi_3(1), \ldots, \psi_3(n_3)$. We then form the array

$$(y_{ijk}) = (\eta_{\psi_1(i)\psi_2(i,j)\psi_3(k)} : i = 1, \ldots, n_1, j = 1, \ldots, n_2, k = 1, \ldots, n_3).$$

By the equivalence above, this is no different from restricting to the *first* n_1 rows, the *first* n_2 subrows within each row, and the *first* n_3 columns of the array (y_{IJK}) having the appropriate permutation distribution, for condition (*) is easily checked in this case. For example, if $I \neq I'$ and $i \neq i'$ then

$$P(\psi_1(i') = I',\ \psi_2(i', j) = J',\ \psi_3(k) = K \mid \psi_1(i) = I,\ \psi_2(i, j) = J,\ \psi_3(k) = K)$$

is equal to $1/(N_1 - 1)N_2$, which is as it should be since Table 3 shows that $k_{\{3\}} = (N_1 - 1)N_2$.

In this case it is clear that the sample $y = (y_{ijk})$ has the same lattice structure as the population: we may therefore use (6) to relate the two sets of parameters. For example, if we write the ξ-parameters for the population as $(\Xi_\alpha : \alpha \in L(\Pi))$ and those for the sample as $(\xi_\alpha : \alpha \in L(\Pi))$, Table 3 shows that

$$\xi_{\{1,3\}} = N_2^{-1}[n_2 \Xi_{\{1,3\}} + (N_2 - n_2)\Xi_{\{1,2,3\}}]. \tag{8}$$

In other words

$$E\left\{\frac{1}{(n_1 - 1)(n_3 - 1)} \sum_i \sum_j \sum_k (y_{i.k} - y_{i..} - y_{..k} + y_{...})^2\right\}$$

$$= \frac{n_2}{N_2(N_1 - 1)(N_3 - 1)} \sum_I \sum_J \sum_K (\eta_{I-K} - \eta_{I--} - \eta_{--K} + \eta_{---})^2$$

$$+ \frac{(N_2 - n_2)}{N_2 N_1 (N_2 - 1)(N_3 - 1)} \sum_I \sum_J \sum_K (\eta_{IJK} - \eta_{IJ-} - \eta_{I-K} + \eta_{I--})^2,$$

where we are using . and – to denote the sample and population averages respectively. If the array η were random with zero mean, we would simply enclose an expectation operator around the right-hand side. Similarly (7) shows that $\xi_{\{1,2\}} = n_3 f_{\{1,2\}} + f_{\{1,2,3\}}$, sample and population f's, and γ's, coinciding because $M = L^* = L(\Pi)$ and neither set of parameters directly involves the sizes of the arrays.

The foregoing example is a *regular* sample in the sense that $U = \prod_p U_p$, where $U_p \subseteq T_p$ for p in Π. Using the results and notation of § 2.2 with $n_p = |U_p|$ and $N_p = |T_p|$ for p in Π, it can be shown that, for a regular sample, the coefficient of ξ_β in the expansion (6) of $\xi_\alpha^{(U)}$ is $b_{\beta\alpha}$, which is equal to

$$(\bar{n}_\beta / \bar{N}_\beta) \prod_{p \in \beta_m \setminus \alpha} (1 - n_p/N_p) \tag{9}$$

if $\alpha \subseteq \beta$, and zero otherwise. This formula is given (in words) by Cornfield & Tukey

(1956) and may be used to derive the coefficients in (8). The first proof was given by Haberman (1975, Th. 2).

Example (cont.). Suppose that our rows and sub-rows are *blocks* and *plots*, respectively, and that our columns correspond to *treatments*. Then we may regard our initial sampling ψ as the selection of $n_1 n_2$ experimental units, in n_1 blocks of n_2 plots per block, together with n_3 treatments. Now put $n_1 = r$ and suppose that $n_2 = n_3 = v$, and allocate treatments to plots in such a way that plots in the same block receive different treatments, all such allocations being equally probable. There are many ways of doing this; we are in effect choosing a complete block design at random. As before, the labelling of the sample by i, j, k is important, and we do not want this labelling to destroy the relationship of having the same treatment. Thus we denote a complete block design by the function θ, where $\theta(i, k) = j$ if treatment k is assigned to plot j of block i. This procedure now defines a doubly-indexed array $\eta_{\psi_1(i)\psi_2(i,\theta(i,k))\psi_3(k)}$, which is a *random* $(1/v)$th *fraction* (subject to certain constraints) of our originally selected sample. However, it is easy to check that this (combined) sampling procedure still satisfies (*): for example,

$$P(\theta(i, k) = j \mid \theta(i', k') = j') = v^{-1}$$

whenever $i \neq i'$, and so

$$P(\psi_1(i') = I',\ \psi_2(i', \theta(i', k)) = J',\ \psi_3(k) = K \mid \psi_1(i) = I,\ \psi_2(i, \theta(i, k)) = J,\ \psi_3(k) = K)$$

is equal to $1/(r-1)v$ whenever $I \neq I'$ and $i \neq i'$. Note that, if we had replaced θ by the more natural function θ^* allocating treatments to plots, so that $\theta^*(i, j) = k$ whenever $\theta(i, k) = j$, and obtained the array

$$\eta_{\psi_1(i)\psi_2(i,j)\psi_3(\theta^*(i,j))},$$

then our combined sampling procedure would not have satisfied (*). This illustrates the care that is necessary in considering a random sample as a labelled subset.

By the equivalence given at the beginning of the section, we may regard the sample as being the *first* r blocks, the *first* v plots within each of these blocks, and the *first* v treatments, with the fraction selected being given by any one complete block design. It is convenient to use the complete block design in which the treatments have the *same* labels as the plots, in every block. Thus, our sampled array is $y^{(U)}$, where $U = \{(i, j, j)$: $i = 1, \ldots r, j = 1, \ldots, v\}$ and $(y_t : t \in T)$ has the permutation distribution based on η. We showed in § 4 that U is a tractable subset of T and that $M = \{\varnothing, \{1\}, \{3\}, \{1, 2, 3\}\}$. Moreover, $R^{(U)}_{\{1,2\}} = R^{(U)}_{\{1,3\}} = R^{(U)}_{\{1,2,3\}}$, whilst $R^{(U)}_{\varnothing}$, $R^{(U)}_{\{1\}}$, $R^{(U)}_{\{3\}}$ and $R^{(U)}_{\{1,2,3\}}$ are distinct: thus $L^* = M$. Equation (7) gives

$$f^{(U)}_{\varnothing} = f_{\varnothing},\quad f^{(U)}_{\{1\}} = f_{\{1\}},\quad f^{(U)}_{\{3\}} = f_{\{3\}},\quad f^{(U)}_{\{1,2,3\}} = f_{\{1,2\}} + f_{\{1,3\}} + f_{\{1,2,3\}}.$$

These identities were first derived by Wilk (1955).

The techniques of this and §§ 3 and 4 allow us to re-derive the results of Throckmorton (1961) and White (1963, 1975) in a unified and direct way which fully exploits the underlying combinatorial structure; but see also the Appendix of Neymann et al. (1935). We make no comments here on the relative merits of these as compared with other approaches to the analysis of experimental data; a discussion which did justice to the topic would take us too far from the main subject of this paper.

6 Prediction

Our final topic is the best (that is, minimum mean-squared error linear) *prediction* of linear combinations of elements of an array $y = (y_t : t \in T)$ with $Ey = 0$, and known Dy in

Factorial Dispersion Models

$V(\Pi, T)$, based upon the observation of a finite subarrary $y^{(U)}$ where $U \subseteq T$. As an illustration from many possible results we shall suppose that $U = \prod_p \{1, \ldots, n_p\}$ is a regular sample from $T = \prod_p \{1, \ldots, N_p\}$ where $n_p \leqslant N_p \leqslant \infty$ for p in Π and even here we shall consider the prediction of only the components $(S_\alpha y_t : \alpha \in L(\Pi),\ t \in U)$ in (4) above. In this case it can be shown that the best predictor of $S_\alpha y_t$ based upon $y^{(U)}$ when $t \in U$ is

$$\xi_\alpha^{(T)} \sum_\beta [\xi_\beta^{(U)}]^{-1} b_{\alpha\beta} S_\beta^{(U)} y_t, \tag{10}$$

where, for $\beta \in L(\Pi)$, $S_\beta^{(U)}$ denotes the matrix introduced in § 2 above for the array $y^{(U)}$. If $t \notin U$ a more complicated expression can be derived. Predictors of more general linear combinations of elements of y are best derived using (10) and linearity. If any of the N_p is infinite we must pass to a limit in (10) by combining $\xi_\alpha^{(T)}$ with the $b_{\alpha\beta}$ which, of course, depend upon the N_p for p in Π.

Example (cont.). Let us compute the best predictor of $y_{i--} - y_{---}$; that is, let us evaluate (10) with Π as in Fig. 1 and $\alpha = \{1\}$. Note that the averages we are predicting are in the population (T-indexed) array, and this will be done using averages from the sample (U-indexed) array. From (9) we find that $b_{\{1\},\{1\}} = n_2 n_3 / N_2 N_3$ whilst $b_{\{1\},\varnothing} = n_2 n_3 (1 - n_1/N_1)/N_2 N_3$ and so

$$\hat{y}_{i--} - \hat{y}_{---} = \xi_{\{1\}}^{(T)} \frac{n_2 n_3}{N_2 N_3} \left[(y_{i..} - y_{...}) / \xi_{\{1\}}^{(U)} + \left(1 - \frac{n_1}{N_1}\right) y_{...} / \xi_{\varnothing}^{(U)} \right].$$

If we let N_1, N_2 and N_3 all tend to infinity and expand the ξ's in terms of f's, this expression simplifies to

$$\frac{n_2 n_3 f_{\{1\}}}{n_2 n_3 f_{\{1\}} + n_3 f_{\{1,2\}} + n_2 f_{\{1,3\}} + f_{\{1,2,3\}}} (y_{i..} - y_{...}) + \frac{n_2 n_3 f_{\{1\}}}{n_1 n_2 n_3 f_\varnothing + n_2 n_3 f_{\{1\}} + n_3 f_{\{1,2\}} + n_1 n_2 f_{\{3\}} + n_2 f_{\{1,3\}} + f_{\{1,2,3\}}} y_{...}.$$

It is clear that the above discussion is essentially what some writers term the *estimation of random effects* (Harville, 1976). Our approach places it firmly within a prediction framework, of unobserved random variables by observed ones, but the two are, of course, equivalent. As we remarked earlier, matters become more complicated in the presence of a structured mean value (fixed effects), and we shall say nothing further about this topic here.

7 Closing remarks

There are many more aspects of this topic which could be addressed if space permitted. For example, concise *rules* for forming analysis of variance tables, that is sums of squares, degrees of freedom and expected mean squares and so on, can be formulated and proved in the above framework. Other randomization analyses can be obtained with a minimum of effort, for example, for split plot and other more complex designs. These topics and a number of others will be expounded together with a fuller exposition of the material we have surveyed in a forthcoming monograph mentioned in § 1.

References

Aigner, M. (1979). *Combinatorial Theory*. New York: Springer Verlag.
Aldous, D.J. (1981). Representations for partially exchangeable arrays of random variables. *J. Mult. Anal.* **11,** 581–598.

T.P. Speed and R.A. Bailey

Bailey, R.A. (1984). Discussion of paper by T. Tjur. *Int. Statist. Rev.* **52,** 65–77.

Bailey, R.A., Praeger, C.E., Rowley, C.A. & Speed, T.P. (1983). Generalized wreath products of permutation groups. *Proc. Lond. Math. Soc.* (3) **47,** 69–82.

Brown, W. (1913). The effects of "observational errors" and other factors upon correlation coefficients in psychology. *Br. J. Psychol.* **6,** 223–238.

Bryant, P. (1984) Geometry, statistics, probability: variations on a common theme. *Am. Statistician* **38,** 38–48.

Cochran, W.G. (1939). The use of the analysis of variance in enumeration by sampling. *J. Am. Statist. Assoc.* **34,** 492–510.

Cornfield, J. & Tukey, J.W. (1956). Average values of mean squares in factorials. *Ann. Math. Statist.* **27,** 907–949.

Cox, D. (1960). Discussion of paper by R. L. Plackett. *J. R. Statist. Soc.* B **22,** 209–210.

Daniels, H.E. (1938). Some problems of statistical interest in wool research (with discussion). *J. R. Statist. Soc. Suppl.* **5,** 89–128.

Daniels, H.E. (1939). The estimation of components of variance. *J.R. Statist. Soc. Suppl.* **6,** 186–197.

Dawid, A.P. (1977). Invariant distributions and analysis of variance models. *Biometrika* **64,** 291–297.

Eden, T. & Yates, F. (1933). On the validity of Fisher's *z* test when applied to an actual sample of non-normal data. *J. Agric. Sci.* **23,** 6–17.

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

Fisher, R.A. (1925). *Statistical Methods for Research Workers.* Edinburgh: Oliver & Boyd.

Haberman, S.J. (1975). Direct products and linear models for complete factorial tables. *Ann. Statist.* **3,** 314–333.

Hannan, E.J. (1965). *Group Representations and Applied Probability.* London: Methuen.

Harville, D. (1976). Extension of the Gauss–Markov theorem to include the estimation of random effects. *Ann. Statist.* **4,** 384–395.

Hooke, R. (1956a). Symmetric functions of a two-way array. *Ann. Math. Statist.* **27,** 55–79.

Hooke, R. (1956b). Some applications of bipolykays to the estimation of variance components and their moments. *Ann. Math. Statist.* **27,** 80–98.

John, P.W.M. (1971). *Statistical Design and Analysis of Experiments.* New York: Macmillan.

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

Lush, J.L. (1931). The number of daughters necessary to prove a sire. *J. Dairy Sci.* **14,** 209–220.

Lush, J.L., Henderson, A.L., Culbertson, C.C. & Hammond, W.E. (1933). The reliability of some measures of the productiveness of individual brood sows. *Record of the Proceedings of the Annual Meeting of the American Society of Animal Production* **26,** 282–298.

MacWilliams, F.J. & Sloane, N.J.A. (1977). *The Theory of Error-Correcting Codes.* Amsterdam: North Holland.

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. Block structure and the null analysis of variance. *Proc. R. Soc. Lond.* A **283,** 147–162.

Nelder, J.A. (1977). A reformulation of linear models (with discussion). *J. R. Statist. Soc.* A **140,** 48–76.

Neyman, J. with co-operation of K. Iwaszkiewicz and St. Kołodziejczyk (1935). Statistical problems in agricultural experimentation (with discussion). *J. R. Statist. Soc. Suppl.* **2,** 107–180. See also *A Selection of Early Statistical Papers of J. Neyman* (1967). Cambridge University Press.

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

Saville, D.J. & Wood, G.R. (1986). A method for teaching statistics using *N*-dimensional geometry. *Am. Statistician* **40,** 205–214.

Scheffé, H. (1956). Alternative models for the analysis of variance. *Ann. Math. Statist.* **27,** 251–271.

Spearman, C. (1910). Correlation calculated from faulty data. *Br.J. Psychol.* **3,** 271–295.

Speed, T.P. (1986). ANOVA models with random effects: An approach via symmetry. In *Essays in Time Series and Allied Processes: Papers in Honour of E.J. Hannan,* Ed. J. Gani and M.B. Priestly, pp. 355–368. Sheffield: Applied Probability Trust.

Speed, T.P. (1987). What is an analysis of variance? *Ann. Statist.* **15.** To appear.

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, pp. 55–74. New York: Marcel Dekker.

Throckmorton, T.N. (1961). Structures of classification data. Iowa State University PhD thesis.

Tjur, T. (1984). Analysis of variance models in orthogonal designs (with discussion). *Int. Statist. Rev.* **52,** 33–81.

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

White, R.F. (1963). Randomization analysis of the general experiment. Iowa State University PhD thesis.

White, R.F. (1975). Randomization and the analysis of variance. *Biometrics* **31,** 552–571.

Wilk, M.B. (1955). The randomization analysis of a generalized randomized block design. *Biometrika* **42,** 70–79.

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

Wilk, M.B. & O. Kempthorne (1956b). Analysis of variance: Preliminary tests, pooling, and linear models, **2**. Wright Air Development Center Tech. Rep. 55-244.
Wilk, M.B. & Kempthorne, O. (1957). Non-additivities in a Latin square design. *J. Am. Statist. Assoc.* **52,** 218–236.
Yates, F. & Zacopanay, I. (1935). The estimation of the efficiency of sampling, with special reference to sampling for yield in cereal experiments. *J. Agric. Sci.* **25,** 545–577.
Youden, W.J. & Mehlich, A. (1937). Selection of efficient methods for soil sampling. *Contrib. Boyce Thompson Inst.* **9,** 59–70.

Résumé

On étudie un ensemble de modèles de structure des covariances pour des tableaux multi-indexés (i.e. indexés par les éléments d'un produit cartésien) de variables aléatoires. Par leurs propriétés au second ordre, ces modèles généralisent les modèles de composantes de la variance, les modèles de randomisation, ainsi que les modèles d'échangeabilité. Ils conduisent de façon naturelle à des techniques générales pour effectuer des décompositions orthogonales, calculer les espérances des carrés moyens et évaluer les autres quantités intervenant dans l'analyse de variance de ce type de tableaux.

[*Received April* 1986, *revised March* 1987]

The Annals of Statistics
2002, Vol. 30, No. 6, 1649–1665

JOHN W. TUKEY'S CONTRIBUTIONS TO ANALYSIS OF VARIANCE

BY T. P. SPEED

University of California, Berkeley

John Tukey connected the theory underlying simple random sampling without replacement, cumulants, expected mean squares and spectrum analysis. He gave us one degree of freedom for nonadditivity, and he pioneered finite population models for understanding ANOVA. He wrote widely on the nature and purpose of ANOVA, and he illustrated his approach. In this appreciation of Tukey's work on ANOVA we summarize and comment on his contributions, and refer to some relevant recent literature.

1. Introduction. Most (9/15) of John Tukey's contributions to analysis of variance (hereafter ANOVA) can be found in Volume 7 of *The Collected Works of John W. Tukey* [17]. Also in that volume are two items which will be of interest to readers of this paper. One is a six-page foreword to the nine collected papers by John Tukey himself. The other is an historical introduction to and remarks on the roles of analysis of variance, and some brief comments on the individual papers by the volume editor, David R. Cox. However, Tukey being Tukey, there is no substitute for reading the papers themselves. Every one of them advances our knowledge, at times dramatically, while seeming to be no more than a lucid exposition from first principles of some well-established part of our subject. There are exceptions to this last statement.

John Tukey's main published contributions to ANOVA were made in a little over a decade, from 1949 to 1961. They constitute approximately 20% of his output over this period, and so about 5% of his total output. In subject matter these papers range from the foundational to the computational, from the algebraic to the interpretational, and contain some strikingly original views of the topics he discusses. How many of us see a clear connection between finite-population simple random sampling as in books on sampling, Fisher's k-statistics and cumulants for calculating moments of sample moments, the moments of mean squares in ANOVA tables and the arithmetic of spectrum analysis? At the same time as he was clarifying the analysis of variance qua variance, he highlighted the importance of scale to the notion of interaction in the analysis of means, and gave us a tool for identifying and removing removable nonadditivity. He also showed us how to analyze a complex multifactorial data set; indeed in no fewer than four of the

Received January 2002; revised March 2002.
AMS 2000 subject classifications. Primary 62J10; secondary 94A20.
Key words and phrases. Odoffna, ANOVA, moments, cumulants, k-statistics, polykays, variances, components of variance, mean squares, factorials, interactions, pigeonhole model.

papers below we get his views on the nature and purpose of ANOVA. It was much broader than the usual one which focusses on testing.

In my opinion much of Tukey's work on ANOVA is underappreciated, and much of that which was appreciated at the time has been forgotten. He laments [17, page lii], wrongly as it turns out, "Perhaps regrettably, I am not aware of very much that extends papers 5, 6, 7, and 9" (of [17]). Some of his work on ANOVA, for example, his "dyadic ANOVA" and his "components in regression," was never followed up. Neither of these titles scores a hit (with Tukey's meaning) in *Current Index to Statistics*. Fashions change, and the foundational worries or solutions of one generation of statisticians can cease to be of interest to a later generation. It is for this reason as well as to celebrate Tukey's genius that it is a real pleasure to be able to remind readers of his wonderful contributions to ANOVA, including creating the abbreviation itself.

2. ODOFFNA. How we will miss Tukey's neologisms. His one degree of freedom for nonadditivity (ODOFFNA) paper [2] is perhaps his best-known and most striking contribution to the analysis of variance and needs little introduction here. Whereas others had paid attention to nonconstancy of the variance or nonnormality of the "errors" in ANOVA, Tukey was concerned with nonadditivity. Explaining his ideas in the context of a singly replicated row-by-column table, he showed how to isolate a single degree of freedom from the "residual," "error" or "interaction" sum of squares ("call it what you will" he said), and so test the null hypothesis of additivity using a statistic which gave power against a restricted class of multiplicative alternatives. The statistic was motivated by the idea of a power transformation; it was illustrated graphically through three examples, and some elegant distribution theory was presented. This is a gem of a paper and amply deserves its place in the texts [see, e.g., Scheffé (1959) or Seber (1977)]. Tukey's later papers [5, 13] on the same topic present no new ideas; rather they illustrate the earlier ideas in more general contexts, something he pointed out was possible in [2].

What has happened to ODOFFNA since the 1960s? These days most people concerned about the possibility that their linear model might better satisfy the standard assumptions of additivity, homoscedasticity and normality of errors after a transformation will make use of the Box and Cox (1964) theory. However, their approach to transformations is not a complete substitute for ODOFFNA, as can be seen in Tukey's [14] discussion of additive and multiplicative fits to two-way tables (see especially [14], Section 10F). It is likely that we will continue to extract ODOFFNA in new contexts in the future, and for more on this, see Tukey's own comments on the follow-up to ODOFFNA in his foreword to [17].

3. Complex analyses of variance: general problems [11]. In [11] Green and Tukey made a number of general points concerning complex analyses of variance in the course of analyzing a specific experimental data set. Some of the points are

familiar, some were new at the time but most are still of interest today. The authors explain that the purpose of their analysis is "to provide a simple summary of the variation in the experimental data, and to indicate the stability of means and other meaningful quantities extracted from the data." They intended their approach to be in opposition to the view that the sole purpose of ANOVA is to provide tests of significance. It was aimed at researchers in psychology and followed a review of the use of ANOVA in that field a few years earlier.

The experiment is from psychophysics and involves six factors: sex (S, two levels), sight (I, two levels), persons (P, eight levels), rate (R, four levels), weight (W, seven levels) and date (D, two levels). All of S, I, R, W and D are crossed, while P is nested in a balanced way within $S \times I$ so we may describe the factor relationships by the formula $((S \times I)/P) \times R \times W \times D$. The response was a difference limen, a kind of threshold of perception, which could be expressed as a difference in weights, a squared difference in weights, a ratio of weights, a logarithm of a ratio of weights or even a response time.

One novel aspect of this paper is that the authors discuss not only what scale to use for the dependent variable; that is, possible transformations, but also just what the dependent variable should be in that context: a difference, a squared difference, a ratio, a log ratio, etc. After an initial analysis with one response variable, they choose another and obtain a new, and to their minds better, analysis. Another novelty at that time was the careful discussion of the nesting and crossing between factors and their implications for the analysis. This was no doubt inspired by the discussion of these matters Tukey and Cornfield gave in [8], which was published some four years before [11].

Perhaps the most interesting part of this paper is the extended section "Variance components and the proper error term" and the section "Variance components in the illustrative example" which follows it. The first of these discusses an example simpler than the actual experiment and draws heavily on material concerning the pigeonhole model in [8]; see Section 4.3. Then they turn to the experiment and things get interesting when they seek to impose a sampling model on the factors. The four levels of rate (50, 100, 150 and 200 g/s) and the seven levels of weight ($100, 150, \ldots, 400$ g) are admitted to present a problem for their pigeonhole model. Are they exhaustive samples from finite populations, that is, fixed; are they small samples from large populations of levels, that is, random; or are they something else? Whereas it was easy for them to view sex and sight (blind or not) as fixed, and person as random, the choice for R and W was far less obvious. After some discussion of various options, including a mention of using polynomials to fit responses to rate and weight, they decide to regard R and W as random "although we recommend against this procedure [for scaled variables] in general." The ideal that one ANOVA theory fits all cases seems hard to live up to, even when you are the creator of the theory.

As soon as all factors are assigned the category fixed or random, it is possible to write out all 39 expected mean square lines of the ANOVA table, and this they

do. Next follows an illuminating discussion of "aggregation and pooling" of lines in the table, which, when implemented with the illustrative data, reduces the 39 lines to 15. They make use of a modified version of a procedure of Paull (1950) which Tukey highlights in his Introduction to [17] and seems to be of interest today. There are two useful graphical representations of the relative contributions of the different sources of variability, one in two dimensions which is especially appealing, but on the whole there is relatively little plotting of the data, a large contrast with Tukey's later work, for example, in [14].

A later analysis of this same data set can be found in Johnson and Tukey [15]. Looking back on this paper after four decades, and bearing in mind all that Tukey wrote on ANOVA before and after that time, one cannot help but be struck by how little use he made in this paper of the processes and procedures he recommended when considering such an analysis. Referring to matters to be discussed in Section 4, he made no attempt to assign standard errors to his estimated variance components, under either normality or any other assumptions, the scientific purpose of the experiment was nowhere mentioned, the situations or populations to which inference was to be made were nowhere mentioned, even the means he calculated and plotted were not assigned any measures of their stability, something that was stated at the beginning of the paper to be one of the major purposes of ANOVA. Granted this was an expository paper with a limited objective, and probably already long by the standards of the journal, but I think the point remains that it is hard to put Tukey's ANOVA theory into practice, even for Tukey himself.

4. Some moment calculations. Tukey wanted to derive average values and variances and later a third moment of consider later. He tells us [17, page liv] that his first attempt at deriving the variance of the between variance component in an unbalanced one-way design took five or six full days "using old-fashioned clumsy methods." He was "convinced that it ought not to be so hard" and so "went looking for better tools, and eventually came out with the polykays." Polykays are generalizations of Fisher's k-statistics and we now outline the main points from the papers in which they were introduced.

4.1. *Some sampling simplified; keeping moment-like computations simple* [3, 6]. In 1929 Fisher introduced k-statistics as unbiased estimators of cumulants and a computational technique which radically simplified much previous research on moments of moments. It would take us too far astray to describe his technique in detail [see Speed (1986a)], but we can describe the simplest of his results in this area as soon as we recall the following well-known facts. If $X_1, \ldots, X_n$ are i.i.d. random variables with common first two cumulants κ_1 and κ_2 (the mean and variance, respectively), then

$$k_1 = \frac{1}{n}\sum X_i = \bar{X} \quad \text{and} \quad k_2 = \frac{1}{n-1}\sum (X_i - \bar{X})^2$$

satisfy

$$\mathbf{E}k_1 = \kappa_1 \quad \text{and} \quad \mathbf{E}k_2 = \kappa_2.$$

Now the k's are Fisher's k-statistics, that is, unbiased estimates of the corresponding cumulants. The key results of Fisher (1929) were the general definition of k-statistics and a procedure for calculating their joint cumulants whose core was a rule for calculating the coefficients of lower order k-statistics in an expansion for the product of two k-statistics. The relationships above are the simplest relevant results: the expected values or first cumulants of the first two k-statistics. Next would come the results which come from replacing $\mathbf{E}$ by var or covar; that is, replacing first by second cumulant in the sample-population calculation.

We all know that $\text{var}(k_1) = \kappa_2/n$, but what about $\text{var}(k_2)$? This result, first derived by Gauss, is not quite so well known, but turns out to be

$$\text{var}(k_2) = \frac{2}{n-1}\kappa_2^2 + \frac{1}{n}\kappa_4.$$

Deriving this last fact is already messy enough to warrant thinking very carefully about the algebraic formulation one adopts, and any desire to obtain more general expressions of the same kind focusses the mind greatly on the same issue. Fisher had his approach, Tukey simplified it as we shall see and it can be simplified yet again; see Speed (1983) and McCullagh (1987).

Tukey's main aim in [3] and [6] was to extend these results (and others like them) to the finite population case. Apparently unknown to Tukey, this task had been begun by Neyman in 1923 [see Neyman (1925)], though far less elegantly or generally. To achieve his aim Tukey extended Fisher's entire machinery. He named the tool he developed polykays—multiply-indexed generalizations of k-statistics—later noting that these same functions had been introduced earlier by Dressel (1940) in a paper that was not noticed at the time. For Tukey polykays of order or weight r are indexed by partitions of the natural number r. For example, there are two of order 2, indexed by $(1, 1)$ and (2); three of order 3, indexed by $(1, 1, 1)$, $(1, 2)$ and (3); four of order 4, indexed by $(1, 1, 1, 1)$, $(1, 1, 2)$, $(1, 3)$ and (4); and so on. Fisher's k-statistics are the single subscript versions of the polykays, (1), (2), (3), (4) etc., hence Tukey's name. In what follows we drop the commas and parentheses in the partition notation, writing $1, 11, 2$, etc.

How are polykays defined in general? To do this Tukey made use of an auxiliary class of symmetric functions also labelled by partitions, which he called symmetric means or, more simply, brackets, denoted by $\langle 1\rangle$, $\langle 11\rangle$, $\langle 2\rangle$, etc. These functions had the appealing property of rather transparently being "inherited on the average," which means that the average of the sample function over simple random sampling without replacement from a finite population was just the corresponding population function. Tukey avoided using the term "unbiased" as (so he said) "there are now so many kinds of unbiasedness!" The sample mean

$$\langle 1\rangle = \frac{1}{n}\sum x_i$$

is clearly inherited on the average, as is

$$\langle 11\rangle = \frac{1}{n(n-1)}\sum_{i\neq j} x_i x_j.$$

The value of brackets lies in the fact that [3, page 111] "every expression which is (i) a polynomial, (ii) symmetric, (iii) inherited in the average, can be written as a linear combination of brackets with coefficients which do not depend on the size of the set of numbers involved." As one illustration we give the following well-known and useful representation:

$$\frac{1}{n-1}\sum(x_i-\bar{x})^2 = \frac{1}{n}\sum x_i^2 - \frac{1}{n(n-1)}\sum_{i\neq j} x_i x_j,$$

where the last two terms are transparently inherited in the average, neatly proving that the first term is also, a standard fact from sampling theory. Tukey would write this last relationship $(2) = \langle 2\rangle - \langle 11\rangle$, and in general he needed a rule giving the coefficients of brackets in the expansion of his parentheses (polykays). As he said ([3], page 124) "the single-index brackets have the coefficients for moments in terms of cumulants (given numerically by Kendall [(1943), Section 3.13] up to the 10th moment). The coefficients of brackets with several indices can be found by formal multiplication."

How do we use all this machinery? Elegant though it is, there is still some hard work: the multiplication tables need to be derived. Tukey derived his own, but by the time of publication of [3, 6] comprehensive tables had independently appeared [Wishart (1952a, b)]. A simple instance of a multiplication rule is

$$(*) \qquad (2)^2 = (22) + \frac{1}{n}(4) + \frac{2}{n-1}(22)$$

Let us see how this leads very painlessly to the main result of Neyman (1925). First, note that the preceding identity has a version connecting population k-statistics which is of the same form, but with n replaced by N. Next recall that the polykays (22), (4), etc. are all "inherited on the average." We now take the expectation (i.e., average) of $(*)$ over all samples and subtract from the result the population version of $(*)$. This leaves us with

$$\operatorname{var}((2)) = \left[\frac{1}{n} - \frac{1}{N}\right](4) + 2\left[\frac{1}{n-1} - \frac{1}{N-1}\right](22),$$

which is the formula Neyman worked hard to obtain. This was indeed "sampling simplified." Note also that if we let $N \to \infty$ (so-called infinite population) and use the easily proved fact that, in this case, (22) is just $(2)^2$, we obtain Gauss' result.

Tukey certainly simplified sampling. He demonstrated clearly that indeed finite populations are simpler to deal with, and more powerful, and he now had the machinery to carry out certain calculations in ANOVA.

Later developments cast Tukey's work in the framework of tensors [cf. Kaplan (1952) and, most recently within the general theory of symmetric functions, Speed (1986a)]. The gains from so doing are modest, but I think definitely worthwhile. One consequence of the tensor formulation is that some of Tukey's formal calculations (e.g., his symbolic o-multiplication) cease to be "tricks." Another is the greater simplicity which comes from allowing all random variables to be potentially different. For example, instead of calculating variances of variances, we calculate covariances of distinct covariances, and obtain variances by appropriately equating arguments. With this slightly greater generality, (*) above becomes [Speed (1986a), page 43]

$$(12) \otimes (34) = (12|34) + \frac{1}{n}(1234) + \frac{1}{n-1}[(13|24) + (14|23)],$$

where 1, 2, 3 and 4 all label distinct variables. This simplification removes certain multiplicity factors and then reveals the coefficients defining polykays to be values of the Möbius function over a partition lattice, which I think is a real step forward; see Speed (1983) and McCullagh (1987).

Where are polykays now? There was a little theoretical development of them after Tukey's work, but he left no major problems unaddressed. I extended them to multiply-indexed arrays in Speed (1986a, b) and Speed and Silcock (1988a), and used the extensions to generalize the calculations of Tukey discussed in the next section. Apart from my own work the most recent references to polykays are Tracy (1973) and, an application of them, McCullagh and Pregibon (1987). To my knowledge there have been no other publications concerning polykays since then. In short, it seems that after about 25 years of life, polykays have been dead or sleeping for 25 years. Apparently they have served their purpose, though I have no doubt that they will be resurrected, awakened or reborn at some time in the future, when another problem comes along for whose solution they are the natural tool.

4.2. *Variances of variance components* [7, 9, 10]. Why did Tukey go to all the trouble of inventing polykays and their calculus, and what did he learn from so doing? Giving as one purpose of the analysis of variance "to estimate the sizes of the various components contributed to the overall variance from the corresponding sources," he wanted "to obtain formulas for the variances of the natural estimates of these variance components." Along with Gauss, Fisher and many others, Tukey wanted to go beyond normality, but almost uniquely he did so in dispensing with infinite populations. He regretted ([7], page 157) that he still had to leave "the customary (and dangerous) independence assumptions" concerning the terms in his linear population models. This answers the question "Why?" Let us now see some of what he learned in a simple case: the balanced single (or one-way) classification. Tukey's model for this takes the form

$$x_{ij} = \mu + \eta_i + \omega_{ij}, \qquad i = 1, \ldots, c, j = 1, \ldots, r,$$

where the $\{\eta_i\}$ are sampled from a population of size n with k-statistics $k_1, k_2, \ldots$, the $\{\omega_{ij}\}$ are from a population of size N with k-statistics $K_1, K_2, \ldots$ and the samplings are independent and order randomized. If we denote by B and W the usual between-class and within-class mean squares, respectively, with expectations k_2 and K_2, then Tukey showed, among other results, that

$$\begin{aligned}
\operatorname{var}(B) &= \left(\frac{1}{c} - \frac{1}{n}\right)k_4 + 2\left[\frac{1}{c-1} - \frac{1}{n-1}\right]k_{22} \\
&\quad + \frac{4}{r(c-1)}k_2K_2 + \frac{2(rc-1)}{r^2c(r-1)(c-1)}K_{22}, \\
\operatorname{cov}(B, W) &= -\frac{2}{rc(r-1)}K_{22}, \\
\operatorname{var}(W) &= \left[\frac{1}{rc} - \frac{1}{N}\right]K_4 + 2\left[\frac{1}{c(r-1)} - \frac{1}{N-1}\right]K_{22}.
\end{aligned}$$

The remainder of [7] consists of more formulae of this kind, derived for other variance component models: row-by-column classifications, Latin squares, balanced incomplete blocks and more general balanced models.

Paper [9] considers the special, more complicated case of an unbalanced one-way classification. One novelty here is that there is no single compelling estimate of the between-class component of variance, and so Tukey considers a class of estimates involving weights which need to be specified. He then derives the variances and covariances as before, generalizing those just given, and presents numerical examples. Lastly, paper [10] does what its title says: it presents the third moment about the mean, that is, the third cumulant of the quantity W given above.

What can we learn from or do with such formulae? In the first place, we can obtain qualitative insights by comparing the general finite population results with the special case of infinite normal populations. There k_4 and K_4 vanish, while k_{22} and K_{22} are k_2^2 and K_2^2, respectively, and of course $N = \infty$. In this case the results are familiar, and the extent to which the normal variances for the estimated variance components are too small or too large could, in principle, be examined. Interestingly, Tukey does not present formulae giving unbiased estimates of either the individual terms in his expressions for the variances of the estimated variance components, or for the variance expression as a whole. I would be very surprised if he did not have such formulae, for example, for k_4 and K_4 and k_{22} and K_{22} above, but he makes no mention of them. Without them, his aim of calculating estimates of the precision of estimated variance components under these more general assumptions must remain unfulfilled.

What has been done since the 1950s in this area? There has been more work on the topic of variances of estimated components and variance; see, for example, Harville (1969), but there, as in all other such cases that I know, the calculations are carried out under an assumption of normality. In some of my own work [Speed

(1986a, b), Speed and Silcock (1988a, b)] I have tried to extend Tukey's work to ANOVA models which are not built up additively from independent components.

4.3. *Average values of mean squares in factorials* [8]. This is an interesting and important paper: broad in coverage, profound in its analysis, beautifully written and elegant in its dealing with messy algebraic details. It is arguably Tukey's most important contribution to ANOVA. By the early to mid-1950s it was becoming clear that the concise description in Eisenhart (1947) of models for ANOVA did not provide a foundation for all uses of ANOVA. The now well-known mixed-model ambiguity concerning the interaction component of variance when (say) rows are "fixed" and columns "random" had emerged: in some linear model formulations this component appeared in the expected mean square line for both rows and columns, while in others it did not. It was apparent to many that the combining of linear models and ANOVA was not as simple as might have seemed at first. Neyman and his Polish colleagues found this out the hard way in 1935, but made no later attempt at a broad synthesis. Kempthorne (1952) in Ames, building on the work of Neyman and co-workers, Fairfield Smith in Raleigh, Tukey in Princeton, Cornfield at the National Institutes of Health in Bethesda and no doubt others elsewhere all sought to devise models of differing breadth and flexibility which would specialize appropriately under different assumptions, and lead to the desired analyses and inferences. Throughout all this, Fisher was silent on the topic, apparently holding to his view that "the analysis of variance is ... a convenient method of arranging the arithmetic."

Anyone who reads the five sections comprising the Initial discussion of [8] quickly realizes that providing a general framework for ANOVA is no mean task. The subsequent six sections spelling out Cornfield and Tukey's approach prior to their presenting any average values shows that theirs is not an easy resolution. So it should come as no surprise when I say that the situation today is hardly any better than it was then in the mid-1950s. Cornfield and Tukey's paper should be essential reading for all those who care about these matters. But it is not read, and neither their approach nor any other has taken root among the legions of users of ANOVA and linear models. No treatment of the issues that prompted them to write that paper has yet gained acceptance; see below.

What are the issues? Although in most of his writings on ANOVA Tukey emphasized estimation of variance components above significance testing, this paper is very much motivated by testing. Expressions for average values of mean squares in factorials are the primary basis for testing: they tell us which mean squares can usefully be compared with which; that is, they dictate the choice of error term. So attention focuses sharply on the model assumptions leading to these averages. As Tukey and Cornfield point out in Section 2 of their paper, the choice among assumptions is important and is not simple. It includes but goes beyond empirical questions about the behavior of the experimental material. Assumptions must also depend on the nature of the sampling and randomization involved in

obtaining the data, and the purpose of the analysis, as expressed by the situations or populations to which one wishes to make statistical inference.

Cornfield and Tukey's way ahead is by the use of what they call a pigeonhole model, in which combinations of experimental factors (rows, columns, etc.) define pigeonholes containing a finite or infinite population. If, like them, we illustrate ideas with the replicated row-by-column classification, then their assumption is that a sample of r rows is drawn from a possible R, and a sample of c columns is drawn from a possible C. These rc intersections define the pigeonholes which are the cells of the actual experiment, and from each of the rc cells a sample of n elements is drawn. "All the samplings—of rows, of columns, and within pigeonholes—are at random and independent of one another." This is their approach. They discuss at considerable length the way in which an equivalent linear model can be defined, making it clear just how different their linear model was from those previously used (and used today). Of particular significance was their notion of "tied" interaction, their avoidance of what they term the "special and dangerous" assumption of independence of the variation of interaction terms of main effects terms.

After their lengthy preliminaries it is almost a relief to get to the algebraic part of the paper: definitions of components of variance and rules for calculating what we now term expected mean squares. They discuss two-way and three-way designs in detail and give rules for designs with factors nested or crossed in arbitrary ways. There is an interesting discussion of the nature of the various proofs then extant of the formulae. At that time there were two types: "Proofs using special machinery or indirect methods (e.g., symmetry arguments and equating of coefficients for special assumptions)," the approach preferred by Tukey, and "proofs using relatively straightforward algebra," which was the preferred way of Cornfield. Neither of these was particularly effective in full generality.

The mathematical content of [8] has been revisited at least twice since 1956. The first time was by Haberman (1975), in a dense paper which does not seem to have been widely read. He makes effective use of the calculus of tensor products of vector spaces to give very concise proofs of the main results. A quite different approach was used in Speed (1985) [see also Speed and Bailey (1987)] (also not widely read), where the discussion was expressed in terms of the eigenvalues of the associated dispersion matrices. Other, less general formulations can be found in books on linear models and ANOVA, for example, Searle, Casella and McCulloch (1992).

As suggested earlier, all attempts at providing a general framework for ANOVA since 1956 should have come to terms with the material in [8]: they should either incorporate it or suggest an alternative approach. There have been many such attempts over the last 45 years, with Nelder (1977) providing the most far-reaching alternative, building on Nelder (1965a, b). This paper and especially the discussion of it are well worth reading, especially today. The most recent discussions of the "mixed models controversy" [see, e.g., Schwarz (1993) and Voss (1999) and

references therein] refer to neither Cornfield and Tukey, Kempthorne, Nelder nor any other of the earlier generation of researchers in this area. Plackett (1960) gives an excellent review of this early work.

Tukey's contribution to the discussion of Nelder (1977) is particularly interesting, in part because it reveals so clearly his distrust in models. It should be read in full, but here are some tantalizing excerpts, all the more relevant when one bears in mind that all recent discussion of this issue is a discussion of models:

> I join with the speaker in hoping for an eventual and agreed-upon description. I hope the present paper will help us approach this ideal state, but I must say that it has not brought us there.
>
> Three types of variability arise in almost any question about a set of comparative measurements, experimental or not: measurement variability, sampling variability and contextual variability.
>
> A major point, on which I cannot yet hope for universal agreement, is that our focus must be on questions, not models.
>
> One conclusion I draw from such examples is this: Models can—and will—get us into deep trouble if we expect them to tell us what the unique proper questions are.

I close this section with some personal comments, but before I do so, I should confess that I too have attempted to publish a description of ANOVA which I had hoped might have become "agreed-upon." It did not even get accepted for publication. However, I think I represent more than myself when I say that, for all my admiration of [8] and what it attempted to do, that solution was simply too far away from the world of linear models most of us inhabit. In my view, and I suspect that of many others, linear models are most readily specified through a model for the expected values and a model for the variances and covariances of the observables. After all, we are simply specifying (apart from the values of certain unknown parameters) the first two moments of our observables. Had their approach been in these terms, I believe it might still be discussed. Nelder (1977) had a related objection when he pointed out that randomization models (involving finite populations but random effects) could not be seen as a special case of the approach in [8]. The matter of providing linear unbiased estimates of quantities of interest figured nowhere in [8], and I believe this reduces many people's willingness to see its solution as general and relevant to their use of linear models and ANOVA. But perhaps the real reason that the description in [8] is not yet agreed upon is this: the majority of statisticians these days (perhaps even 50 years ago) are not interested in the issues that concerned Tukey, Cornfield, Kempthorne, Fairfield Smith and Neyman and co-workers, before them, and Nelder and others, including me, after them. Perhaps it is just too hard, connecting assumptions and models to the subject matter, to the data collection process, to the questions one is asking and the kinds of answers one seeks. "Does it really matter? Does it make any practical difference?" I get asked. It is so much easier discussing models and parameterizations.

5. Other ANOVA papers by Tukey.

5.1. *Dyadic ANOVA* [1]. This paper was based on a talk Tukey gave in November 1946, and is more interesting for what it tells us about the development of his thinking concerning ANOVA than for the material related to its title. Ostensibly about ANOVA for vectors, that is, what we would now call multivariate analysis of variance (MANOVA), the paper also contains a wealth of interesting material only marginally related to that topic. The reason he wrote it, he says, was that other accounts of MANOVA concentrate too much on tests and too little on that which is most useful and revealing in ordinary ANOVA. It is impossible to resist passing on one of his introductory remarks, presumably aimed at the average reader of *Human Biology*. He writes:

> It is a maxim of arithmetic that it is not proper to add 2 oranges to 1 apple; this is good arithmetic but may be poor vector algebra. For
>
> $$(2\,\text{oranges}, 0) + (0, 1\,\text{apple}) = (2\,\text{oranges}, 1\,\text{apple})$$
>
> is a meaningful and useful statement.

Later, he goes on:

> If we are to have an analysis of variance, we must have squares, and the solution is
>
> $$(2\,\text{oranges}, 1\,\text{apple})^2 = \begin{bmatrix} 4\,\text{orange}^2 & 2\,(\text{orange})(\text{apple}) \\ 2\,(\text{orange})(\text{apple}) & 1\,\text{apple}^2 \end{bmatrix}$$

The paper includes a concise discussion of components of variance, initially in the context of Eisenhart's (1947) models, but also including the finite population pigeonhole models which were to play such a big role in his later work. Rather surprisingly in view of his later disdain for F-tests, and his stated motivation for writing the paper, he makes a start on tests of significance for dyadic ANOVA, that is, the distribution of eigenvalues in 2×2 MANOVA. He even attempts to give fiducial intervals for quantities of interest, but concludes that more distribution theory is required.

A topic not obviously related to dyadic ANOVA is what he calls choice of terms, that is, choice of the response variable to be analyzed in a given experiment. He castigates Fisher for not paying more attention to this point, illustrating it dramatically by carrying out the same analysis on some hydrogen spectrum data using both wavelength and its reciprocal, wave number, as responses. In a fascinating analysis foreshadowing the power transformation underpinnings of ODOFFNA, he uses his newly developed dyadic ANOVA to find that linear combination of a response variate and the variate squared which minimizes the ratio of row plus column sums of squares to interaction sum of squares in an unreplicated row-by-column array. Illustrating the method on one of the data sets which he uses in his later paper on ODOFFNA, Tukey shows the considerable

gain in efficiency he achieves with his transformation. The eigenvalue problem he solves is reminiscent of canonical variate analysis, and he ends that discussion with some interesting speculations on alternative criteria to optimize in the definition of discriminant functions.

A further point of interest in this paper can be found in the Appendix, headed "Two identities and a lemma." The lemma gives the variance of the average and the expectation of the sample variance of a set of variates which have different means and different variances, but a common covariance λ, a simple enough variant on the result which is well known for i.i.d. variates. He goes on to apply this result to his pigeonhole models, illustrating once more what was to be a recurrent theme in his statistical research: a desire to weaken standard assumptions wherever possible. He finds that, under these more general assumptions, the formulae are essentially unchanged, with a common variance σ^2 being replaced by the average variance $\sigma_\cdot^2 - \lambda$.

5.2. *Components in regression* [4]. This paper is about simple linear regression when both variates are subject to "error," and the use of instrumental variates in this context. The fields of application discussed include precision of measurement, psychology and econometrics, and, as is so often the case with Tukey, the paper demonstrates the prodigious breadth of his knowledge. The connection with ANOVA is slight, really only arising because he discusses an example in which measurements are taken in replicate. As he says, "We could have avoided mention of variance components ... since we only deal with the simplest sorts ... between-vs-within or regression-vs-balance. However, we have chosen to bring them in for two reasons. Mainly to set the analysis in terms which can easily be carried over to more complicated analyses where the correct procedure might otherwise be a mystery. Secondarily, to stress the analogy with variance components for a single variate." The paper is not easy reading and, since its connection to other material here is not great, we do not discuss it any further.

5.3. *ANOVA and spectral analysis* [12]. As might be expected from its context—the discussion of two papers on the spectral analysis of time series—[12] is much more about spectral analysis than ANOVA. It was placed in one of the time series volumes [21], not in [17], yet I want to mention it here, in part for its influence on me personally. What Tukey makes very clear in this discussion is that spectrum analysis, with a *line* for each frequency, *is* ANOVA. More fully, he says "the spectrum analysis of a single time series is just a branch of variance component analysis." This was one of his inspired connections which proved illuminating in both directions. It is clear from his remarks that Tukey supposed that his statistical audience knew something about ANOVA and could read [8] if they wished, and that this would enlarge their understanding of spectrum analysis, the topic of the papers. What was probably not apparent at the time was that there were people, myself included, for whom spectrum analysis was straightforward,

but variance component analysis a mystery, and that his connection would be helpful to such people in the other direction. For evidence of the impact of this paper on me, see Speed (1987); for a valuable introduction to this paper, see the comments by Brillinger in [21].

5.4. *Toward robust ANOVA* [16]. This paper offers "a recipe for robust/resistant analysis of variance of data from factorial experiments in which all factors have three or more versions." Its motivation is eloquently explained as follows:

> Analysis of variance continues to be one of the most widely used statistical methods. Not only the form of the analysis of variance table with its lines of mean squares and degrees-of-freedom associated with each of several sorts of variation, but the entire analysis, including confidence statements, is classically supposed to be determined by the design—the hierarchical structure, conduct, and the intent of the experiment—alone. The behaviour of the data itself is, classically, not supposed to influence how its description is formatted. Hardly an exploratory attitude.... In this account, rather than using a data-free structure to define our procedure, we provide a further stage of responding to the data's behaviour, one where summarization is based on a robust alternative to the mean.

The recipe is explained by its application to a particular $5 \times 3 \times 8$ array of data from an experiment concerning the hardness of gold alloy fillings. It begins with a *pre-decomposition*, this being a multiway analogue of median polish, and proceeds through the *identification* of so-called exotic entries, to a *re-decomposition* dealing with these, and a *robust analysis of variance* with the familiar sums of squares and degrees-of-freedom calculated from the re-decomposition. Next, a process of *downsweeping* is carried out, this being a variant of the pooling of mean squares which we met in Section 3 above, and the recipe concludes with the calculation of error mean squares, standard errors and confidence statements.

5.5. *Methods, comments, challenges* [18–20]. Tukey expounded and discussed ANOVA in a number of his many overview papers, and I will single out three of these for brief mention.

In [18] he goes over "some methods that form sort of a general core of the statistical techniques" that were used at that time. He aimed "to supply background: statistical, algebraic and perhaps intuitive," and he succeeded admirably. The exposition could hardly be improved upon, indeed is better than most we see today, in that it contains possibly the first instance of the "analysis of variance diagram" mentioned in the discussion of paper [11] in Section 3 above. This diagram surely deserves to be more widely used. Also noteworthy is a remark which may well be the first appearance in print of the abbreviation ANOVA.

In [19] Tukey offers 37 methodological comments about statistics on topics ranging from *exploration versus confirmation*, *re-expression* and *causation*, to *spectrum analysis*, and naturally he has something so say about ANOVA. Relevant comments concern regression and analysis of variance, nonorthogonal analysis and

MANOVA, and can only be described as stimulating and provocative. For example, in seeking a replacement of conventional MANOVA: "We could calculate principal components, but they are not likely to be simply interpretable. So let us not"; and: "Much the same could be said of 'dust bowl empiricists factor analysis'."

In Section 21 of the last of these three overview papers [20], Tukey foreshadows the issues dealt with more fully in [16] discussed above. We see clearly how keen Tukey was to unify his understanding of and approach to ANOVA with his robust/resistant and exploratory data analysis paradigms. While [16] is a fine start, it seems clear that there is much more to be said on this unification.

6. Concluding remarks. John Tukey was an extraordinarily able and creative statistician. He made a number of lasting contributions to ANOVA: to our understanding of what it is and what it can do for us; to the algebraic and computational aspects of the subject; and, perhaps most important and characteristic to showing us how to go beyond the usual assumptions. The impact of all this work on the subject today is less than it should be, perhaps in part because Tukey set his standards rather high. However, his papers are all there for anyone to read, and if this appreciation of them encourages one person who would not otherwise, to do so, its purpose will have been achieved.

Acknowledgments. I thank Gordon Smyth, for reading this manuscript in draft and offering many helpful comments, and David Brillinger, for many valuable references and for his enthusiasm for this project. Thanks also to the referees and editors for their comments and assistance.

JOHN W. TUKEY'S PUBLICATIONS ON ANOVA

[1] (1949a). Dyadic ANOVA. An analysis of variance for vectors. *Human Biology* **21** 65–110. (Paper [21].)

[2] (1949b). One degree of freedom for non-additivity. *Biometrics* **5** 232–242. (Paper [29].)

[3] (1950). Some sampling simplified. *J. Amer. Statist. Assoc.* **45** 501–519. (Paper [38].)

[4] (1951). Components in regression. *Biometrics* **7** 33–69. (Paper [41].)

[5] (1955). Answer to query 113. *Biometrics* **11** 111–113. (Paper [51].)

[6] (1956a). Keeping moment-like sampling computations simple. *Ann. Math. Statist.* **27** 37–54. (Paper [52].)

[7] (1956b). Variances of variance components. I. Balanced designs. *Ann. Math. Statist.* **27** 722–736. (Paper [53].)

[8] (1956c). Average values of mean squares in factorials. *Ann. Math. Statist.* **27** 907–949. (With J. Cornfield. Paper [55].)

[9] (1957a). Variances of variance components. II. The unbalanced single classification. *Ann. Math. Statist.* **28** 43–56. (Paper [57].)

[10] (1957b). Variances of variance components. III. Third moments in a balanced single classification. *Ann. Math. Statist.* **28** 378–384. (Paper [58].)

[11] (1960). Complex analyses of variance: General problems. *Psychometrika* **25** 127–152. (With B. F. Green, Jr. Paper [76].)

[12] (1961). Discussion, emphasizing the connection between analysis of variance and spectrum analysis. *Technometrics* **3** 191–219. (Paper [79].)

[13] (1962). Handout for Meeting of Experimental Station Statisticians "Tests for Non-additivity." ([17], Paper 3).

[14] (1977). *Exploratory Data Analysis*. Addison-Wesley, Reading, MA. (Book [307].)

[15] (1987). Graphical exploratory analysis of variance illustrated on a splitting of the Johnson and Tsao data. In *Design, Data and Analysis by Some Friends of Cuthbert Daniel* (C. Mallows, ed.) 171–244. Wiley, New York. (With E. G. Johnson. Paper [282].)

[16] (2001). Towards robust analysis of variance. In *Data Analysis from Statistical Foundations* (A. K. Md. E. Saleh, ed.) 217–244. Nova Science Publishers, Huntington, NY. (With A. H. Seheult. Paper [213].)

[17] (1992). *The Collected Works of John W. Tukey VII. Factorial and ANOVA: 1949–1962*. Wadsworth, Belmont, CA.

The following are not entirely about ANOVA, but contain relevant material.

[18] (1951). Standard methods of analyzing data. In *Proc. Computation Seminar, December 1949* 95–112. IBM, Armonk, NY. (Paper [216].)

[19] (1980). Methodological comments focused on opportunities. In *Multivariate Techniques in Human Communication Research* (P. R. Monge and J. Cappella, eds.) 489–528. Academic Press, New York. (Paper [194].)

[20] (1982). An overview of techniques of data analysis, emphasizing its exploratory aspects. In *Some Recent Advances in Statistics* (J. Tiago de Oliveira and B. Epstein, eds.) 111–172. Academic Press, New York. (With C. L. Mallows. Paper [273].)

[21] (1984). *The Collected Works of John W. Tukey I. Time Series: 1949–1964*. Wadsworth, Belmont, CA.

REFERENCES

BOX, G. E. P. and COX, D. R. (1964). An analysis of transformations (with discussion). *J. Roy. Statist. Soc. Ser. B* **26** 211–252.

DRESSEL, P. L. (1940). Statistical semi-invariants and their estimates with particular emphasis on their relation to algebraic invariants. *Ann. Math. Statist.* **11** 33–57.

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

FISHER, R. A. (1929). Moments and product moments of sampling distributions. *Proc. London Math. Soc.* **30** 199–238.

HABERMAN, S. J. (1975). Direct products and linear models for complete factorial tables. *Ann. Statist.* **3** 314–333.

HARVILLE, D. A. (1969). Variances of variance-component estimators for the unbalanced 2-way cross classification with application to balanced incomplete block designs. *Ann. Math. Statist.* **40** 408–416.

KAPLAN, E. L. (1952). Tensor notation and the sampling cumulants of k-statistics. *Biometrika* **39** 319–323.

KEMPTHORNE, O. (1952). *Design and Analysis of Experiments*. Wiley, New York.

KENDALL, M. G. (1943). *The Advanced Theory of Statistics* **1**. Griffin, London.

MCCULLAGH, P. (1987). *Tensor Methods in Statistics*. Chapman and Hall, London.

MCCULLAGH, P. and PREGIBON, D. (1987). k-statistics and dispersion effects in regression. *Ann. Statist.* **15** 202–219.

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

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

NELDER, J. A. (1977). A reformulation of linear models (with discussion). *J. Roy. Statist. Soc. Ser. A* **140** 48–76.

NEYMAN, J. S. (1925). Contributions to the theory of small samples drawn from a finite population. *Biometrika* **17** 472–479. [Reprinted from *La Revue Mensuelle de Statistique* **6** (1923) 1–29.]

PAULL, A. E. (1950). On a preliminary test for pooling mean squares in the analysis of variance. *Ann. Math. Statist.* **21** 539–556.

PLACKETT, R. L. (1960). Models in the analysis of variance (with discussion). *J. Roy. Statist. Soc. Ser. B* **22** 195–217.

SCHEFFÉ, H. (1959). *The Analysis of Variance*. Wiley, New York.

SCHWARZ, C. J. (1993). The mixed-model ANOVA: The truth, the computer packages, the books. Part I. Balanced data. *Amer. Statist.* **47** 48–59.

SEARLE, S. R., CASELLA, G. and MCCULLOCH, C. E. (1992). *Variance Components*. Wiley, New York.

SEBER, G. A. F. (1977). *Linear Regression Analysis*. Wiley, New York.

SPEED, T. P. (1983). Cumulants and partition lattices. *Austral. J. Statist.* **25** 378–388.

SPEED, T. P. (1985). Dispersion models for factorial experiments. *Bull. Inst. Internat. Statist.* **51** 1–16.

SPEED, T. P. (1986a). Cumulants and partition lattices. II. Generalized k-statistics. *J. Austral. Math. Soc. Ser. A* **40** 34–53.

SPEED, T. P. (1986b). Cumulants and partition lattices. III. Multiply-indexed arrays. *J. Austral. Math. Soc. Ser. A* **40** 161–182.

SPEED, T. P. (1986c). Cumulants and partition lattices. IV. A.s. convergence of generalized k-statistics. *J. Austral. Math. Soc. Ser. A* **41** 79–94.

SPEED, T. P. (1987). What is an analysis of variance? (with discussion). *Ann. Statist.* **15** 885–941.

SPEED, T. P. and BAILEY, R. A. (1987). Factorial dispersion models. *Internat. Statist. Rev.* **55** 261–277.

SPEED, T. P. and SILCOCK, H. L. (1988a). Cumulants and partition lattices. V. Calculating generalized k-statistics. *J. Austral. Math. Soc. Ser. A* **44** 171–196.

SPEED, T. P. and SILCOCK, H. L. (1988b). Cumulants and partition lattices. VI. Variances and covariances of mean squares. *J. Austral. Math. Soc. Ser. A* **44** 362–388.

TRACY, D. S. and GUPTA, B. C. (1973). Multiple products of polykays using ordered partitions. *Ann. Statist.* **1** 913–923.

VOSS, D. T. (1999). Resolving the mixed models controversy. *Amer. Statist.* **53** 352–356.

WILK, M. B. and KEMPTHORNE, O. (1955). Fixed, mixed, and random models. *J. Amer. Statist. Assoc.* **50** 1144–1167.

WISHART, J. (1952a). Moment coefficients of the k-statistics in samples from a finite population. *Biometrika* **39** 1–13.

WISHART, J. (1952b). The combinatorial development of the cumulants of k-statistics. *Trabajos Estadistica* **3** 13–26.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA, BERKELEY
367 EVANS HALL
BERKELEY, CALIFORNIA 94720-3860
E-MAIL: terry@bilbo.berkeley.edu