MULTIWAY DATA ANALYSIS
R. Coppi and S. Bolasco (Editors)

GENERALIZED CORRESPONDENCE ANALYSIS OF MULTI-WAY CONTINGENCY TABLES AND MULTI-WAY (SUPER-)INDICATOR MATRICES

Peter G.M. van der Heijden & Frits Meijerink*

Department of Psychometrics and Research Methods, Leiden University, Hooigracht 15, 2312 KM Leiden, The Netherlands

Multi-way tables are recoded into two-way tables and subsequently analyzed with correspondence analysis (CA). It is known that for contingency tables CA can be interpreted as giving a decomposition of residuals from a specific loglinear model. A generalization can be used to decompose residuals from less restricted loglinear models; it will be shown that this can lead to average restrictions in the category scores. Loglinear analysis concepts are also introduced in the context of multiple CA of (super-)indicator matrices. This helps understanding which "interactions" are displayed by multiple CA. It also becomes possible to obtain multiple CA solutions with average restrictions to the scores.

1. INTRODUCTION

One way to analyze multi-way tables is by recoding them to two-way tables, and analyzing these with ordinary multivariate techniques. This is the approach that we will adopt in this paper. The specific way of recoding the multi-way table that we will use is stacking, i.e. treating all combinations of categories of original variables as categories of a single composite variable. For example, if we have a three-way table of order 2x3x4, we can create two-way tables of order 6x4, 8x3 and 2x12 by stacking two of the original variables. The techniques discussed in this paper are correspondence analysis (CA) and multiple correspondence analysis (MCA), and the types of table that we will analyze with these techniques are multi-way contingency tables, multi-way indicator matrices, and multi-way super-indicator matrices.

As we just indicated, the variables of a multi-way table can be stacked in more than one way. In general, the way of stacking chosen influences the way in which properties of the data are revealed to us. This is already studied for CA of stacked multi-way contingency tables by relating this approach to loglinear analysis (van der Heijden & de Leeuw, 1985). They showed that CA of stacked tables could be interpreted as giving a decomposition of the residuals from a specific loglinear model. For example, CA of the three-way table indicated

*The second author was supported by the Netherlands Organization of Pure Research (N.W.O.). We gratefully acknowledge the helpful comments of Frits Bareman, Patrick Groenen, Willem J. Heiser and Leo J.Th. van der Kamp on an earlier version.

bove, stacked into a table of order 6x4, decomposes the residuals from the odel in which variable 3 is independent of variables 1 and 2 jointly. A eneralization of CA was introduced to decompose residuals from less restricted oglinear models. Thus a very specific interpretation is given to suggestions o use CA and loglinear analysis complementary to each other (compare Daudin & 'récourt, 1980; Israëls & Sikkel, 1982; Lauro & Decarli, 1982).

In this paper we will show first that, by decomposing residuals from some of he less restricted loglinear models, subsets of category quantifications have reighted averages equal to zero. Then we will show that, by applying some oglinear analysis concepts to MCA of multi-way (super-)indicator matrices, we obtain a better understanding of which "interactions" are revealed by MCA. By ntroducing the generalization of CA in this context, we can obtain average estrictions in MCA. Average restrictions can be used to suppress or eliminate specific aspects in the data, or, what comes to the same, to enhance other aspects. In this sense our work is similar to that of Gifi (1981, ch.10), Nishisato (1984), Yanai (1986), Ter Braak (1986), Heiser (1987) and Escofier (1983, 1987) who all restrict CA in one way or another.

In the sequel we will first summarize the earlier results concerning CA of multi-way tables and loglinear analysis (section 2.). In section 3. we will show the average restrictions that can be obtained by using CA and loglinear analysis jointly. In section 4. and 5. we will discuss the implications of this approach for MCA.

2. CORRESPONDENCE ANALYSIS, AND ITS RELATION TO LOGLINEAR ANALYSIS

2.1. Correspondence analysis and generalized correspondence analysis

For reasons of space, we will not give an extended introduction to CA here, but refer instead to, for example, Nishisato (1980), Gifi (1981) and Greenacre (1984). CA of contingency tables can be represented by showing its reconstitution formula

$$P = D_r(1+R\Lambda C')D_c \tag{1}$$

where P is a two-way matrix with proportions p_{ij} adding up to 1; 1 is a matrix with ones, having the same order as P; D_r and D_c are diagonal with elements p_{i+} and p_{+j} respectively (we replace an index by a '+' when added up over the corresponding variable); R is a matrix with row scores $r_{i\alpha}$ for row i (i=1,...,I) on dimension α; C is a matrix with column scores $c_{j\alpha}$ for column j (j=1,...,J) on dimension α; Λ is a diagonal matrix with singular values λ_α in

$$P-E = D_rR\Lambda C'D_c \tag{2}$$

where E is an independence matrix having elements $e_{ij}=p_{i+}p_{+j}$. This shows that CA decomposes the departure from independence.

Departing from (2), the generalization of CA that we will use in this paper is simply

$$P-Q = D_rR\Lambda C'D_c \tag{3}$$

where Q is a matrix of the same order as P, but not necessarily restricted as E. We will assume that the margins of P are equal to those of Q, i.e. $p_{i+}=q_{i+}$ and $p_{+j}=q_{+j}$. Due to this, we can obtain solutions for generalized CA using ordinary CA programs by analyzing the matrix (P-Q+E), i.e. the independence matrix E added to the difference between the matrices P and Q. This generalization is introduced by Escofier (1983), see also van der Heijden & de Leeuw (1985) and van der Heijden (1987).

2.2. Multi-way contingency tables

When we analyze a multi-way contingency table by means of CA, this table first has to be recoded to a two-way table. We will do this by stacking variables of the multi-way table, thereby reducing the number of ways of the multi-way table to two. Consider the three-way table P with elements p_{ijk} (categories of variable 3 are indexed by k (k=1,...,K)). If we stack variables 1 and 2, we denote the two-way matrix as $P^{(12)3}$, having elements $p_{(ij)k}$. This matrix has IxJ rows and K columns. We call such two-way tables "multiple tables". For a three-way table P we can construct two other multiple tables, namely $P^{(13)2}$ and $P^{(23)1}$. This notation generalizes straightforwardly to cases with more than three variables.

If we perform CA of multiple tables, we have to change our notation for the elements of the matrices slightly (compare (2)): when we analyze the matrix $P^{(12)3}$ with elements $p_{(ij)k}$, elements of $E^{(12)3}$ are

$$e_{(ij)k} = p_{(ij)+}p_{(++)k}. \tag{4}$$

Diagonal matrices D_r and D_c have elements $p_{(ij)+}$ and $p_{(++)k}$ respectively, and matrices R and C have elements $r_{ij\alpha}$ and $c_{k\alpha}$ respectively.

2.3. Loglinear analysis

Loglinear analysis is a statistical method to investigate the structural relations between the categorical variables in a multi-way contingency table. For an introduction we refer to Bishop et al. (1975). In loglinear analysis the logarithm of the frequencies is decomposed with a linear model. For a three-way matrix the general loglinear model can be written as where π_{ijk} is the probability for cell (i,j,k). In order to identify model (5) constraints have to be imposed upon the u-terms, for example to add up to zero for each index. Decomposition (5) shows that we can discern main effects $u_{1(i)}$, $u_{2(j)}$ and $u_{3(k)}$, first-order interactions $u_{12(ij)}$, $u_{13(ik)}$ and $u_{23(jk)}$, and second-order interaction $u_{123(ijk)}$.

The main purpose of loglinear analysis is to describe the proportions of a contingency table in a parsimonious way by reducing the number of u-terms as far as possible. This is done by subsequently deriving formulae for expected proportions under some restricted model, estimating these expected proportions, and testing the differences between the observed and the expected proportions using chi-squared tests. If the differences are not too large, the restricted model is thought to give an adequate description of the data.

There are different estimation procedures available to arrive at estimated expected proportions. An ML-procedure often used is iterative proportional fitting (IPF). Using IPF some of the margins of the observed proportions are fitted into a matrix of expected frequencies. For example, consider the model in which variables 2 and 3 are conditionally independent, given the level of variable 1. This corresponds to the loglinear model in which $u_{23(ij)}=0$ and $u_{123(ijk)}=0$ for each i,j,k. The expected proportions for this model can be derived using

$$\hat{\pi}_{ijk} = p_{ij+}p_{i+k}/p_{i++} \tag{6}$$

In (6) the observed margins p_{ij+} and p_{i+k} are fitted to the estimated values $\hat{\pi}_{ijk}$ so that $\hat{\pi}_{ij+}=p_{ij+}$ and $\hat{\pi}_{i+k}=p_{i+k}$. Therefore loglinear models are often denoted by placing the variables that constitute the highest fitted margins between square brackets: [12][13] for this model. Using IPF we restrict ourselves to so-called hierarchical models, i.e. when some margin is included in the model all margins that can be derived from this margin are also in the model since they are fitted implicitly. For example, by fitting p_{ij+} we have also fitted p_{i++} and p_{+j+}, and therefore the parameters $u_{1(i)}$ and $u_{2(j)}$ that correspond to these margins are also included in the model.

In (6) the estimation procedure only takes one iteration (i.e. the model can be estimated directly). Sometimes the model has to be estimated using more than one iteration, because so-called closed form estimates do not exist, such as is the case for model [12][13][23]. For this model, we fit the sets $\hat{\pi}_{ij+}=p_{ij+}$, $\hat{\pi}_{i+k}=p_{i+k}$, and $\hat{\pi}_{+jk}=p_{+jk}$ one by one until convergence (see Bishop et al., 1975, for a discussion of when closed form estimates do exist, and when they do not).

2.4. Correspondence analysis decomposing residuals from loglinear models

We will now focus on one specific type of relations between loglinear

loglinear models. This approach is discussed in Van der Heijden & De Leeuw (1985), Van der Heijden (1987), De Leeuw & Van der Heijden (1988) and Van der Heijden et al. (in press). For another approach we refer to Goodman (1981, 1986), who showed among others that the scores obtained with CA are under some conditions very similar to those in so-called RC-association models.

Considering equation (4) again, we find that the margins p_{ij+} and p_{++k} are fitted to the matrix $E^{(12)3}$. So $E^{(12)3}$ contains estimates under model [12][3]. Using this, equation (2) shows that CA of $P^{(12)3}$ can be interpreted as a decomposition of residuals from model [12][3]. We can conclude that CA of $P^{(12)3}$ does not show all interactions in the matrix, but only the interaction between 1 and 2 on the one hand, and 3 on the other. The first-order interaction between 1 and 2 is not displayed graphically. So we see the first-order interactions between 1 and 3, 2 and 3, and the second-order interaction between 1, 2 and 3 jointly.

We can use this when we want to choose a proper multiple table: when we are not in the first place interested in the relations between two of the three variables, we have to stack these two variables. When we are particularly interested in the relation of two variables with a third variable, we should not stack this third variable with one of the other two. In general, when we can distinguish explanatory variables and response variables, it follows that we should stack these distinct types of variables separately. In this way we see all interactions between the explanatory variables on the one hand and the response variables on the other hand, whereas the interactions in the set of explanatory variables and in the set of response variables are not displayed.

In dealing with four- or higher-way tables, the relation between CA and loglinear analysis still holds: for example, when we analyze the multiple tables $P^{(12)(34)}$ or $P^{(123)4}$, CA can be interpreted as decomposing the residuals from models [12][34] and [123][4] respectively. Residuals from less restricted loglinear models can be decomposed using the generalization of CA discussed in section 2.1.: in equation (3) we use multiple table $P^{(12)3}$, and $Q^{(12)3}$ contains expected proportions computed under some model less restricted than [12][3].

2.5. Conclusion

We have just introduced an analysis-of-residuals approach to CA of multi-way tables. The main advantage of this approach is that we have a better understanding of CA, in the sense that we know now which interactions in the multi-way table are displayed by CA. The generalization of CA can be used to suppress specific aspects of the data from the ordinary CA solution. For example, first an ordinary CA is performed on a multiple table, and it is noted that some interaction is suppressed from the solution by decomposing residuals from a less restricted loglinear model in which the first-order interaction is contained. For a related approach to the analysis of multi-way tables we refer to Takane (1987, 1989), who proposes to use ideal point discriminant analysis to combine both the assessment of the importance of interactions with loglinear analysis and the display of these interactions with CA into one data analysis procedure.

3. AVERAGE RESTRICTIONS IN CORRESPONDENCE ANALYSIS OF MULTI-WAY TABLES

3.1. Weighted averages equal to zero

We will now show that average restrictions on the category scores can be obtained by applying the analysis-of-residuals approach of CA. Consider the three-way matrix P of order 2x3x4 shown in table 1 (these data are artificial).

TABLE 1
Proportions multiplied by 1000

i	j	k = 1.	2.	3.	4.	Tot.
1	1	67	101	41	24	233
1	2	18	8	114	50	191
1	3	15	9	49	26	100
2	1	34	88	34	34	190
2	2	19	28	51	25	122
2	3	10	77	60	17	164
Tot.		163	311	350	176	1000

Variables 1 and 2 in the rows, variable 3 in the columns.

TABLE 2 Model [12][3], i.e. ordinary correspondence analysis of $P^{(12)3}$ (table 1)

		Table 2a: $E^{(12)3}$ multiplied by 1000				Table 2b: $P^{(12)3}-E^{(12)3}$ multiplied by 1000				Table 2c: $R\Lambda$ (dimension 1)	Table 2d: $D_rR\Lambda$ (dimension 1)
i	j	k=1	2	3	4	1	2	3	4		
1	1	38	73	82	41	29	28	-40	-17	-.462	-.108
1	2	31	59	67	34	-13	-51	47	17	.693	.132
1	3	16	31	35	18	-1	-22	15	8	.480	.048
2	1	31	59	66	33	3	29	-32	0	-.370	-.070
2	2	20	38	43	21	-1	-10	8	3	.200	.025
2	3	27	51	57	29	-16	25	3	-12	-.164	-.026

We will consider this matrix as a multiple table $P^{(12)3}$ of order 6x4. In table 2a we show the matrix $E^{(12)3}$, and in table 2b the matrix $(P^{(12)3}-E^{(12)3})$. We saw in section 2.3. that the proportions in $E^{(12)3}$ follow model [12][3], so that $p_{(ij)+}=e_{(ij)+}$, and $p_{(++)k}=e_{(++)k}$. Therefore the values in table 2b add rowwise and columnwise up to zero. As a result of this we find that, when we perform CA as given in equation (2), the weighted average of R and the weighted average of C is 0 for each dimension. For the first dimension row scores $R\Lambda$ are displayed in table 2c, and weighted values $D_rR\Lambda$ can be found in table 2d. See figure 1 for a two-dimensional plot.

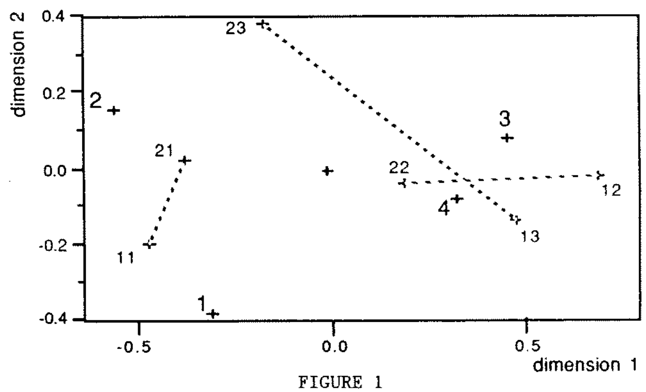

FIGURE 1

Ordinary CA of $P^{(12)3}$. Singular values are .45 (83%), .19 (14%) and .08 (3%). Row points have two numbers: first is category number for variable 1, second for variable 2. Corresponding categories for the second category are connected with dotted lines.

TABLE 3 Model [12][23]

		Table 3a: $Q^{(12)3}$ multiplied by 1000				Table 3b: $P^{(12)3}-Q^{(12)3}$ multiplied by 1000				Table 3c: $R\Lambda$ (dimension 1)	Table 3d: $D_rR\Lambda$ (dimension 1)
i	j	k=1	2	3	4	1	2	3	4		
1	1	56	104	42	32	12	-3	-0	-8	.002	.001
1	2	23	22	101	46	-4	-14	14	5	.166	.032
1	3	10	33	42	16	6	-23	8	9	.504	.051
2	1	45	85	34	26	-12	3	0	8	-.004	-.001
2	2	14	14	65	29	4	14	-14	-5	-.259	-.032
2	3	16	53	68	26	-6	23	-8	-9	-.309	-.051

Now consider the case that we want to decompose residuals from model [12][23], using the same multiple table $P^{(12)3}$. In table 3a we find estimates $Q^{(12)3}$ under [12][23]. Due to IPF $P_{(ij)+}=q_{(ij)+}$, and $P_{(+j)k}=q_{(+j)k}$. As a result of the latter equality the row vectors of $P^{(12)3}-Q^{(12)3}$ add up to zero for each j (see table 3b). When we decompose $P^{(12)3}-Q^{(12)3}$ with generalized CA (equation (3)), this results in scores $R\Lambda$ (for first dimension see table 3c), and weighted values $D_r R\Lambda$ in table 3d. Now we find that $\Sigma_i P_{(ij)+} r_{ij\alpha}=0$, due to the property that $P_{(+j)k}-q_{(+j)k}=0$. This is illustrated in figure 2.

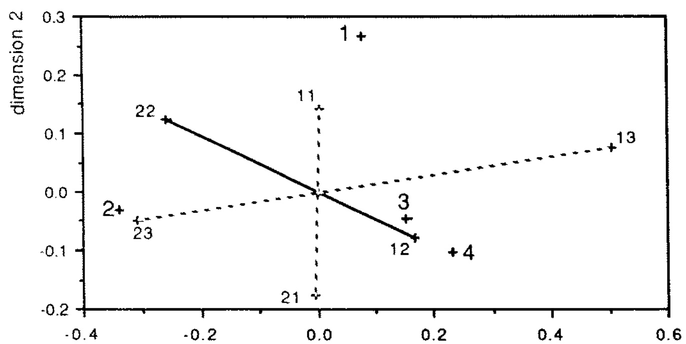

FIGURE 2

CA of departure from [12][23]. Singular values are .22 (76%), .12 (20%) and .05 (4%). Corresponding categories for the second category are connected with lines. The weighted averages fall into the origin.

TABLE 4 Model [12][13][23]

	Table 4a: $Q^{(12)3}$ multiplied by 1000				Table 4b: $P^{(12)3}-Q^{(12)3}$ multiplied by 1000				Table 4c: $R\Lambda$ (dimension 1)	Table 4d: $D_r R\Lambda$ (dimension 1)
i j	k=1	2	3	4	1	2	3	4		
1 1	66	82	49	36	1	2	-8	-12	-.200	-.047
1 2	24	15	106	46	-5	-7	8	14	.101	.019
1 3	11	21	49	19	4	-12	0	8	.272	.027
2 1	35	107	26	22	-1	-19	8	12	.246	.047
2 2	13	21	59	29	5	7	-8	-4	-.158	-.019
2 3	14	65	60	25	-4	12	-0	-8	-.167	-.027

Now consider model [12][13][23]. The estimates $Q^{(12)3}$ are given in table 4a. Due to IPF, $P_{(ij)+}=q_{(ij)+}$, $P_{(i+)k}=q_{(i+)k}$ and $P_{(+j)k}=q_{(+j)k}$. As a result the row vectors of $P^{(12)3}-Q^{(12)3}$, displayed in table 4b, add up to zero both for each i and for each j. When we decompose the difference $P^{(12)3}-Q^{(12)3}$ with generalized CA, this results in scores $R\Lambda$ (for first dimension see table 4c), and weighted values $D_r R\Lambda$ in table 4d. Now not only $\Sigma_i P_{(ij)+} r_{ij\alpha}=0$, but also $\Sigma_j P_{(ij)+} r_{ij\alpha}=0$. The solution is shown in figure 3.

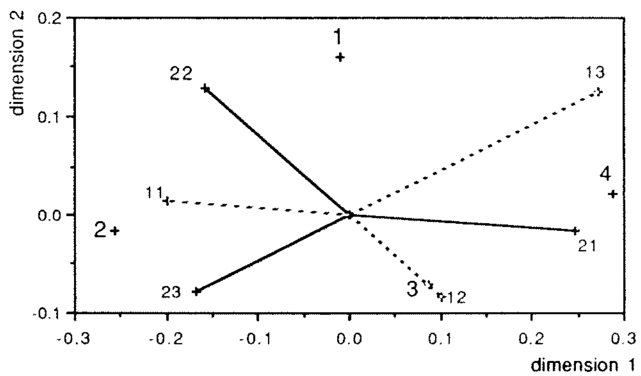

FIGURE 3

CA of departure from [12][13][23]. Singular values are .19 (86%) and .08 (14%). Both corresponding categories for the first variable are connected (a dotted triangle for the first category) as well as corresponding categories for the second variable (dotted line changing in solid line in the origin). The weighted averages for both variables fall into the origin.

3.2. Maximum dimensionality

In ordinary CA, the maximum dimensionality (i.e. the dimensionality to obtain perfect reproduction in general) of the solution for $P^{(12)3}$ is min ((IJ-1),(K-1)), i.e. for the example above the dimensionality is 3 because the four columns of $P^{(12)3}-E^{(12)3}$ add up to 0. By decomposing residuals from [12][23], the dimensionality becomes min ((J(I-1)),(K-1)). This can be easily seen from our example: in table 3b we find row vectors adding up to zero for each j. This results in a dimensionality of 3, due to the four columns adding up to 0 and due to the 6 rows adding up to zero for each j. For model [12][13][23] the dimensionality is min ((I-1)(J-1),(K-1)), being 2 for our example. This is easily verified from table 4b.

3.3. Chi-squared distances

In section 2.1. we presented CA using the reconstitution formula. It is also possible to present CA as a technique that represents profiles of rows and columns as points in a weighted euclidean space (compare Greenacre, 1984). We will explain this in more detail for the rows. The matrix $D_r^{-1}P$ (see table 5a)

TABLE 5
Tables with row profiles: rowwise, the matrices add up to 1

i j	Table 5a: $D_r^{-1}P$ multiplied by 1000 k=1	2	3	4	Table 5b: $D_r^{-1}Q$, [12][23] multiplied by 1000 1	2	3	4	Table 5c: $D_r^{-1}Q$, [12][13][23] multiplied by 1000			
1 1	288	432	176	104	238	447	178	137	282	351	212	155
1 2	95	42	599	264	118	114	528	240	124	78	556	242
1 3	154	94	497	255	97	327	416	161	114	214	494	178
2 1	177	465	181	177	238	447	178	137	185	564	137	114
2 2	154	225	418	203	118	114	528	240	109	169	484	238
2 3	62	469	366	102	97	327	416	161	86	396	368	150
Mean	163	311	350	176								

is the matrix with row profiles, and the IxJ row points are represented in a weighted euclidean space of dimension K with metric D_c^{-1}, where the profile values are used as coordinates of the points. The marginal proportions in D_r define the masses attached to each row point. The thus obtained space can be centered by subtracting from each row profile in $D_r^{-1}P$ the average row profile, i.e. the profile of the column totals of P. This is done by subtracting $D_r^{-1}E$ from $D_r^{-1}P$: all rows of $D_r^{-1}E$ are equal to the profile of the column totals of P (see column margins of table 5a). Thus the origin can be interpreted as the average row profile, and distances of points to the origin reflect the departure of their profiles from the average profile. The solution for the column points shows for which columns an observed profile is larger than the average profile, i.e. when $p_{(ij)k}/p_{(ij)+} > p_{(++)k}$.

Now consider CA in which the departure from [12][23] and [12][13][23] is studied. From section 2.3. we know that these solutions can be obtained with ordinary CA by analyzing the matrices (P-Q+E), where elements of Q follow model [12][23] or [12][13][23]. So one way to interpret the thus obtained solutions is simply by considering chi-squared distances for the matrix (P-Q+E).

However, as Escofier & Drouet (1983) have shown, there is a much more attractive interpretation for the analysis of the departure from [12][23]. First notice that for this analysis the metric for the row points is still D_c^{-1}, and the points have weights defined by D_r. For our example in section 3.1. the first row of matrix $D_c^{-1}Q$ is equal to the fourth row, the second to the fifth, and the third to the sixth (see table 5b). This shows that, by using the matrix $D_r^{-1}(P-Q)$ as coordinates, the rows are centered by placing the weighted average for each cluster of the I categories having j in common into the origin. Compared to ordinary CA, this solution is obtained by translating weighted averages for clusters to the origin (see figure 1 and 2). Thus distances within each cluster remain identical, and distances between points in different clusters become different. For our example, distances between points 11 and 21, 12 and 22, and 13 and 23 remain the same in full-dimensional space, but now these distances are represented better in the first dimensions of CA, because the between cluster differences are eliminated. The origin of this space should be interpreted as the average for each category of variable 2, and distances of points to the origin reflect the departures from this average. From the solution of column points it can be derived for which columns the observed row vectors are smaller or larger than those expected under model [12][23].

For model [12][13][23] the distances (and hence the interpretation) are not related to the ordinary CA in the way the distances for model [12][23] are. When we compare the matrix $D_r^{-1}E$ (profiles for model [12][3]) with the matrix $D_r^{-1}Q$ for model [12][23], we see that the profiles differ in a very simple way. However, there is not an easy way to go from the matrix $D_r^{-1}Q$ for model [12][23] to the matrix $D_r^{-1}Q$ for model [12][13][23] (displayed in table 5c). This is due to the fact that in the latter case the matrix Q has to be fitted iteratively: it is impossible to restrict three margins of Q directly.

On the other hand, the center of the space has a clear interpretation. It is the point of the average for the clusters of the categories of both the first and the second variable. Distances of an individual point to the origin can be interpreted as reflecting differences of this individual profile from these averages. For our example, point 11 in figure 3 shows that 11 occurs more often than average for variable 1 and for variable 2 with category 2 of variable 3 (compare table 4b).

3.4. Tables with more than three ways

For tables with more than three ways, similar relations hold. As an example, let us take a five-way table P, coded as multiple table $P^{(123)(45)}$, having elements $p_{(ijk)(lm)}$. Row scores are denoted as $r_{ijk\alpha}$, column scores as $c_{lm\alpha}$. Ordinary CA decomposes residuals from model [123][45]. Row scores are restricted as $\Sigma_i\Sigma_j\Sigma_k p_{(ijk)++} r_{ijk\alpha}=0$, column scores as $\Sigma_l\Sigma_m p_{(+++)(lm)} c_{lm\alpha}=0$.

When we want to restrict column scores further so that, for example, $\Sigma_l p_{(+++)(lm)} c_{lm\alpha}=0$, we should decompose residuals from model [1235][45], i.e. we have to fit the four-way margin including the three variables that constitute the rows, plus the variable for which we want the weighted averages to fall into the origin. In the same way, when we want the row scores to be restricted as $\Sigma_j\Sigma_k p_{(ijk)(++)} r_{ijk\alpha}=0$, we should decompose residuals from [123][145], etc.. It is also possible to have both restrictions to the rows and to the columns. As an example, by decomposing residuals from [1235][145], we will find the average restrictions $\Sigma_j\Sigma_k p_{(ijk)(++)} r_{ijk\alpha}=0$ and $\Sigma_l p_{(+++)(lm)} c_{lm\alpha}=0$. The idea is simply to construct a multi-way matrix Q in which some margins are equal to those in the matrix P. These margins should be chosen in such a way that, if we construct the multiple tables $P^{(123)(45)}$ and $Q^{(123)(45)}$, and if we add up over the categories of some specific row variable in the matrix $(P^{(123)(45)}-Q^{(123)(45)})$, then we will find values equal to zero.

It should be noticed that, by restricting only the column scores so that $\Sigma_l c_{lm\alpha}=0$, we eliminate more than one type of interaction. In ordinary CA we studied the departure from [123][45], and in this analysis the residuals from [1235][45]. In doing so, we eliminate (compared to ordinary CA) the u-terms u_{15}, u_{25}, u_{35}, u_{125}, u_{135}, u_{235}, u_{1235}. We might, of course, only eliminate u_{34} by studying residuals from [123][34][45]. However, this does not result in a CA with average restrictions. The reason is that we do not find a matrix with values equal to zero when we add up over the categories of either variable 3, or 4, but only if we add up over both 3 and 4. So models like [123][34][45] are not useful for average restrictions, although they are still useful in an analysis of residuals approach of CA.

3.5. Conclusions

By using loglinear models, we can obtain CA with average restrictions on the category points. This is especially useful as a method to suppress some interactions in the data, so that the other interactions become more clearly visible.

We emphasize that, in our opinion, it is a good idea to use chi-square tests to assess the importance of the deviation from the model (i.e. if the assumptions for these tests are satisfied). If the tests are significant, then the model does not fit adequately, and it makes sense to investigate the structure in the residuals of the model. Even if the model seems to fit adequately, CA might still be useful, since the power of the test might be low (see Agresti, 1984, p.82-83).

4. MULTIPLE CORRESPONDENCE ANALYSIS OF MULTI-WAY INDICATOR MATRICES

4.1. Ordinary multiple correspondence analysis

We will now discuss CA of another type of multi-way table, namely the multi-way indicator matrix. Since CA of (super-)indicator matrices is also

known as MCA (see, for example, Greenacre, 1984), we will speak in the sequel of MCA.

Multi-way indicator matrices will be introduced for the three-way case. In a three-way indicator matrix the rows correspond to objects (e.g. persons), the columns to the categories of a single variable, and the layers to the situations (e.g. time-points). We denote this matrix as Z, having elements z_{ijt}. Usually we code $z_{ijt}=1$ if object i (i=1,...,I) falls into category j (j=1,...,J) of situation t (t=1,...,T), and $z_{ijt}=0$ else. A special property of this matrix is that $z_{i+t}=1$. So a three-way indicator matrix summarizes the categorical measures on one variable in different situations. The idea of a three-way indicator matrix was first introduced in Saporta (1981), see also De Leeuw e.a. (1985), Van der Heijden (1987).

One way to do MCA of the three-way indicator matrix is by coding it in two-way form, as we did for ordinary contingency tables. In principle there are three ways to do this. We can stack the objects and the categories, by creating the matrix $Z^{(ij)t}$. However, it is not clear what purpose would be served by an analysis of such a matrix, and we therefore skip this possibility. Secondly, it is possible to stack the objects and the situations, yielding a so-called LONG matrix $Z^{(it)j}$ (Visser, 1985). This matrix, however, has one 1 in each row, the other values being zero, and it is well-known that CA of such a matrix yields a trivial solution with all singular values equal to 1.

What remains therefore is the so-called BROAD matrix $Z^{i(jt)}$, in which each object has a row, and each combination of categories and situations has a column. In fact, this comes down to ordinary MCA of the matrix with T variables (situations). We will consider the analysis of this matrix in some more detail. In order to be able to use the formulas of CA introduced in section 2.1., we replace each 1 in Z by 1/IT, so that $z_{+++}=1$.

Now consider formula (2). The matrix D_r can be replaced by a unit-matrix multiplied by 1/I: each object has the same weight. Another remarkable point is that $E^{i(jt)}$ has a very special structure, due to the fact that $z_{i(+t)}=1/IT$. Therefore values $e_{i(jt)}=z_{i(++)}z_{+(jt)}$ also have the property that $e_{i(+t)}=1/IT$. This is a remarkable point: we do not only find that $z_{i(++)}=e_{i(++)}$ and $z_{+(jt)}=e_{+(jt)}$, but also $z_{i(+t)}=e_{i(+t)}$! This implies that for the column scores $\Sigma_j z_{+(jt)} c_{jt\alpha}=0$ holds (see section 3.1.). In fact, this is well known for MCA, namely that the weighted average of the scores is zero for each variable.

Let us now consider this using loglinear analysis concepts. Of course, we cannot use models in this context to test anything, since it is not natural to consider the values in Z as proportions or frequencies. However, we can use the loglinear concept of interactions to obtain a better understanding of which aspects of the data are displayed in MCA.

Due to fitting $z_{+(jt)}$ and $z_{i(+t)}$, we can think of this MCA as an analysis decomposing the departure from [IT][JT]. Clearly, the margins corresponding to [IT] are uninteresting, but the margins corresponding to [JT] are the marginal frequencies of the variable in the different situations. This is often interesting information, but MCA does not display it. Only the departure of the objects from these margins is displayed. An ordinary MCA of the BROAD matrix only shows the "first-order interaction" between objects and categories (e.g., irrespective of the situation, some objects choose more often category 1 than average for all objects), and the second-order interaction (e.g. some objects choose more often category 1 at situation 2 than the average for all objects). A dominating first-order interaction will result in, for example, a solution in which all categories 1 have negative scores and all categories 3 have positive scores on the first dimension.

As for ordinary contingency tables, we can also decompose the departure from [IJ][IT][JT] for three-way indicator matrices. In doing so, we eliminate the "first order interaction" between objects and categories, that corresponds to the margins $z_{i(j+)}$. By eliminating this interaction, the weighted average for each category will be zero, i.e. $\Sigma_t z_{+(jt)} c_{jt\alpha}=0$.
This might be useful when the relation between objects and categories is dominating the solution. By decomposing the departure from [IJ][IT][JT] it is only shown how objects differ in their use of categories at different situations, irrespective of their preference for specific categories.

4.2. Multi-way indicator matrices

If the indicator matrix has more than three ways, matters become only slightly more complicated. The main principles for decomposing departure from less restricted models are already described in section 3.4.: we use IPF in order to make some margins in multi-way Q equal to those in the observed multi-way Z. Subsequently, ways of Z and Q are stacked in order to get two-way tables. Depending on the margins in Z and Q that are equal, weighted averages will be restricted to be zero. We will discuss one typical example shortly.

When the objects are already stacked, e.g. in couples i having a husband and a wife indexed with s, we can analyze the BROAD matrix with elements $z_{is(jt)}$. It follows from 4.1. that in ordinary MCA the departure from [IST][JT] will be decomposed, since $z_{is(+t)}$ is constant. The generalization can be used in order to suppress specific interactions, for example the relation between the sexes and the categories. This might be useful when interest goes out to differences between couples, but sex differences dominate the first dimension. By decomposing departure from [IST][SJT] the weighted average for each sex becomes zero: $\Sigma_i z_{is(++)} r_{is\alpha}=0$.

4.4. Conclusions

As for ordinary contingency tables, we think that two loglinear analysis concepts can be useful in the context of multi-way indicator matrices: firstly, the interaction concept gives us a better understanding of what types of interaction of the data are displayed by MCA. Secondly, by using less restricted models than the model used for ordinary MCA, weighted averages can be set equal to zero. Thus aspects we are not interested in can be eliminated from ordinary MCA.

5. MULTIPLE CORRESPONDENCE ANALYSIS OF MULTI-WAY SUPER-INDICATOR MATRICES

5.1. Ordinary multiple correspondence analysis

We can code data into a super-indicator matrix when we deal with more than one variable. A super-indicator matrix is a concatenation of indicator matrices, where each variable has its own indicator matrix (compare section 4.1.). A super-indicator matrix has objects i in the rows, and each category of each variable receives a separate column, indexed by j. We speak of a three-way super-indicator matrix Z when there are layers t, indexing the different situations. When the number of variables equals m, we fill in (comparable to what we did in section 4.1.) $z_{ijt}=1/mIT$ if an object falls into a category on a particular occasion, and $z_{ijt}=0$ if not. In doing so, $z_{+++}=1$, so that we can use the CA formulas of section 2.1.. The matrix Z has the special property that $z_{i+t}=1/IT$.

The difference with section 4. is that for MCA of three-way superindicator matrices, we now deal with the situations of more variables, and therefore we can not only analyze BROAD tables by concatenating the layers horizontally, but also analyze LONG tables by concatenating them vertically. Now the LONG matrix does not yield a trivial solution, since there is more than one value larger than zero in each row. The BROAD matrix is denoted again as $Z^{i(jt)}$, and the LONG matrix as $Z^{(it)j}$.

For the BROAD matrix we find that $z_{+(jt)}=e_{+(jt)}$ and $z_{i(+t)}=e_{i(+t)}$ (as in section 4.1.). We know that for each variable in this MCA the weighted average of the category scores equals zero (see, for example, Greenacre, 1984). We can think of MCA as an analysis decomposing the departure from [IT][JT]. As before, the margins corresponding to [IT] are uninteresting, the margins corresponding with [JT] are now the marginal frequencies of all variables at different situations. An ordinary MCA solution of the BROAD matrix only displays the first-order interaction between objects and categories (e.g. irrespective of the occasion, some objects fall more often in category 1 of variable 2 than average), and the second-order interaction (e.g. some objects fall more often in category 1 of variable 2 at occasion 2 than average).

For the LONG matrix $Z^{(it)j}$ the departure from [IT][J] is decomposed. So the only interaction not displayed is the interaction between objects and time-points, but we just saw that this interaction was uninteresting anyway. So all the other interactions are displayed into one solution. In this one solution we find the first order interaction between objects and categories (e.g. some objects use more often category 1 than average), the first order interactions between situations and categories (e.g. at time-point 2 category 1 is used more than at time-point 1), and second order interaction (the relation between categories and time-points differs for different objects).

5.2. Decomposing the departure from less restricted models

It will be clear that the same principles can be used in this context as were used in section 3. and 4. We therefore do not consider any less restricted models in the context of the BROAD matrix, and discuss only one alternative for the LONG matrix.

In analyzing the LONG matrix with ordinary MCA all types of interactions are displayed, since the departure from [IT][J] is decomposed. This is often not a very fortunate situation: the relation between time-points and categories can be so strong that the other relations remain hidden. For this reason it can be useful to eliminate this interaction. This can be done by studying the departure from model [IT][JT]. In this way the weighted average for each time-point is made equal to zero, i.e. $\Sigma_i z_{(it)} + r_{it\alpha} = 0$.

This approach is in fact a special case of conditional MCA in which it is allowed that the objects of the LONG matrix come from different samples, and in which each sample can have an unequal size. In conditional MCA in fact the departure from [I][J] is decomposed in each of the T samples (note that values under [I][J] will be different in each of the T samples). We refer to Escofier (1987) for more details.

In longitudinal data analysis a procedure comparable to the approach adopted here is often applied in the context of principal component analysis or factor analysis of LONG matrices. This procedure is that first in each layer the variables are set in deviations from zero, so that differences in mean between time-points do not influence the correlations between two variables. It will be clear that the approach for MCA presented here is identical to this procedure, when we keep in mind that MCA can be considered as a generalization of principal components analysis (see Gifi, 1981).

5.3. Conclusions

As in section 4., we have pointed out the usefulness of loglinear model concepts in the context of MCA. The main difference with the univariate case discussed there is that in the multivariate case it is possible to analyze the data using the LONG matrix. Again we find that loglinear analysis concepts can be used to obtain a better understanding of the types of interactions that are displayed by MCA, and to obtain average restrictions.

6. GENERAL CONCLUSIONS

As was emphasized in earlier conclusions, we think that by using loglinear model concepts in the context of CA and MCA we get a better understanding of which interactions in the data are displayed by these techniques. Furthermore, this gives us an easy way to impose average restrictions to CA and MCA: we do not need any new computer programs for it, but instead can use ordinary CA programs if we modify the matrix to be analyzed: we just should analyze the matrix (P-Q+E) instead of P. The matrix Q can easily be obtained with IPF algorithms (a fast version is implemented in SAS). For problems with CA programs that do not allow for non-integers or negative entries, we refer to Van der Heijden (1987, appendix A), where some tricks are described that can be used to circumvent these problems.

REFERENCES

Agresti, (1984) The analysis of ordinal categorical data. New York: Wiley.

Bishop, Y.M.M., Fienberg, S.E. and Holland, P.W. (1975) Discrete Multivariate Analysis. Cambridge: M.I.T. Press.

Daudin, J.-J. and Trécourt, P. (1980) Analyse factorielle des correspondances et modèle log-lineaire: comparaison des deux méthodes sur un exemple. Revue de Statistique Appliquée, **26**, 15-29.

De Leeuw, J. and Van der Heijden, P.G.M. (1988) Correspondence analysis of incomplete tables. Psychometrika, **53**, 223-233.

De Leeuw, J., Van der Heijden, P.G.M., and Kreft, I. (1985) Homogeneity analysis of event history data. Methods of Operations Research, **50**, 299-316.

Escofier, B. (1983) Analyse de la difference entre deux mésures sur le produit de deux mêmes ensembles. Cahiers de l'Analyse des Données, **8**, 325-329.

Escofier, B. (1987) Analyse des correspondances multiples conditionnelles. Data analysis and informatics 5. Amsterdam: North-Holland.

Escofier, B. and Drouet, D. (1983) Analyse des différences entre plusieurs tableaux de fréquence. Cahiers de l'Analyse des Données, **8**, 491-499.

Gifi, A. (1981) Non-linear multivariate analysis. Leiden: Leiden University, Department of Data Theory.

Goodman, L.A. (1981) Association models and canonical correlation in the analysis of cross-classifications having ordered categories. Journal of the American Statistical Association, **76**, 320-334.

Goodman, L.A. (1986) Some useful extensions of the usual correspondence analysis approach and the usual log-linear model approach in the analysis of contingency tables. International Statistical Review, **54**, 243-309.

Greenacre, M.J. (1984) Theory and applications of correspondence analysis. New York: Academic Press.

Heiser, W.J. (1987) Joint ordination of species and sites: the unfolding technique. In Developments in numerical ecology (P. and L. Legendre, eds.). Berlin: Springer.

Israëls, A. and Sikkel, D. (1982) Correspondence analysis and comparisons with other techniques. Voorburg: Central Bureau for Statistics.

Lauro, N.C. and Decarli, A. (1982) Correspondence analysis and loglinear models in multiway contingency tables study. Some remarks on experimental data. Rivista Internazionale di Statistica (Metron), **40**, 214-234.

Nishisato, S. (1980) Analysis of categorical data: dual scaling and its applications. Toronto: University of Toronto Press.

Nishisato, S. (1984) Forced classification: a simple application of a quantification method. Psychometrika, **49**, 25-36.

Saporta, G. (1981) Méthodes exploratoires d'analyse de données temporelles. Cahiers de Buro no. 37-38. Paris.

Takane, Y. (1987) Analysis of contingency tables by ideal point discriminant analysis. Psychometrika, **52**, 493-514.

Takane, Y. (1989) Ideal point discriminant analysis: implications for multiway data analysis. (This volume).

Ter Braak, C.J.F. (1986) Canonical correspondence analysis. Ecology, **67**, 1167-1179.

Van der Heijden, P.G.M. (1987) Correspondence analysis of longitudinal categorical data. Leiden: D.S.W.O.-Press.

Van der Heijden, P.G.M., De Falguerolles, A. and De Leeuw, J. (in press). A combined approach to contingency table analysis using correspondence analysis and loglinear analysis. Applied Statistics.

Van der Heijden, P.G.M. and De Leeuw, J. (1985) Correspondence analysis used complementary to loglinear analysis. Psychometrika, **50**, 429-447.

Visser, R.A. (1985) The analysis of longitudinal data in behavioural and social research. Leiden: D.S.W.O.-Press.

Yanai, H. (1986) Some generalizations of correspondence analysis in terms of projection matrices. In Data analysis and informatics IV (E. Diday et al., eds.). Amsterdam: North-Holland.