THE RELATIONSHIP ALGEBRA OF AN EXPERIMENTAL DESIGN

By A. T. James

Division of Mathematical Statistics, C.S.I R.O.[1]

0. Summary. Important properties of an experimental design, including the analysis of variance appropriate to it, are revealed by analysing the structure of an algebra generated by the relationships between the experimental units of the design. As an illustration, the relationship algebra of balanced incomplete blocks is analysed in detail.

1. Introduction. An experimental design consists of a set of N experimental units, which we shall call plots, classified into subsets in various ways. They may be classified according to their position, as for example blocks in a randomized block, rows and columns in a latin square, or the classification may be based upon the treatments applied to the plots, or some other characteristics which certain plots share.

Define a *relationship*, R, between the plots as a set of ordered pairs (i, j) of them. If the ordered pair (i, j) of plots belongs to R, we say that plot i is related to plot j by the relationship R. In a randomized block design, for example, one may define that two plots in the same block bear the relationship, B, to each other, whilst two plots in different blocks do not. Likewise, a relationship T, meaning "same treatment" can be defined.

A relationship R among a set of N plots can be expressed as an $N \times N$ matrix of 0's and 1's:

$$r_{ij} = \begin{cases} 1 & \text{if } i \text{ is related to } j \text{ by the relationship } R \\ 0 & \text{otherwise.} \end{cases}$$

The *relationship matrix* (r_{ij}) will also be denoted by the letter R.

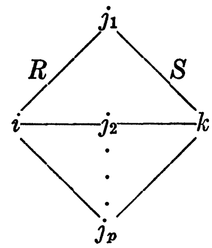
There are two relationships which appear in *any* design: (1) the *identity* relationship of each plot to itself and (2) the *universal* relationship which relates each plot to every plot in the design. The identity relationship corresponds to the matrix I and the universal relationship to a matrix G, all of whose elements are unity.

The matrix product RS of two relationship matrices R and S, gives a matrix whose elements have values 0, 1, 2, 3, $\cdots$. It can be interpreted in terms of

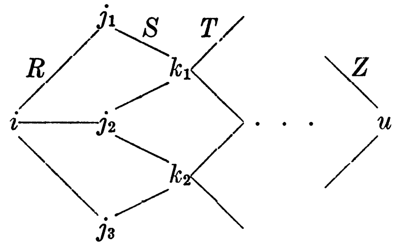
Received January 4, 1957.

[1] Commonwealth Scientific and Industrial Research Organization, University of Adelaide, South Australia.

derived relations. The (i, k)-th element of RS is the number, p, of ways one can combine a relation (i, j) of R with a relation (j, k) of S to connect i and k.



More generally, the (i, u)-th element of a product $RST \cdots Z$ is the number of paths leading from the ith plot to the uth plot via each of the relations R, S, T, $\cdots$, Z successively.



cf. Kendall [2], pp. 49, *et seq.*

Under the operations of matrix multiplication, matrix addition and scalar multiplication, the relationship matrices generate an associative algebra, which we shall call the "relationship algebra of the experimental design."

From the relationship matrices and all their products, a set of linearly independent matrices, R, S, T, $\cdots$, Z, can be chosen such that all the matrices of the algebra can be expressed as linear combinations,

$$\lambda R + \mu S + \cdots ,$$

of them. The set R, S, T, $\cdots$, Z, is called a "basis" of the algebra. Since the product of any two matrices of the algebra can be calculated from the products of the basis matrices, a multiplication table for the basis matrices summarizes the algebra.

For example, the relationship algebra of a randomized block has 4 basis matrices I, B, T, G whose elements are

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise,} \end{cases}$$

$$b_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are in the same block} \\ 0 & \text{otherwise,} \end{cases}$$

$$t_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ have the same treatment} \\ 0 & \text{otherwise,} \end{cases}$$

$$g_{ij} = 1.$$

The multiplication table is

I	B	T	G	
B	tB	G	tG	b blocks
T	G	bT	bG	t treatments
G	tG	bG	btG.	

Each basis matrix is written, once in the first row and once in the first column. The product of the matrix at the beginning of a row, with the matrix at the top of a column is written down in that row and that column.

The algebra is the direct product of two subalgebras:

$$\begin{matrix} I & B \\ B & tB \end{matrix} \times \begin{matrix} I & T \\ T & bT. \end{matrix}$$

For an $n \times n$ latin square, the multiplication table of the basis matrices is

I	R	C	T	G
R	nR	G	G	nG
C	G	nC	G	nG
T	G	G	nT	nG
G	nG	nG	nG	n^2G.

R, C, T, are the relations, same row, same column, and same treatment respectively.

Balanced incomplete blocks have an interesting algebra. The two relations, same block B, same treatment T, satisfy the equation

$$TBT = \lambda G + (r - \lambda)T, \tag{1}$$

which reflects the requirement of balance, namely, that each pair of treatments occur together in λ blocks. r is the number of replicates and k the number of plots per block.

One can verify Eq. (1) by counting the number of paths from plot i to plot j using the relations T, B and T successively. If i and j have different treatments there are exactly λ blocks containing both these treatments, through which the connection can be established. Therefore there are λ paths and the (ij)-th element of TBT is λ. If i and j have the same treatment, the number of paths is clearly r and the (ij)-th element of TBT has this value. Thus Eq. (1) holds.

With this equation one can work out the multiplication table (Table 1).

As would be expected, the algebra of relationships corresponds to the numerical operations involved in the analysis of the design. If the observations $x_1, x_2, \cdots, x_N$ taken on the N plots respectively, are written as a column vec-

TABLE 1

I	G	B	T	BT	TB	BTB
G	NG	kG	rG	krG	krG	k^2rG
B	kG	kB	BT	kBT	BTB	$kBTB$
T	rG	TB	rT	$\lambda G + (r - \lambda)T$	rTB	$\lambda kG + (r - \lambda)TB$
BT	krG	BTB	rBT	$\lambda kG + (r - \lambda)BT$	$rBTB$	$\lambda k^2G + (r - \lambda)BTB$
TB	krG	kTB	$\lambda G + (r - \lambda)T$	$\lambda kG + (r - \lambda)kT$	$\lambda kG + (r - \lambda)TB$	$\lambda k^2G + (r - \lambda)kTB$
BTB	k^2rG	$kBTB$	$\lambda kG + (r - \lambda)BT$	$\lambda k^2G + (r - \lambda)kBT$	$\lambda k^2G + (r - \lambda)BTB$	$\lambda k^3G + (r - \lambda)kBTB.$

tor x, then multiplication by the relationship matrices gives linear transformations of x; e.g., for a block design, the transformation

$$x \to Bx$$

is the operation of replacing each value x_i by the total for the block in which the ith plot occurs. Similarly, viewed in this way, the matrices T and G are operators which replace each observation x_i by the treatment total or the grand total respectively.

When the relationship matrices are thus considered as operators, their products are often obvious—e.g., for the randomized block, clearly,

$$BT = TB = G.$$

2. Structure of the relationship algebra. If j bears the same relation to i as i bears to j, as will usually be the case, the relationship matrices will be symmetric. Note, however, that their products will not necessarily be so; as can be seen in the case of balanced incomplete blocks where

$$(TB)' = B'T' = BT \neq TB.$$

The fact that the algebra can be generated by symmetric matrices has a very important implication in its mathematical analysis.

Let $\mathfrak{B}$ be the vector space of column vectors, x. A subspace $\mathfrak{B}_1$ is invariant under the relationship algebra $\mathfrak{A}$; i.e., $\mathfrak{A}\mathfrak{B}_1 \subset \mathfrak{B}_1$, if and only if it is invariant under the relationship matrices which generate $\mathfrak{A}$. But these are symmetric; hence the orthogonal complement, $\mathfrak{B}_1^{\perp}$ of $\mathfrak{B}_1$, is also invariant under them. Hence $\mathfrak{B}_1^{\perp}$ is invariant under $\mathfrak{A}$; i.e., $\mathfrak{A}$ is a *completely reducible* set of linear transformations of $\mathfrak{B}$. Therefore $\mathfrak{A}$ is a *semi-simple* algebra. According to a theorem of Wedderburn, a semi-simple algebra is isomorphic to a direct sum of complete matrix algebras[2] (see Van der Waerden [3], Chap. XVI).

Hence, *an algebra generated by symmetric relations is isomorphic to a direct sum of complete matrix algebras.*

[2] It may be necessary to extend the field of scalars from the *real* numbers to the *complex* numbers.

Example 1. *The randomized block.* As one can see by inspection of the multiplication table, the algebra is commutative. Now the algebra of all 2×2 matrices or of matrices of higher order is not commutative. Hence the algebra of the randomized block cannot be isomorphic to a direct sum of matrix algebras which contains one of these. Thus it must be isomorphic to a direct sum of 1×1 matrix algebras; i.e., the algebra of the randomized block is isomorphic to the algebra of all diagonal 4×4 matrices,

$$\begin{bmatrix} * & & & \\ & * & & \\ & & * & \\ & & & * \end{bmatrix}.$$

Example 2. *The latin square.* As the algebra is commutative and 5-dimensional, it is isomorphic to the algebra of all diagonal 5×5 matrices.

Example 3. *Balanced incomplete blocks.* The algebra is 7-dimensional and noncommutative. Since it is noncommutative, the direct sum of complete matrix algebras to which it is isomorphic must include a complete matrix algebra of order at least 2×2, but not more than 2×2, because a complete 3×3 matrix algebra is 9-dimensional and our algebra has only 7 dimensions. Hence the algebra of balanced incomplete blocks is isomorphic to the algebra of all matrices of the form

$$\begin{bmatrix} * & & & & \\ & * & & & \\ & & * & & \\ & & & * & * \\ & & & * & * \end{bmatrix}.$$

3. Analysis of the relationship algebra. A direct sum of k matrix algebras can be decomposed into its k component parts; e.g., for $k = 2$,

$$\begin{bmatrix} * & * & * & & \\ * & * & * & & \\ * & * & * & & \\ & & & * & * \\ & & & * & * \end{bmatrix} = \begin{bmatrix} * & * & * & & \\ * & * & * & & \\ * & * & * & & \\ & & & 0 & 0 \\ & & & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & & \\ 0 & 0 & 0 & & \\ 0 & 0 & 0 & & \\ & & & * & * \\ & & & * & * \end{bmatrix}.$$

Each part, e.g., the set of all matrices of the form

$$\begin{bmatrix} * & * & * & & \\ * & * & * & & \\ * & * & * & & \\ & & & 0 & 0 \\ & & & 0 & 0 \end{bmatrix},$$

is a minimum two-sided ideal. The product of two matrices belonging to different parts is clearly zero; i.e., the different parts annihilate each other. Correspondingly, the relationship algebra $\mathfrak{A}$, being isomorphic to a direct sum of k complete matrix algebras, can be expressed as a direct sum of minimum two-sided ideals which annihilate each other:

$$\mathfrak{A} = \mathfrak{A}_1 \dotplus \mathfrak{A}_2 \dotplus \cdots \dotplus \mathfrak{A}_k ; \tag{2}$$

i.e. any element of $\mathfrak{A}$ can be expressed uniquely as a sum of elements belonging respectively to $\mathfrak{A}_1, \cdots, \mathfrak{A}_k$.

In particular, the identity element can be so expressed

$$I = E_1 + E_2 + \cdots + E_k , \tag{3}$$

and the components E_i will be idempotent. Writing the corresponding quadratic forms, we have the decomposition of the sum of squares for the *analysis of variance*

$$\sum x_i^2 = x'x = x'E_1x + x'E_2x + \cdots + x'E_kx. \tag{4}$$

So far, the decomposition is unique. If one of the ideals—e.g., $\mathfrak{A}_1$—is isomorphic to an $r \times r$ matrix algebra, then $x'E_1x$ can be further decomposed into r parts, each on the same number of degrees of freedom, but the decomposition is not unique. The example of balanced incomplete blocks will illustrate this.

4. The analysis of the relationship algebra of the balanced incomplete block design. The algebra can be analysed by the standard procedures. For purposes of illustration the method is given in detail.

The problem is to decompose the algebra into its minimum two-sided ideals and to find the corresponding principal idempotents. These are the unit elements of the ideals. Since. as we have seen, our algebra is isomorphic to the direct sum of three 1×1 matrix algebras and a 2×2 matrix algebra, there must be three 1-dimensional two-sided ideals and one 4-dimensional two-sided ideal, which are respectively isomorphic to the matrix algebras. E_1, E_2, E_3, E_4 will denote the respective principal idempotents. Our first step is to pick out these ideals.

The multiples of G form a 1-dimensional two-sided ideal whose idempotent is $E_1 = \frac{1}{N} G$. The corresponding sum of squares, $x'E_1x$ is just the correction factor, (grand total)^2/N. Let us consider the algebra modulo G. We can take care of G later on by replacing all sums of squares by the corresponding sums of squares about the mean. There is now a 6-dimensional algebra to be analysed. Its multiplication table is obtained by putting $G = 0$ in the original multiplication table. We must look for some more two-sided ideals.

The linear combinations of the basis elements containing T—namely, T, BT, TB, BTB—form a two-sided ideal, because all multiples of these elements again contain T. This must be the 4-dimensional ideal that we are seeking. The principal idempotent, E_4, of this ideal corresponds to the unit element of the 2×2 matrix algebra to which the ideal is isomorphic.

One can set up such a correspondence by finding the left-*regular representation*

—namely by considering how T, BT, TB, BTB are transformed by left multiplication by each of them in turn.

$$[T \quad BT \quad TB \quad BTB] \xrightarrow{T} [T^2 \quad TBT \quad T^2B \quad TBTB]$$

$$= [T \quad BT \quad TB \quad BTB] \begin{bmatrix} r & (r-\lambda) & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & r & (r-\lambda) \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The row $[T\ BT\ TB\ BTB]$ is written formally as a row vector, even though the elements belong to an algebra instead of being numbers. This "vector" and the matrix are to be multiplied by row-column multiplication. Similarly,

$$[T \quad BT \quad TB \quad BTB] \xrightarrow{B} [T \quad BT \quad TB \quad BTB] \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & k & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & k \end{bmatrix}.$$

Each 4×4 matrix is the direct sum of two identical 2×2 matrices, as is to be expected in a regular representation. Hence we can set up an isomorphism between the ideal and the 2×2 matrices:

$$T \leftrightarrow \begin{bmatrix} \mathrm{r} & \mathrm{r} - \lambda \\ 0 & 0 \end{bmatrix}.$$

Although B does not belong to the ideal, we can calculate the matrices isomorphic to the elements BT, TB, BTB, which do belong to the ideal, by using the map

$$B \rightarrow \begin{bmatrix} 0 & 0 \\ 1 & k \end{bmatrix}.$$

Thus

$$BT \leftrightarrow \begin{bmatrix} 0 & 0 \\ r & r-\lambda \end{bmatrix},$$

$$TB \leftrightarrow \begin{bmatrix} r-\lambda & k(r-\lambda) \\ 0 & 0 \end{bmatrix},$$

$$BTB \leftrightarrow \begin{bmatrix} 0 & 0 \\ r-\lambda & k(r-\lambda) \end{bmatrix}.$$

To find the principal idempotent E_4 of the ideal, we must express the matrix corresponding to it, namely the unit matrix, in terms of these matrices:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \frac{1}{(k-1)r+\lambda}$$

$$\cdot \left\{ k \begin{bmatrix} r & r-\lambda \\ 0 & 0 \end{bmatrix} + \frac{r}{r-\lambda} \begin{bmatrix} 0 & 0 \\ r-\lambda & k(r-\lambda) \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ r & r-\lambda \end{bmatrix} - \begin{bmatrix} r-\lambda & k(r-\lambda) \\ 0 & 0 \end{bmatrix} \right\}.$$

Hence,

$$E_4 = (kT + r(r - \lambda)^{-1}BTB - BT - TB)/((k - 1)r + \lambda) \mod G.$$

Having dealt with the 1-dimensional ideal generated by G and the 4-dimensional ideal, we now have to find the idempotents of the other two 1-dimensional ideals. We can obtain an algebra *isomorphic* to the direct sum of these two-sided ideals by taking the whole algebra modulo G, T. If we put $G = 0$, $T = 0$, the multiplication table reduces to

$$\begin{matrix} I & B \\ B & kB. \end{matrix}$$

This 2-dimensional algebra splits into two 1-dimensional ideals whose idempotents are $k^{-1}B$ and $I - k^{-1}B$. But, modulo G, T, the algebra is generated by the two idempotents E_2 and E_3 ; hence we can put

$$\left.\begin{aligned} E_2 &= k^{-1}B \\ E_3 &= I - k^{-1}B \end{aligned}\right\} \mod G,\ T.$$

Dropping the modulo T, we may write

$$\left.\begin{aligned} k^{-1}B &= E_2 + F_2 \\ I - k^{-1}B &= E_3 + F_3 \end{aligned}\right\} \mod G,$$

where F_2 and F_3 are the components of $k^{-1}B$ and $I - k^{-1}B$ belonging to the 4-dimensional two-sided ideal, which was mapped on zero when we worked modulo T by putting $T = 0$. Thus, modulo G,

$$F_2 = F_2E_4 = k^{-1}BE_4 = [k(r - \lambda)]^{-1}BTB,$$

$$\begin{aligned} F_3 = F_3E_4 = (I - k^{-1}B)E_4 &= [(k - 1)r + \lambda]^{-1}(kT + k^{-1}BTB - BT - TB) \\ &= k((k - 1)r^2 + \lambda r)^{-1}(T - k^{-1}BT)(T - k^{-1}TB), \end{aligned}$$

since $E_2E_4 = 0$ and $E_3E_4 = 0$. Therefore,

$$\left.\begin{aligned} E_2 &= k^{-1}B - F_2 = k^{-1}B - [k(r - \lambda)]^{-1}BTB \\ E_3 &= (I - k^{-1}B) - F_3 = I - E_2 - E_4 \end{aligned}\right\} \mod G.$$

Now we have the principal idempotents of all the ideals.

At the same time, we obtain a further decomposition of E_4 :

$$\begin{aligned} E_4 &= (k^{-1}B + (I - k^{-1}B))E_4 \\ &= (E_2 + F_2)E_4 + (E_3 + F_3)E_4 \\ &= F_2 + F_3 . \end{aligned}$$

F_2 and F_3 are idempotents but not *principal* idempotents, because, unlike E_4, they do not correspond to the unit matrix, $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, of the 2×2 matrix algebra,

TABLE 2

Blocks (ignoring treatments)	
Treatment component	$[k(r - \lambda)]^{-1}BTB$
Remainder	$k^{-1}B - [k(r - \lambda)]^{-1}BTB$
Total	$k^{-1}B$
Treatments (eliminating blocks)	$k[(k - 1)r^2 + \lambda r]^{-1}(T - k^{-1}BT)(T - k^{-1}TB)$
Intra-block error	by difference
Total	I

but, *in the appropriate isomorphism,* they correspond to the idempotent matrices $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ and $\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ respectively. This part of the decomposition is not unique, but nevertheless it is appropriate, as may be seen by putting the usual interpretations on the quantities.

The idempotents may be arranged as in Table 2, all of them modulo G.

The quadratic forms, of which the idempotents are the matrices, are the sums of squares in the usual analysis of variance as given in Fisher and Yates [1].

5. Conclusion. Whilst it is too early to see the full implications of the relationship algebra, the following points may be noted:

1. Anyone investigating or proposing a new experimental design can throw considerable light upon it by enumerating the basic relationships set up by the design, and analysing the algebra they generate.

2. The relationship algebra leads to a simple and natural notation for the component sums of squares appearing in an analysis of variance. The sums of squares are specified by their matrices. The table above for balanced incomplete blocks, illustrates this point.

3. The analysis of variance corresponding to the analysis of the algebra into its minimum two-sided ideals, is unique. If the minimum two-sided ideals are one dimensional, no further decomposition is possible. More precisely, the relationships which have generated the algebra will not resolve the sums of squares beyond this point.

All minimum two-sided ideals are one dimensional if and only if the algebra is commutative. Designs possessing such an algebra have a unique analysis whose components are automatically orthogonal. The randomized block and latin square are of this type.

4. When the algebra contains a minimum two-sided ideal isomorphic with an $m \times m$ matrix algebra, the sums of squares corresponding to that two-sided ideal can be decomposed into m components each on the same number of degrees of freedom; but the decomposition is not unique. However, the system of possible decompositions is delimited by the fact that a transformation from one decomposition to another, induces an automorphism of the algebra. This point deserves a more detailed treatment than I can give at the moment.

5. For certain designs, the relationship algebra is the commutator algebra

of the representation of a group expressing the symmetry of the experimental design. Such will be the subject of a further paper.

6. Acknowledgment. The idea of classifying the pairs of plots according to the relationships between them was suggested to the author by Wilkinson [4], who introduced it in connection with his work on missing values.

REFERENCES

[1] FISHER AND YATES, *Statistical Tables*, 4th ed., Oliver and Boyd, London, 1954.
[2] M. G. KENDALL, "Further contributions to the theory of paired comparisons," Biometrics, Vol. 2 (1955), pp. 43–62.
[3] B. L. VAN DER WAERDEN, *Moderne Algebra*, Vol. 2, Springer, Berlin, 1940; English translation, Ungar, New York, 1950.
[4] G. N. WILKINSON, "Estimation of missing values for the analysis of incomplete data" *Biometrics*, Vol. (1956), to be published.