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.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics*. www.jstor.org

(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 $i = 1, \dots, m$ and $j = 1, \dots, n$ 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 $y_{i.} = n^{-1}\Sigma_j y_{ij}$, $y_{..} = m^{-1}\Sigma_i y_{i.}$, 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 $E\{SS_0\} = mn\mu_.^2 + \sigma^2$, $E\{SS_1\} = (m-1)\sigma^2 + n\Sigma_i(\mu_i - \mu_.)^2$ 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 $y_{..}$, $y_{h.} - y_{..}$ and $y_{ij} - y_{i.}$ 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 $y_{h.} - y_{..}$ is $-1/mn$ if $k \neq h$ and $1/n - 1/mn$ if $k = h$, whilst that of y_{kl} in $y_{ij} - y_{i.}$ 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 $h = i$,

$$\operatorname{cov}(y_{h.} - y_{..}, y_{ij} - y_{i.})$$
$$= \sigma^2\left[-\frac{1}{mn}0(m-1)n + \left(\frac{1}{n} - \frac{1}{mn}\right)\left(-\frac{1}{n}\right)(n-1) + \left(\frac{1}{n} - \frac{1}{mn}\right)\left(1 - \frac{1}{n}\right)1\right],$$

which is zero as stated; the case $h \neq i$ is dealt with similarly. Similar calculations prove that $\operatorname{cov}(y_{..}, y_{h.} - y_{..}) = \operatorname{cov}(y_{..}, y_{ij} - y_{..}) = 0$ and, further, that $E\{y_{..}^2\} = \mu_{..}^2 + (1/mn)\sigma^2$, $E\{(y_{h.} - y_{..})^2\} = ((m-1)/mn)\sigma^2 + (\mu_h - \mu_.)^2$ and $E\{(y_{ij} - y_{i.})^2\} = (m(n-1)/mn)\sigma^2$.

It has just been proved that the three components in the sum

$$y_{ij} = y_{..} \quad + \quad y_{i.} - y_{..} \quad + \quad y_{ij} - y_{i.} \tag{2.2}$$

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 $y_{..}$, $y_{i.} - y_{..}$ and $y_{ij} - y_{i.}$ 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

$$(y_{ij}) = (y_{..}) + (y_{i.} - y_{..}) + (y_{ij} - y_{i.}), \tag{2.4}$$

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 $y = (y_t\colon t = 0, 1, \dots, n-1)$ 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

$$\xi_\alpha = \gamma_0 + 2\sum_1^m \gamma_a \cos\left(\frac{2\pi}{n}a\alpha\right), \qquad \alpha = 0, \dots, m, \tag{2.10a}$$

with inverses

$$\gamma_a = \frac{1}{n}\xi_0 + \frac{2}{n}\sum_1^m \xi_\alpha \cos\left(\frac{2\pi}{n}a\alpha\right), \qquad a = 0, \dots, m. \tag{2.10b}$$

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 $y = (y_t\colon t \in T)$ 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 $\Sigma_a p_{\alpha a} q_{a\beta} = n\delta^\alpha_\beta$ and $\Sigma_\alpha q_{a\alpha} p_{\alpha b} = n\delta^a_b$, 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 $\mathrm{cov}(S_\alpha y_t, S_\beta y_u) = (S_\alpha \Gamma S_\beta)(t, u) = \xi_\alpha S_\alpha(t, u)\delta^\alpha_\beta = 0$ if $\alpha \neq \beta$ and so the different terms on the R.H.S. of (3.4) are uncorrelated. Further $\mathrm{var}(S_\alpha y_t) = \xi_\alpha S_\alpha(t, t)$. Next suppose that $\mathrm{var}(y_t) = \sigma^2$ 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 $d_\alpha = \mathrm{rank}(S_\alpha) = \mathrm{trace}(S_\alpha)$, and we can sum the variances in (3.4) obtaining

$$\sigma^2 = \sum_\alpha \phi_\alpha, \tag{3.5}$$

where $\phi_\alpha = n^{-1} d_\alpha \xi_\alpha = \mathrm{var}(S_\alpha y_t)$, 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 $y = (y_t\colon t \in T)$ 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 $\{A_a\colon a \in X\}$ 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, $\{\gamma_a\colon a \in X\}$ 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 $\{S_\alpha\colon \alpha \in Z\}$, 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:

$$S_\alpha = \frac{1}{n}\sum_a q_{a\alpha} A_a, \tag{4.2a}$$

$$A_a = \sum_\alpha p_{\alpha a} S_\alpha, \tag{4.2b}$$

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

(4.3a)
$$\xi_\alpha = \sum_a p_{\alpha a} \gamma_a$$

whilst the entries γ_a of Γ in (4.1) are recoverable from the eigenvalues via

(4.3b)
$$\gamma_a = (1/n) \sum_\alpha q_{a\alpha} \xi_\alpha.$$

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

(4.4)
$$E\{(S_\alpha y_t)(S_\beta y_u)\} = n^{-1} \xi_\alpha q_{a(t,u)\alpha} \delta_\beta^\alpha,$$

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

(4.5)
$$\gamma_a = \sum_\alpha (d_\alpha^{-1} q_{a\alpha}) \phi_\alpha,$$

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 $k_r = 1 = d_0$, $k_a = (n_{a+1} - 1)n_{a+2} \cdots n_r$, $0 \le a < r$,

$d_\alpha = n_1 \cdots (n_\alpha - 1)$, $0 < \alpha \leq r$, 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:

$$y_{t_1 t_2 \cdots t_{r-1} t_r} = y_{\cdot\cdot \cdots \cdot\cdot} + (y_{t_1 \cdot \cdots \cdot\cdot} - y_{\cdot\cdot \cdots \cdot\cdot}) + \cdots + (y_{t_1 t_2 \cdots t_{r-1} t_r} - y_{t_1 t_2 \cdots t_{r-1} \cdot})$$

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 $y = (y_{ij}\colon\ i = 1, \dots, m;\ j = 1, \dots, n)$ 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 $1, \dots, m$ and $1, \dots, n$, 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 $n' \geq n$,

$$\| y_{\cdot\cdot}^{(m,n)} - y_{\cdot\cdot}^{(m',n')} \|^2 = \phi_2 \left[\frac{1}{mn} - \frac{1}{m'n'} \right] + \phi_1 \left[\frac{1}{m} - \frac{1}{m'} \right], \tag{5.10a}$$

$$\begin{aligned} &\| (y_{i\cdot}^{(m,n)} - y_{\cdot\cdot}^{(m,n)}) - (y_{i\cdot}^{(m',n')} - y_{\cdot\cdot}^{(m',n')}) \|^2 \\ &\quad = \phi_2 \left[\left(\frac{1}{n} - \frac{1}{n'} \right) \left(1 - \frac{1}{m} \right) + \frac{1}{n'} \left(\frac{1}{m} - \frac{1}{m'} \right) \right] + \phi_1 \left[\frac{1}{m} - \frac{1}{m'} \right], \end{aligned} \tag{5.10b}$$

$$\| (y_{ij}^{(m,n)} - y_{i\cdot}^{(m,n)}) - (y_{ij}^{(m',n')} - y_{i\cdot}^{(m',n')}) \|^2 = \phi_2 \left[\frac{1}{n} - \frac{1}{n'} \right], \tag{5.10c}$$

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

$$\Gamma = \sum_{a=0}^{a=\infty} \gamma_a A_a. \tag{5.11}$$

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

$$\phi_\alpha^{(r)} - \phi_\alpha^{(r+1)} = (1 - n_\alpha^{-1}) n_{\alpha+1}^{-1} \cdots n_r^{-1} (1 - n_{r+1}^{-1}) (\gamma_e - \gamma_r), \tag{5.12}$$

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

$$\begin{aligned} S_\alpha &= \sum_{a \ge \alpha} A_a - (n_\alpha - 1)^{-1} A_{\alpha - 1} \\ &= \frac{n_\alpha}{n_\alpha - 1} I_{n_1} \otimes \cdots \otimes I_{n_{\alpha-1}} \otimes \left(I_{n_\alpha} - \frac{1}{n_\alpha} J_{n_\alpha} \right) \otimes J_{n_{\alpha+1}} \otimes J_{n_{\alpha+2}} \otimes \cdots . \end{aligned}$$

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

$$S_\alpha^{(r)} y_t = y_{t_1 \cdots t_\alpha \bullet \cdots \bullet t_{r+1} t_{r+2} \cdots} - y_{t_1 \cdots t_{\alpha-1} \bullet \cdots \bullet t_{r+1} t_{r+2} \cdots}$$

as $r \to \infty$. Indeed $\|S_\alpha^{(r)} y_t - S_\alpha^{(r')} y_t\|^2 = \phi_\alpha^{(r)} - \phi_\alpha^{(r')}$ for $1 \le r \le r'$, and by (5.12) this converges to zero as $r, r' \to \infty$ (assuming $n_r \ge 2$ 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 $f_1, f_2, \dots$, 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 $\{A_a\colon a \in L(F)\}$ 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 $\Gamma = \sum_a \gamma_a A_a$, the sum being over $L(F)$, with the parameters $\{\gamma_a\colon a \in L(F)\}$ 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 $\{k_a\}, \{d_\alpha\}$ and the functions $\{s_\alpha(a)\}$ directly. To do this we introduce a second representation of $\mathbf{V}(F,T)$ involving relationship matrices $\{R_b\colon b \in L(F)\}$, 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 $R_b = \sum_{a \supseteq b} A_a$ and the representation we refer to is

$$\Gamma = \sum_b f_b R_b, \tag{6.2}$$

where the parameters $\{f_b\colon b \in L(F)\}$ 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 $\sum \zeta(a,b)\mu(b,c) = \sum \mu(a,b)\zeta(b,c) = \delta(a,c) = 1$ 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:

$$(6.4) \qquad \mu(a,b) = \begin{cases} (-1)^{|b \setminus a|}, & \text{if } b \supset a \text{ and } b \setminus a \subseteq b_m, \\ 0, & \text{otherwise,} \end{cases}$$

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

$$(6.5) \qquad S_\alpha = \sum_a \mu(a, \alpha)\bar{n}_a^{-1}R_a, \qquad \alpha \in L(F)$$

defines a set of pairwise orthogonal symmetric idempotent matrices summing to the identity matrix I over T. Further the formula

$$(6.6) \qquad \xi_\alpha = \sum_b \zeta(\alpha, b)\bar{n}_b f_b$$

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 $(p_{\alpha a})$ and $(q_{a\alpha})$. The following formulas give expressions for the key quantities:

$$(6.7) \qquad d_\alpha = \prod_{f \in \alpha \setminus \alpha_m} n_f \times \prod_{f \in \alpha_m} (n_f - 1),$$

where α_m denotes the set of minimal elements of α,

$$(6.8) \qquad k_a = \prod_{f \in \bar{a} \setminus \bar{a}^m} n_f \prod_{f \in \bar{a}^m} (n_f - 1),$$

where $\bar{a}^m$ denotes the set of maximal elements of $\bar{a} = F \setminus a$, and the common value $s_\alpha(a)$ of $d_\alpha^{-1}q_{a\alpha} = k_a^{-1}p_{\alpha a}$ is

$$(6.9) \qquad s_\alpha(a) = \begin{cases} \displaystyle\prod_{f \in \alpha_m \setminus a} \{-1/(n_f - 1)\}, & \text{if } \alpha \setminus \alpha_m \subseteq a, \\ 0, & \text{otherwise,} \end{cases}$$

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 $U_g y_t = y_{t^g}$, $t \in T$, $g \in G$ 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 $w = (w_t\colon t = 0, \dots, n-1)$ 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 $A_a = B_a + B_a'$, $a = 1, \dots, m$. The $(m+1) \times (m+1)$ structural matrices $P = (p_{a\alpha})$ and $Q = (q_{a\alpha})$ are best described by the equations

$$k_a^{-1} p_{a\alpha} = d_\alpha^{-1} q_{a\alpha} = \cos\left(\frac{2\pi}{n} a\alpha\right), \tag{8.1}$$

where $k_0 = d_0 = 1$, $k_a = d_\alpha = 2$, $1 \le a$, $\alpha \le m$. 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

$$T_\alpha = (1/n) \sum_1^m t_{b\alpha}(B_b - B_b'), \qquad \alpha = 1, \dots, m. \tag{8.3}$$

In these terms we have

$$B_b = S_0 + \tfrac{1}{2} \sum_1^m (p_{\alpha b} S_\alpha + l_{\alpha b} T_\alpha), \qquad b = 1, \dots, m, \tag{8.4}$$

which, incidentally, agrees with our earlier notation since

$$A_a = B_a + B_a' = 2S_0 + \sum_1^m p_{a\alpha} S_\alpha = \sum_0^m p_{a\alpha} S_\alpha, \qquad a = 1, \dots, m.$$

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 $T'_\alpha = -T_\alpha$, $T_\alpha^2 = -S_\alpha$, $\alpha = 1, \dots m$, and with all these preliminaries we can write the real form of the spectral decomposition of Γ^{xy} as

$$\Gamma^{xy} = c_0^{xy} S_0 + \sum_1^m \left(c_\alpha^{xy} S_\alpha + q_\alpha^{xy} T_\alpha \right), \tag{8.5}$$

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 $c_\alpha^{xy} = \operatorname{Re}(\xi_\alpha^{xy})$ and $q_\alpha^{xy} = -\operatorname{Im}(\xi_\alpha^{xy})$, $\alpha = 0, 1, \dots, m$, where ξ_α^{xy}, $\alpha = 0, \dots, n$, 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

$$\Gamma = A_e \otimes \Gamma_e + \sum_1^m \left[B_b \otimes \Gamma_b + B_{n-b} \otimes \Gamma_{n-b} \right], \tag{8.8}$$

which is

$$\Gamma = S_0 \otimes C_0 + \sum_1^m \left[S_\alpha \otimes C_\alpha + T_\alpha \otimes Q_\alpha \right]. \tag{8.9}$$

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:

$$\operatorname{cov}(S_\alpha x_t, S_\alpha y_t) = n^{-1} d_\alpha c_\alpha^{xy}, \qquad \operatorname{cov}(T_\alpha x_t, S_\alpha y_t) = n^{-1} d_\alpha q_\alpha^{xy}. \tag{8.11}$$

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

$$(8.14)\qquad \begin{aligned}\Gamma = &\sum_{\alpha}^{(1)} S_\alpha \otimes C_\alpha + \sum_{\alpha}^{(2)} [S_\alpha \otimes C_\alpha + T_\alpha \otimes Q_\alpha] \\ &+ \sum_{\alpha}^{(3)} [S_\alpha \otimes C_\alpha + T_\alpha \otimes Q_\alpha + U_\alpha \otimes D_\alpha + V_\alpha \otimes E_\alpha],\end{aligned}$$

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 $T = (t_{b\alpha})$ 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

$$(8.15)\qquad w_t = \sum_\alpha S_\alpha w_t,$$

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

$$(8.16)\qquad \Gamma_e = \sum_\alpha \Phi_\alpha,$$

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