RR – 85-19

QUASI-CORRESPONDENCE ANALYSIS

Jan de Leeuw
Department of Data Theory

Peter G.M. van der Heijden
Department of Psychometrics and Research Methodology

University of Leiden

QUASI-CORRESPONDENCE ANALYSIS

Jan de Leeuw* & Peter G.M. van der Heijden**

Abstract

It is a well-known fact that correspondence analysis can be described as a technique which decomposes the departure from independence in a two-way contingency table. In this paper a technique is decomposed with which the departure from quasi-independence can be decomposed. We name this technique quasi-correspondence analysis. Quasi-correspondence analysis seems to be a good alternative to correspondence analysis in cases that the use of the latter should not be recommended, e.g. in case of structural zeros. It is shown that one form of quasi-correspondence analysis is formally identical to Nora's reconstitution of order zero. Furthermore, it is shown how quasi-correspondence analysis can be performed using existing correspondence analysis programs. We discuss several examples.

Keywords: Correspondence analysis, data analysis, quasi-independence, reconstitution of order zero, structural zeros.

*Dept. of Data Theory, Faculty of Social Sciences, University of Leiden, Middelstegracht 4, 2312 TW Leiden, The Netherlands. Tel: 071-148333-ext. 2255.

**Dept. of Methodology, Subfaculty of Psychology, Faculty of Social Sciences, University of Leiden, Hooigracht 15, 2312 KM Leiden, The Netherlands. Tel: 071-148333-ext. 6381

1. Introduction

In this paper we introduce a modification of correspondence analysis (CA, from now on) which can be used in combination with the quasi-independence models familiar from log-linear analysis. The technique we propose decomposes the residuals that are left after fitting a quasi-independence model. The decomposed residuals are represented geometrically. Thus our paper interprets CA as a technique which can be used complementary to log-linear analysis. A similar approach has been adopted by Daudin & Trécourt (1980), Israëls & Sikkel (1982), and Lauro & Decarli (1982). It is also possible to think of CA as a model in its own right. This is the approach taken by Goodman (1985), for example.

The French approach to CA, originated by Benzécri (1973, 1980), and described in considerable detail by Greenacre (1984), interprets CA as a multi-dimensional scaling technique which makes pictures of data matrices. No model is involved. Although we think that the model-free interpretation of CA is in many cases the most natural one, we also think that combination and comparison with current modelling approaches for frequency tables is quite useful. This was illustrated in Van der Heijden (1985, 1986) for transition matrices, and in Van der Heijden and De Leeuw (1985) for multi-way tables.

In the complementary interpretation of CA we study it as a technique to represent residuals of a log-linear analysis in a picture. Both the geometrical and the modelling aspects are present in this approach, but clearly the modelling is predominant. We only apply CA on the variation that is left after the model is fitted. A model with a good fit leaves very little variation, and thus CA will be quite useless in such cases. This is more or less true by definition: a model fits well if there is no systematic variation in the residuals. As a consequence CA is most useful in combination with models that do not fit well. Thus we must combine the use of CA with the use of fairly restrictive models. Classical CA is, in our interpretation, complementary to the complete independence model, which is of course highly restrictive.

The technique in this paper is called quasi-correspondence analysis

(QCA) for two reasons. In the first place it analyses residuals from the quasi-independence model. In the second place there are purists who argue that CA is a very specific technique, and that techniques which differ from it are consequently not CA. If we use such strict criteria, then our technique is not really CA, but it has a number of important features in common with it. Thus it is perhaps better to use a name such as QCA for this reason too.

2. Quasi-independence models

Before we proceed with a discussion of CA, and our generalisation of it, we briefly outline the quasi-independence model for two-way tables. For a much more complete discussion we refer to Caussinus (1965), Mosteller (1968), Goodman (1968), Bishop, Fienberg and Holland (1975, pp. 177-210), and Haberman (1979, pp. 444-486).

The quasi-independence model is a generalization of the complete independence model to incomplete tables. Tables can be incomplete for various reasons. In the first place, of course, we may not know entries of some cells. This often happens in secondary analysis. It can also happen because some data are only available in some regions, or for some years, or for some groups, and not for others. Thus data are missing. In principle they could be collected, or could have been collected, but for some reason or another they were not. Since we cannot use the empty cells in our calculations, we need some form of adaptation of the usual statistical analysis. The "usual" statistical analysis, by the way, is computing the chi-square statistic for testing independence of the row- and column-classification.

Tables can also be incomplete because observations cannot possibly occur in given cells, because these cells pertain to events that cannot logically occur. If we cross 'age-at-first-marriage' with 'current age', then cells for which age-at-first-marriage is larger than current age are obviously empty. If we analyze import-export tables between countries, then the diagonal cells of the table are, by definition, empty. Such empty cells are often called structural zeros. They must be distinguished both from observed zeros, which happen to

be zero because the sample is not large enough, and from missing data, which happen to be zero because relevant data were not collected. Again classical chi-square techniques cannot be used.

A third possibility is that 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. Thus the table need not be incomplete in this case. This happens in transition matrices, in confusion matrices, and in social mobility tables, in which the diagonal elements usually require separate parameters. Interaction matrices with a structural zero diagonal, such as input-output tables, or migration tables, are very similar, because in that case we also do not model the (structural) zero.

A final application, related to the previous one, is that we decide a posteriori that some cells will not be restricted. Thus we first perform a classical chi-square analysis, for example, and discover a large residual which influences the results in a dramatic way. We eliminate the outlying cell by not requiring independence for this cell, and repeat the analysis. This is, of course, a form of 'data snooping', and it should be used with great care. As a technique to eliminate dominating residuals it seems fairly promising, however.

The mathematical form of the quasi-independence model is very simple. Suppose that a complete table p_{ij} is observed. We want to fit this table with a model π_{ij} which assumes that $\pi_{ij} = \alpha_i\beta_j$ for all (i,j) in a given set of index pairs K. The π_{ij} with (i,j) not in K are unrestricted. The multinomial likelihood equations are simply

$$(1a)\quad \Sigma\{p_{ij} \mid j \in J(i)\} = \Sigma\{\pi_{ij} \mid j \in J(i)\},$$

$$(1b)\quad \Sigma\{p_{ij} \mid i \in I(j)\} = \Sigma\{\pi_{ij} \mid i \in I(j)\}.$$

Here we have written J(i) for those j for which (i,j) is in K, and I(j) for those i for which (i,j) is in K. Thus (1) tells us that the marginals over the restricted cells of the observed and expected table must be the same. For the (i,j) not in K we find simply $\pi_{ij} = p_{ij}$, which means that they are estimated by substituting observed cell

entries. In the case of structural zeros or missing data we do not estimate π_{ij} if (i,j) is not in K.

The likelihood equations (1a) and (1b) can be written in yet another way. Define a matrix R with $r_{ij} = p_{ij}$ for all (i,j) in K, and with $r_{ij} = \alpha_i\beta_j$ for all (i,j) not in K. Then

$$(2)\quad \Sigma_{j=1}^{m}\, r_{ij} = \Sigma\{p_{ij} \mid j \in J(i)\} + \alpha_i\Sigma\{\beta_j \mid j \notin J(i)\}.$$

Suppose we replace indices over which we have summed by a '+'. Then substituting from (1a) we find that the maximum likelihood estimate $r_{i+} = \alpha_i\beta_+$, and in the same way $r_{+j} = \alpha_+\beta_j$. This can also be written as

$$(3)\quad r_{ij} = r_{i+}r_{+j}/r_{++},$$

which must be true for all (i,j) not in K. Conversely if we have an R which satisfies (3) for all (i,j) not in K, and which satisfies $r_{ij} = p_{ij}$ for all (i,j) in K, then we can define $\pi_{ij} = r_{i+}r_{+j}/r_{++}$ for all (i,j) in K and prove easily that these π_{ij} satisfy (1a) and (1b).

3. Algorithms for computing maximum likelihood estimates

The 3 x 3 matrix in table 1 is taken from Reynolds (1977, p. 25). Columns indicate vote of husband in the 1968 elections, rows indicate the vote of his spouse. The chi-square for independence is 5353.33, with 4 degrees of freedom. We have the feeling that this mainly reflects the very large diagonal entries of the tables, and we want to investigate whether votes of husbands and wifes are independent if it is known that they vote differently. Thus we fit a quasi-independence model, in which K is the set of all six index pairs corresponding with off-diagonal cells.

The first algorithm we discuss is of the 'iterative proportional fitting' type. It is suggested directly by the form (1) in the likelihood equations. Start with a table in which the π_{ij} for (i,j) not in K are equal to their observed values, and the π_{ij} for (i,j) in

Table 1: Relationship between Respondents' Votes and Spouses' Votes in 1968 Presidential Election.

		Respondent's spouse vote		
		Nixon	Humphrey	Wallace
respondents vote	Nixon	1586	117	49
	Humphrey	103	1540	40
	Wallace	34	17	359

K are equal to $\alpha_i\beta_j$ for some choice of α and β. Now, for each i, multiply all elements π_{ij}, with $j \in J(i)$, by a constant in such a way that (1a) is satisfied for row i. If this has been done for all i, rows add up to the correct numbers, but columns will not. Repeat the same procedure for columns. This will undo the correct sums of the rows again, so we renormalize rows as in the next step. And so on, until convergence.

In table 2 we have given iterations 0, 1, 5, 10 of this procedure, together with the table to which it converges. In general convergence is slow, but sure. The chi-square for quasi-independence is 2.53, with one degree of freedom.

There is a second, somewhat less familiar, algorithm for computing the maximum likelihood estimates. It iterates on the unrestricted or unknown elements of the table, and not on the pairs (i,j) in K. We start with a matrix R_0 in which the r_{ij} for (i,j) in K are equal to their observed values. The other r_{ij} are arbitrary. We then iterate by $r_{ij}^{(m+1)} = r_{i+}^{(m)} r_{+j}^{(m)} / r_{++}^{(m)}$ for all (i,j) not in K. The r_{ij} for (i,j) in K remain fixed at their observed values. Note that we use the subscript + here for sums over all indices in a row or column.

Table 3 gives selected iterates of this algorithm, together with the point of convergence. Convergence of this method is about equally fast

Table 2: Selected iterations of off-diagonal algorithm

iterate 0		
1586.00	1.00	1.00
1.00	1540.00	1.00
1.00	1.00	359.00
iterate 1		
1585.00	102.51	47.81
100.98	1540.00	41.19
36.02	31.49	359.00
iterate 5		
1586.00	112.47	52.49
107.43	1540.00	36.51
29.57	21.53	359.00
iterate 10		
1586.00	112.81	53.12
107.19	1540.00	35.88
29.81	21.19	359.00
optimum		
1586.00	112.83	53.17
107.17	1540.00	35.83
29.83	21.17	359.00

Table 3: Selected iterations of diagonal algorithm

iterate 0		
1586.00	117.00	49.00
103.00	1540.00	40.00
34.00	17.00	359.00
iterate 1		
785.10	117.00	49.00
103.00	732.73	40.00
34.00	17.00	47.77
iterate 5		
193.07	117.00	49.00
103.00	113.90	40.00
34.00	17.00	7.19
iterate 10		
160.32	117.00	49.00
103.00	76.96	40.00
34.00	17.00	9.84
optimum		
159.04	117.00	49.00
103.00	76.03	40.00
34.00	17.00	9.97

as that of the previous one, the program is somewhat simpler. Convergence of both algorithms is proved easily by majorization methods, such as those used in proving convergence of the EM-algorithm, together with general results on the uniqueness of the maximum likelihood estimates (Haberman, 1974). It may not be apparent from tables 1 and 2 that they converge to the same point, but we must realize that table 1 computes the off-diagonal elements of $\alpha_i\beta_j$ while table 2 computes the diagonal elements. Thus the ordinary chi-square test for independence, applied to table 2, gives the value 2.53.

4. Correspondence analysis

What is it that we generalize? In order to discuss this properly, we first define CA in terms of the Fisher-Lancaster decomposition of an observed table. This is sometimes called the canonical analysis of a contingency table (for instance in Kendall and Stuart, 1967, chapter 33), while the French call it the reconstitution formula. Suppose P is the observed table, with entries that add up to one. The diagonal matrix D contains row marginals, E contains the column margins, u is a vector with all elements equal to one. Then we can find X and Y such that $u'DX = 0$, $u'EY = 0$, $X'DX = I$, $Y'EY = I$, and

$$(4)\quad P = D(uu' + X\Omega Y')E,$$

with Ω diagonal. The proof is simple. Let

$$(5)\quad Z = D^{-\frac{1}{2}}(P - Duu'E)E^{-\frac{1}{2}},$$

and suppose $Z = K\Omega L'$ is the singular value decomposition of Z. Let $X = D^{-\frac{1}{2}}K$ and $Y = E^{-\frac{1}{2}}L$. It is easy to show that (4) is satisfied. We see, moreover, that the sum of squares of the elements of Ω is equal to the sum of squares of the elements of Z, which is Pearson's index of mean square contingency. If P is based on a sample of size n, then n times this coefficient of contingency is equal to the chi-square statistic for testing independence. Thus we can say that CA, if interpreted as computing the Fisher-Lancaster decomposition (4), studies the deviations from the independence model.

In the introduction we said that CA gave a geometrical representation of the residuals, in this case of the residuals from independence. This can be explained most easily by introducing the Benzécri-distances between the rows of P (Benzécri calls them the chi-square distances). They are

$$(6)\quad \delta^2_{ik} = (e_i - e_k)'D^{-1}PE^{-1}P'D^{-1}(e_i - e_k),$$

where the e_i and e_k are unit vectors (with exactly one element +1, the others zero). If we substitute (4) in (6) we find

$$(7)\quad \delta_{ik}^2 = (e_i-e_k)'(uu'+X\Omega Y')E(uu'+Y\Omega X')(e_i-e_k) =$$

$$= (e_i-e_k)'(uu'+X\Omega^2X')(e_i-e_k) =$$

$$= (\bar{x}_i-\bar{x}_k)'(\bar{x}_i-\bar{x}_k).$$

Here x_i and x_k are rows of $\bar{X} = X\Omega$. We see that the Benzécri-distance between rows i and j of P is equal to the ordinary Euclidean distance between rows i and j of $\bar{X}$. Thus we can represent the rows of $\bar{X}$, and we can approximate the Benzécri-distance by considering only the r columns of $\bar{X}$ corresponding with the largest singular values. This approximates the Benzécri-distances from below (De Leeuw & Meulman, 1985). It is clear that dually we can also define distances between columns of P, and approximate them from below by ordinary Euclidean distances between rows of $\bar{Y} = Y\Omega$.

For many applications having two separate plots, one for the rows and one for the columns, is not very convenient. We would like to have a joint plot or biplot (Gabriel, 1971). It is possible to plot both row and column objects in a single plot by using the centroid principle. From (3) we have

$$(8a)\quad D^{-1}PY = \bar{X},$$

$$(8b)\quad E^{-1}P'X = \bar{Y}.$$

In (8a) we see that representing Y and $\bar{X}$ in a joint plot allows a simple interpretation. The row points $\bar{X}$ are centroids (weighted averages, conditional expectations) of the column points Y. In (8b) it is the other way around. We already know, of course, that $\bar{X}$ and $\bar{Y}$ approximate the Benzécri-distances from below. This is the basic geometry of CA. There are two different joint plots $(\bar{X},Y)$ and $(X,\bar{Y})$, depending on the choice of the centroid principle (8a) and (8b). In the first plot the row points are inside the convex hull of the column points, in the second plot it is the other way around.

For reasons of symmetry some authors plot $(X\Omega^{\frac{1}{2}},Y\Omega^{\frac{1}{2}})$, but in this plot there is no clear distance and centroid interpretation. We can say

that in the symmetric interpretation we represent the residuals $D^{-1}PE^{-1}-uu'$ by scalar products between vectors. This interpretation follows from (4) as well, but unfortunately most people find it difficult to think in terms of scalar products.

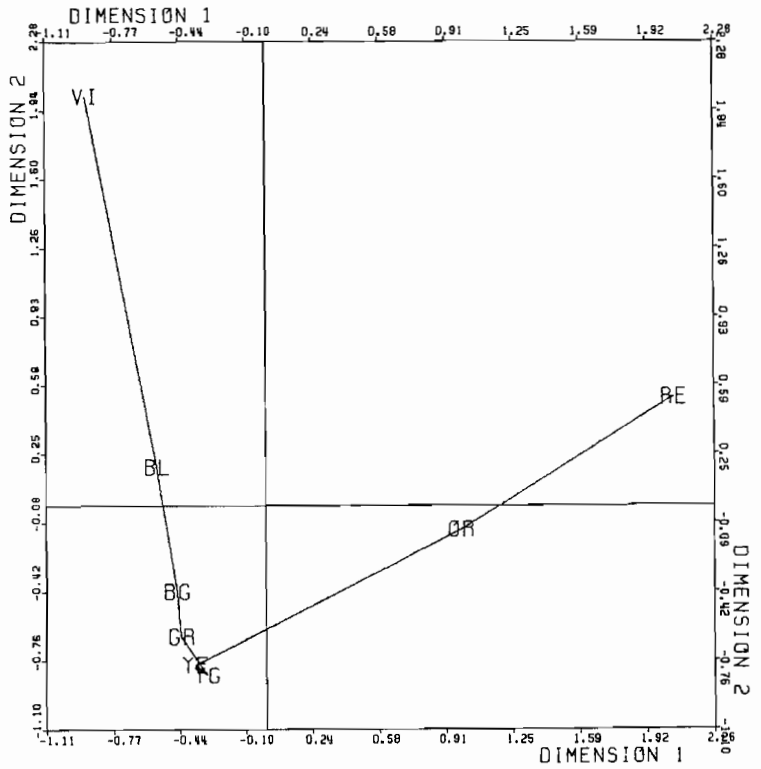
In the French CA literature it is quite customary to make joint plots of the pair $(X\Omega, Y\Omega)$ (Baccini, 1984). This has some rather serious disadvantages, because distances between row- and column-points cannot be interpreted in terms of the centroid principle. Moreover the inner products of row- and column-vectors do not reproduce residuals any more. There are also some advantages, mentioned for instance by Israels (1985). Both the distances between different row-points and the distances between different column-points approximate the Benzécri-distances, while the distance of any point to the origin approximates its contribution to the total inertia (chi-square). Although the choice of normalization can be quite important, it is not really necessary to decide which one is the best. The important thing is to realize that it is not possible to get all the desirable geometric properties in one plot, so a choice is necessary.

We illustrate CA by analyzing a confusion matrix taken from Benzécri (1970, p.9). It resulted from a learning experiment, in which subjects had to learn to associate colours with keys of a piano. The data in table 4 give the number of response confusions between the colours

Table 4: Confusion matrix between colour stimuli, columns correst responses, rows actual responses.

	RE	OR	YE	YG	GR	BG	BL	VI
red	415	45	2	8	7	4	4	3
orange	32	373	16	17	8	11	12	8
yellow	10	12	343	70	22	20	13	10
yellow-green	6	19	50	303	31	23	18	6
green	6	12	23	36	305	71	29	8
blue-green	10	10	15	32	91	274	38	19
blue	8	11	14	6	17	60	356	36
violet	3	5	22	13	11	13	24	403

Figure 1: CA of table 4
Row points; symmetric normalization



(columns are correct responses, rows actual responses) in trial 20, aggregated over subjects. Because of the learning effect the diagonal of the table is very dominant. The chi-square is 12782.3. CA finds the dominant singular values to be .87 and .79. The first two dimensions account for 42% of chi-square. In figure 1 we have plotted $X\Omega^{\frac{1}{2}}$, which is virtually identical to $Y\Omega^{\frac{1}{2}}$ in this example. It shows the colors in their spectral order, along the familiar curved dimension sometimes called the horse-shoe (Schriever, 1985, Heiser, 1986, Van Rijckevorsel, 1985). Because of the strong diagonal dominance in this example the singular values are not nicely separated, and a very large part of the inertia remains unaccounted for.

5. Quasi-correspondence analysis

Now suppose P and Q are two contingency tables. We suppose P and Q

have the same marginals, collected in the diagonal matrices D and E. The interpretation we have in mind is to take P as the observed data matrix and Q as the maximum likelihood estimates under quasi-independence. The technique we discuss is somewhat more general, however, because Q could also be maximum likelihood estimates under models such as quasi-symmetry or the RC-model. This is discussed in more detail in Van der Heijden and De Leeuw (1985), who also give references to some of the relevant French literature. The idea of using a model to generalize correspondence alaysis is due to Escofier (1983, 1984).

If we start with the singular value decomposition

$$(9)\quad D^{-\frac{1}{2}}(P-Q)E^{-\frac{1}{2}} = K\Omega L',$$

we find, analoguous to (3), that

$$(10)\quad P = Q + DX\Omega Y'E.$$

We loose the connection with chi-square, because the sum of squares of the singular values is equal to

$$(11)\quad \Sigma\Sigma(p_{ij} - q_{ij})^2/d_i e_j,$$

which is not the weighting of residuals we need for a chi-square distribution. Nevertheless, we still decompose residuals, of course.

The interpretation in terms of Benzécri-distances can still be maintained, because

$$(12)\quad \delta_{ij}^2 = (e_i - e_j)'D^{-1}(P - Q)E^{-1}(P' - Q')D^{-1}(e_i - e_j) =$$

$$= (\bar{x}_i - \bar{x}_j)'(\bar{x}_i - \bar{x}_j),$$

with $\bar{X} = D^{-\frac{1}{2}}K\Omega$, as before. The centroid principle occurs in a somewhat different, but still geometrically very easily understood way. From (10) we find

(13a) $D^{-1}PY - D^{-1}QY = \bar{X}$,

(13b) $E^{-1}P'X - E^{-1}Q'X = \bar{Y}$.

This makes it interesting to look at a joint plot which contains Y, $D^{-1}PY$, $D^{-1}QY$, and $\bar{X}$. Thus each row is represented by two centroids and their difference.

It is clear from our results so far that if the quasi-independence model fits well, then P - Q is small. Thus the singular values are small, and $\bar{X}$ will be small. We can also say that if the model fits well, then $D^{-1}PY$ and $D^{-1}QY$ will be very similar, and consequently $\bar{X}$ will be small. This brings us back to the point mentioned in the introduction: if the fit of the model is too good, then there will be no interesting variation left for CA. Because structural zeros or nonrestricted cells do not contribute to P - Q, this means that we will need a fair percentage of restricted cells in the analysis.

By comparing (5) and (9), it is easy to see that one can find a quasi-correspondence analysis solution using usual correspondence analysis programs. All one has to do is to take the matrix (P-Q+Duu'E) as input matrix.

6. Alternative weighting schemes

In the previous section we have defined quasi-correspondence analysis by using as diagonal weighting matrices D and E the observed marginals of P and Q. This however, is not the only possible choice. There are at least two alternatives.

As a second possibility we can define D and E by

(14a) $d_{ii} = \Sigma\{p_{ij} \mid j \in J(i)\}$,

(14b) $e_{jj} = \Sigma\{p_{ij} \mid i \in I(j)\}$.

It is clear that this coincides with our previous definition of D and

E in the case of structural zeros, but in general (14) will define matrices with diagonal elements which are strictly smaller. Thus singular values will be higher.

The third possibility is to choose d_{ii} equal to $\ddot{\alpha}_i$ and e_{jj} equal to $\ddot{\beta}_j$. In this case the sum of squares of the singular values becomes

$$(15)\quad \Sigma\Sigma\{(p_{ij} - q_{ij})^2/q_{ij} \quad (i,j) \in K\},$$

which is, of course, the chi-square statistic for testing quasi-independence. This is a considerable advantage over other forms of scaling. There is another advantage, which is somewhat less obvious. Suppose R is the matrix with r_{ij} equal to p_{ij} for all (i,j) in K and with r_{ij} equal to $\ddot{\alpha}_i\ddot{\beta}_j$ for all (i,j) not in K. Thus R is the matrix that the second maximum likelihood algorithm converges to. For the presidential choice data R is given as the last matrix in table 3. It is now not difficult to see that quasi-correspondence analysis of P and Q, with weights $d_{ii} = \ddot{\alpha}_i$ and $e_{jj} = \ddot{\beta}_j$, is identical to ordinary CA of R. This alternative interpretation is quite useful, and does not apply to the other normalizations.

For all three possibilities the matrices P and Q are, of course, the same. Because $(P - Q)u = 0$ and $(P' - Q')u = 0$ it follows that $E^{\frac{1}{2}}u$ and $D^{\frac{1}{2}}u$ are singular vectors corresponding with a singular value equal to zero. Thus X and Y have one column equal to u, no matter how we choose D and E, and all other columns satisfy $u'Dx = u'Ey = 0$. Moreover formulas (12) and (13) do not depend on the choice of D and E and remain valid. On the other hand, in case of the second scaling, formula (13) cannot be interpreted in terms of centroids. But this is easily remedied. If P_0 is P with all p_{ij} for which (i,j) is not in K replaced by zero, and Q_0 is defined in a similar way, then $P - Q = P_0 - Q_0$. Matrices D and E in the second scaling are the marginals of P_0 and Q_0, and thus (13) can be rewritten as

$$(16a)\quad D^{-1}P_0Y - D^{-1}Q_0Y = \bar{X},$$

$$(16b)\quad E^{-1}P_0'X - E^{-1}Q_0'X = \bar{Y}.$$

These are the centroid principles for the second normalization. For the third normalization both (13) and (16) are not centroids, and we have to use yet another interpretation. Let S = Ruu'R/u'Ru. Thus S is the matrix of expected values on the hypothesis of independence if R are the observed values. Now R - S = P - Q, and D and E in this case are the marginals of R and S. The centroid principle is now simply

(17a) $D^{-1}RY = \bar{X}$,

(17b) $E^{-1}R'X = \bar{Y}$.

The centroid principle in this case shows why the third normalization reduces to ordinary CA of R (compare formula 8). Thus it is possible to perform quasi-correspondence analysis using a correspondence analysis program, namely by using R as the input matrix.

Thus, from a mathematical point of view, all three normalizations have their own centroid principles and Benzécri-distances. Normalization three has the advantage of simplicity, and the advantage of the nice relationship with the chi-square for quasi-independence. From the interpretational point of view it is difficult to give convincing reasons to prefer one normalization over another. Van der Heijden (1985) discusses some considerations. In case of structural zeros, in which the first two normalizations are identical, it seems not logical to fill in the empty cells. Thus the third normalization does not seem appropriate here. In case of data which are simply missing (but not logically impossible) we could use normalization three, or its generalizations which are discussed in the next section. In the case in which all elements are observed, but the model is only fitted to a subset, all three choices can be defended.

We illustrate the developments in this section by analyzing an example. The presidential election data are not very suitable for this, because almost no variation is left after the quasi-independence model is fitted. The colour-confusion data are somewhat more promising in this respect.

If we use the maximum likelihood algorithm to adjust diagonal

elements, we find the following values. Red: 4.71, Orange: 10.92, Yellow: 23.44, Yellow-Green: 29.29, Green-Blue: 51.29, Blue: 20.89, Violet: 7.27. Thus colours at the end of the spectrum are confused a great deal less that colours in the middle. The remaining chi-square for quasi-independence is 656.64, with 41 degrees of freedom. Thus a considerable amount of inertia is still left. The quasi-independence model fits rather poorly.

Our first analysis is the quasi-correspondence analysis which uses the marginals of the original, unadjusted table as weights D and E. Its first two singular values are .14 and .11, explaining 61% of the total inertia (which is not chi-square in this case!). In figure 2 we have plotted X, normalized by X'DX = I. These are the scores for the actual responses. In figure 3 we have plotted $Y_1 = E^{-1}Q'X$ and $Y_2 = E^{-1}P'X$. We know that $\bar{Y} = Y_2 - Y_1$. Instead of plotting $\bar{Y}$ we have drawn arrows from the points of Y_1 to corresponding points of Y_2. This also shows what $\bar{Y}$

Figure 2: QCA of table 4
Row points; normalization X'DX=I
First choice of weights

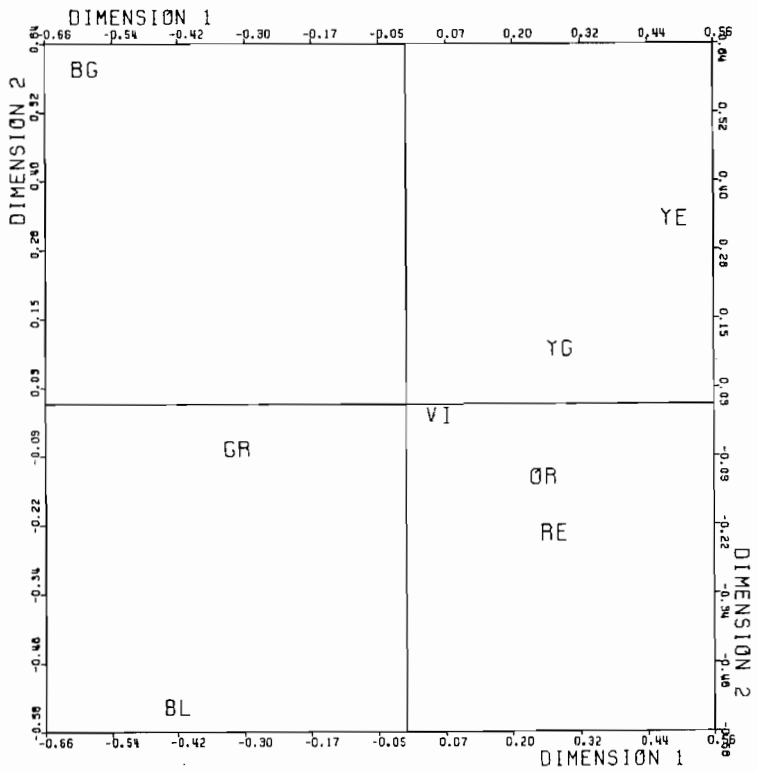
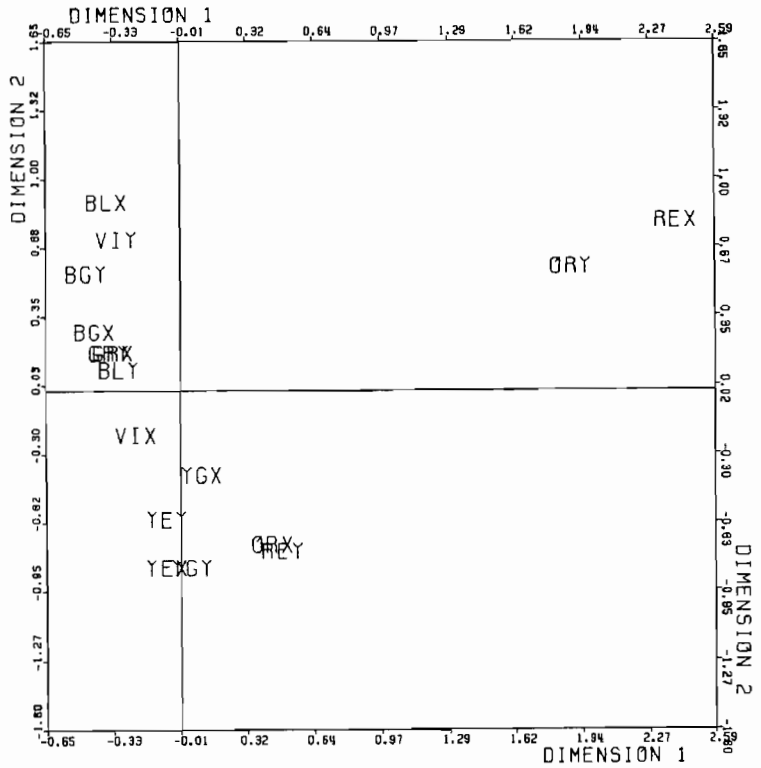


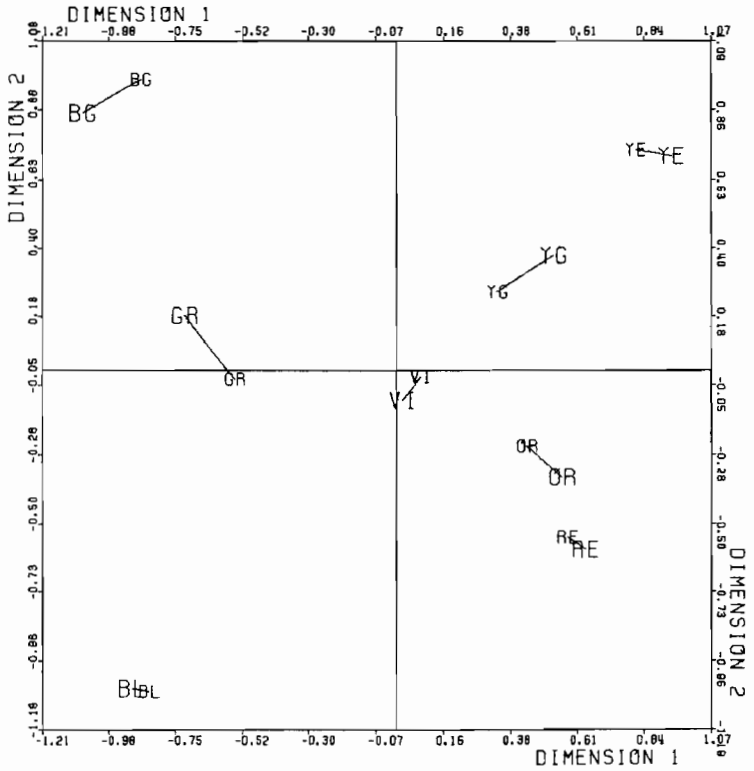
Figure 3: QCA of table 4
Points Y_1 (small) and Y_2 (large)



looks like, because $\bar{Y}$ is obtained by translating all arrows to the origin. The analysis shows the cluster structure ((OR,RE),(YE,YG)),(BL,BG,GR) in the residuals, with the residuals for violet either very small or largely unexplained. Comparing X and $\bar{Y}$ shows that green and blue-green residuals behave differently in rows and columns.

The next analysis uses the third normalization, i.e. it is a CA of R. Figure 4 is a joint plot of $(X\Omega^{\frac{1}{2}}, Y\Omega^{\frac{1}{2}})$, which could be called the biplot or inner product representation. In this plot there is no direct interpretation in terms of Benzécri distances or in terms of centroid principles. We see that the horse-shoe-representation of the spectrum is still there. A dominant feature of the plot is the interchanging of the positions of orange and red in the row and column plots, indicating relatively large (orange,red) and (red,orange) residuals with positive signs. Observe that the two orange vectors, the two red vectors, and the two violet vectors are roughly

Figure 4: QCA of table 4
Row and column points, symmetric normalization
Third choice of weights



orthogonal. This is only natural, because diagonal residuals are by definition zero. The singular values are .43 and .33. They account for 60% of the inertia, which is the chi-square for quasi-independence in this case.

7. Correspondence analysis for incomplete tables

A technique for CA of incomplete tables has been proposed by Nora (1975). It is also discussed in Benzécri et al. (1980, Vol. 2, chapter III, no. 8), and by Greenacre (1984, pp. 236-244). First choose the dimensionality h. Then reconstitution of order h is the iterative process

$$(18)\quad r_{ij}^{(m+1)} = r_{i*}^{(m)} r_{*j}^{(m)} \left(1 + \sum_{s=1}^{h} \omega_s^{(m)} x_{is}^{(m)} y_{js}^{(m)}\right) / r_{**}^{(m)},$$

which is applied for all (i,j) not in K. For (i,j) in K we simply set $r_{ij}^{(m)} = p_{ij}$ for all m. The solution will, in general, depend on the choice of the dimensionality h. Benzécri himself seems to favor iterated reconstitution of order zero, i.e. for all (i,j) not in K we set

$$(19)\quad r_{ij}^{(m+1)} = r_{i*}^{(m)} r_{*j}^{(m)} / r_{**}^{(m)}.$$

This is, of course, exactly identical to one form of quasi-correspondence analysis. Iteration (19) is the second algoritm to compute maximum likelihood estimates. We have seen that it converges to a matrix R, and that quasi-correspondence analysis with the third choice of weights D and E is identical to CA of R. Thus, in this sense, one particular form of quasi-correspondence analysis has already been described in the literature as CA with iterative reconstitution of order zero.

8. Examples

There are various ways in which possible examples could be categorized. We have chosen for the categorization distinguishing square matrices, where special attention has to be given to the diagonal, vs. non-square matrices, where other cells seem to cause problems. The square matrix we analyzed, a migration table, will be discussed as an example of a case in which we are not interested in some cells, i.e. the diagonal cells. Here diagonal frequencies indicate the number of migrations in which the subject moved but remained in the same suburb of Paris. We refer to Van der Heijden (1985,1986) for an analysis of a transition matrix containing diagonal and off-diagonal structural zeros, and off-diagonal cells which are 'eliminated' because they dominate the solution. Two non-square examples are treated which are interesting for different reasons. First we show the analysis of a triangular matrix, because it is typical for a lot of applications. Secondly we discuss the analysis of a flattened three-way matrix in which a triangle is deleted. In this last example we indicate how QCA can be seen to be related to loglinear models containing structural zeros, thereby generalizing

results on relations between ordinary CA and loglinear analysis reported by Van der Heijden & De Leeuw (1985).

Example 1: a migration table

Square matrices often provide us with a case in which the use of QCA can be helpfull for the understanding of the structure of the off-diagonal cells. CA is not appropriate here, because of the diagonal cells: these are often not defined, as might be the case in transition matrices and import-export tables, or they are not the primary point of interest, such as in confusion matrices, migration tables, etc. In the French literature CA of import-export and related tables has been given considerable attention. An import-export table is a square matrix with importing areas for the rows, and exporting areas for the columns. The diagonal elements are either very high (being the trade in that area), or not defined. Various proposals are made to fill in values for the diagonal cells, to make it possible to analyze this type of matrix with CA. Burtchy (1984) reviews the various approaches that have been used in combination with CA. They either replace the diagonal with values chosen on theoretical grounds, or they adjust the diagonal (Stemmelen, 1977), or they complete the diagonal by iterative reconstitution. Burtchy has many additional references on adjusting and completing input-output table (see also Le Foll & Burtchy, 1983). The approach proposed by Burtchy to complete the diagonal by iterative reconstitution (compare section 7, and also Greenacre, 1984, ch. 8) is criticized by Foucart (1985) because this method produces modified margins, and therefore modified weights, for which there is no substantial justification. Foucart proposes to modify the diagonal elements of the matrix, and subsequently analyse the symmetricized matrix $(p_{ij}+p_{ji})$ and plot the rows and column of p_{ij} as passive points in the resulting solution.

Of a different nature is the proposal of Escofier (1984), who decomposes the departure from models using her generalization of CA. In these models the expected diagonal elements are equal to observed diagonal elements. However, her proposal has the drawback that the margins of the observed and expected frequencies are not equal (see Van der Heijden & De Leeuw, 1985, for a discussion of these drawbacks).

As before, we will study the decomposition of the departure from quasi-independence. For the diagonal cells we take $p_{ij} = \pi_{ij}$, and for the off-diagonal cells $\pi_{ij}=\alpha_i\beta_j$. Compared with Foucart, our way of dealing with the analysis of square matrices has the advantage that we do not have to fill in some artificial values for the diagonal cells. Compared with Escofier, in our case the marginal frequencies of the observed and expected frequencies are equal.

Figure 5: QCA of table 5, dimension 1 and 2
Second choice of weights (compare section 6)
Singular values .658 (.35), .467 (.18), .384 (.12), .364 (.11)

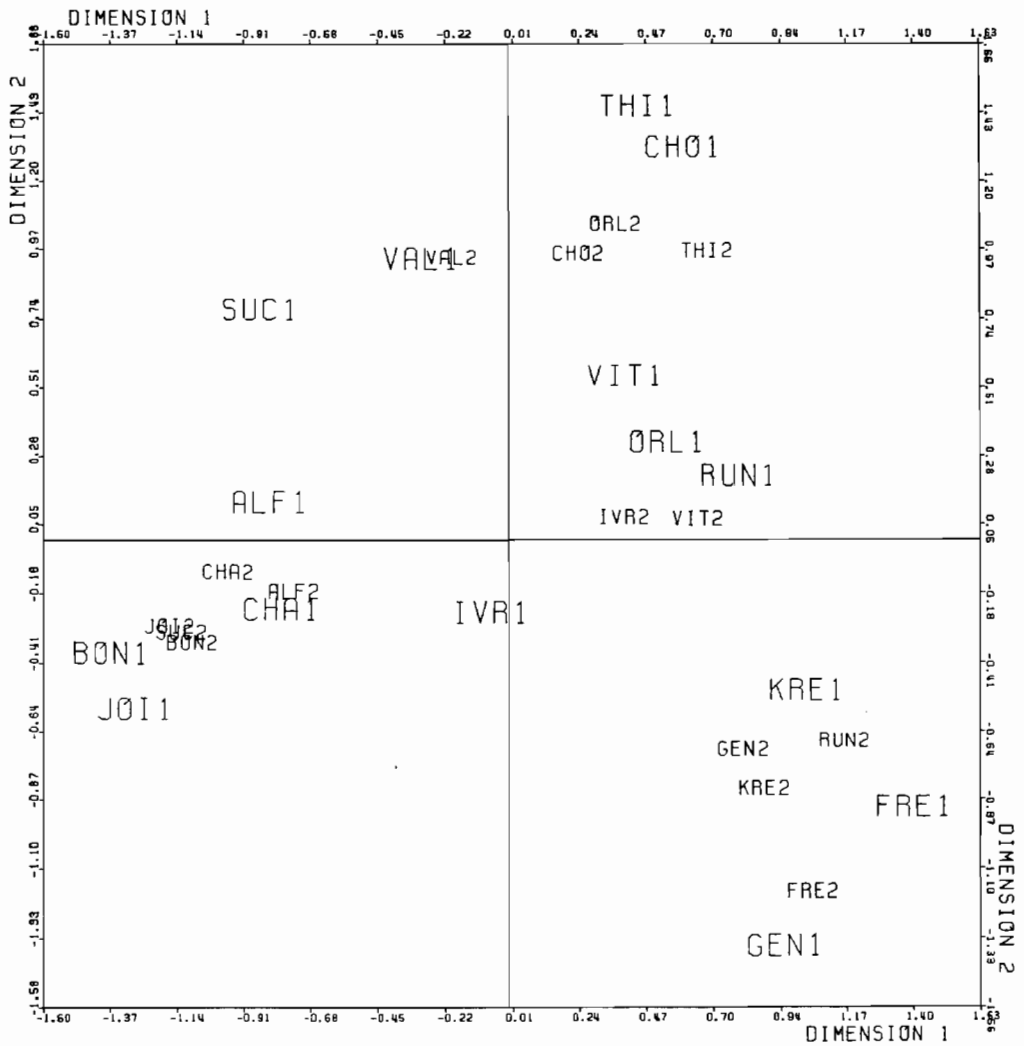


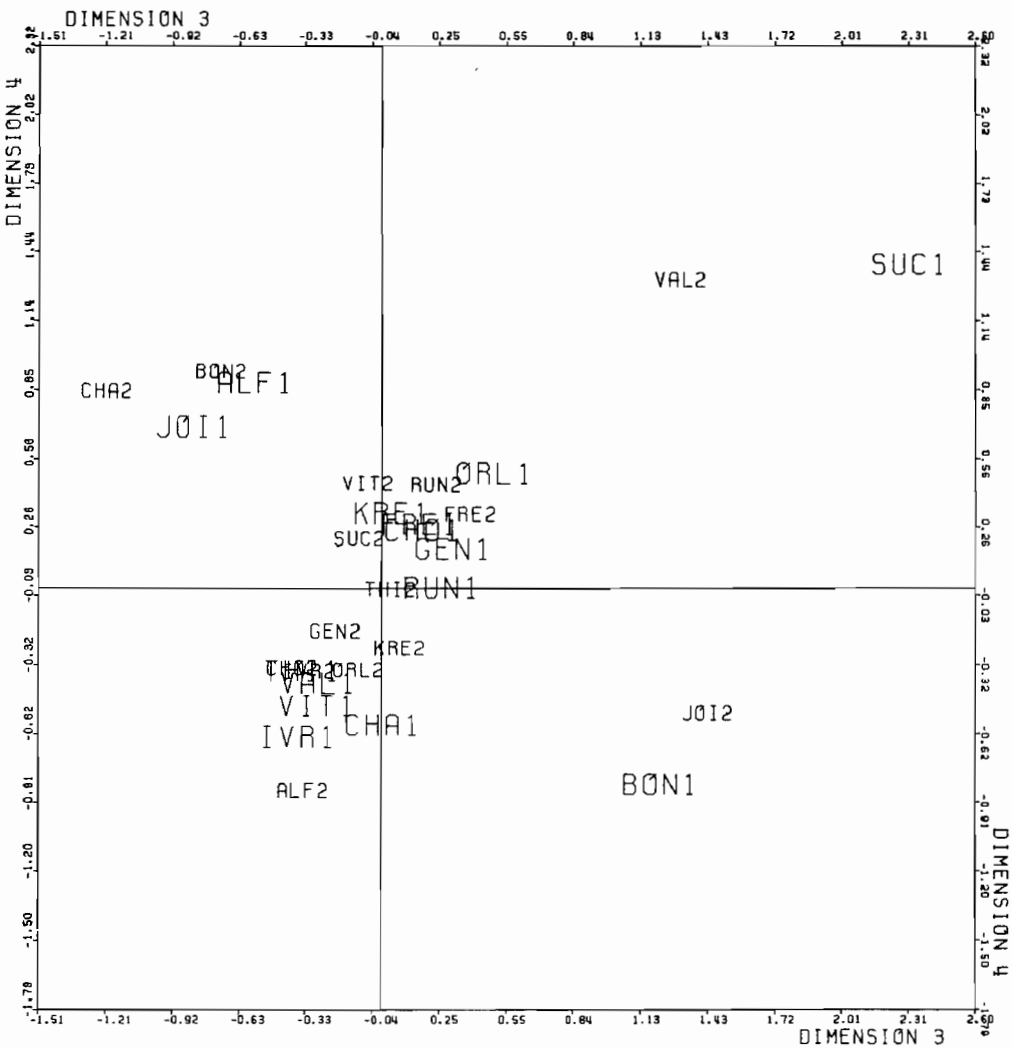
Table 5: Migration in the suburbs of Paris; rows are destinations, columns are origins.

	CHA	IVR	KRE	GEN	VIT	ALF	CHO	BON	VAL	ORL	RUN	FRE	THI	JOI	SUC
Charenton	6238	269	45	14	204	824	57	250	70	76	16	36	0	403	189
Ivry	270	11268	1113	1113	257	2483	530	708	166	878	166	205	281	457	174
Kremlin	34	585	11353	1001	1493	32	143	62	133	207	327	549	226	133	0
Gentilly	0	106	1389	10695	425	100	99	220	27	111	215	1037	26	152	117
Vitry	186	667	894	281	11263	1009	1577	148	123	1021	154	265	860	314	90
Alfort	713	258	134	75	632	16420	595	1675	563	250	29	0	118	507	297
Choisy	0	181	78	41	763	148	5590	24	396	964	104	38	745	25	87
Bonneuil	51	81	68	0	133	1094	109	9235	107	92	0	28	39	1831	491
Valenton	31	34	34	28	34	316	271	148	6161	628	0	0	59	83	228
Orly	14	108	492	177	353	104	528	209	568	6461	315	408	551	191	130
Rungis	0	21	160	83	81	33	23	20	64	248	1455	110	106	21	0
Fresnes	0	53	310	260	156	0	0	0	0	82	481	3889	131	0	0
Thiais	0	66	21	0	151	40	421	24	43	248	26	0	1498	25	0
Joinville	327	43	0	63	206	801	42	1362	0	40	54	90	35	17045	774
Sucy	0	0	0	26	26	20	28	159	591	102	0	0	0	403	5624

As an example we show the analysis of a migration table published in Foucart (1985) (see table 5). In a cell of this matrix a frequency gives the number of persons which have moved from one suburb of Paris to another. As weights we have taken the sum of the off-diagonal frequencies, so the diagonal frequencies are neglected completely. Furthermore, a point represents the difference between the <u>profiles</u> of the off-diagonal cells in p_{ij} and π_{ij}.

The first four singular values, with their proportion of the total

<u>Figure 6</u>: QCA of table 5, dimension 3 and 4
Second choice of weights (compare section 6)



inertia (which is not chi-square), are .658 (.348), .467 (.176), .384 (.118), and .364 (.107). .749 of the total inertia is decomposed in the first four dimensions, which is 7% more compared with Foucalt. A plot of the first two dimensions is shown in figure 5. A horse-shoe-like curve can be seen, with JOI, BON, CHA, ALF and SUC on the left, going to KRE, GEN, RUN, and FRE on the right. Foucarts plot is about the same. However, in his plot distances between corresponding 'migration to' (large point) and 'migration from' (small point) suburbs are much smaller.

In figure 6 we see the third and fourth dimension. Interpreting the elements with the larger contributions to these dimensions, we trace some asymmetries between suburbs on the left part of the first dimension: BON to ALF happens more then ALF to BON; VAL to SUC more then SUC to VAL; ALF to IVR more then IVR to ALF, etc. Migrations in the pairs BON-JOI, CHA-ALF and CHA-JOI seem equally strong. Also, from points lying far apart, we can conclude that CHA to SUC happens more then SUC to CHA, SUC to BON occurs more, and ALF to IVR happens more than the reverse.

Example 2: whorls vs. small loops

CA of triangular matrices is not meaningfull. However, QCA, where the quasi-independence model is fitted to the non-missing values, can be helpfull here to obtain an interpretable plot. To show this, we discuss the analysis of the a triangular matrix taken from Goodman (1968). The matrix is shown in table 6. In each cell two values can be found: the observed frequency and the expected frequency, following the quasi-independence model. In the rows we find the number of whorls, in the columns the number of small loops in finger prints of the right hand. Since the sum of these two numbers cannot exceed 5, the lower triangle of this matrix is void.

Comparing profiles does not seem useful here, since the number of cells for each line is unequal. Therefore we have taken the values α_i and β_j as weights. The four singular values are .365 (.64), .239 (.27) .108 (.06) and .087 (.04). The first two dimensions are shown in figure 7. (The points for 5 whorls and 5 small loops are not shown,

Table 6: Number of whorls vs. number of small loops
Observed and expected frequencies (between brackets)
Expected frequecies rounded to integers.

		Small Loops					
		0	1	2	3	4	5
Whorls	0	78 (201)	144 (167)	204 (167)	201 (150)	179 (131)	45 (45)
	1	106 (122)	153 (102)	126 (101)	80 (92)	32 (80)	0 (0)
	2	130 (86)	92 (71)	55 (71)	15 (64)	0 (0)	0 (0)
	3	125 (64)	38 (53)	7 (53)	0 (0)	0 (0)	0 (0)
	4	104 (71)	26 (59)	0 (0)	0 (0)	0 (0)	0 (0)
	5	50 (50)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)

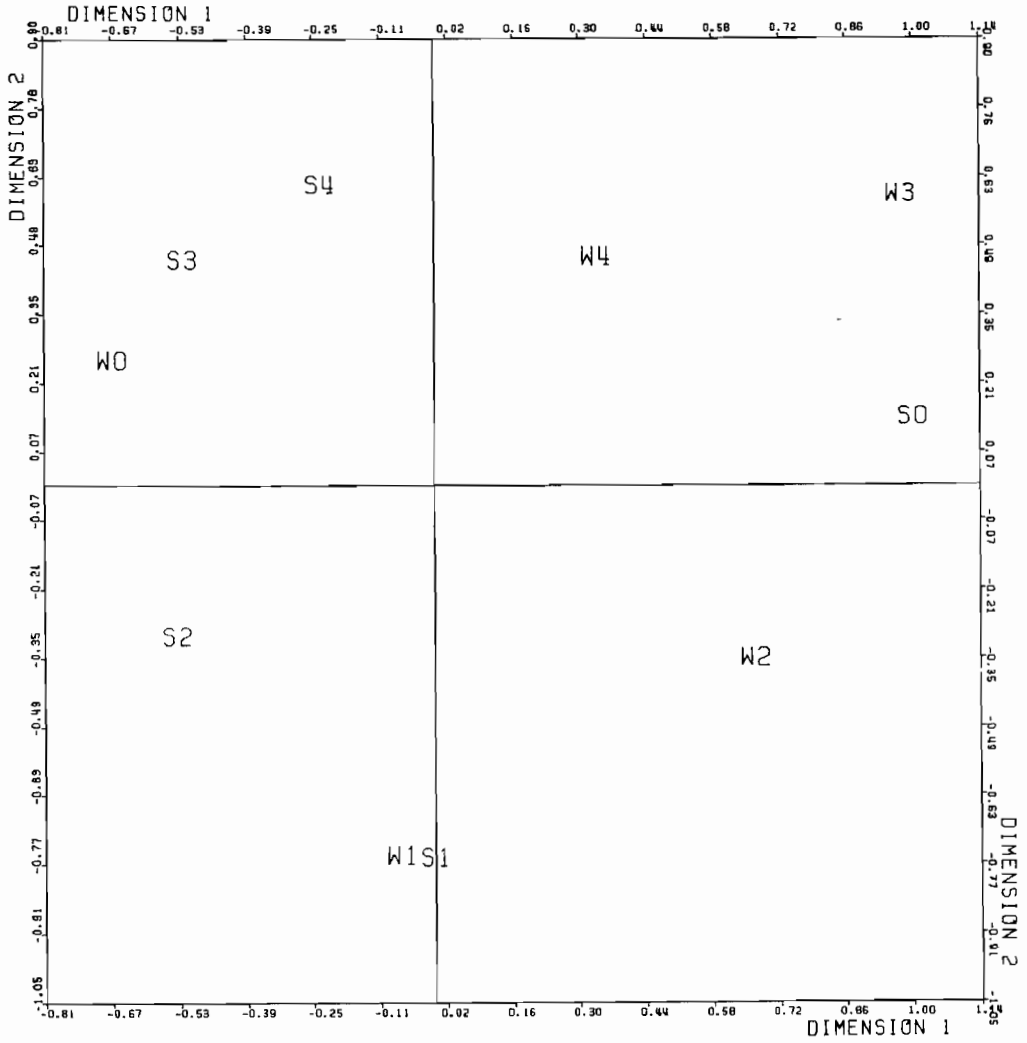
because for all cells of these lines $p_{ij}=\pi_{ij}$, and so there is no difference to be decomposed). Interpretation will be clear: 0 whorls occur more often then expected with 2 or 3 small loops and the other way around. The location of the other points should be interpreted in the same way.

Example 3: Current age x age at first marriage: structural zeros

The third example we will discuss is taken from Haberman (1979, pp. 455-471). There are three variables: age at first marriage, current age, and sex. We deal with structural zeros since the age at first marriage cannot exceed the current age. Therefore, these two variables are related in a trivial way. To investigate whether there is any other relationship, we fitted a quasi-independence model to the data, defining all cells for which current age is smaller than the age at first marriage as structural zeros.

For the moment we decided that the first-order interaction between current age (C) and sex (S) did not interest us, therefore we treated the three-way table as a two-way table of rank 4x16. See table 7. The

Figure 7: QCA of table 6, dimension 1 and 2
Third choice of weights (compare section 6)
Singular values .365 (.64), .239 (.27), .108 (.06), .087 (.04)



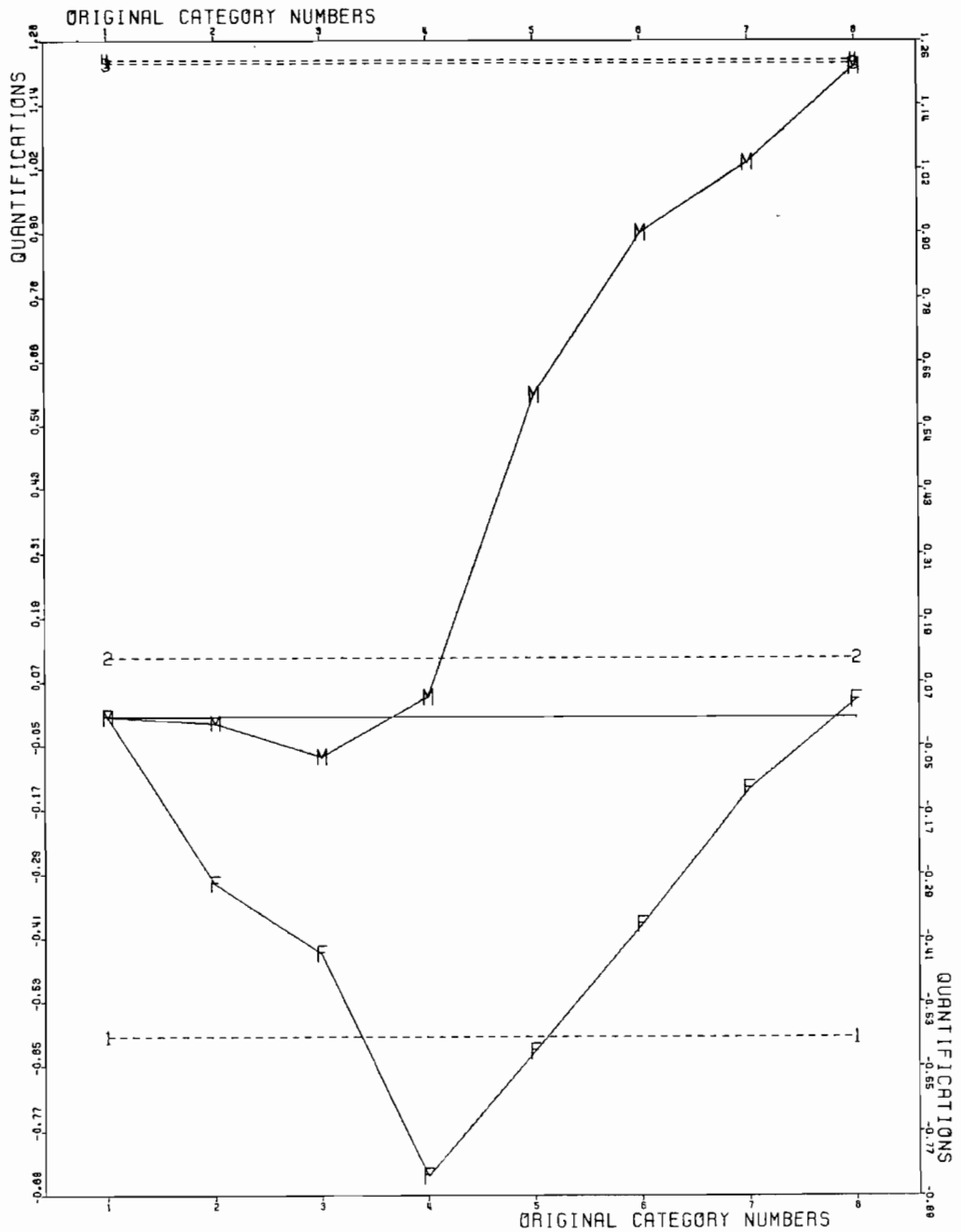
chi-squared value equals 245. Fitting a quasi-independence models to this two-way table comes to the same as fitting the loglinear model [A][CS] (including structural zeros) to the three-way table. Therefore the correct number of degrees of freedom is 33: the difference between the observed and expected frequencies is significant. We use QCA to decompose the departure from model [A][CS] (compare Van der Heijden & De Leeuw, 1985, in which it was shown that classical CA decomposes the departure from the loglinear model [A][CS] in case it did not contain structural zeros.).

Table 7: Age at first marriage x current age x sex observed and expected frequencies for quasi-independence model

	Current age	Age at first marriage ≦20	21-25	26-30	≧31
Female	≦20	9 (9)	- -	- -	- -
	21-25	43 (32)	20 (31)	- -	- -
	26-30	51 (41)	40 (41)	3 (12)	- -
	31-40	103 (65)	53 (64)	4 (19)	1 (12)
	41-50	68 (49)	45 (48)	5 (14)	3 (9)
	51-60	65 (50)	43 (50)	7 (15)	9 (10)
	61-70	39 (32)	24 (32)	12 (9)	4 (6)
	≧71	22 (24)	26 (24)	7 (7)	4 (5)
Male	≦20	2 (2)	- -	- -	- -
	21-25	24 (24)	23 (23)	- -	- -
	26-30	21 (26)	34 (25)	3 (8)	- -
	31-40	30 (43)	61 (42)	10 (13)	4 (8)
	41-50	22 (40)	49 (40)	20 (12)	10 (8)
	51-60	19 (45)	50 (44)	27 (13)	15 (9)
	61-70	16 (38)	38 (38)	23 (11)	17 (7)
	≧71	11 (24)	19 (24)	19 (7)	11 (5)

Figure 8 shows the category quantifications for the first dimension, in which 81% of the total inertia is decomposed (a two-dimensional plot shows the well-known horse-shoe). We have taken row and column margins as weights, so that a point represents the difference between

Figure 8: QCA of table 7, original category numbers vs. first quantifications. First choice of weights (is equal to second choice here). Current age-line is solid, age at first marriage-categories are dotted lines. Singular values .408 (.81), .174 (.15), .089 (.039)



the profile in the matrix of the observed and the matrix of expected frequencies. Column points for age younger than 20 are given quantification 0, since for these profiles the observed frequencies are equal to the expected frequencies for all cells. The plot shows us that men marry more than expected at an age older than 25, while women marry more than expected when they are younger than 25. Furthermore, the older the male respondent is, the higher the probability that he is married for the first time at an age older than 25; the women line shows a dip: female respondents of an age from younger than 20 to 31-41 married more often than expected when they are younger then 20, and for older respondents this relation becomes weaker.

Conclusion

Quasi-correspondence analysis seems to be a good alternative to correspondence analysis in all cases where the study of departure from quasi-independence seems more logical, or appropriate than from independence. It was shown that quasi-correspondence analysis is a very versatile technique: firstly, given a cross-table, it is possible to construct several quasi-independence models; secondly, there is more than one alternative for the choice of weights. Important for the applicability of quasi-correspondence analysis is that all forms can be performed using computer programs for ordinary correspondence analysis. The user should first estimate expected frequencies, and subsequently construct the proper input matrix for the correspondence analysis program.

References

Baccini, A. (1984) Etude comparative des représentations graphiques en analyses factorielle des correspondances simples et multiples. Toulouse: Laboratoire de Statistique et Probabilités. Université Paul Sabatier.

Benzécri, J.P. (1970). Sur l'analyse des matrices de confusion. Revue de Statistique Appliquée, 18(3), 5-63.

Benzécri, J.P. et al. (1973). Analyse des données (2 vols.). Paris: Dunod.

Benzécri, J.P. et al. (1980). Pratique de l'analyse des données (3 vols.). Paris: Dunod.

Bishop, Y.M.M., Fienberg, S.E. & Holland, P.W. (1975) Discrete multivariate analysis: theory and practice. Cambridge: MIT-press.

Burtchy, B. (1984). Analyse factorielle des matrices d'échanges. In: E. Diday et al. (eds.) Data analysis and informatics, III. Amsterdam: North-Holland.

Caussinus, H. (1965). Contribution à l'analyse de la corrélation de deux caractères qualitatifs. Annales de la Faculté des Sciences de l'Université de Toulouse, 29, 77-182.

Daudin, J.J. & Trécourt, P. (1980). Analyse factorielle des correspondances et modéle log-lineaire: comparaison des deux méthodes sur un exemples. Revue de Statistique Appliquée, 1, 524.

de Leeuw, J. & Meulman J. (1985) Principal components analysis and restricted multi-dimensional scaling. In: Schrader, M. & Goul, W. (eds.) Classification as a tool of research. Amsterdam: North-Holland.

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. (1984). Analyse factorielle en reference á un modèle; application à l'analyse de tableaux d'échanges. Revue de Statistique Appiquée, 32(4), 25-36.

le Foll, Y. & Burtchy, B. (1983). Représentations optimales des matrices imports-exports. Revue de Statistique Appliquée, 31(3), 57-72.

Foucart, T. (1985) Tableaux symmétriques et tableaux d'échanges. Revue de Statistique Appliquée, 33(2), 37-54.

Gabriel, K.R. (1971). The biplot-graphic display of matrices with application to principal component analysis. Biometrika, 58, 453-467.

Goodman, L.A. (1968). The analysis of cross-classified data: independence, quasi-independence, and interactions in contingency tables with or without missing entries. Journal of American Statistical Association, 63, 1091-1131.

Goodman, L.A. (1985). The 1983 Henry L. Rietz memorial lecture. 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. The Annals of Statistics, 13, 1069.

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

Haberman, (1974). The analysis of frequency data. Chicago: The University of Chicago Press.

Haberman (1979). Analysis of qualitative data (2 vols.). New York: Academic Press.

Heiser, W. (1986). Undesired nonlinearities in nonlinear multivariate analysis. In: Diday, E. (ed), Data analysis and informatics, IV. Amsterdam: North-Holland.

Israëls (1985). Untitled. (Unpublished research note).

Israëls, A.Z. & Sikkel, D. (1982). Correspondence analysis and comparisons with other techniques. Voorburg: Centraal Bureau voor Statistiek.

Kendall, D.G. & Stuart, A. (1967). The advanced theory of statistics (Vol. 2, 2nd. ed.). London: Griffin.

Lauro, N.C. & Decarli, A. (1982). Correspondence analysis and log-linear models in multiway contingency tables study. Some remarks on experimental data. Metron (Rivista internazionale di statistica), 15(1,2), 213-234.

Mosteller, F. (1968). Association and estimation in contingency tables. Journal of American Statistical Association, 63, 128.

Nora, C. (1975). Une méthode de reconstitution et d'analyse de données incomplétes. Unpublished Thèse d'Etat, Université P. et M. Curie, Paris VI'.

Reynolds, H.T. (1977). The analysis of cross-classifications. Glencoe, Illinois: The Free Press.

Schriever, B.F. (1985) Order dependence (unpublished thesis). Amsterdam: Free University.

Stemmelen, E. (1977) Tableaux d'échanges: description et prévision. Paris: Bureau Universitaire de Recherche Opérationnelle.

van der Heijden, P.G.M. (1985). Correspondence analysis of transition matrices. Kwantitatieve Methoden.

van der Heijden, P.G.M. (1986). Transition matrices, model fitting, and correspondence analysis. In: E. Diday (ed.) Data analysis and informatics, IV. Amsterdam: North-Holland.

van der Heijden, P.G.M., & de Leeuw, J. (1985). Correspondence analysis used complementary to loglinear analysis. Psychometrika.

van Rijckevorsel, J. (1985). About horseshoes in multiple correspondence analysis. In: Schrader, M. & Goul, W. (Eds.) Classification as a tool of research. Amsterdam: North-Holland.