Appl. Statist. (1989)
38, *No.* 2, *pp.* 249–292

A Combined Approach to Contingency Table Analysis using Correspondence Analysis and Log-linear Analysis

By PETER G. M. VAN DER HEIJDEN,†

University of Leiden, The Netherlands

ANTOINE DE FALGUEROLLES

Université Paul Sabatier, Toulouse, France

and JAN DE LEEUW

University of California at Los Angeles, USA

[*Read before* The Royal Statistical Society *on Wednesday, October 26th, 1988, a Vice-President,* Dr G. H. Freeman, *in the Chair*]

SUMMARY

Correspondence analysis and one of its generalizations are presented as tools that can be used for the analysis of residuals from log-linear models. By recognizing relations between correspondence analysis and log-linear analysis a better understanding of correspondence analysis is obtained. Furthermore, it is shown how these relations can be used to arrive at a combined approach to contingency table analysis using both log-linear analysis and correspondence analysis.

Keywords: Association models; Asymmetry; Correspondence analysis; Exploratory data analysis; Log-linear analysis

1. Analysis of Categorical Data

Correspondence analysis (CA) has quite a long history as a method for the analysis of categorical data. The starting point of this history is usually set in 1935, and since then CA has been reinvented several times. In this history, we can distinguish simple CA (CA of contingency tables) and multiple CA (CA of so-called indicator matrices). At present simple CA is known under different names, such as reciprocal averaging, dual (or optimal) scaling, canonical correlation analysis and simultaneous linear regressions. Greenacre (1984) discusses the differences in approach between simple CA and these methods (see also Nishisato (1980), ch. 3). Multiple CA is also known under other names such as homogeneity analysis and dual (or optimal) scaling (see Tenenhaus and Young (1985)).

The term CA originates from France, where it is quite popular. This is probably due to Benzécri and his associates. They developed CA in the 1960s, and their work

†*Address for correspondence*: Department of Psychometrics and Research Methods, University of Leiden, Hooigracht 15, 2312 KM Leiden, The Netherlands.

 0035–9254/89/38249 $2.00

culminated in the standard references Benzécri *et al.* (1973) and the series *Pratique de l'Analyse des Données* (Benzécri *et al.*, 1980, 1986). Furthermore, since 1976 there has been a special journal dedicated almost exclusively to (the applications of) CA, called *Les Cahiers de l'Analyse des Données*. CA places a heavy emphasis on geometrical representations, and probably this is one of the reasons why this approach became so popular, at least in France. One reason for the lag in development outside France is perhaps the language problem. However, recent publications in English have done much to alleviate this situation (see, for example, Nishisato (1980), Gifi (1981), Jambu and Lebeaux (1983), Greenacre (1984) and Lebart *et al.* (1984); see also the papers of Hill (1974), Healy and Goldstein (1976) and Deville and Malinvaud (1983)), and it is our impression that CA is becoming rapidly popular outside France (see the bibliography of Nishisato (1986)). A second reason might be that CA is often introduced without any reference to other methods of statistical treatment of categorical data which have proven their usefulness and flexibility.

A major difference between CA and most other techniques for categorical data analysis lies in the use of models. In log-linear analysis (LLA), for example, a distribution is assumed under which the data are collected, then a model for the data is hypothesized and estimations are made under the assumption that this model is true, and finally these estimates are compared with the observed frequencies to evaluate the model. In this way it is possible to make inferences about the population on the basis of sample data.

In CA it is claimed that no underlying distribution has to be assumed and no model has to be hypothesized, but a decomposition of the data is obtained to study the 'structure' in the data. The primary interest is in the presentation of the structure in the observed data in some optimal way. A rationale for this approach is that the observed data matrix is approximated in a least squares sense by a data matrix of lower rank. Strictly, conclusions about the data may not be generalized to the population (although in practice it is difficult to prevent people from doing this).

In what follows we shall concentrate on simple CA, i.e. CA of contingency tables, and whenever we speak of CA, we mean simple CA.

The first approach to bridge the gap between CA and model-based approaches was to show under what conditions CA results (i.e. row and column scores, singular values) are similar to those of the log-multiplicative row–column (RC) model (see Goodman (1979, 1981a, b); in Goodman (1985, 1986) this model is discussed under the name RC association model). Recently, the gap between CA and model-based approaches has been bridged further by Goodman (1985, 1986) and Gilula and Haberman (1986), who introduced a version of CA as a model.

In this paper we summarize work that sets off at a different angle, namely that of CA as a tool for residual analysis. In this approach CA is viewed as a technique that can be used for the exploration of residuals from log-linear models. The usage of a generalization of CA plays an important role in this approach. It will be shown that CA and LLA are closely related. This will lead us to two general conclusions.

(a) A better understanding of CA procedures is obtained by recognizing the relationships between CA and LLA. This facilitates the understanding of CA for those who are more accustomed to LLA.
(b) It is possible to combine CA and LLA into one approach of contingency table analysis. Crudely, in this approach LLA is used for the assessment of the

importance of interaction effects and then CA can be used to study the residuals of models.

The distinction between points (a) and (b) may seem artificial. However, our impression is that experienced users of log-linear and related models often do not seem to feel a need for CA, and therefore the second point does not seem relevant to them, but the first point may still be of interest to them.

In what follows we shall first give a short introduction to (generalized) CA and LLA. In Section 3 we present some results indicating relationships between CA and LLA. In Section 4 we illustrate a combined approach for the analysis of a square contingency table.

2. Tools of Analysis: Correspondence Analysis, Generalized Correspondence Analysis and Log-linear Analysis

2.1. Correspondence Analysis

We shall discuss CA here with an emphasis on its geometrical aspects. For details and proofs, we refer to Benzécri *et al.* (1973, 1980, 1986), Nishisato (1980), Gifi (1981), Greenacre (1984) and Lebart *et al.* (1984) and the references given there.

CA is a technique by which it is possible to find a multidimensional representation of the dependence between the rows and columns of a two-way contingency table. This representation is found by allocating scores to the row and column categories and displaying the categories as points, where the scores are used as co-ordinates of these points. These scores can be normalized in such a way that distances between row points and/or between column points in Euclidean space are equal to so-called chi-square distances.

Consider a two-way contingency table P with proportions p_{ij}, having I rows ($i = 1, \ldots, I$) and J columns ($j = 1, \ldots, J$). An index is replaced by '+' when summed over the corresponding variable, e.g. $\Sigma_j p_{ij} = p_{i+}$. Chi-square distances can be computed between rows as well as between columns. We proceed by considering chi-square distances between rows. These distances are computed on the profiles of the rows of a matrix, where the profile of row i is the vector of conditional proportions p_{ij}/p_{i+}. The chi-square distance between rows i and i' is defined by

$$\delta^2(i, i') = \sum_j \frac{(p_{ij}/p_{i+} - p_{i'j}/p_{i'+})^2}{p_{+j}} . \tag{1}$$

Formula (1) shows that $\delta^2(i, i')$ is a measure for the difference between the profiles of rows i and i'. When i and i' have the same profile, $\delta^2(i, i') = 0$. The difference between profiles i and i' for column j is divided by p_{+j}, thus giving less influence to differences for column categories having large margins. The configuration of I row points is located in a Euclidean space of dimension $I - 1$. In this space co-ordinates X are used such that $d^2(x_i, x_{i'}) = \delta^2(i, i')$. The profile of column proportions p_{+j}, being the mean row profile, is the weighted average of the row points, where the p_{i+} are used as weights. This weighted average is located at the origin.

The chi-squared distance concept is used in the interpretation of the configuration of points. When two rows are close together, their profiles must be similar, and these rows are related in roughly the same way to the columns. When two rows are far apart, they are related in different ways to the columns. When a row point is near the

centre, its profile is similar to the profile of column proportions p_{+j}. When two row points are in opposite directions from the centre, they deviate in opposite ways from the profile of column proportions.

So far we have only discussed the situation for the row categories. CA is symmetric in the sense that similar results hold for the columns: it is also possible to construct a $(J - 1)$-dimensional space in which the mean column profile is placed at the weighted average of the other points, where the p_{+j} are used as weights. Chi-square distance interpretations can be used similarly to those already discussed.

The CA solution can be found as follows: let P be the matrix to be analysed; let D_r and D_c be the diagonal matrices whose diagonal entries are respectively the marginal row proportions p_{i+} and column proportions p_{+j}, where it is assumed that $p_{i+} > 0$ and $p_{+j} > 0$; let $E = D_r tt' D_c$, where t is a unit vector, the length of which depends on the context. Elements of E have the form

$$e_{ij} = p_{i+} p_{+j}. \tag{2}$$

We then compute the singular value decomposition (SVD) of the matrix $D_r^{-1/2}(P - E)D_c^{-1/2}$. Elements of this matrix have the value $(p_{ij} - e_{ij})/e_{ij}^{1/2}$; they are proportional to standardized residuals. These residuals are decomposed as follows:

$$D_r^{-1/2}(P - E)D_c^{-1/2} = U\Lambda V' \tag{3}$$

where $U'U = I = V'V$ and Λ is a diagonal matrix with singular values λ_α in descending order. When we define $D \leqslant \min(I - 1, J - 1)$, U is of order $I \times D$, V is of order $J \times D$ and Λ is of order $D \times D$ and is non-singular. The row and column scores are normalized as follows:

$$R = D_r^{-1/2} U \tag{4a}$$

and

$$C = D_c^{-1/2} V. \tag{4b}$$

Thus $R'D_r R = I = C'D_c C$ and, reflecting the fact that the row and column sums of $P - E$ vanish, $t'D_r R = 0 = t'D_c C$, i.e. the row scores are uncorrelated, as are the column scores, while for each dimension both these scores have zero mean and unit variance.

The relation between the row and column points is specified by the 'transition formulae'

$$\tilde{R} = D_r^{-1} PC \tag{5a}$$

and

$$\tilde{C} = D_c^{-1} P'R \tag{5b}$$

where $\tilde{R} = R\Lambda$ and $\tilde{C} = C\Lambda$. Equations (5a) and (5b) show that row points $\tilde{R}$ are the weighted averages of the column points C while column points $\tilde{C}$ are the weighted averages of the row points R. In this context the profile values are used as weights. Using $\tilde{R}$ as co-ordinates for row points and $\tilde{C}$ as co-ordinates for column points, distances between row points and distances between column points are chi-square distances. For this reason in French publications usually simultaneous plots of row and column points are made using the pair $(\tilde{R}, \tilde{C})$. Other choices made are to use the

pairs $(R, \tilde{C})$, $(\tilde{R}, C)$ or $(R\Lambda^{1/2}, C\Lambda^{1/2})$. We refer to Israëls (1987) for a discussion of advantages and drawbacks of these possibilities.

Equations (5a) and (5b) are used to interpret distances between row and column points. When a row profile is equal to the average row profile, equation (5a) shows that the row point will be at the weighted average of the columns, i.e. the origin. When for some column j the row profile value p_{ij}/p_{i+} is larger than the average p_{+j}, this column will attract the row point in its direction. Multiplying both values by p_{i+}, we see that when $p_{ij} > e_{ij}$ (i.e. when the residual is positive), row i will be attracted by column j, and, as is shown by equation (5b), vice versa. In general, the larger the difference $p_{ij} - e_{ij}$, the nearer i and j will be. However, because the criterion optimized in CA is not stated in terms of between-set distances but only in terms of within-set distances (namely $\tilde{R}'D_r\tilde{R} = \Lambda^2 = \tilde{C}'D_c\tilde{C}$), it is advisable to be careful with such interpretations.

By substituting equation (4) into equation (3), we find

$$P = E + D_r R\Lambda C' D_c = D_r(tt' + R\Lambda C')D_c \tag{6}$$

which is known as a 'reconstitution formula'. Equation (6) shows that CA decomposes the departure from independence in matrix P. It follows that CA only makes sense when these residuals are not merely a result of random variation from independence (in practice this is very often not checked for). Whether this is the case or not can be tested by using the Pearson X^2 statistic:

$$X^2 = n \sum_i \sum_j (p_{ij} - e_{ij})^2/e_{ij} \tag{7}$$

where n is the sample size. The relation between X^2 and the squared singular values in Λ^2 follows from equations (3) and (7):

$$\operatorname{tr}\Lambda^2 = X^2/n. \tag{8}$$

In French publications $\operatorname{tr}\Lambda^2$ is often called the 'total inertia'. In English publications X^2/n is often referred to as Pearson's index of mean-square contingency. Equation (8) shows that CA decomposes the chi-square value for testing independence in the matrix.

Concluding, we can summarize CA as follows: a naive model, the independence model, is chosen and the residuals of this model are decomposed. In this way a picture of the interaction between the row and column variables is obtained.

2.1.1. *Example*

As an example we shall analyse the data displayed in Table 1. The example deals with shop-lifting (see Israëls (1987) for an extensive analysis of this data set using many different techniques). A frequency denotes the number of people of specific sex (S) and age (A) that were suspected of stealing specific goods (G) in the Netherlands in 1978 and 1979. The table is a three-way table, but for the moment we shall consider this table as an ordinary two-way table of order $(2 \times 9) \times 13$ by coding sex and age interactively. In Section 3.2 we shall discuss the higher way aspect of this table in detail.

The CA solution has first three singular values (with proportion of chi square) 0.59 (0.58), 0.35 (0.20) and 0.27 (0.12), so the first two dimensions display 78% of the departure from independence. We shall study the two-dimensional solution displayed

TABLE 1
People suspected of shop-lifting by sex, age and kind of stolen goods

Sex	*Age (years)*	*Clothing*	*Clothing accessories*	*Provisions, tobacco*	*Writing materials*	*Books*	*Records*	*Household goods*	*Sweets*	*Toys*	*Jewelry*	*Perfume*	*Hobbies, tools*	*Other*	*Total*
Male	< 12	81	66	150	667	67	24	47	430	743	132	32	197	209	2845
	12–14	138	204	340	1409	259	272	117	637	684	408	57	547	550	5622
	15–17	304	193	229	527	258	368	98	246	116	298	61	402	454	3554
	18–20	384	149	151	84	146	141	61	40	13	71	52	138	252	1682
	21–29	942	297	313	92	251	167	193	30	16	130	111	280	624	3446
	30–39	359	109	136	36	96	67	75	11	16	31	54	200	195	1385
	40–49	178	53	121	36	48	29	50	5	6	14	41	152	88	821
	50–64	137	68	171	37	56	27	55	17	3	11	50	211	90	933
	65+	45	28	145	17	41	7	29	28	8	10	28	111	34	531
Female	< 12	71	19	59	224	19	7	22	137	113	162	70	15	24	942
	12–14	241	98	111	346	60	32	29	240	98	548	178	29	58	2068
	15–17	477	114	58	91	50	27	41	80	14	303	141	9	72	1477
	18–20	436	108	76	18	32	12	32	12	10	74	70	14	67	961
	21–29	1180	207	132	30	61	21	65	16	12	100	104	30	157	2115
	30–39	1009	165	121	27	43	9	74	14	31	48	81	36	107	1765
	40–49	517	102	93	23	31	7	51	10	8	22	46	24	66	1000
	50–64	488	127	214	27	57	13	79	23	17	26	69	35	64	1239
	65+	173	64	215	13	44	0	39	42	6	12	41	11	55	715
Total		7160	2171	2835	3704	1619	1230	1157	2018	1914	2400	1286	2441	3166	33101

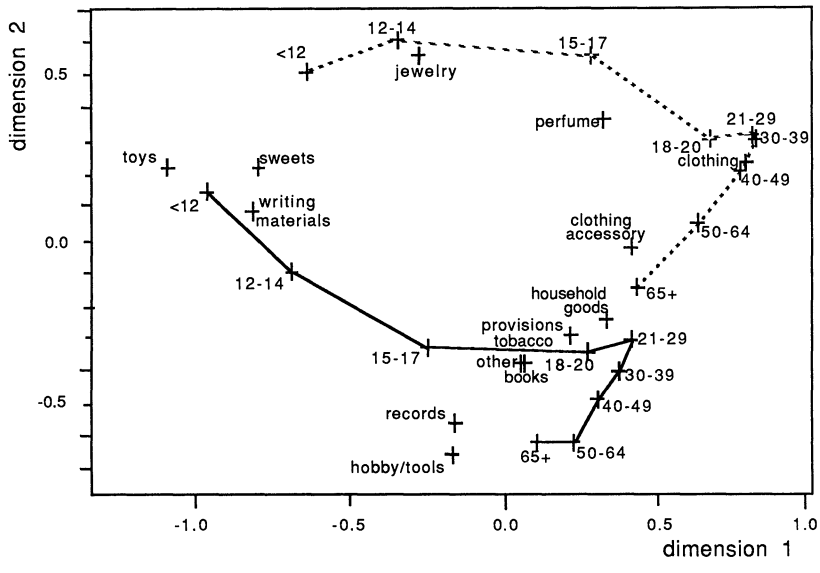


Fig. 1. CA for Table 1: ——, for males; – – –, for females

in Fig. 1, keeping in mind that some information may be hidden in higher dimensions. For both sexes, the age categories are ordered from younger than 12 on the left-hand side to 21–29 on the right-hand side, and subsequently bend downwards for older ages. Using equation (1), we can deduce that the profiles at older ages are distinct from the profiles at young age. We notice larger distances between the younger age groups than between older age groups, indicating that the profiles change more drastically between younger age groups. On the second dimension, males (full curve) are clearly separate from the females for identical age groups, indicating that the stealing behaviour of males and females, as reflected by their profiles, differs considerably. Using equation (5), we can relate row and column points, thus interpreting which way the individual groups deviate from the average (the proportions of the total group, that is placed at the origin). Young children steal toys, sweets and writing materials more than average. Young girls steal jewelry more than average, young adult girls perfume, adult women clothing. Young adult males and old females steal clothing accessories, household goods, provisions and tobacco more than average. Old males and males of 15–17 steal records, hobby accessories/tools, books and other goods more than average. In the third dimension a distinction can be found between young boys and women starting from 18 on the one hand, stealing toys more than average, and young girls and males starting from 15 on the other hand, stealing jewelry more than average. The fourth dimension shows that people older than 50 steal food and tobacco more than average. We come back to this analysis of the data in Section 3.2.

2.2. *Generalization of Correspondence Analysis*

We are not always most interested in a decomposition of residuals from independence, but prefer to study the residuals from less restricted models. To be able to do this, while retaining CA properties, we can make use of a generalization of CA, proposed by Escofier (1983, 1984). It is easiest to discuss this generalization using its

reconstitution formula

$$P = Q + S_r R \Lambda C' S_c \tag{9}$$

where Q is an arbitrary matrix, having the same order as P, S_r and S_c are two diagonal matrices of arbitrary weights for row and column points respectively under the restriction that all weights are larger than zero and tr S_r and tr S_c equal unity. The matrices R, C and Λ in equation (9) can be derived using equations (3) and (4), when we replace E by Q, D_r by S_r and D_c by S_c.

A comparison of equation (9) with equation (6) reveals that CA is generalized in two ways. Firstly, it is possible to decompose the departure from other matrices than the independence matrix E; secondly, the weights S_r and S_c of the row and column points are not necessarily defined by the margins of P. We shall not discuss any properties of this generalization here, but refer the reader to the literature mentioned in the sections that follow.

Escoufier (1985) also discusses other alternatives of generalizing CA, but we shall only use the particular generalization given in equation (9), because it is convenient for the decomposition of residuals from various log-linear models.

2.3. *Log-linear Analysis*

LLA is a well-known method for studying structural relationships between variables in a contingency table. Here we introduce only some notation for the two-variable case. With a two-way matrix, the unrestricted log-linear model has the form

$$\log \pi_{ij} = u + u_{1(i)} + u_{2(j)} + u_{12(ij)} \tag{10}$$

where π_{ij} denotes the probability for cell (i, j) and the u parameters have to be constrained to identify the model. The generalization to notation for higher way tables is straightforward. We follow the convention of denoting log-linear models by placing the variables that constitute the highest fitted margins in square brackets, e.g. [1] [2] for the independence model.

The interpretation of individual u parameters is sometimes difficult, especially if their number is very large, which may be the case when the number of categories is large and when there are (higher order) interactions that cannot be neglected. It is sometimes helpful to attempt an interpretation by representing the parameters graphically or by using odds ratios. An alternative is to place restrictions on the interaction parameters (see Section 3.1).

3. Correspondence Analysis and Log-linear Analysis as Complementary Tools of Analysis

In this section we show how CA and LLA can be related. This leads to a better understanding of CA. We also indicate how this can lead to a combined use of both methods.

3.1. *Relationships between Correspondence Analysis and Models with Restrictions on the Interactions*

One way to overcome the problem of interpreting a large number of log-linear parameters is to restrict the interaction parameters in some form or another, for

example, to have a product form. This is done in the RC association model (Andersen, 1980; Goodman, 1979, 1981a, b, 1985, 1986). When the number of categories is large, the number of parameters to be interpreted can be considerably reduced in this way.

It is well known that CA and the RC association model are related in the following way. The CA representation in k dimensions can be written in an adapted version of equation (6) as

$$m_{ij} = p_{i+}p_{+j}\left(1 + \sum_{\alpha=1}^{k} \lambda_\alpha r_{i\alpha} c_{j\alpha}\right) \tag{11}$$

where m_{ij} is the reconstituted proportion for cell (i, j). Escoufier (1982) noted that if

$$x = \sum_{\alpha=1}^{k} \lambda_\alpha r_{i\alpha} c_{j\alpha}$$

is small compared with unity (so that $\log(1 + x) \approx x$) we can rewrite equation (11) as

$$\log m_{ij} \approx u + u_{1(i)} + u_{2(j)} + \sum_{\alpha=1}^{k} \lambda_\alpha r_{i\alpha} c_{j\alpha}. \tag{12}$$

where $u = 0, u_{1(i)} = \log p_{i+}, u_{2(j)} = \log p_{+j}$. Since $r_{i\alpha}$ and $c_{j\alpha}$ are normalized such that

$$\sum_i p_{i+} r_{i\alpha}^2 = 1 = \sum_j p_{+j} c_{j\alpha}^2,$$

Escoufier's condition roughly reduces to the situation that the departure from independence is not too large (compare equation (8)).

Another result is that, if $k = 1$ and the proportions come from a discretized bivariate normal distribution (or a distribution that is bivariate normal after a suitable transformation of the rows and columns), then no matter how large the singular value λ_1 is, equation (12) is closely related to the RC association model

$$\log \pi_{ij} = u + u_{1(i)} + u_{2(j)} + \phi u_{1(i)}^* u_{2(j)}^* \tag{13}$$

(Goodman, 1981a). In this expression $\phi \approx \lambda_1$, $r_{i1} \approx u_{1(i)}^*$ and $r_{j1} \approx u_{2(j)}^*$, where $u_{1(i)}^*$ and $u_{2(j)}^*$ are normalized in the same way as $r_{i\alpha}$ and $c_{j\alpha}$. The difference between equation (12) and equation (13) is that equation (12) is an approximation involving k factors, whereas equation (13) has only one factor. Compared with the saturated model (10), model (13) can be interpreted as a model allowing for interaction, where $u_{12(ij)} = \phi u_{1(i)}^* u_{2(j)}^*$.

A multifactor RC association model was introduced by Goodman. For more details and related material we refer to Goodman (1985, 1986). For log-linear models with restrictive interactions in higher dimension tables we refer to Clogg (1982), de Leeuw (1983), Agresti (1984) and Pannekoek (1985).

From this we can conclude that in CA the interaction is decomposed approximately in a log-multiplicative way: the graphical CA displays show approximations of log-multiplicative parameters. This also shows that CA can give indications for restricting the (unrestricted) log-linear interaction parameters. We do not pursue this issue any further but refer for examples to Worsley (1987), van der Heijden (1987) and van der Heijden and Worsley (1988).

3.2. *Correspondence Analysis as a Technique for Decomposing Residuals from Log-linear Models of Higher Way Tables*

In Section 2.1 it was emphasized that simple CA decomposes residuals from the independence model. Recently it has been shown that CA is also related to LLA in the case of higher way contingency tables (van der Heijden and de Leeuw, 1985). We summarize these results in this section.

For a two-way table with proportions p_{ij} it is clear from Section 2.1 that CA can be interpreted in terms of the difference between models [12] and [1][2]. Results are more interesting for the analysis of higher way tables by means of so-called 'multiple tables', i.e. two-way tables constructed by concatenating 'slices' of the higher way tables, thus creating 'interactive' variables. First consider the three-way table P with proportions p_{ijk}. To do CA, the two-way multiple table $P^{i[jk]}$ is constructed by coding the categories of variables 2 and 3 interactively. We denote the elements of $P^{i[jk]}$ and $E^{i[jk]}$ as $p_{i[jk]}$ and $e_{i[jk]}$. Values $e_{i[jk]}$ are equal to maximum likelihood estimates of expected proportions under model [1][23]: $\hat{\pi}_{ijk} = p_{i++}p_{+jk} = p_{i[++]}p_{+[jk]} = e_{i[jk]}$. So, when the three-way matrix P is flattened to the multiple table $P^{i[jk]}$, CA can be interpreted in terms of the difference between the log-linear models [123] and [1][23]. When multiple tables $P^{j[ik]}$ or $P^{k[ij]}$ would have been constructed, values in E would follow model [2][13] or [3][12] respectively.

Generally, when we have a higher way table P, we can construct a multiple table $P^{[a][b]}$, where a and b index the elements of two groups of variables A and B coded interactively. A and B together constitute the higher way table, where $A \cap B = \varnothing$. In this case CA can be interpreted in terms of the difference between models $[AB]$ and $[A][B]$: elements e_{ab} of E are equal to $e_{ab} = p_{a+}p_{+b} = \hat{\pi}_{ab}$.

Thus in the analysis of multiple tables, ordinary CA can be interpreted as a tool to analyse residuals from rather restrictive log-linear models. However, in many applications these models are not the models that we are most interested in. More interesting models allow only conditional independence, or an absence of higher order interaction. The generalization of CA discussed in Section 2.2 makes it possible to analyse the difference between such less restricted log-linear models and the saturated model. For this estimated proportions $\hat{\pi}_{ijk}$ are taken as the elements of Q (see equation (9)), and the margins of multiple tables of P (being equal to those for Q) are taken as the diagonal elements of S_r and S_c. In this application generalized CA solutions can be obtained using ordinary CA programs, namely by feeding the program with the matrix $P - Q + E$, i.e. a matrix with raw residuals $P - Q$ to which an independence matrix E is added. This is easily seen by comparing equation (6) with equation (9).

The use of less restricted log-linear models has interesting geometrical implications (see also Escofier (1983) and van der Heijden (1987)). Consider the analysis of multiple table $P^{i[jk]}$. We denote the column scores yielded by its analysis as the set of scores $c_{jk\alpha}$. Now, in ordinary CA residuals from [1][23] are decomposed, and the scores are normalized so that

$$\sum_j \sum_k p_{+[jk]} c_{jk\alpha} = 0,$$

i.e. for each dimension the column scores have weighted average zero. This was due to the fact that in the residual matrix $p_{i[++]} - e_{i[++]} = 0$ (see Section 2.2). When we decompose residuals from [12][23], a restricted CA solution is found, namely one in

which

$$\sum_k p_{+[jk]} c_{jk\alpha} = 0,$$

i.e. for each category j the weighted average is zero. This is because, owing to fitting marginal frequencies p_{ij+}, in the residuals $p_{i[j+]} - e_{i[j+]} = 0$. In fact, the weighted averages of the J clouds of K points are all translated to the origin. This immediately shows that the chi-squared distances within each cloud j remain the same, and the chi-squared distances between points in different clouds are changed in a way that is easy to understand. The singular value decomposition performed takes care of the optimal projection of the new full dimensional configuration to a lower dimensional space.

The principle of translations of subclouds to the origin easily generalizes to less restricted hierarchical log-linear models or to the case of more than three variables. To give an example, in decomposing the residuals from [12][13][23] using the multiple table described earlier, not only

$$\sum_k p_{+[jk]} c_{jk\alpha} = 0,$$

but also

$$\sum_j p_{+[jk]} c_{jk\alpha} = 0.$$

We conclude the following. Firstly, we know more about CA now. For higher way tables we conclude that CA of multiple tables results in a solution showing the difference between two log-linear models. The results also show that CA does not cover all forms of dependence in the data matrix. For example, for the multiple table $P^{i[jk]}$, the first-order interaction between variables 2 and 3 is not displayed.

This leads to our second point, namely that we can use these results to come to a combined approach using both CA and LLA. For example, when the most suitable multiple table must be chosen to analyse a higher way table, these findings give us guidelines. Let us consider again the three-way table. Firstly, when some first-order interaction does not seem particularly interesting, the corresponding two variables should be coded interactively. Secondly, when we are especially interested in the relation of some variable with the other two, this variable should not be coded interactively with another variable. Ordinary CA shows the first-order interactions of this variable with the other two, and the second-order interaction. The generalization of CA can be used to suppress other first-order interactions.

Another implication is that in some cases it is possible to use CA and LLA in a complementary way. In van der Heijden and de Leeuw (1985) it was proposed to use the two techniques to find answers to different questions. In the first step LLA can answer the question of which variables are related by detecting the important interactions; in the second step CA can be used in two ways. First, when a model in which we are particularly interested does not fit well, CA can be used to explore the residuals. Secondly, by choosing a model that fits badly, CA can be used in place of an interpretation of interaction parameters. We think that in both contexts CA will be specifically useful when the matrix under study is large: in this case a graphical representation can simplify the interpretation.

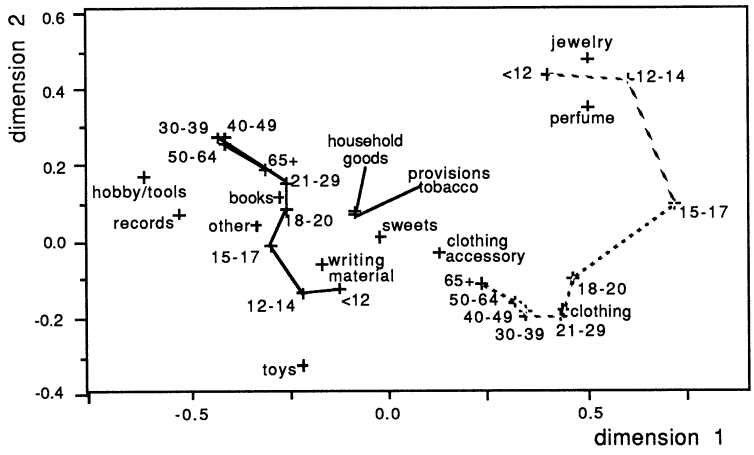


Fig. 2. Generalized CA decomposing residuals from [AS][AG] (for each age the weighted average of the sexes is zero): ——, for males; – – –, for females

3.2.1. *Example*

When we discussed ordinary CA of the shop-lifting data in Section 2.1, we neglected the fact that the table analysed was actually a three-way table. The results in this section show us that the analysis in Section 2.1 can be understood as a decomposition of residuals from the model [AS][G], because age A and sex S were coded interactively. The chi-square value for this model was $X^2 = 19\,940$ with 204 degrees of freedom. When we take another look at Fig. 1, we see that, roughly, the first dimension is dominated by the first-order interaction between age and goods, whereas the second dimension displays the first-order interaction between sex and goods.

Now consider the case where we are interested in the log-linear model [AS][AG], i.e. sex and goods are independent for a given age. This model fits badly: the chi-square value is $X^2 = 5252$ with 108 degrees of freedom. To understand why this model does not fit, we can study the residuals of the model with CA. We can also motivate this CA from the idea that we want a graphical representation of all the interactions in which both sex and goods are involved. The corresponding solution has first three singular values (with proportion of explained inertia) 0.37 (0.73), 0.19 (0.20) and 0.09 (0.04), so a two-dimensional solution displays 93% of the information (Fig. 2). Now the first dimension displays mainly the first-order interaction between sex and goods, whereas the second dimension displays mainly the second-order interaction. For each age group the weighted average over the sexes is zero. So, compared with the solution in Fig. 1, in Fig. 2 the weighted averages for each age group are translated to be at the origin, thus eliminating the first-order interaction between age and goods from the solution. In the full dimensional solution the chi-squared distances between a male and female point for each age group, for example the distance between females younger than 12 and males younger than 12, has not been changed by this translation. We can therefore interpret Fig. 2 in the same way as we interpreted Fig. 1, keeping in mind that weighted averages for each age group are translated to the origin and that therefore the origin should now be interpreted as the average age–goods relation. For example, adult females steal more clothing than average for adults, and consequently adult males steal less than average for adults.

The first-order interaction is that jewelry, perfume, clothing and clothing accessories are stolen more than average by females, whereas the other goods are stolen more than average by males, sweets being in between. The second-order interaction displays itself by the fact that the age lines are roughly reflected through the origin. This shows that jewelry and perfume are stolen more than average by young females, while they are stolen less than average by young males; clothing is stolen more than average by adult females while it is stolen less than average by adult males, etc.

3.3. *Correspondence Analysis of Incomplete Tables*

Now we shall use the generalization of CA to study the departure from independence in incomplete tables (see van der Heijden (1986, 1987) and de Leeuw and van der Heijden (1988)). The log-linear models used in this context are quasi-independence models. We shall consider incomplete two-way tables only, although our results generalize to independence models for the higher way tables discussed in Section 3.2 such as [1][23], [2][13] and [3][12] or, more generally, models $[A][B]$ for groups of variables A and B. Quasi-independence models are useful in several situations. Firstly, some entries in the table might be missing, for example they are available for some regions (years, groups), but not collected for others although they could have been. Secondly, they can be structurally zero, i.e. the cells pertain to events that cannot logically occur, such as the diagonal cells in import–export tables. Thirdly, we may decide that we want to apply the independence model to some cells but not to others. The remaining cells are observed, but we do not want to model them. As an example, in a social mobility table, it might be that we are only interested in the cells that reflect changes.

The mathematical form of the model is very simple. Compared with the saturated model (10), this model defines independence for only a limited number of cells, collected in S, while the values for the cells (i, j) not in S are not restricted:

$$\begin{aligned} u_{12(ij)} &= 0 \text{ when } (i, j) \text{ is in } S, \\ \pi_{ij} &= p_{ij} \text{ when } (i, j) \text{ is not in } S. \end{aligned} \tag{14}$$

When the departure from model (14) is significant, it makes sense to study the residuals from model (14) using the generalization of CA.

However, starting from the generalization of CA as defined in equation (9), we still have several possibilities. It will be clear that for Q we take the matrix of quasi-independent proportions. For S_r and S_c, the diagonal matrices of weights, we first reformulate π_{ij} in model (14) as $\pi_{ij} = a_i b_j$ if cell (i, j) is in S. Irrespective of the normalization chosen (necessary to make the model identified), if we take the resulting a_i and b_j as diagonal elements of S_r and S_c respectively, then the sum of squares of the singular values becomes

$$\mathrm{tr}\Lambda^2 = \sum_i \sum_j \frac{(p_{ij} - \hat{\pi}_{ij})^2}{\hat{\pi}_{ij}} = \frac{X^2}{n}. \tag{15}$$

This property also holds for ordinary CA (compare equations (7) and (8)) but was lost for the generalization to its most general form. Here each dimension can again be interpreted as a decomposition of part of the X^2 statistic, which is the criterion used to decide whether the residuals contain interesting information.

Further insight into this choice of S_r and S_c can be derived from the fact that it leads to a CA procedure for dealing with missing values. This procedure is proposed by Nora (1975). See also Greenacre (1984), pp. 236–244. In this procedure values are imputed for the missing cells. These imputations have the property that, for a specific dimensionality, say h, these values are perfectly represented by the CA solution. To find these imputations, first choose the dimensionality h. Then so-called 'reconstitution of order h' is the iterative process that starts with the ordinary CA solution for the matrix ($m = 0$) and subsequently finds imputations using iteratively

$$p_{ij}^{(m+1)} = p_{i+}^{(m)} p_{+j}^{(m)} \left(1 + \sum_{\alpha=1}^{h} \lambda_\alpha^{(m)} r_{i\alpha}^{(m)} c_{j\alpha}^{(m)}\right) \tag{16}$$

for the cells (i,j) not in S (compare formula (6)). For (i,j) in S we simply set $p_{ij}^{(m)} = p_{ij}$ for all m. The solution will, in general, depend on the choice of the dimensionality h. Iterative reconstitution of order zero simplifies to

$$p_{ij}^{(m+1)} = p_{i+}^{(m)} p_{+j}^{(m)} \tag{17}$$

for all (i,j) not in S. This is exactly identical with our form of CA of incomplete tables. Iterating equation (17) converges to 'independent' values $a_i b_j$, computed for the cells not in S. For these cells, fitting the independence model on $p_{ij}^{(m)}$ gives residuals equal to zero.

We deduce from this that CA of incomplete tables can be computed using classical CA programs. We just have to fill in values computed with equation (17). This makes this usage of the generalization easily interpretable: it reduces to ordinary CA of the matrix $P^{(m)}$. An example will be given in Section 4.2.

We conclude that by including the quasi-independence model in the CA approach we have broadened the scope of data analysis for CA. Furthermore a better understanding of the CA approach to reconstitution of order zero is obtained.

3.4. *Conclusions*

In this section we have presented some useful relations between CA and LLA. Firstly, the scores obtained with CA are often approximately the same as parameters from the RC association model. Secondly, we have seen that CA of higher way tables by means of multiple tables can be interpreted in terms of log-linear models, and we have indicated that CA can replace the interpretation of interaction parameters or an ordinary study of residuals. Thirdly, the CA procedure for treating missing cells is closely related to the quasi-independence model.

In the following sections we shall illustrate a combined approach of data analysis for square tables, using both LLA and CA. Furthermore, some results will be presented regarding the analysis of asymmetry using CA.

4. Example for Square Matrix, with some New Results

Usually, square matrices are analysed by means of log-linear models. Here we discuss some of the more important models and see what CA has to offer for the analysis of residuals from these models. The models are the independence model, the quasi-independence model, the symmetry model and the quasi-symmetry model (see Bishop *et al.* (1975)). For all of these, the complementary CAs have special properties,

either in the computations, in the solutions or in both (see also van der Heijden (1987)).

4.1. *Decomposing Residuals from Independence*

A starting point in the analysis of square matrices is often the comparison of observed proportions and proportions expected under the assumption that the independence model holds. Compared with the saturated model (10), the independence model has the restriction

$$u_{12(ij)} = 0 \tag{18}$$

and estimates of expected proportions are computed as $\hat{\pi}_{ij} = p_{i+}p_{+j}$. The difference between p_{ij} and $\hat{\pi}_{ij}$ can be tested with $(I - 1)(J - 1)$ degrees of freedom. We now study the residuals from this model using CA.

4.1.1. *Example*

The example is taken from Harshman *et al.* (1982) and deals with car changing. In 1979 recent new car buyers were surveyed to collect information on their old and new car. The cars are categorized in 16 segments, yielding a transition matrix of 16 × 16 (Table 2). The abbreviation of the segments is explained in Table 3. For more details see Harshman *et al.* (1982). The row and column margins of the same cars are sometimes very different, indicating a loss or gain in market share.

Testing for independence, we find $X^2 = 1\,357\,827$, with 225 degrees or freedom. This exorbitant value is due to the extreme sample size. The CA solution has as first five singular values (with proportions of chi square) 0.542 (0.366), 0.390 (0.189), 0.332 (0.137), 0.273 (0.093) and 0.232 (0.067). A plot of the first two dimensions is shown in Fig. 3, where the cars disposed of have the larger lettering and the new cars have the smaller lettering. We can interpret distances between rows and columns using equation (1) and distances between rows and columns using equation (5) (see Section 2.2). A striking feature is that the row and column points of LUXI and LUXD are far apart from the other points. Row LUXD is far apart from all other row points because the LUXD row profile is very dissimilar from the other row profiles. Since the column for LUXD is the only column near row LUXD, this must be due to the diagonal cell for LUXD, for which the observed value is much larger than expected. The same can be said for LUXI. The tendency to stick to the type of luxury car which one already has overshadows information on the relationships between other cars. However, on the second dimension a line of points can be seen from standard cars (STDL and STDM), via domestic midsize cars, domestic compact cars, domestic small specialities and domestic subcompact cars to the import cars. As an example, when we consider the row point for STDL, equation (1) tells us that row profiles of midsize cars are very similar, and this similarity becomes less when we go to (sub)compact and import cars; when we relate row STDL to the columns using equation (5), we see that for standard, midsize and compact varieties the observed values with STDL are much larger than expected, and this tendency disappears when we go to the import cars. Similar reasoning for the other points shows that standard cars are interchanged more than expected with domestic midsize and compact cars, which in turn are interchanged more than expected with domestic subcompact cars and small specialities. These are interchanged more than expected with import cars.

TABLE 2
1979 car changing data†

	SUBD	*SUBC*	*SUBI*	*SMAD*	*SMAC*	*SMAI*	*COML*	*COMM*	*COMI*	*MIDD*	*MIDI*	*MIDS*	*STDL*	*STDM*	*LUXD*	*LUXI*	*Total*
SUBD	23272	1487	10501	18994	49	2319	12349	4061	545	12622	481	16329	4253	2370	949	127	110708
SUBC	3254	1114	3014	2656	23	551	959	894	223	1672	223	2012	926	540	246	37	18344
SUBI	11344	1214	25986	9803	47	5400	3262	1353	2257	5195	1307	8347	2308	1611	1071	288	80793
SMAD	11740	1192	11149	38434	69	4880	6047	2335	931	8503	1177	23898	3238	4422	4114	410	122539
SMAC	47	6	0	117	4	0	0	49	0	110	0	10	0	0	0	0	343
SMAI	1772	217	3622	3453	16	5249	1113	313	738	1631	1070	4937	338	901	1310	459	27139
COML	18441	1866	12154	15237	65	1626	27137	6182	835	20909	566	15342	9728	3610	910	170	134778
COMM	10359	693	5841	6368	40	610	6223	7469	564	9620	453	9731	3601	5498	764	85	67919
COMI	2613	481	6981	1853	10	1023	1305	632	1536	2738	1005	990	454	991	543	127	23282
MIDD	33012	2323	22029	29623	110	4193	20997	12155	2533	53002	2140	61350	28006	33913	9808	706	315900
MIDI	1293	114	2844	1242	5	772	1507	452	565	3820	3059	2357	589	1052	871	595	21137
MIDS	12981	981	8271	18908	97	3444	3693	1748	935	11551	1314	56025	10959	18688	12541	578	162714
STDL	27816	1890	12980	15993	34	1323	18928	5836	1182	28324	938	37380	67964	28881	6585	300	256354
STDM	17293	1291	11243	11457	41	1862	7731	6178	1288	20942	1048	30189	15318	81808	21974	548	230211
LUXD	3733	430	4647	5913	6	622	1652	1044	476	3068	829	8571	2964	9187	63509	1585	108236
LUXI	105	40	997	603	0	341	75	55	176	151	589	758	158	756	1234	3124	9162
Total	179075	15339	142259	180654	616	34215	112978	50756	14784	183858	16199	278226	150804	194228	126429	9139	1689559

†Rows denote cars disposed of, columns denote new cars. The abbreviations are defined in Table 3.

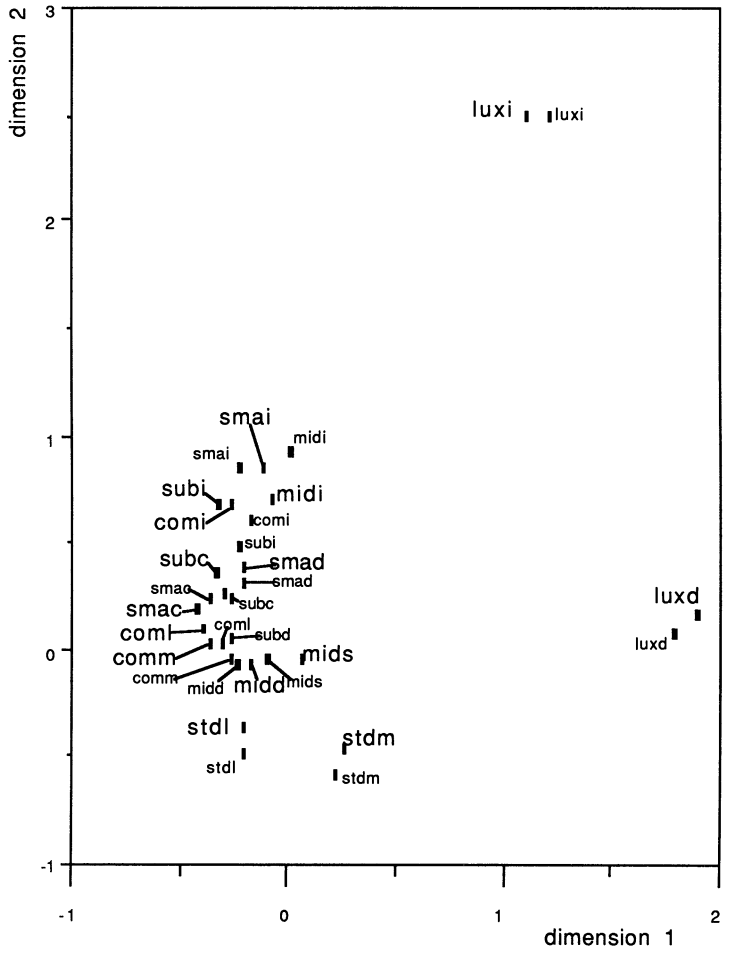


Fig. 3. Ordinary CA of car changing data: the larger lettering denotes cars disposed of; the smaller lettering denotes new cars (the abbreviations are defined in Table 3)

4.2. *Decomposing Residuals from Quasi-independence*

Another independence model with which transition matrices are often analysed is the quasi-independence model (see Section 3.3). In the context of square matrices this model can be useful when substantive interest is for some reason restricted to the off-diagonal cells. Then $\hat{\pi}_{ii} = p_{ii}$, and the residuals are zero for these cells. Therefore diagonal cells do not contribute to the total inertia $\Sigma_\alpha \lambda_\alpha^2$ (see equation (15)). This is an elegant way to remove the usual dominant influence of the diagonal elements in classical CA.

4.2.1. *Example*

In Section 4.1 we found that the luxury cars dominated the solution because of their high diagonal proportions. Now the diagonal proportions do not influence

TABLE 3
Definitions of the abbreviations in Table 2†

SUBD	Subcompact/domestic
SUBC	Subcompact/captive imports
SUBI	Subcompact/imports
SMAD	Small speciality/domestic
SMAC	Small speciality/captive imports
SMAI	Small speciality/imports
COML	Low price compacts
COMM	Medium price compacts
COMI	Import compact
MIDD	Midsize domestic
MIDI	Midsize imports
MIDS	Midsize speciality
STDL	Low price standard
STDM	Medium price standard
LUXD	Luxury domestic
LUXI	Luxury import

†A more elaborate description can be found in Harshman *et al.* (1982).

the solution any longer. The quasi-independence model fits much better than the independence model: $X^2 = 235\,914$ compared with $X^2 = 1\,357\,827$.

The first two dimensions of CA are shown in Fig. 4. The larger lettering denotes the cars disposed of. The first five singular values are 0.250 (0.361), 0.206 (0.245), 0.139 (0.112), 0.126 (0.091) and 0.090 (0.047). We have drawn clusters around the corresponding car types, to make the main structure easily discernible. Note that Fig. 4 may only be interpreted in terms of changes of car type.

Roughly, the first dimension distinguishes luxury/standard/midsize cars from (sub)compact/small specialities. The second dimension discriminates the import cars from the domestic cars. When we interpret Table 2 starting from row STDL, we see that MIDD, COMM and COML have similar row profiles, and the similarity becomes less when we go farther away from row STDL. Relating row STDL to the columns, we know that we should only interpret the relation to other car types, since the influence of the diagonal cells is eliminated. Doing this, we see that the observed values with STDM, MIDS, MIDD, COMM, COML and SUBD are larger than expected. On the whole, it seems that, when people buy a new car of a different type, they remain more than average in the same cluster, or go to a neighbouring cluster.

4.3. *Decomposing Residuals from Symmetry*

Recently much attention has been given to asymmetry in square matrices. An important line of contributions restricts attention to the skew symmetric part of the asymmetric matrix. This skew symmetric part can be found as

$$P = M + N \tag{19}$$

where P is the asymmetric square matrix of observed proportions, M is a symmetric matrix with values

$$m_{ij} = m_{ji} = (p_{ij} + p_{ji})/2 \tag{20}$$

and N has elements $n_{ij} = -n_{ji} = (p_{ij} - p_{ji})/2$, and is a skew symmetric matrix. A justification for decomposition (19) is that values m_{ij} are equal to maximum likelihood

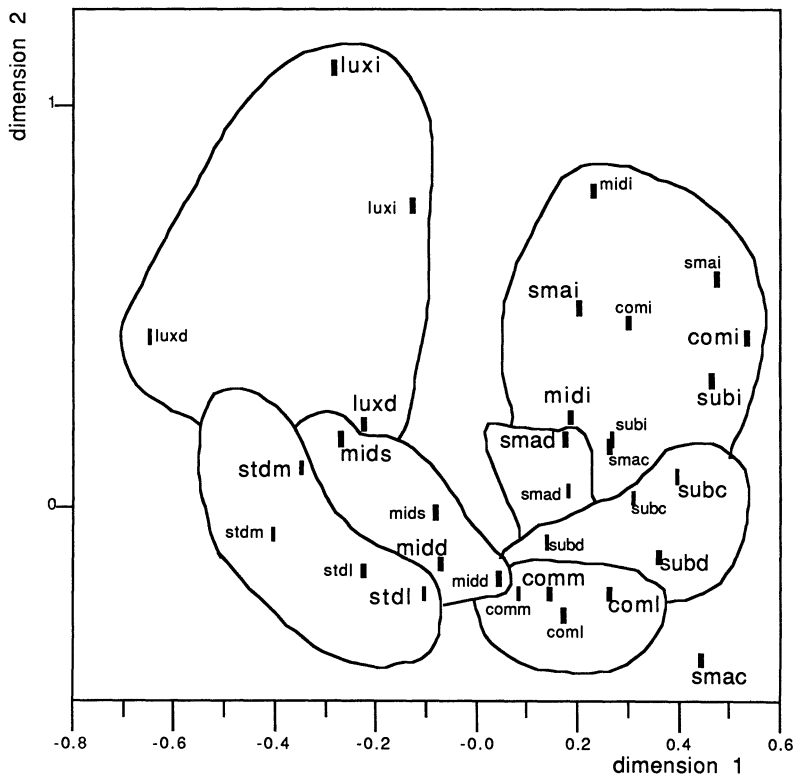


Fig. 4. CA of an incomplete table (diagonal cells are ignored by decomposing residuals from quasi-independence): the larger lettering denotes cars disposed of; the smaller lettering denotes new cars

estimates for the symmetry model

$$\log \pi_{ij} = u + u_{1(i)} + u_{2(j)} + u_{12(ij)} \tag{21}$$

where $u_{1(i)} = u_{2(i)}$ and $u_{12(ij)} = u_{12(ji)}$ (compare with the saturation model (10)). So N is a matrix of residuals. The difference between p_{ij} and π_{ij} can be tested with $I(I-1)/2$ degrees of freedom.

The attention to skew symmetry is due to Gower (1977) and Constantine and Gower (1978), who show that the SVD of N has the form

$$N = U\Lambda V' = U\Lambda JU', \tag{22}$$

where J is a block diagonal matrix with blocks

$$\begin{matrix} 0 & 1 \\ -1 & 0 \end{matrix}$$

and the singular values are ordered in pairs $\lambda_1, \lambda_1, \lambda_2, \lambda_2, \ldots$. If the number of rows I of N is odd, $\lambda_I = 0$ and $J_I = 0$. Thus, when we make a plot of singular vectors U and UJ for paired dimensions, the configurations of row points is the same as the configuration of column points, up to a rotation of 90°. A parsimonious plot can be made by showing only the row points and giving the direction of the rotation.

When rotations from row to column points are made clockwise, Gower (1977) and Constantine and Gower (1978) show that in the plot of row points elements n_{ij} and

n_{ji} are approximated by twice the area of the triangle formed by row i, row j and the origin (this triangle will further be denoted OIJ). Element n_{ij} is positive if the rotation from i to j is counter-clockwise and n_{ij} is negative when the rotation is clockwise. (These relations hold approximately when a paired dimension gives only an approximation of N.) Points on a line through the origin have no asymmetric relation because area OIJ is zero. Points near the origin have small asymmetries. Points on different sides from the line OI have opposed signs in fitted skew symmetries. Finally, if all points are at one side of the paired dimension (compared with the origin), circular triads are absent in the data.

The decomposition of skew symmetry has the following relation with CA. Decomposition (22) is a special form of the generalization of CA (compare equation (9)): when we take $Q = M$, then $N = P - Q$, and $S_r = S_c = I$. However, it is standard to weight individual rows and columns in the CA approach. To bring down the influence of the margins of the table, weighting of raw residuals is also necessary in this context. To keep the properties of the SVD of skew symmetry, the weights for the rows have to be equal to those for the columns: $S_r = S_c = S$. This can be done by taking the margins of M as the elements of S: for these margins $\pi_{i+} = (p_{i+} + p_{+i})/2 = \pi_{+i}$. Matrix $S^{-1/2}NS^{-1/2}$ is then skew symmetric, and therefore the SVD of this matrix has the properties required. The CA decomposition becomes

$$P = Q + SR\Lambda JR'S. \tag{23}$$

However, since the margins of P are not equal to those of Q, we may find the situation that all points are on one side of the space, showing that circular triads are absent.

4.3.1. *Example*

Testing for symmetry in the car data gives $X^2 = 213\,837$, with 120 degrees of freedom. It follows that a study of the residuals from symmetry might reveal interesting information. The first two dimensions have singular values 0.217 and display 84% of the total inertia. The restrictions for model (21) show that in principle two types of asymmetry can be found, namely asymmetry due to the margins and asymmetry not due to the margins. Fig. 5 shows mainly asymmetry due to the margins. All car types are placed at one side of the origin, showing that (in the first two dimensions) circular triads are absent. This implies that it is possible to reorder the rows and columns of N in such a way that the upper triangular matrix contains only positive elements and the lower triangular matrix only negative elements. The order to be taken is the order going clockwise from SMAC to MIDI in Fig. 5.

Counter-clockwise from MIDI to SMAC we find cars that lost a relatively large amount of their market share, compared with cars that won (see the margins of Table 2). So people who have a MIDI model more often buy another car than vice versa. People who have a SUBD model more often buy a small speciality, a SUBI or a MIDS, than vice versa. They less often buy a luxury car or a standard car (STDM and STDL) than vice versa, etc. People who have another car than the SMAC more often buy an SMAC than vice versa.

4.4. *Decomposing Residuals from Quasi-symmetry*

In the residuals from the symmetry model the asymmetry is partly due to unequal margins of P and partly due to aspects other than unequal margins. The quasi-symmetry model (Caussinus, 1965; Bishop *et al.*, 1975) is a model with which

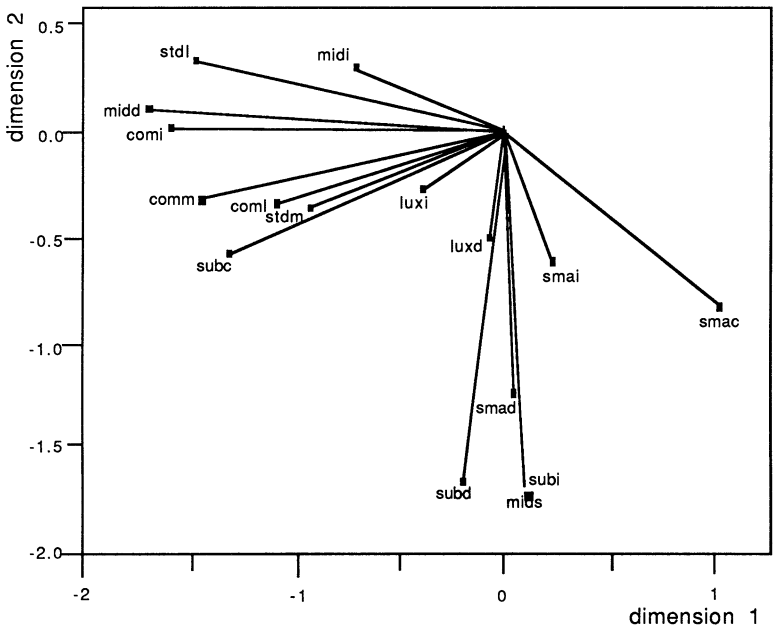


Fig. 5. Generalized CA decomposing residuals from symmetry: counter-clockwise, the (approximated) residuals are positive

it is possible to study whether the latter asymmetry, not due to different margins, can be neglected. This model can be formulated as

$$\log \pi_{ij} = u + u_{1(i)} + u_{2(j)} + u_{12(ij)}, \tag{24}$$

where $u_{12(ij)} = u_{12(ji)}$ (compare with the saturated model (10)). The model has $(I - 1)(I - 2)/2$ degrees of freedom. Compared with the symmetry model (21) the restriction $u_{1(i)} = u_{2(i)}$ is dropped. If $p_{i+} = p_{+i}$, the quasi-symmetry model simplifies to the symmetry model.

It will be clear that the matrix Π, following model (24), is not necessarily symmetric. However, the residual matrix $P - M = N$ is skew symmetric. We want to study the residual matrix with CA, while also retaining the properties of an SVD of skew symmetry. As for the residuals from symmetry, this can be accomplished by taking the same set of weights for the rows and the columns, such as again taking $(p_{i+} + p_{+i})/2$ as elements of S. Now we have the same situation as in the previous section. The reconstitution formula can be written as in equation (23), with the difference that $M = Q$ is not symmetric.

4.4.1. *Example*

Here we shall discuss symmetry in the car changing data, to the extent that this asymmetry is not a result of a higher or lower market share. The quasi-symmetry model has $X^2 = 35\,048$, with 105 degrees of freedom. This departure is considerable, so it makes sense to study the residuals.

The first pair of dimensions has singular values 0.068 (0.548), the second pair 0.048 (0.270) and the third pair 0.029 (0.102). The first pair of dimensions is shown in Fig. 6. Compared with Fig. 5, asymmetry due to the margin is suppressed from the solution. Compared with Fig. 2, symmetric aspects are suppressed from the solution.

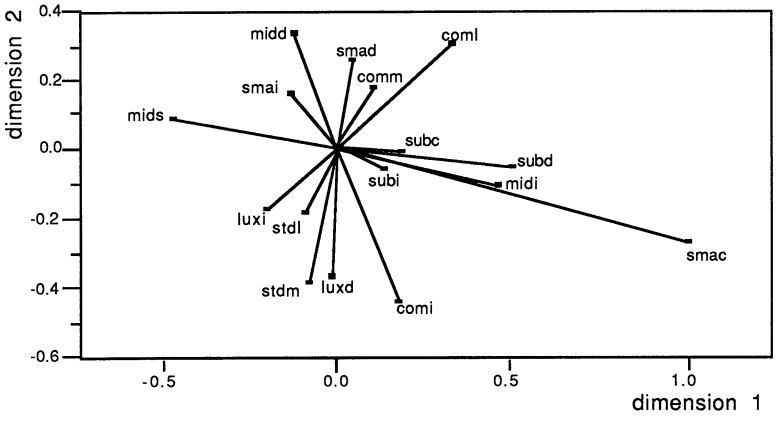


Fig. 6. Generalized CA decomposing residuals from quasi-symmetry: counter-clockwise, the (approximated) residuals are positive

Restricting attention to points with a contribution to the singular values of more than 5%, only domestic types of cars remain. Because the residuals are positive when we go counter-clockwise, we see that, corrected for marginal frequencies, SUBD is followed more often by COML and MIDD than vice versa and similarly for COML to MIDD and MIDS, MIDD to MIDS, MIDS to STDM/LUXD and STDM/LUXD to SUBD. Considering the price of these types, there is a tendency for moving up the scale (buyers going from SUBD, COML, MIDD and MIDS to STDM/LUXD), with a sudden downward movement from STDM/LUXD to SUBD. The order STDM/LUXD to SUBC might be interpreted as upward movement to the extent that it reflects that people buy a cheap second car. Note that the SMAC type of car has very large triangles with other cars, but because of its lower margins it does not have a large influence on the determination of the factorial axes (remember that $\tilde{R}'S\tilde{R} = \Lambda^2$, so that points with larger margins contribute more to the eigenvalues, which are subsequently maximized). The second pair of dimensions is determined by asymmetries between the import types of cars in the order SUBI to MIDI via SMAI. In conclusion, there are asymmetric relationships mainly in two groups: between domestic cars and between import cars.

4.5. *Conclusion*

We have shown how CA can be used for the analysis of square matrices. CA can help to find the structure in the residuals. CA can also provide a clear view of the important asymmetries in the square matrix.

Implicit in our discussion we have worked through the idea of suppressing information from a CA solution. Starting with the decomposition of residuals from independence (Fig. 3), we first suppressed the information about the diagonal by using the quasi-independence model, thus concentrating attention on the off-diagonal cells (Fig. 4). The information in the off-diagonal cells was already contained in the CA decomposing residuals from independence but was not optimally shown in the first dimensions. Going from quasi-independence to quasi-symmetry, we suppress the symmetric information from the solution, thus having an optimal view of the asymmetry (that is not due to the margins; Fig. 6). So we have the line of nested models independence—quasi-independence—quasi-symmetry. Another line is the

line symmetry—quasi-symmetry. Decomposing residuals from the symmetry model displays both asymmetry due to the margins and asymmetry not due to the margins (Fig. 5). By decomposing residuals from quasi-symmetry we have suppressed the asymmetry due to the margins from the solution, thus taking an optimal view of the asymmetry not due to the margins (Fig. 6).

5. General Discussion and Conclusions

Our approach has been to sketch a residual analysis interpretation of CA. However, it is important to use CA only in those cases where the residuals contain meaningful information: the log-linear model under study should not fit very well. Furthermore, CA makes sense only when there are many cells for which the observed values depart from the expected values. CA will not be very useful if there is only one outlying cell. Nor will it be useful if the total number of cells is small: in this case the structure in the departure might as well be studied directly (i.e. without CA).

The main advantage of a residual analysis interpretation of CA is twofold. First, we obtain a better understanding of CA by relating it to LLA. This holds mainly for the results displayed in Section 3. Secondly, we have shown how CA can be used for the analysis of residuals from various log-linear models. In this respect CA might either replace the cumbersome interpretation of a large number of parameters, or replace the interpretation of residuals. LLA is used for the fitting and testing of specific models; CA is used for constructing meaningful geometrical representations of specific aspects of the data.

In the context of CA, log-linear models can play the following role (see de Falguerolles and van der Heijden (1987)). First, some LLA models may have a central place in the analysis, and CA can then be used to supplement LLA by an analysis of the residuals from this model. Thus CA will only be useful if the departure from the model is significant and if it makes sense to try to study structure in the residuals. Secondly, the use of log-linear models makes it possible to suppress the dominant interaction (that might be uninteresting from a substantive point of view) in CA solutions. This means that we are not interested in parameters of the model but only in residuals. Thirdly, as was briefly indicated in Section 3.1, CA can be used as an exploratory tool to find a log-linear model that is parsimonious and easily interpretable.

As was discussed in Section 1, Goodman's results indicate in what way ordinary CA is related to model-based approaches, by introducing the RC association and RC correlation models (see Goodman (1985, 1986)). Relations between the generalized CA decompositions and models are not yet known, but this is a subject that is worth investigating.

Acknowledgements

The first author received a grant from the Netherlands Organization for the Advancement of Pure Research. The authors gratefully acknowledge the helpful comments of the referees.

References

Agresti, A. (1984) *Analysis of Ordinal Categorical Data.* New York: Wiley.

Andersen, E. B. (1980) *Discrete Statistical Models with Social Science Applications.* Amsterdam: North-Holland.

Benzécri, J.-P. *et al.* (1973) *Analyse des Données*, vols 1 and 2. Paris: Dunod.
—— (1980) *Pratique de l'Analyse des Données*, vols 1–3. Paris: Dunod.
—— (1986) *Pratique de l'Analyse des Données en Economie*, vol. 5. Paris: Dunod.
Bishop, Y. M. M., Fienberg, S. E. and Holland, P. W. (1975) *Discrete Multivariate Analysis*. Cambridge: Massachusetts Institute of Technology Press.
Caussinus, H. (1965) Contribution à l'analyse statistique des tableaux de corrélation. *Ann. Fac. Sci. Univ. Toul.*, **29**, 77–182.
Clogg, C. C. (1982) Some models for the analysis of association in multi-way cross-classifications having ordered categories. *J. Amer. Statist. Ass.*, **77**, 803–815.
Constantine, A. G. and Gower, J. C. (1978) Graphical representation of asymmetric matrices. *Appl. Statist.*, **27**, 297–304.
Deville, J.-C. and Malinvaud, E. (1983) Data analysis in official socio-economic statistics. *J. R. Statist. Soc.* A, **146**, 335–361.
Escofier, B. (1983) Analyse de la différence entre deux mesures sur le produit de deux mêmes ensembles. *Cah. Anal. Donn.*, **8**, 325–329.
——(1984) Analyse factorielle en référence à un modèle: application à l'analyse de tableaux d'échanges. *Rev. Statist. Appl.*, **32**, 25–36.
Escoufier, Y. (1982) L'analyse des tableaux de contingence simples et multiples. *Riv. Int. Statist.*, **40**, 53–77.
——(1985) L'analyse des correspondances: ses propriétés et ses extensions. *Bull. Int. Statist. Inst.*, **51**, 28.2-1–28.2-16.
de Falguerolles, A. and van der Heijden, P. G. M. (1987) Sur l'analyse factorielle des correspondances et quelques unes de ses variantes (with discussion). *Rev. Statist. Appl.*, **35**, 7–12, 83–94.
Gifi, A. (1981) *Non-linear Multivariate Analysis*. Department of Data Theory, University of Leiden.
Gilula, Z. and Haberman, S. J. (1986) Canonical analysis of contingency tables by maximum likelihood. *J. Amer. Statist. Ass.*, **81**, 780–788.
Goodman, L. A. (1979) Simple models for the analysis of association in cross-classifications having ordered categories. *J. Amer. Statist. Ass.*, **74**, 537–552.
——(1981a) Association models and canonical correlation in the analysis of cross-classifications having ordered categories. *J. Amer. Statist. Ass.*, **76**, 320–334.
——(1981b) Association models and the bivariate normal for contingency tables with ordered categories. *Biometrika*, **68**, 347–355.
——(1985) The analysis of cross-classified data having ordered and/or unordered categories: association models, correlation models, and asymmetry models for contingency tables with or without missing entries. *Ann. Statist.*, **13**, 10–69.
——(1986) Some useful extensions of the usual correspondence analysis approach and the usual log-linear models approach in the analysis of contingency tables. *Int. Statist. Rev.*, **54**, 243–309.
Gower, J. C. (1977) The analysis of asymmetry and orthogonality. In *Recent Developments in Statistics*. (eds J. Barra *et al.*). Amsterdam: North-Holland.
Greenacre, M. J. (1984) *Theory and Applications of Correspondence Analysis*. New York: Academic Press.
Harshman, R. A., Green, P. E., Wind, Y. and Lundy, M. E. (1982) A model for the analysis of asymmetric data in marketing research. *Marktng Sci.*, **1**, 205–242.
Healy, M. J. R. and Goldstein, H. (1976) An approach of the scaling of categorized attributes. *Biometrika*, **63**, 219–229.
van der Heijden, P. G. M. (1986) Transition matrices, model fitting, and correspondence analysis. In *Data Analysis and Informatics IV* (ed E. Diday). Amsterdam: North-Holland.
——(1987) *Correspondence Analysis of Longitudinal Categorical Data*. Leiden: DSWO.
van der Heijden, P. G. M. and de Leeuw, J. (1985) Correspondence analysis used complementary to loglinear analysis. *Psychometrika*, **50**, 429–447.
van der Heijden, P. G. M. and Worsley, K. J. (1988). Comment on Correspondence analysis used complementary to loglinear analysis (by P. G. M. van der Heijden and J. de Leeuw). *Psychometrika*, **53**, 287–291.
Hill, M. O. (1974) Correspondence analysis: a neglected multivariate method. *Appl. Statist.*, **23**, 340–354.
Israëls, A. (1987) *Eigenvalue Techniques for Qualitative Data*. Leiden: DSWO.
Jambu, M. and Lebeaux, M.-O. (1983) *Cluster Analysis and Data Analysis*. Amsterdam: North-Holland.
Lebart, L., Morineau, A. and Warwick, K. M. (1984) *Multivariate Descriptive Statistical Analysis: Correspondence Analysis and Related Techniques for Large Matrices*. New York: Wiley.

de Leeuw, J. (1983) Models and methods for the analysis of correlation coefficients. *J. Econometr.*, **22**, 113–137.

de Leeuw, J. and van der Heijden, P. G. M. (1988) Correspondence analysis of incomplete tables. *Psychometrika*, **53**, 223–233.

Nishisato, S. (1980) *Analysis of Categorical Data: Dual Scaling and its Applications*. Toronto: University of Toronto Press.

——(1986) *Quantification of Categorical Data: a Bibliography 1975–1986*. Toronto: Microstats.

Nora, C. (1975) Une méthode de reconstitution et d'analyse de données incomplètes. *Doctoral Dissertation*, Université Paul et Marie Curie, Paris.

Pannekoek, J. (1985) Log-multiplicative models for multiway tables. *Sociol. Meth. Res.*, **14**, 137–153.

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

Worsley, K. J. (1987) Un exemple d'identification d'un modèle log-linéaire grâce à une analyse des correspondances (with discussion). *Rev. Statist. Appl.*, **35**, 13–20, 83–94.

Discussion of the Paper by van der Heijden, de Falguerolles and de Leeuw

J. C. Gower (Rothamsted): In France they sometimes speak of Anglo-Saxon statistics. Here we have the opportunity of discussing some Franco-Dutch statistics. These national distinctions are more real than one might think and in welcoming the authors to the Royal Statistical Society, I particularly welcome their paper as a contribution towards a better understanding between British statisticians and our colleagues in the rest of Europe.

Correspondence analysis (CA) has been enthusiastically developed in France and widely adopted in other continental countries but has had a more cautious reception in Britain. In part this has been a consequence of claims that CA is a descriptive method and not model based; Anglo-Saxons are said to like models and especially their inferential consequences. Links between CA and log-linear analysis (LLA) have helped CA to gain more acceptance in Britain, and perhaps for LLA to gain more acceptance abroad. Interpreting the unknown in terms of the known is useful but does not in itself give grounds for favouring one method more than the other. I suspect not the mixed model muddle that Nelder (1977) discussed but a mixed-up model muddle. I see opportunity for confusion between what is and what is not a model, between different criteria for fitting models, between different ways of expressing interaction, between using graphical methods for display and diagnosis and between definitions of residuals. I know that all three authors are very clear-headed people, so what seem to me to be muddles are probably manifestations of different philosophical outlooks. Let me therefore become solidly Anglo-Saxon and see what response I receive from the authors.

Much of what I have to say does not focus on the special needs of contingency tables but many of the issues concerning CA may be viewed in the more general framework which I adopt.

Advocates of CA rely heavily on geometrical displays which indeed have attractions, as anyone can see by looking at the analyses given in the paper. However, to interpret these diagrams needs a thorough understanding of chi-square distance, different choices of scaling, the transition formulae and most of Section 2.1; this includes a knowledge of equation (11) which surely represents a model (in fact a bilinear model) in which interaction is expressed as a sum of k product terms. Bilinear models lend themselves to geometrical representations because of the relationship between inner products and the cosine formula. Convincing displays need two dimensions (i.e. $k = 2$), so interaction is expressed as the sum of *two* product terms. For $k = 1$, main effects and a single multiplicative term can be represented in three dimensions, the points representing rows lying in a plane orthogonal to another plane containing the points representing columns. In general, additive models, especially those with interactive terms, do not have nice geometrical representations. So part of the muddle is whether we put more weight on the model or more on the ability to express results diagrammatically; another part is the extent of our interest in main effects.

However, the muddle is more complex than this for a glance at equation (12) shows that the approximation $\log(1 + x) \sim x$ leads not to a log-linear model but to a log-*bi*linear model. I have long felt that generalized linear models (GLMs) could be beneficially further generalized to include bilinear terms. This would not be difficult to do and the product terms so estimated could be displayed very much as in CA. Such an extension to GLMs would combine the advantages of likelihood estimation with the

displays of CA—another confusing disparity is the use of (weighted) least squares for CA estimation and of maximum likelihood for GLMs. A good mathematical reason for preferring to model interaction as a product is that the Eckart–Young theorem then guarantees the best least squares approximation of given rank; any additive interaction model of the same rank cannot do better and will usually do much worse. However, the consequences of this result in terms of parsimony are not so clear cut as might appear, for the degrees of freedom associated with k product terms are the successively decreasing terms of $\Sigma_{i=1}^{k}(I + J - 1 - 2i)$ whereas each *a priori* additive interaction, expressed in terms of orthogonal polynomial coefficients (where these are appropriate) or other orthogonal contrasts, has a single degree of freedom. A serious criticism of these purely mathematical considerations is that a deep knowledge of the data and the mechanisms that generate them may provide better reasons for deciding between additive and multiplicative forms of interaction. Fisher and Mackenzie (1923), the first reference I know to a multiplicative model, try out both forms of model but seem to have no *a priori* preference, even though there is no attempt to examine the data on any transformed scale. If things are difficult to judge on the original scale of measurement, how much more so on a logarithmic or other transformed scale?

Another part of the mix-up is the use of Escofier's generalizations of CA as expressed by equation (9). Here again we seem to have a clear model but rather than using *ad hoc* estimates of Q, S_r, S_c and CA to display what remains, I would prefer to fit the whole model *ab initio*; again this does not preclude producing a CA-type diagram for the estimate of that part of the model expressed multiplicatively. Perhaps when I say a clear model, I should say a fairly clear model. It is not absolutely clear whether S_r and S_c and perhaps Q are essential parameters in some fully specified model, or whether they play more the role of weights to allow for unequal variances, or whether they express conditioning on some aspect of the data such as given margins. I suspect that under different circumstances all three usages are acceptable.

Another muddle is to decide whether the display is purely for presentation or is to be used as a diagnostic tool to suggest further terms that might be included in a model. This diagnostic use seems to be in the authors' minds, because we are repeatedly told that the various 'residuals' they consider are only to be analysed by CA if there is evidence that they contain interesting structure among terms so far not accounted for. To Anglo-Saxons, residuals are usually expected to reflect random variation and are examined for evidence of departures from the probabilistic assumptions of a model rather than from its systematic part.

LLA seems to score heavily over CA when we proceed to three- or more way tables. LLA, as for all GLMs, takes this in its stride; CA does not. Indeed the correspondence analyst resorts to curiosities such as pretending that an $I \times J \times K$ table is a pseudo two-way table of form $I \times (J \times K)$ or $(I \times J) \times K$ or $J \times (I \times K)$. I am grateful to the authors for showing exactly what these differing forms imply in terms of fitting two-factor multiplicative interactions, but I believe that we are now beginning to pay too much for our attractive diagrams. This seems to be recognized by the correspondence analysts themselves who are actively engaged in generalizing the bilinear models. This is necessary because, although bilinear models can be easily extended to cope with multiway tables, they can express no more than two factor interactions, hence an interest in trilinear and higher order models. Two of the most important of these generalizations are the CANDECOMP model of Carroll and Chang (1970)

$$y_{ijk} = \sum_{s=1}^{t} a_{is} b_{js} c_{ks}$$

and models for three-way components analysis introduced by Tucker (1964), and further studied by Kroonenberg (1983), which include

$$y_{ijk} = \sum_{p,q,r} a_{ip} b_{jq} c_{kr} w_{pqr}$$

where the ranges of p, q and r are (hopefully) much less than the ranges of i, j and k. Special cases of these models are certainly useful, especially when y_{ijk} lends itself in part to an inner product formulation, but, in general, interpretation is difficult and graphical displays are complex or lost altogether. The use of such terms in even more general models which also include additive and bilinear terms is something that I would prefer others to contemplate; the rewards would have to be very great to pay for the loss of simplicity of linear, or at a stretch, bilinear models.

I am very grateful to the authors for giving the Society an opportunity to discuss these issues. Although I have taken an Anglo-Saxon point of view, I have to admit that I find much that is attractive in graphical methods of data analysis. I hope that this paper will mark the beginning of further exchanges of ideas with our European neighbours.

I have much pleasure in proposing the vote of thanks.

Dr Frank Critchley (University of Warwick): I welcome the paper principally because of its insistence on establishing a constructive interplay between two broad approaches to analysing data. On the one hand we have the more formal, essentially confirmatory, modelling approach, and on the other hand the less formal, essentially exploratory, graphical approach. In the particular case of contingency table analysis, the authors advocate a productive dialogue between the log-linear analysis (LLA) modelling approach and the graphical approach of correspondence analysis (CA).

The appropriate context within which to review the paper is that of the general relationship between the modelling and graphical approaches to data analysis and I would therefore like to begin by offering some observations about this debate. Clearly, to attempt a complete account of this vast subject would be impossible. At the same time, if we identify these two approaches with the left and right hands respectively, then various biblical allusions summarize well much of what has been said from several positions taken up by statisticians on this issue.

Consider first the graphical explorer who, convinced of the futility of fitting inappropriate models, prefers instead a good picture of his data both as being of great value in itself and as a potential aid to building an appropriate model. Followers—not to say sheep—of this school of thought point us to *Ecclesiastes* 10:2 'A wise man's heart inclines him toward the right, but a fool's heart toward the left'.

Conversely, both the modeller keen to avoid any possible accusation of fiddling his results to fit a particular data set and the proponent of the traditional text-book advice that we should specify precisely which hypotheses we are going to test *before* seeing the data, would draw support from *Matthew* 6:3 'Do not let your left hand know what your right hand is doing'. More extreme types would point out that there can be no clearer warning of the dangers of data prying than *Matthew* 5:30 'If your right hand causes you to sin, cut if off and throw it away'.

Mercifully, the authors take a far more balanced, tolerant and mature view of data analysis. Their message is clear. In a minor adaptation of *1 Corinthians* 12:21, they remind us that 'Neither hand can say to the other, "I have no need of you" '.

I would like to finish by offering the following more or less unrelated remarks specific to the paper.

Related work

Whereas the proposed joint use of LLA and CA for contingency table data advocated in the present paper is undoubtedly original, closely related suggestions have appeared elsewhere. For example, Bradu and Gabriel (1978) proposed the biplot as a diagnostic tool for a particular multiplicative model for row-and-column centred contingency tables. They went on to point out that certain collinearities in this plot are suggestive of restrictions on the parameters of the model. This alternative approach illustrates that even for the same type of data there is nothing unique about the choice of (model, graphic) pair. What is needed, however, is a happy marriage between the chosen model and graphic. In a different context Caliński *et al.* (1987a, b) propose such a marriage between (multivariate) analysis of variance techniques on the one hand, and canonical variates and principal components on the other. Their problem relates to the analysis of a series of experiments repeated at several places over a period of years.

Higher way tables

The analysis of three- and higher way tables poses genuine problems. The log-linear modeller is faced with a proliferation of parameters. The correspondence analyser (and even the multiple-correspondence analyser) is faced with the fact that his is a (joint) bivariate technique, i.e. he has to carry out one of many different possible preliminary reductions of a multiway table to a two-way table. The authors make several valuable suggestions for a combined approach in Section 3.2. Much closely related and, I believe, important work is contained in Coppi and Bolasco (1989) on the analysis of multiway tables. One strand in this book is the switch from matrix to tensor algebra.

Caveats

Several researchers have raised caveats about the use of (multiple-) correspondence analysis and, in particular, about the interpretation of the resulting plots.

(a) Healy and Goldstein (1976) and Goldstein (1987) have pointed out that the row and column scores obtained are highly dependent on the constraints imposed for identifiability.

(b) In their discussion of equation (5) in Section 2.1 the authors suggest a general interpretation for row-to-column distances, although they remark that care is needed. Greenacre and Hastie (1987), p. 441, take a much harsher line. They state baldly that 'row-to-column distances are meaningless'.

Asymmetry (Section 4.3)

The authors justify the decomposition (19) of asymmetric data into its symmetric and skew parts on the basis that the former contains the maximum likelihood estimates for the symmetry model (21). I would like to point out another, more general, justification. The set of all real square matrices of given order is a real vector space in which the subsets of skew and symmetric matrices appear as complementary vector subspaces with trivial intersection. If we endow the whole space with the usual inner product

$$\langle A, B \rangle = \mathrm{tr}(A^{\mathrm{T}}B)$$

then these two subspaces are orthogonal and equation (19) is then the corresponding orthogonal decomposition. Otherwise said, M is the least squares symmetric fit to P and N the least squares skew symmetric fit. However, the usual inner product is not the only one with the property that it renders the skew and the symmetric subspaces orthogonal. There is a whole class of such inner products which have been characterized in Critchley (1988).

In conclusion, it gives me great pleasure to second the vote of thanks.

The vote of thanks was passed by acclamation.

Professor M. A. Moran (University College Cork): The paper is to be welcomed in giving further exposure to correspondence analysis (CA), which can be regarded as the natural counterpart to principal component analysis for categorical data, and deserves to be as well known. That said, I am not convinced that the way in which the authors use it to represent residuals from a log-linear model is the most appropriate.

A surprising omission from the references is the Gabriel (1971) biplot. Both the biplot and CA concern a reduced rank representation of a matrix **Y**. The best such approximation in terms of least

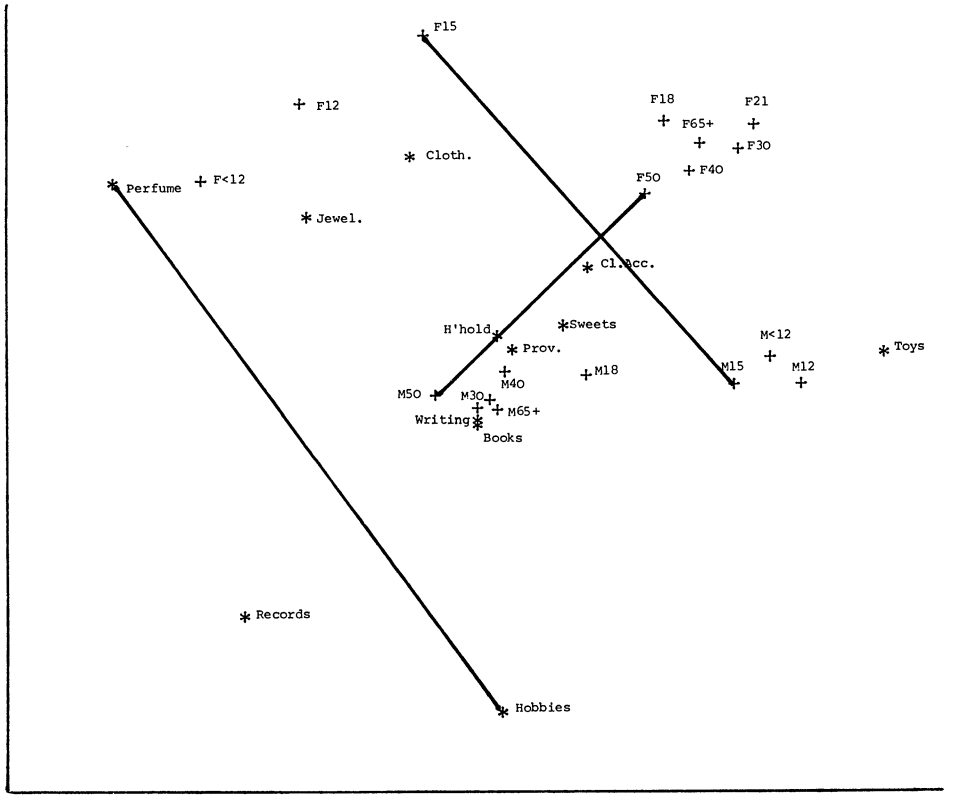


Fig. 7. Biplots of residuals $\log(p_{ij}/\hat{\Pi}_{ij})$ for model [AS] [AG]

squares is provided by the singular value decomposition $\mathbf{Y} = \mathbf{U}\Lambda\mathbf{V}^t$. In CA the elements of $\mathbf{Y}$ are the Pearson residuals $(p_{ij} - p_{i\cdot}p_{\cdot j})/\sqrt{(p_{i\cdot}p_{\cdot j})}$. The sum of squares to be 'explained' is Pearson's X^2 and a scaling for $\mathbf{U}$ and $\mathbf{V}$ is adopted which leads to the useful representation of distances between profiles as described in the paper.

The biplot uses an alternative scaling such as $\mathbf{G} = \mathbf{U}\Lambda^{1/2}$ and $\mathbf{H} = \mathbf{V}\Lambda^{1/2}$ to give $\mathbf{Y} = \mathbf{G}\mathbf{H}^t$. Thus individual elements of $\mathbf{Y}$ are represented by scalar products

$$y_{ij} = \mathbf{g}_i^t\mathbf{h}_j = \|g_i\| \, \|h_j\| \cos\theta$$

where $\mathbf{g}_i$ and $\mathbf{h}_j$ are (row) vectors of $\mathbf{G}$ and $\mathbf{H}$. Only the first two elements of each vector are used. If we define a residual as $\log p_{ij} - \log \hat{\Pi}_{ij}$, then for each two-way table formed by rows (i, i') and columns (j, j') we have

$$\log\left(\frac{\text{observed odds}}{\text{fitted odds}}\right) = (\mathbf{g}_i - \mathbf{g}_{i'})^t(\mathbf{h}_j - \mathbf{h}_{j'})$$

with an obvious geometrical interpretation. Consider the biplot representation of these residuals for the log-linear model [AS], [AG] for the data of Table 1. The result of an unweighted biplot approximation is shown in Fig. 7. As the model assumes that, given age, sex and goods are independent, it is appropriate, for each age group, to join the points for the two sexes.

It is immediately apparent that

(a) there is a substantial lack of independence between goods and sex in specific cases, e.g. perfume versus hobbies by sex for the younger age groups and
(b) there is a substantial three-factor interaction resulting from the radically different relationship of goods to sex for the three younger age groups compared with that for the six older groups.

Log-linear analysis involves contrasts of log-odds and this approach seems a more natural complement than that proposed in the paper, particularly for large frequencies where standardized residuals with their emphasis on statistical significance have limited relevance.

I had difficulty, in any case, with the interpretation of the residuals considered in Section 3.2.

Apart from the trivial case where rows and columns are fitted with saturated models the residuals would not be proportional to the Pearson residuals and thus CA would no longer provide a decomposition of the chi-square statistic for goodness of fit. Has the emphasis on profiles been pushed too far?

Dr M. Green (Lancaster University): The foundation of correspondence analysis (CA) on singular value decomposition (SVD) is both its strength and its weakness. The strength is that CA is a computationally efficient technique allowing rapid data exploration. The weakness is that SVD restricts CA to a two-dimensional representation of a multiway table when studying the log-multiplicative term. The alternative approach of generalizing the row–column (RC) association model to multiway tables by allowing more than one log-multiplicative term provides a class of models of great flexibility. Such models can be readily (if slowly) fitted by the technique described in Green (1988). For example consider the shop-lifting data. The RC association model for the equivalent analysis to that in Section 2.1.1 is

$$\log \mu_{ijk} = u_0 + u_{12}(i, j) + u_3(k) + \sum_\alpha \phi_\alpha v_{12}(i, j, \alpha) v_3(k, \alpha) \tag{25}$$

where i, j and k refer to AGE, SEX and GOODS respectively.

A display of the v_{12} and v_3 parameters produces a plot almost identical with Fig. 1 and suggests the possibility that the pattern over age is the same for both sexes but shifted by a constant amount. A suitable model would assume that the score for a combination of sex and age is the sum of a sex score and an age score, leading to the model

$$\log \mu_{ijk} = u_0 + u_{12}(i, j) + u_3(k) + \sum_\alpha \phi_\alpha \{v_1(i, \alpha) + v_2(j, \alpha)\} v_3(k, \alpha). \tag{26}$$

Fitting these models gives the following deviances:

Model (0)	null model	19111

The null model is the log-linear model, AGE × SEX + GOODS.

	Association term	$\alpha = 1$	$\alpha = 2$
Model (25)	(AGE × SEX) ∧ GOODS	5456 (71.5%)	1930 (89.9%)
Model (26)	(AGE + SEX) ∧ GOODS	5549 (71.0%)	2158 (88.7%)
Model (27)	(AGE + SEX + FEMT) ∧ GOODS	5543 (71.0%)	1964 (89.7%)

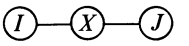
where the $\wedge$ operator corresponds to a product in the log-multiplicative terms. In terms of proportion of residual deviance explained (figures in parentheses in the table above), model (26) is almost as good as model (25) indicating that the assumption that the sex effect is constant over age is a good first approximation. However, the deviance difference of 228 is statistically significant and it is of interest to identify the major sources of deviation from model (26). A visual comparison of the displays of $v_{12}(i, j)$ from model (25) and $v_1(i) + v_2(j)$ from model (26) suggests that teenage girls in age groups 12–14 and 15–17 deviate most. An extra factor, FEMT, is defined that isolates this group. It takes the value t in the model formula defined by $t = 2$, for the group of teenage girls, $t = 1$, otherwise. This leads to the model

$$\log \mu_{ijk} = u_0 + u_{12}(i, j) + u_3(k) + \sum_{\alpha} \phi_\alpha \{v_1(i, \alpha) + v_2(j, \alpha) + v_{12}(t, \alpha)\} v_3(k, \alpha). \quad (27)$$

This extra parameter explains 194 of the 228 deviance difference between models (25) and (26) and results in a model hardly distinguishable from model (25). Inspection of the biplot suggests a preference of this group for jewelry and perfume items. That this preference is stronger than that expected from the combination of sex effect and age effect would have an important sociological interpretation. In conclusion can the authors forsee any way that the CA technique could be further generalized to incorporate this type of analysis?

Joe Whittaker (University of Lancaster): I should like to point out a connection between the Goodman row–column (RC) model and a mixed graphical model, described in a forthcoming paper by Lauritzen and Wermuth (1989). The mixed graphical model considered has a single conditionally Gaussian random variable X and two discretely valued random variables I and J. It is assumed that

(a) $X \mid (I, J) \sim N(\mu(i, j), \sigma^2)$, the usual model for a two-way analysis of variance (ANOVA), and
(b) I and J are independent given, X, $I \perp\!\!\!\perp J \mid X$, so that the conditional independence graph is



Then Lauritzen and Wermuth (1989) show $\mu(i, j) = \alpha(i) + \beta(j)$, so that the two sets of parameters $\{\alpha(i)\}$ and $\{\beta(j)\}$, associated with the two edges (I, X) and (J, X) respectively, are intepretable as additive effects in the ANOVA model. By integration, Lauritzen and Wermuth (1989) show that the marginal distribution of I and J is of the form

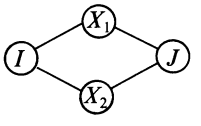
$$\log f_{I,J}(i, j) = a(i) + b(j) + \phi \alpha(i) \beta(j),$$

exactly the Goodman RC model, where $\phi = \sigma^{-2}$. Hence, the Goodman model has a latent variable interpretation, which, firstly, may suggest a physical mechanism and, secondly, always allows an interpretation for the parameters in terms of the interaction with an unobservable continuous variable. Though this representation appears to suggest that the EM algorithm is an appropriate way to solve the maximum likelihood problem, the mixed graphical model with complete data on (X, I, J) has no direct estimates.

Now consider a generalization to two continuous variables:

(a) $X_r \mid (I, J) \sim N(\mu_r(i, j), \sigma_r^2)$, for $r = 1, 2$,
(b) $I \perp\!\!\!\perp J \mid (X_1, X_2)$ and
(c) $X_1 \perp\!\!\!\perp X_2 \mid (I, J)$,

so the conditional independence graph is



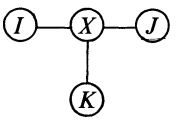
The marginal distribution of (I, J) is now

$$\log f_{I,J}(i, j) = a(i) + b(j) + \phi_1 \alpha_1(i) \beta_1(j) + \phi_2 \alpha_2(i) \beta_2(j).$$

The correspondence analysis diagrams of the paper are essentially the same as given by plotting a point representing the category $I = i$ with co-ordinates $(\alpha_1(i), \alpha_2(i))$ and one representing the category $J = j$ with co-ordinates $(\beta_1(j), \beta_2(j))$, the very coefficients of the additive effects in the ANOVA. The strengths of the two components X_1 and X_2 are estimated by the ϕ parameters and interpretable as the inverse of the corresponding variances. The independence graph is the notorious chordless four cycle, which has no direct estimates even when the variables are entirely discrete or entirely continuous, and partly explains why fitting the Goodman model is a non-trivial exercise.

Extensions of the model to include a third continuous variable are immediate, and then, by dropping the conditional independences between the elements of (X_1, X_2, X_3), we may allow interesting rotations of the co-ordinates, in the same manner as factor analytic rotations.

A generalization in a different direction is to contemplate graphs for three observed discrete variables: I, J and K. For example, the graph



leads to models of the form

$$\log f_{I,J,K}(i, j, k) = a(i) + b(j) + c(k) + \phi\{\alpha(i)\beta(j) + \alpha(i)\gamma(k) + \beta(j)\gamma(k)\},$$

which allow a parsimonious form of three-way interaction between I, J and K. This model can then be extended to include a second component as will be discussed in a forthcoming paper.

Dr Willem J. Heiser (University of Leiden): It is certainly not true that correspondence analysis (CA) —or multiple CA—is a single method that 'has been reinvented several times' and 'is known under different names': some reinventions and names refer to conceptually different starting points. The diversity of conceptions is frequently obscured by the omnipresence of least squares, the exponential family of distributions and the associated algebra of quadratic forms. Fortunately, today it is precisely 'generalized CA'—or the 'Escofier generalization' (see Benzécri *et al.* (1973), who explain that various now famous theorems were discovered and formulated by her)—that shows us clearly this diversity. For if CA is canonical analysis of residuals from independence, then it is natural to generalize it to other cases like residuals from quasi-independence, symmetry and quasi-symmetry. But if we regard CA as a seriation technique that happens to use reciprocal averaging then the Escofier generalization is much less convincing, for why would we discard the idea of taking convex combinations? And if we regard CA as canonical correlation analysis of two sets of indicator variables, then the only idea involved here is filtering out the effect of a covariate by introducing it in both sets and ignoring the first few trivial axes (a point made by Gifi (1981), who refers for this idea to Roy (1957)).

With regard to CA as a log-linear model with bilinear restrictions, not only CA can provide a simplified interpretation of interaction parameters, especially not in the two-way case. Well-known alternatives from psychometrics are the Shepard–Luce model that explains the interaction in terms of distance and the feature–choice model that in turn restricts the distance in terms of an *a priori* given set of features (see Takane and Shibayama (1986)). When a specific model is absent there are no pressing reasons for particularly using CA here, rather than any other general purpose multidimensional scaling method, or the biplot.

Dr van der Heijden's paper demonstrates exploratory analysis under the guise of the confirmatory. The sample size in the car data is so enormous that any model other than the saturated model is doomed to fail a naive statistical test. I would not blindly follow the rule 'if the X^2 value is significant then it makes sense to study the residuals'. Conversely, as remarked very shrewdly by Lebart (1976), the X^2 value of the whole table may be insignificant while the inertia accounted for by the first CA axis is significant.

There is still another way to use CA in relation to a multidimensional contingency table. This is by regarding each cell of the multidimensional table as a row object of the correspondence table P and by packing each row with an indicator coding of the response pattern multiplied by the cell frequency. In Gifi's (1981) terminology this approach is called ANAPROF, and its use is indicated when there are quite a few variables with very few categories—a situation complementary to that just discussed, where there are very few variables with quite a large number of categories.

Dr J. L. A. van Rijckevorsel (TNO Institute of Preventive Health Care, Leiden): On the second page van der Heijden *et al.* refer to 'structure' in the data that correspondence analysis (CA) supposedly highlights. To some extent they discuss what this structure is. However, more is needed to be said to reduce the disreputableness, unclear interpretation or informality of CA (Gilula and Haberman (1988), p. 760). The rehabilitating effect on the status of CA by linking it to log-linear analysis (LLA) will be further enhanced by including order relationships in the definition of structure to be revealed by CA or by LLA.

An example of an unclear interpretation is connecting category points with each other in a way that is based on *a priori* information, as in Fig. 1 and Fig. 2 with the respective age categories. Connecting category points with each other in the same order as the canonical scores is statistically justified if the

contingency table is order dependent for some order. Order dependence is a mathematical property of a contingency table related to total positivity and regression dependence that, for instance, causes horseshoes and the like (van Rijckevorsel, 1987). Ordering properties in CA are thoroughly discussed by Schriever (1983, 1986). See also Goodman (1981, 1986) and Gilula and Haberman (1988). Gilula *et al.* (1988) define order dependence for CA as stochastic order extremity for correlation models. Real data, like rating scales, ecological data, seriation data or any kind of gradient data are frequently order dependent or nearly order dependent. The existence of order dependence and the way that it is reflected in CA is certainly not informal or unclear, nor are the corresponding interpretations.

Another related observation is that several discretized bivariate distributions, including the bivariate normal, the bivariate binomial, the bivariate Poisson, the bivariate gamma, the bivariate trinomial, the bivariate F, the bivariate Pareto, the bivariate logistic and the bivariate exponential distributions, generate order-dependent probability tables whose order is reflected in the canonical scores of these tables (Schriever (1986), p. 52). Would it not be interesting to know whether the close relationship between the CA model (12) and the row–column association model (RC–AM) (13) of Goodman is related to the occurrence of order-dependent probability tables? Goodman (1981), p. 325, has already observed that, if variables are bivariate normals or can be transformed into bivariate normals, RC–AM scores are approximately equal to the canonical scores. We know that several discretized bivariate distributions are order dependent and that RC–AM and CA scores of such tables are identically ordered (Schriever (1986), p. 54). My conjecture is that CA and log-multiplicative decomposition generate approximately equal scores if the variables are order dependent for some order.

Dr J. A. Nelder (Imperial College, London): I welcome this paper as showing the connection between correspondence analysis and log-linear analysis. I make two points: first because correspondence analysis depends on the singular value decomposition of a matrix it is essentially limited to two-way classifications; this is a severe limitation, not shared by log-linear models, which generalize easily to an arbitrary number of classifying factors. Any attempt to force multiply classified data into correspondence analysis by combining the factors arbitrarily and then ignoring their structure is likely to lead to uninterpretable pictures.

The second point concerns techniques in general which depend on eigenvalue or singular value decomposition. Both can be interpreted as finding a direction removing maximum variation in some sense, then a second direction orthogonal to the first removing maximum remaining variation and so on. The principal directions depend partly on the systematic structure of the data and partly on irrelevant 'random' errors. It is common in factor analysis to look for directions with relatively simple-structured components which span a space close to those from the eigen or other analysis. I suggest that there is scope for a similar activity when using the methods of this paper.

The following contributions were received in writing after the meeting.

Professor Murray Aitkin (Tel Aviv University): The authors refer to the increasing use of correspondence analysis (CA) published in English. I hope that there will be a corresponding increase in the use of modelling published in French: to further this end, Dr de Falguerolles and I and our coworkers have published the results of our comparative analyses in several papers (see Aitkin *et al.* (1987)).

I endorse the authors' view that CA is most useful in analysing the residuals from an inadequately fitting model, and that there are explicit statistical models with the multiplicative interaction structure of CA which can be fitted by maximum likelihood as an alternative to CA. This was the conclusion of our project.

The authors' Table 1 shows the type of data structure for which CA has most promise: a high dimensional unstructured multinomial response factor, with simple age and sex cross-classifying factors, and a large sample size. Conventional log-linear modelling cannot do much with this, especially if there is a significant three-way interaction. Plotting the response and (crossed) explanatory factor in the space of the leading components of the residual matrix from a simple model can be very helpful in identifying a simple structure of association (heterogeneity), frequently a subgrouping of response categories with different patterns, or a few outlying points. This leads naturally to better statistical modelling, even if this consists only of excluding some poorly fitting points from a simple model for the remainder of the data. In theory, inspection of interaction parameter estimates or of residuals should be able to provide the same information, but scanning long columns of high order interactions or residuals for patterns is rarely fruitful.

In more complex survey data with many categorical response variables and a complex explanatory structure, CA procedures lose their usefulness, and typically produce bewilderingly complex graphics, because of the failure to model the data by taking out the main structure before plotting the residuals. Explicit statistical modelling is essential in such cases, and there is a need for efficient programs for multinomial logit and mixture modelling with structured responses and continuous explanatory variables. This is not possible in GLIM, nor in many other standard statistical packages. Software development is needed here.

Professor Clifford C. Clogg (Pennsylvania State University, University Park): I have a few comments on the fundamentals of correspondence analysis (CA).

It is odd to construct such an elaborate algebraic methodology around the Pearson statistic. Statisticians interested in testing goodness of fit, particularly in sparse data situations, are often led to consider other statistics, including Neyman's statistic, the likelihood ratio statistic, or modifications of them (Haberman, 1988). The likelihood ratio statistic is minimized when maximum likelihood (ML) estimation is used, and ML is the most common method for log-linear models. A more valid comparison between CA and log-linear models might be possible if the likelihood ratio statistic were used, but this would change the algebra of CA considerably.

One of the main justifications for log-linear models is that interactions are measured in terms of cross-ratios which do not depend on the marginal distributions. In contrast, the score parameters, the eigenvalues and the geometric representations of CA build in a strong dependence on the marginals. When analysing a single contingency table this may be of little consequence. However, if results from one table (shop-lifting in the Netherlands) are compared with results from another table (shop-lifting in the USA), the comparisons will be invalid if marginal distributions differ (Becker and Clogg, 1989). Because of this problem, CA represents a regression as a method either to measure or to plot interaction.

CA still suffers from insufficient attention to sampling variability and general issues of smoothing data through the use of statistical models. For two-way tables, ML methods are available in papers by Goodman and Gilula and Haberman and would almost always be preferable to the simple singular value decomposition of residuals. For data in the form of Table 1, ML methods in Gilula and Haberman (1988) or Becker and Clogg (1989) can be used. When one-dimensional models fit the data, the geometric representations of CA are irrelevant, and when more than two dimensions are required, they do not reveal everything.

Inference procedures should be developed to analyse the clustering of points in the two-dimensional representations. How can we determine whether 'subclouds' exist? We need to have rigorous methods for determining the number of clusters, for assigning rows and columns to clusters, for comparing within-cluster variability to between-cluster variability and for combining row categories and/or column categories. It is perplexing that such natural inferential questions receive so little attention.

Sir David Cox (Nuffield College, Oxford): This is a very welcome paper with two intrinsically interesting examples. A difficulty here in assessing the relative merits of alternative analyses is the absence of a moderately well-specified purpose: are the data background to some administrative action, are they to test some sociological 'theory' about say shop-lifting or to aid in the formulation of such a theory? In the last two cases measurement and inclusion of further explanatory variables are called for and it is a little difficult to see what the role of correspondence analysis would be. While the distinction between confirmatory and exploratory analysis is useful, its importance has been considerably exaggerated.

Y. Escoufier (Université des Sciences et Techniques du Languedoc, Montpellier): Many researchers around the world are now working on extensions of correspondence analysis (CA). Section 4 of the paper is a very valuable contribution to the debate.

I am not so enthusiastic about Section 2. The CA solution given in equation (3) as a singular value decomposition (SVD) is true but the preceding statements on the chi-square distance are not sufficient to justify it. The matrix of the scalar products between the row profiles (centred as indicated in the paper) corresponding to the chi-square distance is

$$D_r^{-1}(P - E)\,D_c^{-1}(P - E)'D_r^{-1}$$

which leads to the SVD of $D_r^{-1}(P - E)\,D_c^{-1/2}$ and not to equation (3).

Working with the metric D_c^{-1} on the centred data $D_r^{-1}(P - E)$ is equivalent to using the identity matrix on the transformed centred data $D_r^{-1}(P - E)\,D_c^{-1/2}$. If the rows of this matrix receive the weights

given by the diagonal of D_r, the variance–covariance matrix is

$$D_c^{-1/2}(P - E)'D_r^{-1}(P - E)D_c^{-1/2}$$

with normalized eigenvectors V. The co-ordinates of the rows on the principal components will be given by

$$D_r^{-1}(P - E)D_c^{-1/2}V = D_r^{-1}(P - E)C = D_r^{-1}PC = \tilde{R}.$$

With an SVD program we could compute V and R from the SVD of $D_r^{-1/2}(P - E)D_c^{-1/2}$ but we must explain why. From a practical point of view, a consideration of the weights is very important: the method tries to favour the representation of the rows with high weights and because of the properties $t'D_r\tilde{R} = 0$ and $R'D_r\tilde{R} = I$ the co-ordinates of all the rows are affected.

The authors claim that CA can be used as a technique for decomposing residuals for log-linear models. It has been shown in my discussion contribution with S. Junca of Goodman (1986) or in Escoufier (1988) that the SVD approach can be used to estimate the parameters of the row–column association model if we accept that we can substitute a least squares criterion for the maximum likelihood criterion. This method is specifically useful when the matrix under study is large and can be extended to other models.

Readers interested in the analysis of square matrices should read Escoufier and Grorud (1980) who use an SVD approach for simultaneous representations of M and N.

Professor Zvi Gilula (Hebrew University and University of Chicago): The argument regarding the use of correspondence analysis (CA) is somewhat self-defeating. The authors say that CA does not assume an underlying model and that its primary goal is to study 'structure' in the data. They emphasize that 'conclusions about the data may not be generalized to the population'.

How then should we understand results regarding patterns of association among variables in the examples analysed? It is not clear whether the data analysed pertain to a *sample* or to an entire population.

They exclusively use a *two*-dimensional plot, accompanied by a ratio called the 'chi-square proportion'. Although no explicit explanation of the connection between this ratio to the two-dimensional plot is given, it is clearly used to assess the *goodness of approximation* (GOA) of the residual matrix by another matrix of a lower rank. In five different analyses we find ratios of 0.78, 0.93, 0.55, 0.71 and 0.82 without a relation to the differences between these ratios and their meaning regarding the plot. Does the *interpretation of the results* depend on the degree of GOA or is it just our *confidence* in the interpretation linked to this ratio? We must conclude, for instance, that a degree of GOA of 0.55 indicates that the plot should consist of more than two dimensions. A criterion is missing on the degree of GOA and its impact on the number of dimensions needed to describe the data reasonably well. The final sentence on p. 261 amplifies the confusion regarding this ratio. The ratio, through its relation to X^2, is meant to help us to 'decide whether the residuals contain interesting information'. It is not clear how X^2 is linked to what is 'interesting'.

As a descriptive tool only, there is no difference between CA and Gabriel's biplot where significant interpretation is given to 'distances' between rows (columns) and to the angles between vectors describing these rows (columns). If CA should be linked to a model-based approach then it can be used as a *model exploratory* tool. Since CA is based on a canonical form of bivariate distributions, I have shown (Gilula, 1986) that whether or not some of Goodman's association models are saturated depends on the rank of the matrix describing the underlying bivariate distribution.

Finally, CA does *not* have to be a *secondary* tool with respect to a primary underlying model but can be meaningful in its own right. I have shown (Gilula, 1984) that a *functional association* exists between CA (of any rank) and latent class models.

Professor Michael Greenacre (University of South Africa, Pretoria): The authors put too severe a restriction on correspondence analysis (CA) when they say that 'CA only makes sense when these residuals are not merely a result of random variation from independence'. Benzécri's philosophy is that CA is a data re-expressing (or information transforming) instrument. Just as a histogram presents a batch of data in a different form, so CA transforms a data matrix into a graphical image, which is valid

irrespective of the statistical weight of evidence in favour of any feature in the data. I know many instances of CA pictures of contingency tables where features which are obviously non-random reveal themselves in spite of the failure of the omnibus X^2 test to reject independence.

What characterizes CA is the cardinal importance placed on the geometric properties of the row and column scores. The geometric interpretation is by no means simple (see Greenacre and Hastie (1987)), especially in the typical case where row and column points are plotted simultaneously. The authors advise against interpreting row-to-column distances in plots such as Fig. 1, but their interpretation of the figure shows that they have fallen into this trap. For example, they single out young adult males as stealing clothing accessories, household goods, provisions and tobacco more than average. This deduction can only be made because these male age groups lie close to these items in the plot. The data of Table 1, however, when looked at in row percentage form, show that household goods, provisions and tobacco are stolen generally by *older* males more than average, while clothing accessories are stolen more than average by females in general as opposed to males, with scant reason for singling out young adult males. The interpretation of such pictures is the greatest problem that people have with CA. Greenacre (1987) makes a specific attempt to define the allowable limits of the geometric interpretation.

I have considerable difficulty with the geometric interpretation of CA pictures when residuals from more complex log-linear models are analysed, e.g. Fig. 2. I cannot see that the 'chi-squared distances between points in different clouds are changed in a way that is easy to understand'. Joint row and column displays from CA are only acceptable when residuals from independence are analysed. Only then can we easily defend the inherent metric, for example, and here we at least know what the full space geometry involves before the points are projected on to a principal plane. In the present situation of generalized residual analysis a full space geometry certainly exists, but it is so esoteric that it defeats the object of producing a graphical display. What advantages does a (generalized) CA of log-linear residuals hold over a simple biplot display of the residuals? The biplot is based on the ordinary singular value decomposition of the residual matrix and has a much less complicated interpretation in this case (e.g. Gabriel (1981)).

A. Z. Israëls (Netherlands Central Bureau of Statistics, Voorburg): For the method presented in Section 3.2, log-linear analysis (LLA) can handle more than two variables, whereas *simple* correspondence analysis (CA) is limited to only two variables. By interactive coding of variables, the authors make simple CA suitable for three or more variables also. It might be worthwhile to consider whether other forms of CA or optimal scaling can be used in the analysis of residuals of (or differences between two) log-linear models, to include three or more variables. We may consider here, for instance, multiple CA or CA applied to a Burt matrix, both resulting in the same (normalized) category scores. (In the shop-lifting example, the Burt matrix is defined as the two-way table $(A + S + G) \times (A + S + G)$ of order $(9 + 2 + 13) \times (9 + 2 + 13)$. It contains all bivariate contingency tables (even twice) as well as three block diagonal matrices with univariate cell frequencies as diagonal elements.)

Finally, I want to make a suggestion with respect to the Burt matrix in connection with quasi-independence. When replacing the elements of the block diagonal matrices by the product of the corresponding marginals, divided by n, CA of this 'modified Burt matrix' results in the same (normalized) category scores, but its singular values can be interpreted more easily than in CA of the unmodified Burt matrix, as the block diagonal elements no longer contribute to the total inertia (or X^2/n) (see equation (8)). As a result, the chi-square value of the modified table is equal to the sum of the chi-square values of the bivariate tables. This neutralizing of the block diagonal matrices can, technically, be seen as a quasi-independence assumption for the Burt matrix. In this way, it could be possible to neutralize subtables other than the block diagonal subtables.

Professor N. C. Lauro (University of Naples): I shall confine myself to two aspects of the paper that should be examined further: the choice of the two-way multiple table to consider and the role played by the categorical variables in the analysis.

Although the authors give some advice for choosing the two-way multiple table the fact remains that their procedure does not lead to a unique and objective representation of interaction parameters. Therefore a suitable statistical criterion should be introduced as the initial step in the analysis.

However, the choice of a two-way multiple table implies a non-symmetrical treatment of the categorical variables that is not taken into account by using classical correspondence analysis (CA).

Both aspects are bypassed when non-symmetrical CA (NSCA) and logistic models are considered in an integrated strategy to analyse the structure of dependence between categorical variables (Lauro and Siciliano, 1988). Instead of CA that decomposes Pearson's ϕ^2 association index, NSCA (Lauro and

D'Ambra, 1984; D'Ambra and Lauro, 1988) is based on the τ predictability criterion of Goodman and Kruskal.

For instance, given the estimation of a saturated logistic model for a three-way table

$$\ln(f_{ijk}/f_{.jk}) = \beta_i + \beta_{ij} + \beta_{ik} + \beta_{ijk} \qquad (i \in I, j \in J, k \in K),$$

this pertains specifically to the table $I \times (JK)$. In the non-symmetrical case, multiple or partial NSCA can be used to analyse the residuals between the saturated model and one of the models

$$\ln(f_{ijk}/f_{.jk}) = \beta_i \quad \text{or} \quad \ln(f_{ijk}/f_{.jk}) = \beta_i + \beta_{ij}.$$

Furthermore, the interaction representation can be alternatively obtained by considering the first-order approximation to an asymmetric association model (i.e. logistic model), i.e.

$$\ln(f_{ijk}/f_{.jk}) = \ln f_{i..} + (1/f_{i..}) \sum_m l_m v_{im} w_{jkm} \qquad (m \in \bar{M} \leqslant M),$$

where l_m, v_{im} and w_{jkm} are respectively eigenvalues, rows and columns scores related to the mth principal axis in the reconstitution formula of NSCA on the table $I \times (JK)$. Furthermore, an analysis of variance type of decomposition of $w_{jkm} = \delta_{jm} + \delta_{km} + \delta_{jkm}$ allows us to obtain separate plots based on the scores v_{im} versus δ_{jm}, v_{im} versus δ_{km} and v_{im} versus δ_{jkm}, which are in some way related to the interaction parameters (β_{ij}, β_{ik} and β_{ijk}) of a logistic model. Similar results also hold for the symmetrical case.

Peter McCullagh (University of Chicago): I find it surprising that the authors can contemplate an analysis of the data in Table 1 without first indicating the aims of the study or the purposes of the analysis. How were the data collected? By whom and for what purpose are they to be used? These questions may seem churlish, but the issues that confront the proprietor of a stationery shop or the head of security at a large department store are only indirectly relevant to a sociologist studying adolescent behaviour or to the headmaster of a boys' school. It is at best an error of strategy to think that a single analysis is capable of satisfying the needs of all who might use such data. To quote Jeffreys (1939) 'It is sometimes considered a paradox that the answer depends not only on the observations but on the question; it should be a platitude'.

These comments are not to be construed as criticisms either of log-linear models or of correspondence analysis, either of which might conceivably be helpful depending on the purpose of the analysis. However, it is undesirable to adopt a strategy of analysis that requires neither model nor assumptions, focuses on no prespecified questions and promises answers to all of them. After all, to compute is not, in itself, to analyse.

Dr R. L. Plackett (Newcastle upon Tyne): Correspondence analysis (CA) is seldom discussed at meetings of the Society, and this interesting paper provides a welcome opportunity to assess the method in the light of current research. I must begin by rejecting the hypothesis that the language problem accounts for the lag in development outside France. Language has not been a barrier to the recognition of French scientific achievements in the past. If there has been a lag, then the reasons must be sought elsewhere, and perhaps arise from the impression that CA is a partly qualitative method concerned only with large tables in two dimensions. This view seems to be confirmed from the preliminary analysis of Table 1, where age and sex are rolled together in defiance of intuition, and the conclusions are such as might be obtained by looking at the data and making simple calculations. However, the authors then proceed to show that the method is much more general by moving their base from simple independence to any standard model, log-linear or otherwise. This is a useful development, although if CA is applied to study the residuals from such models then there ought to be a comparison with the methods already available for analysis of residuals. Overall, the position of CA when compared with alternative procedures depends on the degree of success attained with the interpretation of data, so that the choice of examples is important. With this in mind, Table 1 is somehow marginalized, because shop-lifters may steal two or more goods simultaneously; and with Table 2, the absence of economic variables suggests that the conclusions have limited value.

K. J. Worsley (McGill University, Montreal): Multiplicative models can also be useful for exploratory data analysis. Consider the shop-lifting example and the plot of parameters from a multiplicative model given in Fig. 1. For the age–sex parameters, the horizontal axis is a function only of age, possibly quadratic in the ordered age categories. The vertical axis largely distinguishes between the sexes. This

suggests the following *linear* model for the log E(frequencies), written in GLIM notation,

$$\text{AGE} * \text{SEX} + \text{GOODS} + (A_1 + A_2 + \text{SEX}).\text{GOODS}, \tag{28}$$

where AGE, SEX and GOODS are factors with 9, 2 and 13 levels each, A_1 is a covariate taking the values 1 for age less than 12, 2 for age 12–14, . . . , 9 for age 65 and above, and $A_2 = A_1^2$. The terms AGE * SEX + GOODS describe the independence model. The terms $(A_1 + A_2)$.GOODS account for a quadratic effect of age on proportion of goods stolen. The term SEX.GOODS accounts for the different proportions of goods stolen by men and women. When this model is fitted by GLIM using the Poisson likelihood, it explains 94% of chi squared, greater than the 78% explained by Fig. 1, but since three sets of goods coefficients are being estimated this model should be compared with the three-dimensional correspondence analysis (CA), which explains 90% of chi squared.

If we compare the coefficients of A_1.GOODS with those of A_2.GOODS by plotting one against the other (Fig. 8) we find that they are strongly linearly related with a slope of -0.083, estimated by least squares. Fig. 9 now plots the female.GOODS coefficients (the male.GOODS coefficients are aliased and set to zero by GLIM) against the A_1. GOODS coefficients; this plot is remarkably similar to the CA plot of the goods coefficients in Fig. 1. Almost the same plot is produced using the A_2.GOODS coefficients instead of the A_1.GOODS coefficients. Below the horizontal axis is $X = A_1 - 0.083A_2$ for the various age categories, which resembles the plot of age coefficients on the CA plot.

Furthermore, we can replace the term $(A_1 + A_2)$.GOODS in the model by X.GOODS, treating X as a fixed covariate. We can then create a variable G say, taking as values the estimated coefficients of X.GOODS for the corresponding levels of the GOODS factor, and then define two new covariates $G_1 = A_1G$ and $G_2 = A_2G$. Replacing X.GOODS by $G_1 + G_2$ allows us once again to estimate the

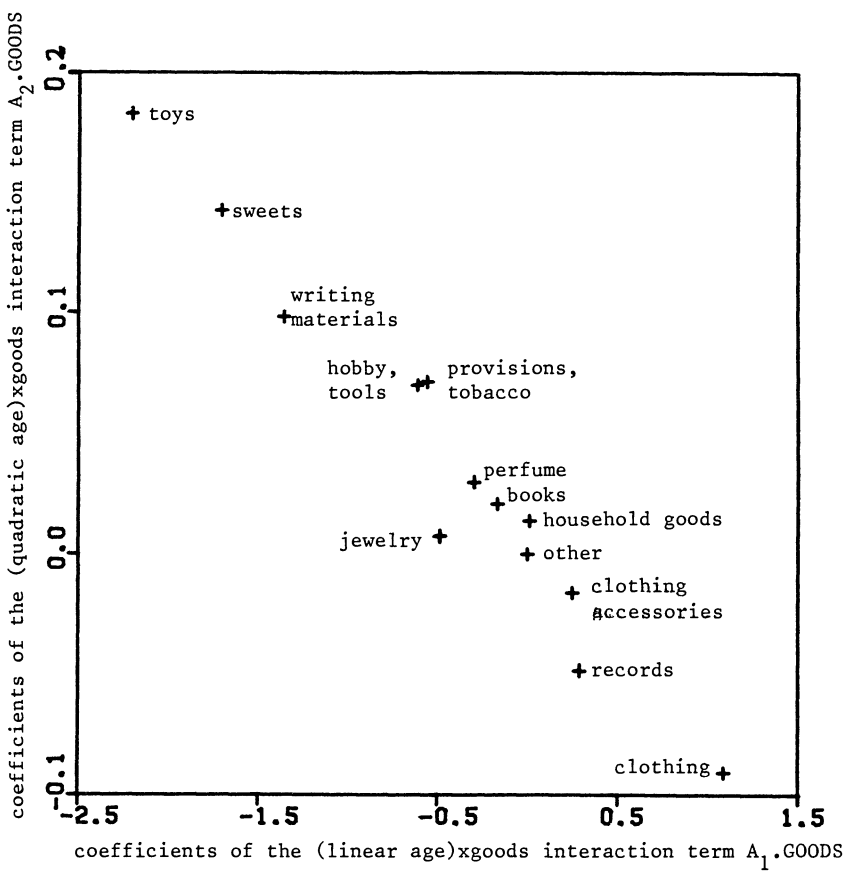


Fig. 8

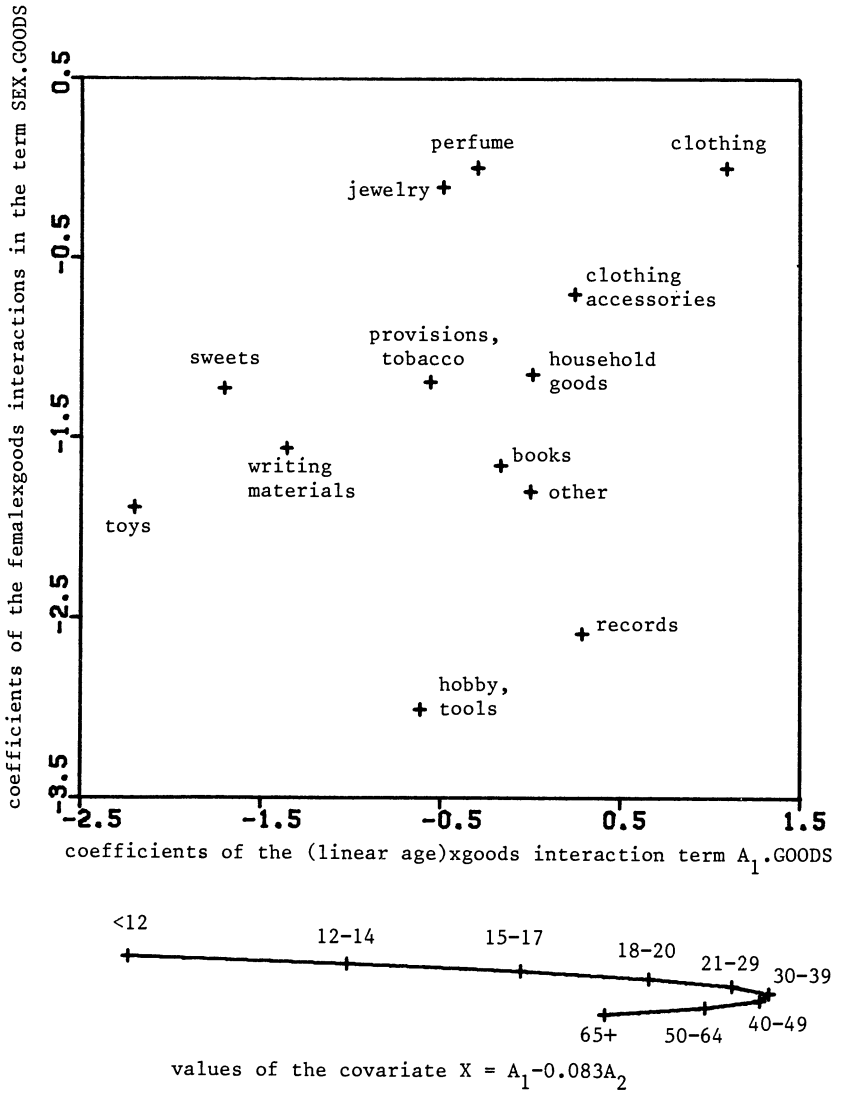


Fig. 9

quadratic age effect, to replace X by this quadratic and to iterate. We are in fact fitting a *multiplicative* term of the form $(i + \alpha i^2)\gamma_j$, where i and j index categories of age and goods respectively and α, $\gamma_1, \ldots, \gamma_{13}$ are parameters to be estimated. After two more iterations the process converges to give $\hat{\alpha} = -0.079$ with 84% of chi squared explained. This is even better than the two-dimensional CA since it explains more of chi squared with fewer than half the parameters, 25 compared with 54. This is not a contradiction since we have chosen to model log(expected frequencies) rather than the expected frequencies.

The **authors** replied later, in writing, as follows.

Mixed-up model muddle

John Gower refers to the mixed-up model muddle. We would like to comment on a particular aspect of this muddle. In our context a model is usually a low dimensional non-linear surface in the space B of all bivariate contingency tables. The distance of the observed contingency table to the model is

measured by defining some metric on the space B and by projecting the observed table on the model. Many aspects of this procedure are quite arbitrary. For the choice of metric, statisticians depend on general statistical theory, and for the choice of the model they depend on prior knowledge of the field, or, as Sir David Cox and Peter McCullagh emphasize, on clearly defined questions from the field. However, very often this paradigm simply does not apply. There are no questions from the field other than to find out what is there, and there is very little prior theory to relate to. The choice of model becomes problematic. This is indicated in the contributions of Mick Green and Keith Worsley, and to a lesser extent of Joe Whittaker, who search through a very complicated set of models until they find one which suits their purposes. From the modelling point of view ordinary correspondence analysis (CA) is more modest. It searches, linearly, through the relatively simple surfaces defined by the singular value decomposition (SVD) in 1, 2, . . . , $\min(R, C) - 1$ dimensions.

Although it may be useful to interpret CA as a technique for fitting a sequence of restrictive models, this interpretation is mostly a device to bridge the conceptual gap between various parts of statistics. It is not what the people had in mind who invented CA. Pearson wanted to investigate the sensitivity of the correlation coefficient to scoring of the categories. Hirschfeld wanted to linearize the regressions, Fisher wanted to find appropriate scores for categorical variables to perform regression and discriminant analysis on those scores. No models were involved here. Benzécri, whose point of view is similar to that of Kalman (1983), did not want to impose prejudices in the form of models on the data analysis. He was interested in describing bivariate tables in terms of suitably chosen co-ordinates, i.e. only in saturated models, in the same way as the system theorist is interested in representing multivariate time series in state space form. A CA simply computes these co-ordinates, and the next step is to make graphical representations of some aspects of this new representation. Although we can interpret the activities of the correspondence analyst in terms of fitting a two-dimensional SVD model, such an interpretation is misleading and can easily lead to communication problems. Our contribution perhaps adds to the model muddle, because we find co-ordinates for a derived table and the residuals after fitting a model. Thus we combine philosophies, although it is clear that our emphasis is on the CA, not on the model fitting.

Using chi-square values

Zvi Gilula, Michael Greenacre and Willem Heiser comment on the use of chi-square values for deciding when performing CA is useful. We shall try to make our position somewhat clearer. It is not useful to study *random* departures from a model. If the chi-square value indicates that the model under study should be rejected, then CA can be helpful. However, it is also possible that the chi-square value for some model is *not* significant while the departure from the model is *not* random. In this case the *power* of the chi-square test of independence against the *unrestricted* alternative was not sufficiently high to detect the non-random departure. The power is higher for tests against *restricted* alternatives (if these alternatives are true). Lebart's results (see Willem Heiser) can be explained in this way (see Agresti (1984) for a similar discussion in the context of log-linear analysis). So the 'if' in 'if the X^2 value is significant then it makes sense to study the residuals' should not be replaced by 'if and only if'.

Another problem with chi-square values is that they are linearly related to sample size so that, if the sample is sufficiently large, all restrictive models should be rejected if we use the chi-square values in a strict sense, i.e. the power goes to unity if the sample size goes to infinity. We would tend to accept restrictive models if in the saturated model sets of u parameters depart so little from zero that this departure is not substantially interesting.

The usefulness of chi-square values as a criterion to decide on doing CA can be reduced even further by the fact that in much research the assumptions of the model will be violated to some extent. Observations in samples are regularly not independent and identically distributed (it is not difficult to think of reasons why the observations in Tables 1 and 2 would not be independent, for example). Also, because of non-response, samples are regularly not random samples of the population that one had in mind when collecting the data. Perhaps this is an explanation for the 'perplexing' (as Clifford Clogg calls it) state of affairs that inferential questions receive so little attention. For some people this argument suggests the contrary: in their opinion it is perplexing that inferential questions receive so *much* attention. Most people will agree that, *if we know* that the observations are independent and identically distributed from some distributions, *then* it would be unwise not to use this information. People differ in their willingness to assume that they know this.

Ordinary correspondence analysis

Yves Escoufier shows clearly how we make the step from chi-squared distances in equation (1) to the SVD of standardized residuals in equation (3), and Michael Greenacre correctly indicates that we were

sloppy in the interpretation of Fig. 1. Contrary to what Michael Greenacre says, we do not advise against interpreting *row-to-column distances*, but instead we show that row-to-column distances can be interpreted using the transition formulae (5a) and (5b) (see the text above equation (6)). The reason that the interpretation of Fig. 1 is sloppy is *not* due to interpretation of row-to-column distances, but instead due to the fact that small distances in the first two dimensions can be much larger in full dimensional space. This holds not only for row-to-column distances but also for row-to-row and for column-to-column distances.

This brings us to Zvi Gilula's second point. The *degree of goodness of approximation* can be derived from equation (8), i.e. the proportion of the total inertia for dimension α is $\lambda_\alpha^2/\Sigma_\alpha\lambda_\alpha^2$. The normalizations $\tilde{R}'D_r\tilde{R} = \Lambda^2 = \tilde{C}'D_c\tilde{C}$ show that this proportion can also be interpreted as the proportion of weighted (chi-squared) distances that is displayed in dimension α. By adding these proportions we can decide how much of the departure from the model is displayed in the first few dimensions and how much is hidden in higher dimensions. It will be clear that in principle we are interested in studying as much departure from the model as possible with CA, but, in contrast, the geometrical representations are less attractive when higher dimensional solutions need to be studied. If the sum of the first few proportions still departs considerably from unity, then we use CA only to obtain a representation of the most important departure from the underlying model. In this case the quality of representation of individual rows and columns is also to be taken into account. In this paper we restricted ourselves to two-dimensional solutions only because we thought that two dimensions sufficed to explain our objectives.

Graphical representations

Michael Greenacre and Professor Moran are doubtful about the way that we make our graphical representations. CA can be considered as a very specific biplot method in the sense that joint plots of rows and columns can be constructed. There are alternatives to our CA biplots. To shed some light on another muddle described by John Gower, we have used a generalization of CA to make biplots that have good *geometrical* properties. Another approach would have been to decompose matrices with residuals that have good *statistical* properties (see, for example, Caussinus and de Falguerolles (1987) and Novak and Hoffman (1989)). It is useful to make biplots of *scaled* residuals (compared with 'raw' residuals) to adjust for the fact that cells with high frequencies are more likely to have large *raw* residuals than cells with low frequencies. Therefore we prefer to study standardized residuals, or adjusted residuals and the like above a decomposition of raw residuals as is suggested by Michael Greenacre.

Moran poses the question whether the emphasis on profiles has been taken too far: one of the good properties of ordinary CA is that it gives a decomposition of the chi-square statistic, and in the solutions for conditional independence (Fig. 2) and (quasi-)symmetry (Figs 5 and 6) this property is lost. Except for quasi-independence, making solutions with good geometrical properties happened at the expense of the decomposition of the appropriate chi-square value. For example, we have chosen to plot the residuals from conditional independence in Fig. 2 in the same metric as the metric for independence (ordinary CA) because in this case the solution is easily interpretable in terms of chi-squared distances; however, now the relation with the Pearson chi-square value does not hold anymore. Also, the *standardized* residuals from (quasi-)symmetry are not skew symmetric, and therefore we have chosen to scale the residuals in another way.

In answer to Clifford Clogg's first point: it is possible, in principle, to use other metrics for CA. The prime example, and one which many people would probably like, is spherical factorial analysis proposed by Domenges and Volle (1979, 1980). It uses the Hellinger distance between multinomials. This simplifies the SVD, because no weights are needed, but it destroys the centroid principles, which are perhaps the main attraction of CA. Another possibility, indicated by Moran, is to take logarithms, to remove additive effects and to perform an unweighted SVD on the residuals. Although these proposals are interesting, and may be attractive for some purposes, they lose much of the simple geometry that makes CA attractive.

Higher way tables

In the paper we have restricted ourselves to *simple* CA to analyse higher way tables. Willem Heiser suggests that we analyse multiway tables with the computer program ANAPROF, which is a program for *multiple* CA (MCA). In the framework of this paper, MCA can be explained as a method that concentrates on bivariate marginal dependence in the data and that decomposes the departure from bivariate marginal independence, as is described by Abby Israëls (see also van der Heijden (1987)). Second-order interactions in the data are ignored, and sometimes this may be considered as a drawback

of MCA compared with simple CA. In circumstances where the number of observations is much lower than the number of cells in the multiway table (for example, when the number of variables is very large), simple CA solutions become rather unstable, whereas MCA solutions are still stable because the bivariate margins are still reasonably well filled. There are also substantial reasons for preferring MCA to CA, but it was beyond the scope of the paper to discuss these (see, for example, Tenenhaus and Young (1985) for different interpretations of MCA).

Abby Israëls describes MCA as a simple CA of the Burt matrix and showed that the diagonal submatrices of the Burt matrix can be neutralized using the quasi-independence model (see also Israëls (1987)). He also suggests that we use quasi-independence models to neutralize other subtables. This is an interesting idea: in a similar way it is probably also possible to suppress the influence of one or more structurally zero cells if these cells are each defined by the joint occurrence of two categories.

In the paper we do not give much attention to the problem of how to code frequencies of a higher way table into a specific two-way table. This choice for a specific two-way table is not completely arbitrary, as is suggested by John Nelder. We gave some guidelines in Section 3.2. These guidelines are very easy to follow when we can split our variables into explanatory variables and response variables: in this case we code the explanatory variables interactively and the response variables interactively, since the main interest is in the relation *between* the explanatory and the response variables and not in the relations *within* the sets of explanatory and response variables. This procedure is also suggested recently in Gilula and Haberman (1988) for the analysis of multiway tables with row–column (RC) correlation models and RC association models, and this approach is also adopted in logit analysis, where the interactions between the explanatory variables are usually not studied. The interpretation of the interactive variables can be simplified by performing analysis of variance types of decomposition on interactive row or column variables, as is suggested by Lauro.

We have no practical experience with non-symmetrical CA (NSCA) (see Carlo Lauro). NSCA seems more natural than CA when interest extends to the predictability criterion of Goodman and Kruskal, because this criterion is decomposed in NSCA. However, the logit models that Lauro describes are precisely equivalent to the log-linear models that we use. The fit of a logit model can be evaluated with the Pearson chi-square value, which corresponds to the association index ϕ^2. Would the use of logit models not suggest doing ordinary CA on the appropriate two-way coding instead of doing NSCA?

Clifford Clogg describes the problem that it is difficult to compare two CA solutions, partly because CA solutions depend on the margins, which is a serious problem. For example, if we have two contingency tables, one for males and one for females, we can find different solutions even if the second-order interaction in the constructed three-way table is absent, due to different metrics for the solution for the males and for the females. As a partial solution to this problem we suggest the following. Say the two variables under study are denoted 1 and 2, and sex 3. If we are interested in studying two separate contingency tables, we are probably interested in the parameters $u_{12(ij)}$ (i.e. describing the interaction between 1 and 2 that both sexes have in common) and the parameters $u_{123(ijs)}$ (i.e. describing how the interaction between 1 and 2 differs for each of the sexes). This could be studied in one CA solution by studying one of the tables $P^{i[js]}$ or $P^{j[is]}$, while decomposing the departure from model [13] [23] (see Section 3.2). However, even when the second-order interaction is zero, the corresponding points for males and females do not have to coincide (see van der Heijden (1987), chapter 3).

Contrary to what some of the contributors suggest, SVD is not limited to two-way tables. We agree with John Gower that the interpretation of solutions from higher way generalizations of SVD is much more difficult. Pioneering work in this field has been performed by Kroonenberg (1983, chapter 15; 1989), who proposes three-way generalizations of CA.

Modelling approaches

We welcome the contributions of Michael Green and Keith Worsley as good examples of what we aimed to show at the end of Section 3.1, namely that graphical CA displays indicate how to restrict the interaction parameters (this is also suggested in the third remark of Zvi Gilula). In answer to Mick Green's question we are not sure of the fruitfulness of building in constraints into the CA technique. For example, if we could build in a constraint that the male and female lines in Fig. 1 have exactly the same form, what have we gained with such a CA solution? The modelling approach adopted by Mick Green and Keith Worsley has the advantage that, if we are willing to assume that the assumptions for maximum likelihood (ML) are not violated, we can test whether the difference between the unconstrained analysis and the constrained analysis is significant. We would simply use CA in this context only as a diagnostic tool that is easy to compute.

John Gower asks whether we propose to use CA for presentation or as a diagnostic tool. CA can be useful for both purposes. Apart from these uses, we also agree with Zvi Gilula that CA is useful in its own right, for example to understand the relation between variables. CA is also useful in other contexts than the context of the paper; see, for example, the interesting contribution of Jan van Rijkevorsel. We are quite sure that his conjecture is false in the general terms in which he states it. Order dependence only generates ordinal restrictions on the scores, so statements about approximate equality will have to be interpreted in terms of monotonicity.

Instead of doing log-linear analysis and analysing residuals with Escofier's generalization of CA, John Gower 'would prefer to fit the whole model *ab initio*' (see also Murray Aitkin). Some of Clifford Clogg's remarks are similar in meaning. We can imagine that there are circumstances in which this is preferable to what we have done, but modelling procedures are not available for all the decompositions that we performed, and the fitting of these models is computationally quite difficult. Simple graphical displays can suffice in many situations.

Clifford Clogg states that the ML methods described by Goodman and Gilula and Haberman would almost always be preferable to ordinary CA since CA suffers from insufficient attention to sampling variability and general issues of smoothing data. We agree that the extra information that these ML models give is very useful when the assumptions for deriving this information are valid. His suggestion to study clustering of points is very interesting. There might be an impression in the paper that there are no inferential results available for ordinary CA. Greenacre (1984) gives an overview of results concerning the significant departure from zero of singular values, and we refer to Gifi (1981), who uses the delta method to derive confidence intervals for CA scores and singular values. Software for these confidence intervals will become available in a CA user's procedure in SPSS-X in the near future. For the generalized CA applications no inferential properties are known.

We shall show here why in some instances ordinary CA is preferable to the RC correlation model (an ML version of ordinary CA, see Goodman (1985) and Gilula and Haberman (1988)). The RC correlation model can be written as

$$\pi_{ij} = \pi_{i+}\pi_{+j}\left(1 + \sum_{\alpha}^{h} \gamma_\alpha u_{i\alpha} v_{j\alpha}\right)$$

and when $h = \min((I - 1), (J - 1))$, we have the *saturated* RC correlation model, for which the estimates are the same as the results found in CA if the scores $u_{i\alpha}$ and $v_{j\alpha}$ are normalized in the same way as the RC correlation estimates. When $h = 1$ we find one of the *restricted* RC correlation models. This model is logically inconsistent, because the row scores and column scores are chosen in such a way that the correlation γ_1 is maximized, but the scores and correlation may not be chosen in such a way that the term $1 + \gamma_1 u_{i1} v_{j1}$ is negative for cell (i, j). In such cases ordinary CA and the RC correlation model can differ considerably. We prefer ordinary CA because it gives us the highest possible correlation with corresponding category quantifications.

The examples

Both Sir David Cox and Peter McCullagh comment on the absence of a well-specified purpose or question from the analysis. This is, to a certain extent, unavoidable. CA examples are often taken from areas where the questions are not at all specific. Moreover we know as much about shop-lifting as anyone. From our point of view it is better to illustrate a technique, especially a graphical technique, with real data than with artificial data. However, it is true that, in a sense, we are treating real data as though they were artificial. In real situations in which CA is used there is more interaction with the person providing the data. In such cases the analysis may not always provide an answer to a well-specified question, but it often produces a feeling of satisfaction in both data analyst and investigator. This feeling of satisfaction should not be underestimated.

We have some more details about the examples. For the first data set we cite Israëls (1987): 'This survey was held among about 350 Dutch stores and big textile shops. An inquiry form was filled in for each case of shoplifting. Apart from the non-response, the shops in the inquiry form a complete population.' 'Cases of shoplifting in which more than one kind of good was stolen, or in which more than one person was suspected, have been omitted.' For the car changing data, we cite Harshman *et al.* (1982): 'Each year Rogers National Research, a marketing consulting firm in Toledo, Ohio, collects car trade-in data for a large sample (tens of thousands) of U.S. buyers of new cars. The data are collected by mail questionnaires asking recent car buyers of new cars to indicate both the newly purchased model and the old model (if any) disposed of at the time of purchase.' Another analysis of the car changing data can be found in Okada and Imaizumi (1987).

References in the Discussion

Agresti, A. (1984) *Analysis of Ordinal Categorical Data.* New York: Wiley.

Aitkin, M., Francis, B. and Raynal, N. (1987) Une étude comparative d'analyses de correspondances ou de classifications et des modèles de variables latentes ou de classes latentes. *Rev. Statist. Appl.*, **35**, 53–81.

Becker, M. P. and Clogg, C. C. (1989) Analysis of sets of contingency tables using association models. *J. Amer. Statist. Ass.*, **84**, in the press.

Benzécri, J.-P. *et al.* (1973) *L'Analyse des Données*, vol. IA, no. 2, pp. 15–60. Paris: Dunod. (*Honolulu*, Fr. transl. of Statistical analysis as a tool to make patterns emerge from data. In *Methodologies of Pattern Recognition* (ed. S. Watanabe), pp. 35–74. New York: Academic Press, 1969.)

Bradu, D. and Gabriel, K. R. (1978) The biplot as a diagnostic tool for models of two-way tables. *Technometrics*, **20**, 47–68.

Caliński, T., Czajka, S. and Kaczmarek, Z. (1987a) A model for the analysis of a series of experiments repeated at several places over a period of years, I: Theory. *Biul. Oc. Od.*, **12**, 7–33.

—— (1987b) A model for the analysis of a series of experiments repeated at several places over a period of years, II: Example. *Biul. Oc. Od.*, **12**, 35–71.

Carroll, J. D. and Chang, J. J. (1970) Analysis of individual differences in multidimensional scaling via an *n*-way generalization of "Eckart–Young" decomposition. *Psychometrika*, **35**, 283–319.

Caussinus, H. and de Falguerolles, A. (1987) Tableaux carrés: modélisation et méthodes factorielles. *Rev. Statist. Appl.*, **35**, 35–52.

Coppi, R. and Bolasco, S. (eds) (1989) *The Analysis of Multiway Data Matrices.* Amsterdam: North-Holland.

Critchley, F. (1988) On characterisations of the inner-products on the space of square matrices of given order which render orthogonal the symmetric and the skew-symmetric subspaces. *Warwick Statistics Research Report 157.* University of Warwick, Coventry.

D'Ambra, L. and Lauro, N. C. (1989) Non symmetrical analysis for three-way contingency tables. In *Analysis of Multiway Data Matrices* (eds R. Coppi and S. Bolasco). Amsterdam: North-Holland.

Domenges, D. and Volle, M. (1979) Analyse factorielle sphérique: une exploration. *Ann. INSEE*, **35**, 3–84.

—— (1980) L'analyse factorielle sphérique. In *Data Analysis and Informatics* (eds E. Diday *et al.*). Amsterdam: North-Holland.

Escoufier, Y. (1988) Beyond correspondence analysis. In *Classification and Related Methods of Data Analysis* (ed. H. H. Bock). Amsterdam: Elsevier.

Escoufier, Y. and Grorud, A. (1980) Analyse factorielle des matrices carrées non-symétriques. In *Data Analysis and Informatics* (eds E. Diday *et al.*). Amsterdam: North-Holland.

Fisher, R. A. and MacKenzie, W. A. (1923) Studies in crop variation, II: The minimal response of different potato varieties. *J. Agricult. Sci.*, **13**, 311–320.

Gabriel, K. R. (1971) The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, **54**, 453–467.

—— (1981) Biplot display of multivariate matrices for inspection of data and diagnosis. In *Interpreting Multivariate Data* (ed. V. Barnett), pp. 147–174. Chichester: Wiley.

Gifi, A. (1981) *Non-linear Multivariate Analysis.* Department of Data Theory, University of Leiden.

Gilula, Z. (1984) On some similarities between canonical correlation models and latent class models for two-way contingency tables. *Biometrika*, **71**, 523–529.

—— (1986) Grouping and association in contingency tables: an exploratory canonical correlation approach. *J. Amer. Statist. Ass.*, **81**, 773–779.

Gilula, Z. and Haberman, S. J. (1988) The analysis of multivariate contingency tables by restricted canonical and restricted association models. *J. Amer. Statist. Ass.*, **83**, 760–771.

Gilula, Z., Krieger, A. M. and Ritov, Y. (1988) Ordinal association in contingency tables: some interpretive aspects. *J. Amer. Statist. Ass.*, **83**, 540–545.

Goldstein, H. (1987) The choice of constraints in correspondence analysis. *Psychometrika*, **52**, 207–215.

Goodman, L. A. (1981) Association models and canonical correlation in the analysis of cross-classifications having ordered categories. *J. Amer. Statist. Ass.*, **76**, 320–334.

—— (1985) The analysis of cross-classified data having ordered and/or unordered categories: association models, correlation models, and asymmetry models for contingency tables with or without missing entries. *Ann. Statist.*, **13**, 10–69.

Green, M. (1988) A modelling approach to multiple correspondence analysis. *Proc. COMPSTAT 88.* Vienna: Physica.

Greenacre, M. J. (1984) *Theory and Applications of Correspondence Analysis.* New York: Academic Press.

—— (1989) The Carroll–Green–Schaffer scaling in correspondence analysis: a theoretical and empirical appraisal. *J. Marktng Res.*, to be published.

Greenacre, M. J. and Hastie, T. J. (1987) The geometric interpretation of correspondence analysis. *J. Amer. Statist. Ass.*, **82**, 437–447.

Haberman, S. J. (1988) A warning on the use of chi-squared statistics with frequency tables with small expected cell counts. *J. Amer. Statist. Ass.*, **83**, 555–560.

Harshman, R. A., Green, P. E., Wind, Y. and Lundy, M. E. (1982) A model for the analysis of asymmetric data in marketing research. *Marktng Sci.*, **1**, 205–242.

Healy, M. J. R. and Goldstein, H. (1976) An approach to the scaling of categorised attributes. *Biometrika*, **63**, 219–229.

van der Heijden, P. G. M. (1987) *Correspondence Analysis of Longitudinal Categorical Data*. Leiden: DSWO.
Israëls, A. (1987) *Eigenvalue Techniques for Qualitative Data*. Leiden: DSWO.
Jeffreys, H. (1939) *Theory of Probability*. Oxford: Clarendon.
Kalman, R. E. (1983) Identifiability and moding in econometrics. In *Developments in Statistics* (ed. P. R. Krishnaiah), vol. 4. New York: Academic Press.
Kroonenberg, P. M. (1983) *Three-mode Principal Component Analysis*. Leiden: DSWO.
—— (1989) Singular value decomposition of interactions in three-way contingency tables. In *Analysis of Multiway Data Matrices* (eds R. Coppi and S. Bolasco). Amsterdam: North-Holland.
Lauritzen, S. L. and Wermuth, N. (1989) Graphical models for associations between variables, some of which are qualitative and some quantitative. *Ann. Statist.*, to be published.
Lauro, N. C. and D'Ambra, L. (1984) L'analyse non symmetrique des correspondences. In *Data Analysis and Informatics III* (eds E. Diday *et al.*), pp. 433–446. Amsterdam: North-Holland.
Lauro, N. C. and Siciliano, R. (1988) Correspondence analysis and modelling for contingency tables: symmetrical and non-symmetrical approaches. *3rd Workshop on Statistical Modelling, Vienna.*
Lebart, L. (1976) The significancy of eigenvalues issued from correspondence analysis of contingency tables. In *Proc. Computational Statistics*, pp. 38–45. Vienna: Physica.
Nelder, J. A. (1977) A reformulation of linear models. *J. R. Statist. Soc.* A, **140**, 48–77.
Novak, T. and Hoffman, D. (1989) Residual scaling: an alternative to correspondence analysis for the graphical representation of residuals from log-linear models. *Multiv. Behav. Res.*, to be published.
Okada, A. and Imaizumi, T. (1987) Nonmetric multidimensional scaling of asymmetric proximities. *Behaviormetrika*, **21**, 81–96.
van Rijckevorsel, J. L. A. (1987) *Fuzzy Coding and Horseshoes in Multiple Correspondence Analysis*. Leiden: DSWO.
Roy, S. N. (1957) *Some Aspects of Multivariate Analysis*. New York: Wiley.
Schriever, B. F. (1983) Scaling of order dependent categorical variables with correspondence analysis. *Int. Statist. Rev.*, **51**, 225–238.
—— (1986) *Order Dependence*. Amsterdam: Centre for Mathematics and Computer Science.
Takane, Y. and Shibayama, T. (1986) Comparison of models of stimulus recognition data. In *Multidimensional Data Analysis* (eds J. de Leeuw *et al.*), pp. 119–138. Leiden: DSWO.
Tenenhaus, M. and Young, F. W. (1985) An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis and other methods for quantifying categorical multivariate data. *Psychometrika*, **50**, 91–119.
Tucker, L. R. (1964) The extension of factor analysis to three-dimensional matrices. In *Contributions to Mathematical Psychology* (eds N. Frederikson and H. Gulliksen), pp. 109–127. New York: Holt, Rinehart and Winston.