FISHER'S OPTIMAL SCORES AND MULTIPLE CORRESPONDENCE ANALYSIS

John C. Gower

Department of Data Theory

University of Leiden

FISHER'S OPTIMAL SCORES AND MULTIPLE CORRESPONDENCE ANALYSIS

SUMMARY

Fisher's method of optimal scores (FOS) and Guttman scaling are often cited among the origins of Multiple Correspondence Analysis (MCA). It is shown that FOS is more strongly linked to canonical correlation analysis (CCA) and that although Fisher's method, as described in the seventh edition of Statistical Methods for Research Workers, essentially leads to the MCA results, this is something of an accident. If there he had not been concerned with a balanced two-way table, the relationship between the two methods would have been weaker. Similarly, Fisher (1940) refers mainly to simple Correspondence Analysis (CA) and cannot be strongly linked to MCA. Nevertheless Fisher had all the essential elements not only of CA but also of MCA.

KEYWORDS: Multiple Correspondence Analysis, Correspondence Analysis, Fisher's method of optimal scores, Guttman scaling, Categorical variables, Response variables, Analysis of variance, Canonical Correlation.

1. INTRODUCTION

R.A.Fisher's contributions to multivariate analysis are well-known, especially his fundamental work on distribution theory and discriminant analysis, but when preparing a contribution for the meeting of the British Region of the Biometric Society, organised to celebrate the centenary of Fisher's birth on February 17th 1890, a less well-known strand in his contributions emerged. This links his work on multiplicative models (Fisher & Mackenzie, 1923) with his method of optimal scores, first described in the seventh edition of Statistical Methods for Research Workers (Fisher, 1938) and then in subsequent editions; this is denoted by SMRW in the following and Fisher's method of Optimal Scores by FOS. The 1923 paper included three major innovations - (i) the introduction of multiplicative models, (ii) the (implicit) use of dummy variables and (iii) the first extended analysis of variance (Fisher, 1918, had given a simple analysis of variance). So far as I am aware, and I have made quite extensive enquires and searches, these were all new. A method for fitting the multiplicative model by least-squares anticipated the fundamental matrix-approximation theorem of Eckart and Young (1936) and gave the equations underlying the reciprocal averaging algorithm, often used in Correspondence analysis (CA). CA, however, is concerned with the analysis of a two-way contingency table and hence with categorical rather than continuous variables. After lying dormant for nearly 50 years, the multiplicative model has achieved some popularity, especially in studies of genotype-environment interaction (see e.g. Kempton and Talbot, 1988 and papers cited therein). Later Fisher successfully used dummy variables in his development of discriminant analysis. Hirschfeld (i.e. H.O.Hartley) (1935) gave one of the first (see deLeeuw, 1983, for an account of Karl Pearson's near miss) developments of the CA of a two-way contingency table, in which he sought scores that maximised the correlation between the two categorical variables, thus presaging the scoring aspects of FOS and the correlational aspects of Hoteling's (1936) canonical correlation analysis (CCA); Fisher (1940) combined the two notions. FOS is often recognised, along with Guttmans's (1941) scaling procedure, as one of the origins of Multiple Correspondence Analysis (MCA) -see e.g. Tenenhaus and Young (1985). However it is usually Fisher (1940) that is cited rather than the SMRW work of Fisher (1936a). Starting with Bartlett (1938) there have been many expositions of how the generality of CCA subsumes many multivariate methods as special cases (see e.g. Gittins, 1985 and Gifi, 1990).

Of course, all these methods, being expressable as optimising ratios of quadratic forms, lead to similar eigenvalue problems and are formally equivalent at the algebraic and computational levels. However these equivalences only serve to obscure essential statistical differences. In the following I aim to clarify things by exploring the links between Fisher's work and MCA, which are shown to be essentially different but, almost by accident, share some common ground. Section 2 recapitulates some of the methodology of MCA and shows its analysis of data used by Fisher in his FOS exposition in SMRW. Section 3 gives the algebraic formulation of FOS for a two-way table and shows how this overlaps with the MCA results. In section 4 the relationship is explored more deeply and the results are linked in with CCA and Guttman's scaling procedure. The concluding section 5 discusses further historical material, especially relationships with the closely related area of discriminant analysis, summarises the findings and points to more recent developments.

2. MULTIPLE CORRESPONDENCE ANALYSIS OF FISHER'S DATA

Table 61.9 of SMRW is reproduced below as Table 1. It gives the responses of the blood cells of twelve people to twelve sera, where the response is coded into five (ordered) categories denoted by -,?,w,(+),+ which correspond to no reaction, a trace, weak, positive and strongly positive.

TABLE 1

NON-NUMERICAL TWO-WAY TABLE OF SERIOLOGICAL READINGS.

		Sera											
		1	2	3	4	5	6	7	8	9	10	11	12
	1	w	w	w	(+)	w	(+)	?	w	w	(+)	w	w
	2	?	w	?	w	w	w	?	w	w	w	w	?
	3	w	w	w	w	w	w	w	w	w	w	w	w
	4	w	w	w	w	w	w	-	w	w	w	w	?
	5	w	(+)	w	(+)	w	w	?	(+)	w	(+)	w	w
Cells	6	w	w	(+)	(+)	w	w	?	w	w	(+)	w	w
	7	(+)	(+)	(+)	(+)	+	+	w	(+)	w	(+)	(+)	w
	8	w	+	(+)	(+)	w	(+)	w	(+)	w	(+)	(+)	w
	9	w	(+)	(+)	(+)	w	(+)	w	(+)	w	(+)	w	w
	10	?	?	w	w	w	w	?	w	w	w	w	?
	11	w	w	(+)	w	w	w	?	w	w	w	w	w
	12	w	(+)	+	(+)	(+)	(+)	w	(+)	w	(+)	+	w

Table 1 may be regarded as giving data on three categorical variables -- persons (with 12 levels), sera (with 12 levels), and responses (with 5 levels). The whole could be presented as a 144x29 indicator-matrix **G**, the first 12 columns of which correspond to the different persons, the second twelve columns to the sera and the final five columns to response. Every row of **G** is zero except for the three columns corresponding to the actual person, serum and response occurring, each of which is scored one. Because every row corresponds to an observation for a single person, a single serum and the resulting unique response, all rows must sum to three, i.e. $\mathbf{G1} = 3\mathbf{1}$, where **1**, in the following, represents any vector of units, length being understood from context - here lengths are 29 and 144, respectively. Also $\mathbf{1'G}$ is a vector giving the number of occurrences of each level; it is convenient to gather these into a diagonal matrix **D**, and thus $\mathbf{D} = \text{diag}(\mathbf{1'G})$. In the remainder of this section, notation is established and some of the basic results of MCA are derived so that the links with the work of Fisher and others can be explored in a uniform manner. For an encyclopaedic account of CA, see Greenacre (1984) and especially his chapter 5 on MCA.

The most direct form of MCA requires the ordinary two-way correspondence analysis of **G**, i.e. the singular decomposition expressed by:

$$3^{-1/2}\,\mathbf{G}\mathbf{D}^{-1/2} = \mathbf{U}\Sigma\mathbf{V}'. \tag{1}$$

The first columns of **U** and **V** are constant (or essentially so, see below) and of little interest; it is the second column of **V** which gives the optimal scores corresponding to the 29 category levels, although in the context of ordination the others are also of interest. The result of Eckart and Young (1936) shows that setting all but the r largest values of the diagonal matrix Σ to zero, gives the best least-squares rank r approximation to the matrix on the left-hand-side of (1). Because, as here, **G** usually has many more rows than columns, it is more convenient to operate on the scalar-products matrix derived from equation (1) to give:

$$3^{-1}\mathbf{D}^{-1/2}\mathbf{G}'\mathbf{G}\mathbf{D}^{-1/2} = \mathbf{V}\Sigma^2\mathbf{V}'.$$

Thus to get the columns of **V** it is only necessary to solve the 29×29 eigenvalue problem:

$$\mathbf{D}^{-1/2}\mathbf{G}'\mathbf{G}\mathbf{D}^{-1/2}\mathbf{v} = 3\lambda\mathbf{v} \tag{2}$$

where the eigenvalue $\lambda = \sigma^2$, the square of the corresponding singular value.
Setting $\mathbf{B} = \mathbf{G}'\mathbf{G}$, equation (2) may be written as a two-sided eigenvalue problem

$$\mathbf{B}(\mathbf{D}^{-1/2}\mathbf{v}) = 3\lambda\mathbf{D}(\mathbf{D}^{-1/2}\mathbf{v}) \tag{3}$$

or $$\mathbf{B}\mathbf{w} = 3\lambda\mathbf{D}\mathbf{w}. \tag{4}$$

Because of the orthogonality of **V** in the singular value decomposition (1), (3) and (4) imply that $\mathbf{w} = \mathbf{D}^{-1/2}\mathbf{v}$ must be normalised so that $\mathbf{w}'\mathbf{D}\mathbf{w} = \mathbf{v}'\mathbf{v} = 1$.

Note that $\mathbf{w} = \mathbf{1}$ is always a solution to (4) corresponding to $\lambda = 1$, for $\mathbf{G1} = 3\mathbf{1}$ implies that $\mathbf{G}'\mathbf{G1} = 3\mathbf{G}'\mathbf{1} = 3\mathbf{D1}$. Returning to the equivalent form (2), $\lambda = 1$ corresponds to the solution $\mathbf{v} = \mathbf{D}^{1/2}\mathbf{1}$ which is a non-negative vector. Also the left-hand side of (2) is a positive definite non-negative matrix, so it follows from the fundamental Frobenius theorem on non-negative matrices that all other eigenvalues are smaller than 1, i.e. $0 \le \lambda < 1$. From (4) it follows also that because **1** is a solution, for any other solution **w** we must have that $\mathbf{1}'\mathbf{D}\mathbf{w} = 0$, which is often described as a constraint on MCA solutions but here has been found as a consequence of the formulation of MCA as the least-squares approximation to the matrix on the left-hand side of (1). Healy and Goldstein (1976) have questioned the inevitability of this "constraint". I believe the problem can be resolved, and in the following there are several pointers how, but I shall not discuss the issue further here. It suffices for the purposes of this paper to accept the "constraint", for this is standard practice in MCA. The matrix **B** is known as a Burt matrix and, in the case of Fisher's example, has the form:

$$\mathbf{B} = \begin{pmatrix} 12\mathbf{I} & \mathbf{11'} & \mathbf{R} \\ \mathbf{11'} & 12\mathbf{I} & \mathbf{C} \\ \mathbf{R'} & \mathbf{C'} & \mathbf{W} \end{pmatrix} \tag{5}$$

where **R** is a 12×5 contingency table giving the numbers of each type of response for each person and **C** is a 12×5 contingency table giving the numbers of each type of response for each serum. We also have that $\mathbf{D} = \text{diag}\mathbf{B}$, which is composed of the elements of the diagonal matrices $12\mathbf{I}, 12\mathbf{I}, \mathbf{W}$ where $\mathbf{W} = \text{diag}(1,13,89,36,5)$ gives the total number of responses of each type. The special consideration of (5) will be discussed in section 2.1. It turns out that everything of interest that needs to be said about problems with more than two categorical variables can be said for three variables, so in the remainder of this section, and elsewhere, **B** will continue to be treated as a general Burt matrix for three categorical variables.

Thus **B** is a 3×3 symmetric block matrix with elements $\mathbf{B}_{ij}$ (i,j = 1,2,3). The diagonal blocks $\mathbf{B}_{ii}$ (i = 1,2,3) are themselves diagonal matrices whose diagonal values give the number of occurrences of the different levels of the ith categorical variable. Writing $\mathbf{G}_i$ for the indicator matrix for the ith variable, then $\mathbf{G} = (\mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3)$ and $\mathbf{B}_{ij} = \mathbf{G}_i'\mathbf{G}_j$, whence:

$$\mathbf{1}'\mathbf{B}_{ij} = \mathbf{1}'\mathbf{G}_j = \mathbf{1}'\mathbf{B}_{jj} \text{ for } i,j = 1,2,3. \tag{6}$$

Partitioning **w** conformably and expanding the first row of (4) gives:

$$\mathbf{B}_{11}\mathbf{w}_1 + \mathbf{B}_{12}\mathbf{w}_2 + \mathbf{B}_{13}\mathbf{w}_3 = 3\lambda\mathbf{B}_{11}\mathbf{w}_1$$

premultiplying by **1**' and using (6) gives:

$$\mathbf{1}'\mathbf{B}_{11}\mathbf{w}_1 + \mathbf{1}'\mathbf{B}_{22}\mathbf{w}_2 + \mathbf{1}'\mathbf{B}_{33}\mathbf{w}_3 = 3\lambda\mathbf{1}'\mathbf{B}_{11}\mathbf{w}_1$$

i.e. $\mathbf{1}'\mathbf{D}\mathbf{w} = 3\lambda\mathbf{1}'\mathbf{B}_{11}\mathbf{w}_1$.

Now for $\lambda \neq 1$, $\mathbf{1}'\mathbf{D}\mathbf{w} = 0$ and hence, unless $\lambda = 0$ or 1, then $\mathbf{1}'\mathbf{B}_{11}\mathbf{w}_1 = 0$ and generally:

$$\mathbf{1'B}_{11}\mathbf{w}_1 = \mathbf{1'B}_{22}\mathbf{w}_2 = \mathbf{1'B}_{33}\mathbf{w}_3 = 0. \tag{7}$$

Thus for all interesting solutions not only are the weighted means of all scores zero but also the weighted means of the scores for each categorical variable. Again these are not arbitrary constraints but merely consequences of the Eckart-Young theorem when approximating the matrix on the left-hand-side of (1). These results extend to any number of categorical variables.

2.1 *Details of the Multiple Correspondence Analysis of a balanced two-way table*

Normally equation (1),(2) or (4) would be solved directly but because **B** here is highly structured, and for comparison with Fisher's method which is discussed in section 3, a more detailed analysis is necessary. We therefore return to the special form (5) of the Burt matrix that arises in Fisher's example, but consider the case of p rows, q columns and r responses; as well as giving greater generality, this removes a potential confusion arising from the example in SMRW where p happens to equal q. Equation (4) with **B** given in a matrix of form (5) becomes:

$$\begin{pmatrix} q\mathbf{I} & \mathbf{11'} & \mathbf{P} \\ \mathbf{11'} & p\mathbf{I} & \mathbf{Q} \\ \mathbf{P'} & \mathbf{Q'} & \mathbf{W} \end{pmatrix}\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \\ \mathbf{z} \end{pmatrix} = 3\lambda\begin{pmatrix} q\mathbf{I} & & \\ & p\mathbf{I} & \\ & & \mathbf{W} \end{pmatrix}\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \\ \mathbf{z} \end{pmatrix} \tag{8}$$

where the vector **w** has been partitioned into components **x,y,z** corresponding to rows, columns and response-scores; here it is the latter that are of primary interest, though normally in MCA all would be of equal interest. In (8) the roles of p and q may appear to be interchanged. This is because although $\mathbf{G}_1$ has p columns, it has q units in each column; similarly $\mathbf{G}_2$ has p units in each column. Expanding (8) gives:

$$\left.\begin{aligned} q\mathbf{x} + \mathbf{11'y} + \mathbf{Pz} &= 3\lambda q\mathbf{x} \\ \mathbf{11'x} + p\mathbf{y} + \mathbf{Qz} &= 3\lambda p\mathbf{y} \\ \mathbf{P'x} + \mathbf{Q'y} + \mathbf{Wz} &= 3\lambda\mathbf{W}\,\mathbf{z} \end{aligned}\right\}. \tag{9}$$

The condition (6) gives:

$$\mathbf{1'P} = \mathbf{1'Q} = \mathbf{1'W} \tag{10}$$

and $$\mathbf{P1} = q\mathbf{1}, \quad \mathbf{Q1} = p\mathbf{1}. \tag{11}$$

Also (7) gives that $\mathbf{1'x} = \mathbf{1'y} = \mathbf{1'Wz} = 0$. The equations (9) now simplify to:

$$
\left.\begin{aligned}
\mathbf{q}\mathbf{x} \quad + \mathbf{P}\mathbf{z} &= 3\lambda\mathbf{q}\mathbf{x} \\
\mathbf{p}\mathbf{y} \quad + \mathbf{Q}\mathbf{z} &= 3\lambda\mathbf{p}\mathbf{y} \\
\mathbf{P}'\mathbf{x} + \mathbf{Q}'\mathbf{y} + \mathbf{W}\mathbf{z} &= 3\lambda\mathbf{W}\,\mathbf{z}
\end{aligned}\right\} \tag{12}
$$

from which **x** and **y** may be eliminated to give:

$$(\mathbf{W}\text{-}\mathbf{P}'\mathbf{P}/\mathrm{q}\text{-}\mathbf{Q}'\mathbf{Q}/\mathrm{p})\mathbf{z} = 3\lambda(2\text{-}3\lambda)\mathbf{W}\mathbf{z} \tag{13}$$

a two-sided eigenvalue problem for the response scores. Unless $3\lambda = 1$ (in which case see 2.2) the scores for rows and columns are:

$$\mathbf{x} = \frac{1}{(3\lambda-1)}\frac{\mathbf{P}\mathbf{z}}{\mathrm{q}}, \qquad \mathbf{y} = \frac{1}{(3\lambda-1)}\frac{\mathbf{Q}\mathbf{z}}{\mathrm{p}}. \tag{14}$$

Thus in this structured case, the full MCA analysis is unnecessary. It suffices to handle the r×r matrices of (13) and obtain the other scores, if they are required, from (14). The solutions obtained from Fisher's data are given in Table 2. These have the expected ordering although no constraint has been put on the analysis to ensure this. The corresponding eigenvalue is $3\lambda = 1.8418$.

TABLE 2

RESPONSE SCORES OBTAINED BY MCA FOR FISHER'S DATA.

Response	-	?	w	(+)	+
Score	-.168	-.118	-.018	.079	.089

2.2 *The full set of solutions*

It is instructive, and relevant to the comparison with Fisher's analysis, to consider all p + q + r = 29 solutions to (8). It is easy to verify that when $3\lambda \neq 0$, 1, or 3 then if $(3\lambda; \mathbf{x},\mathbf{y},\mathbf{z})$ is a solution then so is $(2\text{-}3\lambda; -\mathbf{x},-\mathbf{y},\mathbf{z})$. These account for 2(r-1) = 8 solutions. The trivial solution $3\lambda = 3$ together with two null solutions (one less than the number of categorical variables) account for three more. This leaves p + q - (r + 1) = 18 further solutions. These all correspond to $3\lambda = 1$ and to $\mathbf{z} = 0$ and clearly satisfy (12) only if $\mathbf{P}'\mathbf{x} + \mathbf{Q}'\mathbf{y} = \mathbf{0}$. Thus these solutions lie in the null-space of $(\mathbf{P}',\mathbf{Q}')$ which has rank r = 5 and p + q = 24 columns. Its nullity is therefore 19, one more than is required. The extra solution is $\mathbf{x} = \mathbf{1}$, $\mathbf{y} = -\mathbf{1}$, $\mathbf{z} = \mathbf{0}$ for which $\lambda = 0$ and therefore lies in the space already accounted for and determined by the two zero eigenvalues. From the point of view of MCA, all solutions for which $0 < 3\lambda < 1$ are of potential interest but one might query whether solutions for which

$3\lambda = 1$, and hence $\mathbf{z} = \mathbf{0}$, can have any useful interpretation. The question raises itself clearly in the case of a balanced two-way table but MCA is usually applied in unbalanced cases and with more than three categorical variables. Then, although the equations (4) do not simplify to (12), nevertheless solutions comparable with those for which $3\lambda = 1$ and $\mathbf{z} = \mathbf{0}$ are likely to arise but in a less clearly identifiable form.

3. FISHER'S OPTIMAL SCORES ANALYSIS

In SMRW Fisher assigns dummy numerical values 0,x,y,z,1 to the five responses and effectively fits the usual additive model for a two-way table. In the context of an additive model it is clear that these scores are independent of an arbitrary origin, and in the context of the variance-ratios used by Fisher it is clear that their scaling is arbitrary too. Thus, to fix the scale, Fisher was able to choose one score to be zero, and another to be unity. For the moment I shall assume unconstrained scores represented by a vector $\mathbf{z}$ as in section 2. The analysis of variance then has the following form:

	Sums of squares
Rows	$\dfrac{p\mathbf{z}'\mathbf{P}'\mathbf{P}\mathbf{z} - (\mathbf{1}'\mathbf{W}\mathbf{z})^2}{pq}$
Columns	$\dfrac{q\mathbf{z}'\mathbf{Q}'\mathbf{Q}\mathbf{z} - (\mathbf{1}'\mathbf{W}\mathbf{z})^2}{pq}$
Residual	$\mathbf{z}'\mathbf{W}\mathbf{z} - \dfrac{\mathbf{z}'\mathbf{P}'\mathbf{P}\mathbf{z}}{q} - \dfrac{\mathbf{z}'\mathbf{Q}'\mathbf{Q}\mathbf{z}}{p} + \dfrac{(\mathbf{1}'\mathbf{W}\mathbf{z})^2}{pq}$
Total	$\mathbf{z}'\mathbf{W}\mathbf{z} - \dfrac{(\mathbf{1}'\mathbf{W}\mathbf{z})^2}{pq}$

The scores $\mathbf{z}$ would give an exact additive model if the residuals were all zero. Fisher therefore sought the additive model with the smallest residual relative to the total variance. The minimisation of this ratio of quadratic forms leads directly to the two-sided eigenvalue problem:

$$(\mathbf{W} - \frac{\mathbf{P}'\mathbf{P}}{q} - \frac{\mathbf{Q}'\mathbf{Q}}{p} + \frac{\mathbf{W}\mathbf{1}\mathbf{1}'\mathbf{W}}{pq})\mathbf{z} = \nu(\mathbf{W} - \frac{\mathbf{W}\mathbf{1}\mathbf{1}'\mathbf{W}}{pq})\mathbf{z}. \qquad (15)$$

Now it was seen in section 2, equation (13), that for MCA

$$(\mathbf{W} - \frac{\mathbf{P}'\mathbf{P}}{q} - \frac{\mathbf{Q}'\mathbf{Q}}{p})\mathbf{z} = \nu\mathbf{W}\mathbf{z}$$

has non-zero solutions for ν for which $\mathbf{z} \neq \mathbf{1}$ and $\mathbf{1'Wz} = 0$. These solutions are also solutions to (15). Hence Fisher's solutions are also MCA solutions, apart from the question of scaling which is resolved in the next paragraph.
Because $\mathbf{1'W1} = pq$ we have:

$$(\mathbf{W} - \frac{\mathbf{P'P}}{q} - \frac{\mathbf{Q'Q}}{p} + \frac{\mathbf{W11'W}}{pq})\mathbf{1} = (\mathbf{W} - \frac{\mathbf{W11'W}}{pq})\mathbf{1} = 0,$$

so that for $\mathbf{z} = \mathbf{1}$ both sides of (15) are zero. Hence for values of $\mathbf{z}$ other than $\mathbf{1}$, the solution to Fisher's equations may be replaced by $\mathbf{z} + \alpha\mathbf{1}$ for arbitrary values of α (i.e. as is required by the underlying additive model, $\mathbf{z}$ is identified up to an arbitrary translation) and for any value of ν. Fisher's solution, deriving as it does from a variance-ratio, also has arbitrary scaling. Thus $\mathbf{z}$ may be replaced by $\mathbf{z}^*$ in which $z^*_i = (z_i - z_1)/(z_r - z_1)$ thus ensuring that $z^*_1 = 0$ and $z^*_r = 1$. This transformation of Table 2 then gives the scores found in SMRW and reproduced in Table 3.

TABLE 3

MCA RESPONSE SCORES ADJUSTED TO GIVE FISHER'S RESULTS.

Response	-	?	w	(+)	+
Score	0	.193	.584	.958	1

Furthermore from (13) $\nu = 3\lambda(2 - 3\lambda)$

$= 1.8418 \times 0.1582$

$= 0.2914$

and $1 - \nu = 0.7086$ the value given in SMRW. That Fisher obtained $1 - \nu$ rather than ν is because, for convenience, he worked with maximising the "treatment" rather than minimising the "residual" sum-of-squares relative to the total.
To complete Fisher's analysis the additive constants $\mathbf{x}^*$ and $\mathbf{y}^*$ for rows and columns would be estimated in the usual way from the row and column means of the two-way table to give:

$$\mathbf{x}^* = \mathbf{Pz}^*/q \quad \text{and} \quad \mathbf{y}^* = \mathbf{Qz}^*/p.$$

Using (11) and (14) gives:

$$\mathbf{x}^* = \mathbf{P}(\mathbf{z} - z_1\mathbf{1})/q(z_r - z_1)$$
$$= \frac{(3\lambda - 1)\mathbf{x}}{(z_r - z_1)} - \frac{z_1\mathbf{1}}{(z_r - z_1)} \qquad (16)$$

the constant term of which would normally be absorbed into the mean. Thus $\mathbf{x}$, and similarly $\mathbf{y}$, of MCA give values proportional to the estimated additive effects of the conventional model for rows and columns. Note however that the factor of proportionality includes the term $(3\lambda-1)$ which is not required when relating $\mathbf{z}^*$ to $\mathbf{z}$.
Finally notice that although $\mathbf{z} = \mathbf{1}$ certainly satisfies (15), it satisfies it for any value of ν thus showing that setting $\mathbf{1}'\mathbf{W}\mathbf{z} = 0$ here is merely a convenient way of handling the usual problem of unidentifiability of parameters in a linear model.

4. THE RELATIONSHIP EXPLORED

At first sight this equivalence of MCA to FOS might seem surprising, because (i) in MCA a nested rather than cross-classified set-up is envisaged and (ii) MCA does not distinguish response variables, such as Fisher's reaction to serum, from other variables, and (iii) because simple two-way correspondence analysis, which is used in section 2 to develop MCA, depends fundamentally on a multiplicative rather than an additive model. These apparent inconsistencies are confusing but are fairly easily resolved by considering the precise relationship between the two representations of the same information given by the indicator matrix $\mathbf{G}$ and the two-way table 61.9 of SMRW (see Table 1). The following paragraphs explore this relationship in some detail. During the course of this exploration it will emerge that there are links with canonical correlation analysis and these too are clarified.

4.1 *Nesting and crossing*

In a general approach to the MCA of p categorical variables described in an indicator matrix $\mathbf{G}$, the rows of which correspond to individuals, scores $\mathbf{x}$ may be required for the category levels which minimise the total variation of the scores within individuals. This approach seems first to have been made explicit by Guttman (1941), although Fisher (1936) in SMRW clearly recognised a similar general principle and applied it successfully in the special case described above. There was however a difference in outlook between Fisher and Guttman. Fisher wished to make his model as additive as possible, thus maximising the "treatment" sum-of-squares, or equivalently minimising the "residual" sum-of-squares, relative to the total. If $\mathbf{G} = (\mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3)$ is partitioned into the indicator matrices for rows, columns and responses respectively, Fisher effectively considers the linear model :

$$\mathbf{G}_3\mathbf{z} = \mu\mathbf{1} + \mathbf{G}_1\mathbf{x} + \mathbf{G}_2\mathbf{y} + \text{error}.$$

Fisher never gave a general exposition of FOS, but it is clear that he has in mind the general linear model:

$$\mathbf{G_3z} = \mathbf{X}\beta + \text{error} \tag{17}$$

where **X** is a matrix of independent variables (assumed centred at their means and reparameterised to ensure that **X** has full column-rank) and the elements of β are constants which may be estimated by least-squares in the usual way to give:

$$\hat{\beta} = (\mathbf{X'X})^{-1}\mathbf{X'G_3z}$$

with a residual sum of squares $\mathbf{z'G_3'PG_3z}$, where in this section **P** is the usual least-squares projection matrix $\mathbf{I} - \mathbf{X(X'X)^{-1}X'}$. From the zero sum of estimated residuals it follows from (17) that $\mathbf{1'Wz} = \mathbf{1'G_3'G_3z} = \mathbf{1'G_3z} = \mathbf{1'PG_3z} = 0$, showing that the "constraint" $\mathbf{1'Wz} = 0$ is merely an expression of the zero sum of the estimated residuals when using linear least squares. So far **z** is unknown, but following Fisher can be estimated by minimising the ratio of quadratic forms, giving:

$$\mathbf{G_3'PG_3z} = \rho\mathbf{G_3'G_3z} \tag{18}$$

which may be compared with (13) and with (15).

Guttman (1946) describes the rationale for his approach: "Our basic principle in deriving numerical values -- let us call them x-values -- for the things being compared requires that the x-values of things that a given person judges higher than other things should be as different as possible from the x-values of the things he judges to be lower than other things. This will be achieved if we make the x-values of things judged higher as homogeneous as possible among themselves, for each individual. In the language of the analysis of variance, our principle calls for *minimizing the variation within individuals*, compared with that within the group as a whole." Guttman worked within the psychometric tradition and cited the work of Thurstone (as well as a passing reference to Fisher) and it seems that his search for homogeneity may have been influenced by concepts such as that of "general intelligence"; it is for this reason that MCA is sometimes termed Homogeneity Analysis -- see Gifi(1990). To summarise, Fisher was looking for additivity and Guttman for homogeneity.

Whatever the motivation, Fisher and Guttman needed to solve the same algebraic problem. In Guttman's sense, and with m categorical variables, the group (i.e. individual) means are given by $\mathbf{Gw}/m$ so that the within-group sum-of-squares is $\mathbf{w'Dw} - (\mathbf{w'G'Gw})/m$. The details need not be spelled out but clearly this leads to equations which are satisfied by the MCA solutions; the thing to note is that the analysis is of **G** treated as a two-way table and is of nested form. This approach is consistent with Fisher's cross-classification analysis because when his 12×12 table is expressed in the extended 144×29 form of **G**, and Guttman scores estimated, then the score for individual i, with serum j and response k is given by $x_i + y_j + z_k$. Thus despite the appearance of fitting a nested model, a cross-classification is implicit but it is a cross-classification which includes the response variable among the classification variables.

In the hope of greater enlightenment, but at the risk of confusion, a further example may be considered. Suppose we have a two-way contingency table **S** but no response variable. Then it might be interesting to know, as was required by Fisher (1940), what scores for the column-classes best separate the row-classes (the obvious connection with discriminant analysis is discussed in section 5). This requires an FOS analysis of **S**, whose structure is to be regarded as nested within rows, which differs from the previously described cross-classification analysis of Fisher. The FOS procedure is directly applicable and may be compared with the MCA, which may be regarded as a nested analysis of the corresponding indicator-matrix **G** for two categorical variables The Burt-matrix is:

$$\begin{pmatrix} \mathbf{R} & \mathbf{S} \\ \mathbf{S'} & \mathbf{W} \end{pmatrix}$$

where **R** is a diagonal matrix giving the replication for each level of the row-classification. Thus **R** and **S** correspond to $\mathbf{G_1'G_1}$ and $\mathbf{G_1'G_3}$ respectively. The scores **z** of MCA can be derived similarly to the steps leading to (13) and (14) and found to satisfy:

$$\mathbf{S'R^{-1}Sz} = (2\lambda - 1)^2\mathbf{Wz} \qquad (19)$$

and those satisfying (18) for Fisher's optimal scores give:

$$(\mathbf{W} - \mathbf{S'R^{-1}S})\mathbf{z} = \nu(\mathbf{W} - \mathbf{W11'W}/n)\mathbf{z} \qquad (20)$$

where n = **1'W1** = **1'R1** is the total number of observations. Of course, with only two categorical variables, the MCA (19) gives essentially a simple CA of S. Arguments similar to those of section 3 show that the same **z** satisfies both (19) and (20). The point is that, following Fisher, (20) represents a within-groups analysis of **S** while, following Guttman, (19) represents a within-groups analysis of **G**.

An asymmetry in the treatment of categorical and quantitative variables to note is that two sets of dummy variables code for only two categorical variables while two sets of quantitative variables can encompass any number. There is no reason why the independent variables **X** of (17) should represent dummy variables. When **X** represents continuous variables, (20) becomes the popular method of canonical variate analysis, first explicitly discussed by Rao (1948), and which is normally used with unequal replication within groups. In this case , the unequal replication is given in the Burt matrix by the diagonal matrix **R** and in **G** by having $s_{i,j}$, the (i,j)th element of **S**, coded for by $s_{i,j}$ identical rows. In general forms of MCA there is another kind of unequal replication. This is when values are missing from the indicator -matrix **G** - this occurs, for example, when the indicators represent answers to questions, some of which are unanswered. As with any nested analysis, (17) and (18) do not require that there should be equal replication within groups.With unequal replication the analysis goes through. In **G** this inequality is reflected in missing cells so that its rows no longer sum to a constant. Thus now $\mathbf{G1} = \mathbf{S1} = \mathbf{N1}$ where **N** is a diagonal matrix giving the number of replicates for every row of **S**. Now $3^{-1/2}$ in (1) is replaced by $\mathbf{N}^{-1/2}$ but, because of the within-rows nature of the analysis, the MCA of the so-adjusted **G** remains valid (Meulman, 1982).

4.2 *The response variable*

MCA gives no special recognition to response variables. Fisher's analysis clearly recognises reactions to the sera as a response, effectively the dependent variable in a multiple regression, and he provides an analysis of variance for that variable. Yet MCA gives the same scores as Fisher obtained for the levels of reaction, apart from the scaling and translation adjustments. MCA also gives scores for the twelve people and the twelve sera which equation (16) has shown to be proportional to the additive effects of the conventional model. However, in Fisher's example **z** has fewer levels than do **x** and **y** which, together with the relationship $\mathbf{1'Wz} = 0$, reduces the five possible scoring systems to only four of potential interest. Equation (14) shows that the scores **x** and **y** are also confined to possibilities of rank four. In section 2.2 it was shown that MCA gave four pairs of solutions, each pair corresponding to eigenvalues 3λ and $2 - 3\lambda$. These are nicely combined in Fisher's optimal scores because his method leads to a single set of eigenvalues $\nu = 3\lambda(2 - 3\lambda)$ each with one set of scores $\mathbf{x}^*$ for **x** given by (16), and similarly for **y**. Thus Fisher has one null-solution and four others; the remaining 24 solutions found by MCA are not reproduced in any form. In particular, in the lead up to (18) it was seen that the counterpart of discarding the trivial solution of MCA manifests itself, with FOS, as an expression of zero-sum of residuals, for which there can be few misgivings. Therefore FOS and MCA cannot have any general equivalence. Indeed it should be clear that only in special cases, such as the balanced two-way table of SMRW, will equation (17) reduce to (13) of MCA; details are given in section 4.4. Thus even the scores for the response variable in MCA will not normally have an equivalent FOS interpretation. In the special case of balance, the algebra has indeed shown that no special action need be taken for the response variable, but I still find the result rather remarkable.

4.3 *Additivity and multiplicativity*

This final inconsistency is the most easily explained. The correspondence analysis of a general two-way table depends on its singular-value decomposition and hence on a multiplicative model. However when MCA is treated as a correspondence analysis of the two-way indicator matrix **G**, it should be recognised that **G**, having zero/one values and constant row-sums, is a very special two-way table.The scores **Gz** obtained by Guttman or equally those obtained from Fisher's additive model, are indisputedly of additive form. Yet the MCA depends on the multiplicative decomposition:

$$m^{-1/2}\mathbf{G}\mathbf{D}^{-1/2} = m^{1/2}\mathbf{1}\mathbf{1}'\mathbf{D}^{1/2} + \sum_{i=2}^{m}\sigma_i\mathbf{u}_i\mathbf{v}_i' \tag{21}$$

where the Fisher additive scores **z** are given by $\mathbf{z} = \mathbf{D}^{-1/2}\mathbf{v}_2$. Post-multiplying (21) by $\mathbf{v}_2$ gives:

$$m^{-1/2}\mathbf{G}\mathbf{z} = \sigma_2\mathbf{u}_2$$

showing that the vector $\mathbf{u}_2$ has values that are proportional to the mean-scores estimated for each individual. Thus both components of the multiplicative term $\mathbf{u}_2\mathbf{v}_2'$ have simple additive interpretations.

4.4 *The links with canonical correlation*

The multiple regression formulation (17) leading to (18) makes it clear that FOS must be very closely related to a CCA in which the scores **z** are the loadings for the dependent dummy-variables and β are the loadings associated with all the independent dummy-variables. The only difference is whether the least-squares estimate $\hat{\beta}$ given by (18) is used, or $\rho^{-1}\hat{\beta}$ as given by the canonical equations; because in SMRW Fisher is interested only in estimating **z**, this difference is irrelevant for his problem.

It has been shown that both the optimal scores analysis of SMRW and that for a two-way contingency table, coincide with MCA and it has been asserted that this result is not a general one. If it were generally true, then MCA also would have to be equivalent to CCA. With two categorical variables this equivalence is established but, because CCA handles only two sets of variables, this equivalence must break down with more than two, except perhaps in special cases. To see this with three variables, consider again the notation of section 2 where the Burt-matrix is represented in 3×3 symmetric block form with elements $\mathbf{B}_{ij}$ (i,j = 1,2,3). From (6) and (7) it follows that $\mathbf{B}_{ij}$ may be replaced everywhere by

$$\mathbf{B}^*_{ij} = \mathbf{B}_{ij} - \mathbf{B}_{ii}\mathbf{1}\mathbf{1}'\mathbf{B}_{jj}/n$$

which is the sums-of-squares-and-products corrected for the mean. Thus $\mathbf{w}'\mathbf{B}^*\mathbf{w}$ is the variance-covariance matrix of the scores. Equation (4) may now be written:

$$\mathbf{B}^*\mathbf{w} = 3\lambda\mathbf{D}^*\mathbf{w} \qquad (22)$$

with $\mathbf{D}^* = \text{bdiag}(\mathbf{B}^*_{11}, \mathbf{B}^*_{22}, \mathbf{B}^*_{33})$, where bdiag implies block-diagonal, so that $\mathbf{D}^*$, unlike $\mathbf{D}$, is not a diagonal matrix. This two-sided eigenvalue problem has precisely the same solutions as for the previous formulation of MCA except that now when $\mathbf{w} = \mathbf{1}$ both sides of (22) are identically zero. If now ($\mathbf{x}'$, $\mathbf{y}'$) are regarded as a single set of scores ξ', and $\mathbf{B}^*$ is accordingly partitioned into 2×2 block form:

$$\mathbf{B}^* = \begin{pmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{C}' & \mathbf{B}^*_{33} \end{pmatrix}$$

where $\mathbf{C}' = \begin{pmatrix} \mathbf{B}^*_{31} & \mathbf{B}^*_{32} \end{pmatrix}$ and $\mathbf{A} = \begin{pmatrix} \mathbf{B}^*_{11} & \mathbf{B}^*_{12} \\ \mathbf{B}^*_{21} & \mathbf{B}^*_{22} \end{pmatrix}$, then (22) becomes:

$$\begin{pmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{C}' & \mathbf{B}^*_{33} \end{pmatrix}\begin{pmatrix} \xi \\ \mathbf{z} \end{pmatrix} = 3\lambda\begin{pmatrix} \text{bdiag}(\mathbf{A}) & \mathbf{0} \\ \mathbf{0} & \mathbf{B}^*_{33} \end{pmatrix}\begin{pmatrix} \xi \\ \mathbf{z} \end{pmatrix}.$$

This gives the same values of $\mathbf{z}$ as obtained by canonical correlation between ξ and $\mathbf{z}$ if and only if

$$\mathbf{A} = \text{bdiag}(\mathbf{A}), \qquad (23)$$

when the scores ξ will also agree with CCA; the canonical correlations will have values $3\lambda - 1$.

In the example under discussion (23) requires that $\mathbf{B}^*_{12} = \mathbf{0}$ which is satisfied for the balanced example of SMRW; it is also true for simple CA where the second classification does not exist and hence $\mathbf{A} = \mathbf{B}^*_{11}$, and $\mathbf{C} = \mathbf{B}^*_{13}$, thus confirming the results (19) and (20). Note that if the scores $\mathbf{x}$ (say) had been treated as responses, then the MCA results would be unaffected but they would not now agree with the canonical analysis unless $\mathbf{B}^*_{23} = \mathbf{0}$, which is certainly not the case with the SMRW example.

In the above $\mathbf{A}$ and all $\mathbf{B}^*_{ij}$ are singular matrices, so that ordinary matrix inverses must be avoided. For MCA the equivalent form (4) is available. For canonical correlation one of the usual devices suffices, such as striking out the first row and column of each block-matrix, to find scores corresponding to identifiable contrasts.

The above discussion for three categorical variables, leading to (23), extends easily to any number of variables, and shows that the general condition for the CCA to agree with MCA is when the variables can be partitioned into two sets such that:

$$\mathbf{B}^* = \begin{pmatrix} \mathbf{A}_1 & \mathbf{C} \\ \mathbf{C}' & \mathbf{A}_2 \end{pmatrix} \quad \text{where} \quad \mathbf{D}^* = \begin{pmatrix} \mathbf{A}_1 & \mathbf{0} \\ \mathbf{0}' & \mathbf{A}_2 \end{pmatrix}, \tag{24}$$

so that $\mathbf{A}_1$ and $\mathbf{A}_2$ are both in block-diagonal form. Note that any partition into two such sets that may exist, need not be an interesting partition from the point of view of canonical correlation. FOS naturally focuses on two sets, the dependent variable(s) and the independent variable(s).

5. CONCLUSION

Fisher's development of discriminant analysis is firmly based on a regression approach in which the response (dependent) variable is dummy. Fisher's first published paper on the subject (Fisher, 1936) was concerned only with the two-group case, but an earlier paper, Barnard (1934), had been based on Fisher's advice and had involved four groups of pre-dynastic Egyptian skulls defined by a single dummy variable with four values, taken to be roughly proportional to the time between burials. In SMRW, Fisher states "When only a single component is to be maximised relative to the rest, the equations are linear and the procedure of multiple regression may be used". This statement immediately precedes the section on optimal scores and leads to the more general possibility of maximising ratios of quadratic forms. In his discriminant work, Fisher always seems to have been concerned with "a single component". The evidence strongly suggests that Fisher saw his optimal scores and discrimination as being closely related, as in many ways they are. Fisher(1940) in his paper "The Precision of Discriminant Functions" refers to Maung's compilation of Tocher's data to give a contingency table for 5387 children classified by Eye-colour (blue, light, medium, dark) and Hair-colour (fair, red, medium, dark, black). He uses this example to illustrate "the complications that arise when more than one degree of freedom is maximised", that is when the regression approach to discrimination is replaced by canonical analysis. The subsequent discussion is puzzling in several ways. To ensure the full column-rank of $\mathbf{X}$ in (17) he uses three dummy variables x_1,x_2 and x_3 to differentiate between the four eye-colours and then writes "We may then ask for what eye colour scores, i.e. for what linear function of x_1,x_2,x_3 are the five hair classes most distinct". This seems to indicate a clear application of the variant of the optimal scores procedure discussed at the end of section 4.1, though why eye-colour should be regarded as a response to hair-colour is unclear. This is recognised by Fisher, for he goes on to write "The answer may be found in a variety of ways. For example, by starting with arbitrarily chosen scores for eye colour, determining from these average scores for hair colour, and using these latter to find new scores for eye colour". This is the reciprocal-averaging algorithm first described by Fisher and Mackenzie (1923) which clearly does not single out either variable as a response. The analysis at the end of 4.1 shows that in this case both optimal scores and reciprocal-averaging give the same results and that these results are those of a simple correspondence analysis of Maung's contingency table, with possible minor differences in scaling and origin. Fisher uses the same scaling as he did in SMRW and this is fully justified because he is concerned only with a one-dimensional set of scores and because of the correlational interpretation of correspondence analysis given by Hirschfeld (1935). Fisher is aware of the correlational interpretation, though apparently not of Hirschfeld's paper, writing "Hotelling has called pairs of functions of this kind canonical components". This must be a reference to Hotelling (1936) on canonical correlation, but Fisher is concerned with the correlation between linear combinations of dummy rather than quantitative variables; the mathematics is, of course, identical. Subsequently Maung (1941) discussed and illustrated Fisher's method in detail. He established the equivalence of (i) CA by reciprocal averaging with (ii) determining scores for rows and columns that maximise correlation (see Hirschfeld) and (iii) a between/within discriminant method. The latter is concerned with discriminating between hair-colour classes using optimal scores for eye-colour and so relates to the equivalence of (19) and (20), above. An interesting passage occurs at the foot of page 200 of Maung's paper where he mentions that "Prof. R.A. Fisher has pointed out ...." a result which is essentially the singular-value decomposition (21) associated with the CA of a two-way contingency table. It seems to me that although Fisher (1940) might properly be cited as an early example of the correspondence analysis of a two-way contingency table, the ideas described there fall short of a description of MCA. He certainly has the concept of indicator matrices but for MCA one needs the correspondence analysis of a general indicator matrix **G**, or one of its many equivalents. Fisher (1940) does not have this generalisation but section 3, above, has shown that SMRW makes implicit use of more general indicator matrices and adopts an optimal scoring method and that Fisher (1940) does introduce a general indicator matrix for a single categorical variable. I believe that it is Fisher (1938) in SMRW, rather than Fisher (1940) that comes closest to MCA and should be cited as his main contribution to the early literature on MCA and related methods.

Many of the results discussed above are not new (see Greenacre, 1984) but I am not aware that there has been any previous comprehensive attempt to disentangle the many, and sometimes confusing, relationships between FOC, CCA and MCA. Therefore it may be useful to summarise the main equivalences that have been established, either in this paper or by other authors. This is done in Table 4, where the term *equivalence* does not imply *the same in all respects* but only that the principal eigenvectors are the same up to possible scaling and translation; similarly eigenvalues may not be identical but may be simply related. The main text gives details of several such equivalences.

TABLE 4

THE MAIN RELATIONSHIPS BETWEEN FISHER'S OPTIMAL SCORES (FOS), CANONICAL CORRELATION ANALYSIS (CCA), MULTIPLE CORRESPONDENCE ANALYSIS (MCA), AND CORRESPONDENCE ANALYSIS (CA)

(i) FOS is equivalent to CCA, but Fisher only discussed special cases.

(ii) CA is a special case of CCA.

(iii) The FOS analysis of a two-way contingency table is equivalent to CA - this follows from (i) and (ii).

(iv) CA is equivalent to a special case of MCA.

(v) The FOS analysis of a balanced two-way table with a categorical response variable is equivalent both to a special case of MCA and by (i) to a special case of CCA.

(vi) The FOS analysis of a set of continuous variables measured on samples grouped into disjoint classes, is equivalent to canonical variate analysis and by (i) is a special case of CCA.

(vii) It is not true that MCA is generally equivalent to CCA even when CCA is confined to categorical variables.

(viii) MCA is equivalent to CCA and FOS when all variables are categorical and can be partitioned into two sets such that condition (24) is satisfied.

Thus there are two main families of methods: MCA and CCA. These have an intersection containing CA and some special cases of optimal scores discussed by Fisher. FOS is the same as CCA, but Fisher himself seems to have considered only special cases and to have exploited only the dominant canonical solution.

Although MCA and CCA are distinct, in the sense just described, this does not exclude both being special cases of some more general formulation. Indeed, as is very well known, both are special cases of optimising ratios of quadratic forms. This is an important algebraic unifying principle that, among other things, is helpful to writers of general-purpose software. Further, it has been seen that the quadratic forms in question are similar, requiring only the condition (24) for equivalence. In one of the many generalisations of canonical correlation to handle more than two sets of variables, van de Geer (1986) has proposed a method that requires the solution to $\mathbf{Aw} = \lambda \text{bdiag}(\mathbf{A})\mathbf{w}$ which clearly gives

MCA when **A** is a Burt-matrix and gives CCA when there are only two sets of variables. Yet another extension that encompasses both methods is the generalisation of canonical correlations (GCCA) to several sets of variables proposed by Carroll (1968). Tenenhaus and Young (1985) show that, for categorical variables, GCCA is equivalent to MCA and Nierop (1989) shows how a constrained version includes CCA, and goes on to describe even further generalisations which include GCCA. DeLeeuw, Young and Takane (1976) adopt Fisher's notion of additive scores and extend it to a rich class of models by allowing constraints (including ordinal, linear and bounded) to be imposed on any combination of the scores **x**,**y**,**z**. These kinds of extension are fruitful areas of current research.

REFERENCES

Barnard, M.M. [1935]. The secular variations of skull characters in four series of Egyptian skulls. *Annals of Eugenics 6,* 352-371.

Bartlett, M.S. [1938]. Further aspects of the theory of multiple regression. *Proceedings of the Cambridge Philosophical Society 34*, 33-40.

Carroll, J.D. [1968}. A generalization of canonical correlation analysis to three or more sets of variables. *Proceedings of the 76th annual convention of the American Psychological Association 3,* 227-8.

DeLeeuw, J. [1983]. On the prehistory of correspondence analysis. *Statistica Neerlandica 37,* 161-4.

DeLeeuw, J., Young, F.W. and Takane, Y.[1976]. Additive structure in qualitative data: an alternating least squares method with optimal scaling features. *Psychometrika 41,* 471-503.

Eckart, C. and Young, G. [1936]. The approximation of one matrix by another of lower rank. *Psychometrika 1*, 211-218.

Fisher, R.A. [1918]. The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh 52,* 399-433.

Fisher, R.A. and Mackenzie W.A. [1923]. Studies in crop variation: II The manurial response of different potatoe varieties. *Journal of Agricultural Science 13*, 311-320.

Fisher, R.A. [1936]. The use of multiple measurements in taxonomic problems. *Annals of Eugenics 7,* 179-188.

Fisher, R.A. [1938]. *Statistical Methods for Research Workers, 7th Edition.*.Oliver and Boyd, Edinburgh.

Fisher, R.A. [1940]. The precision of discriminant functions. *Annals of Eugenics 10,* 422-9.

Gifi, A. [1990] *Non-linear Multivariate Analysis.* J. Wiley and Son, New York.

Gittins, R. [1985]. *Canonical Analysis: a Review with Applications in Ecology.* Springer-Verlag, Berlin.

Greenacre, M.J. [1984]. *Theory and Applications of Correspondence Analysis*. Academic Press, London.

Guttman, L. [1941]. The quantification of a class of attributes: a theory and method of scale construction. In P Horst et al.(eds.) *The Prediction of Personal Adjustment.* The Social Science Research Council, New York. *Bulletin no. 48*, 319-48.

Guttman, L. [1946]. An approach for quantifying paired comparisons and rank order. *Annals of Mathematical Statistics* 17, 144-63.

Healy, M.J.R. and Goldstein, H. [1976]. An approach to the scaling of categorised attributes. *Biometrika 63*, 219-29.

Hirschfeld, H.O. [1935]. A connection between correlation and contingency. *Proceedings of the Cambridge Philosophical Society 31*, 520-4.

Hotelling, H. [1936]. Relations between two sets of variates. *Biometrika 28,* 321-377.

Kempton, R.A. and Talbot, M. [1988]. The development of new crop varieties. *Journal of the Royal Statistical Society, A 151*, 327-341.

Maung, K. [1941]. Measurement of association in a contingency table with special reference to the pigmentation of hair and eye colour in Scottish school children. *Annals of Eugenics 11,* 189-223.

Meulman, J. [1982]. *Homogeneity Analysis of Incomplete Data*. DSWO Press, Leiden.

Nierop, D. {1989]. Generalized set component analysis: a toolbox for multiple sets analysis. Department of data theory, University of Leiden, *RR-89-04,* 1-19.

Rao, C.R. [1948]. The utilization of multiple measurements in problems in biological classification. *Journal of the Royal Statistical Society, B 10*, 159-203.

Tenenhaus, M. and Young, F. [1985]. An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis and other methods for quantifying multivariate categorical data. *Psychometrika 50,* 91-119.

van de Geer, J.P. [1986]. *Introduction to Linear Multivariate Data Analysis*. DSWO Press, Leiden.