KM 19(1985)
pag 49 - 72

CORRESPONDENCE ANALYSIS OF TRANSITION MATRICES

Peter G.M. van der Heijden *

Abstract

Transition matrices are frequently analysed using probabilistic models. Two basic models are the independence model and the quasi-independence model. In the context of transition matrices this last model is used to eliminate the influence of diagonal elements with a very high or very low frequency. Usually the independence models do not fit very well. In this paper we propose to analyze the residuals from these models with classical correspondence analysis and a generalization of correspondence analysis recently suggested by Escofier. Several advantages of this approach are discussed. Furthermore, a solution is proposed for the more general problem that in correspondence analysis single cells can dominate the solution. Two examples are given.

Keywords: Correspondence analysis, data analysis, multi-dimensional scaling, quasi-independence, transition matrix, longitudinal data, structural zeros, sequential analysis

*Dept. of Research Methods and Psychometrics, Subfaculty of Psychology, Faculty of Social Sciences, University of Leiden.
Adress: Hooigracht 15, 2312 KM Leiden, The Netherlands.
Tel. 071-148333-6381/5126.

The author is indebted to Pieter M. Kroonenberg, Jan de Leeuw and Ron A. Visser for their stimulating and helpful comments.

1.0 Introduction

Longitudinal categorical data are frequently represented or summarized in transition matrices with elements f_{ij}, where f_{ij} is the number of individuals in state i at time t and in state j at time t+1. In this paper we propose to analyse matrices of this kind with correspondence analysis. Correspondence analysis can be used to construct a multi-dimensional representation of the departure from independence of row and column variables of a matrix. When the departure from the independence model is not of primary interest, a generalization of correspondence analysis can be used to study the departure from other models, e.g. the quasi-independence model. Correspondence analysis has proved a useful tool to analyse ordinary contingency tables. We will show that correspondence analysis is also very suitable for broadening one's understanding of the structure in a transition matrix, especially when the number of categories is large. The present method can also be considered as a solution to a more general problem in correspondence analysis, i.e. that a solution is sometimes dominated by a single cell, or a few cells. It is shown that by using quasi-independence models together with a generalization of correspondence analysis this problem can be solved.

A transition matrix is constructed from a series of observations of a variable. When this variable is observed at T time points (1,.,t,.,T), it is possible to construct a T-way contingency table by defining this variable at each time point as a different variable. A cell of this matrix corresponds for each time point with one category. In this cell a frequency represents the number of objects for which the corresponding combination of categories was scored.

The total frequency in the T-way transition matrix is n, the number of observed objects. The total number of cells equals $\underline{I}^T$, where $\underline{I}$ is the number of categories, or states, of the variable that is observed. When T=2, we deal with a two-way transition matrix. When T>2, it is often found that the number of cells is larger than the number of objects. This is a very unfortunate situation in contingency table analysis: test statistics do not follow a known theoretical distribution, and solutions of exploratory analyses are very unstable. Therefore, in this situation the k-way matrix is often reduced to a two-way matrix. In such a two-way matrix a frequency f_{ij} represents the number of occasions that category i at time t is followed by

category j at time t+1. When one subject is measured at say 20 time points, there will be 19 transitions. An implicit requirement for the correctness of such reductions in the data is that the transition process is stationary, i.e. that the transition probabilities p(j/i) do not change over time.

Usually, transition matrices are analysed with loglinear models. In this paper we propose to analyse these matrices with correspondence analysis. We will show that this has several advantages for applications in ethology and social mobility respectively, especially in the case that the number of categories is not too small.

We distinguish these two different fields of application because these fields have different traditions in the analysis of transition matrices with which correspondence analysis should be connected. In social mobility table analysis, where generally the sample size n is large and the number of time points T = 2, there is a tradition of model fitting: dozens of models are proposed to account for the structure in these tables. The use of a multi-dimensional scaling technique such as correspondence analysis is uncommon in this context, but can very well supplement the original analyses. In ethology, where often n=1 and T is large, factor analysis is sometimes used to study the transition matrix. It is hoped that only a few tendencies (or motives, drives, etc.) are found that account for the diverse exposed behavior. Factor analysis is used in order to find these tendencies. However, in this context factor analysis is criticized for several reasons. In this paper correspondence analysis is proposed to circumvent some of these criticisms.

In the sequel we will first discuss some of the usual methods to analyse transition matrices. In section 3 correspondence analysis will be explained briefly. In section 4 we consider the two fields, and discuss an example from each field.

2.0 Usual ways to analyse transition matrices.

In this section we give a brief sketch of the most commonly used methods to analyse transition matrices. Such matrices are usually analyzed with the independence model, whereby the scores on the variable for t and t+1 are independent. Mostly, this model fits badly. We will discuss a less restrictive model, the quasi-independence model, which allows special attention to be given to the diagonal elements of the transition matrix, i.e. to the cases in which an object remains in the same state. We will conclude with a brief discussion of Markov chain models, which are treated here as a means to evaluate whether the data can adequately be described by a two-way transition matrix.

A starting point in the analysis of sequences of observations (also called 'sequential analysis') is often the comparison between observed transition frequencies and transition frequencies expected under the assumption that the independence model holds, i.e. when the state on time t+1 does not depend on the state on time t. Under the independence model expected frequencies m_{ij} have the form

$$(1)\quad \hat{m}_{ij} = m_{i+}m_{+j}/m_{++}$$

where a '+' indicates summation over the corresponding index. The difference between the observed frequencies f_{ij} and expected frequencies is tested with Pearson's chi-square statistic

$$(2)\quad \underline{X}^2 = \Sigma\Sigma(f_{ij}-\hat{m}_{ij})^2/\hat{m}_{ij}$$

or with the likelihood ratio statistic

$$(3)\quad G^2 = 2\Sigma\Sigma f_{ij}\log(f_{ij}/\hat{m}_{ij}).$$

When the independence model does not hold (which is usually the case) one either fits a less restrictive model to all cells, or one tests individual cells for significant departure from the expected value under the assumption that the independence model holds. This last approach is often used in sequential analysis.

Although the testing of all cells individually is often done in applications, it should not be recommended. From a statistical point of view, it should be remarked that the tests are not independent. By studying individual cells one may loose sight of the relationship which may exist between significant cells or categories.

In ethological literature often the less restrictive quasi-independence model is fitted. This model states that for some cells in the matrix the expected frequencies are to be equal to the observed frequencies, and for other cells a sort of independence model should hold. This model can be useful when the elements on the diagonal of

the matrix are either extremely large, or extremely low. Extreme diagonal elements can be the result of the sampling strategy that is employed: for example when time sampling is used (e.g. the observed behavior is recorded every 3 seconds) or when only the transitions to other states are registered (in which case the diagonal is empty by design). When for the diagonal elements the observed frequencies are equal to the expected frequencies, the quasi-independence model can be written as

$$(4)\quad m_{ij} = f_{ij} \text{ when } i=j$$
$$m_{ij} = a_i b_j \text{ when } i\neq j,$$

where a_i and b_j can be estimated iteratively (cf. Bishop, Fienberg & Holland, 1975, p.188-202). The difference between observed and expected frequencies is usually tested using (2) or (3). Snijders (1975) developed a test in which it is explicitly taken into account that a transition matrix, and not a contingency table, is being analyzed. The quasi-independence model can also be used when there are off-diagonal structural zeros (e.g. in Slater & Ollason, 1972, where certain behavior states cannot be followed by each other by design) or when there are outliers in the matrix. In this context an outlier is a cell which departs a great deal from the specified restricted model, while other cells do not (Gokhale & Kullback, 1979). For example in communication research, the state 'question asked' is in most cases followed by 'question answered'. In section 4.2 we will analyse an example with an outlier. When the quasi-independence model does not fit - which is almost always the case in social mobility applications - less restrictive models can be fitted. We will come back to these models in section 4.2, when we discuss the social mobility literature. Of course, it is possible to test individual cells for departure from the quasi-independence model, but generally, this is not done.

In applications, it is sometimes investigated whether the two-way transition matrix is an appropriate way to summarize the T-way transition matrix (cf. section 1). This can be done using Markov chain models. When a first-order stationary Markov chain holds, at any time the information on the state at time (t-1) is sufficient to predict the state at time t. Space limitations withhold us from treating this matter more thoroughly. Feller (1968) gives an overview of properties of Markov chains. Bishop et al. (1975, p. 261-279) show that loglinear models can be used to fit Markov chain models.

In Markov chain models transition probabilities play an important role. In the sequel these probabilities, which add row-wise up to one, will be called 'profiles'. As is also the case in usual contingency tables, profiles are important: when the state at time t is known, the profile specifies the probabilities that, given this known state, some other state will follow at time t+1. When for the transition matrix the independence model holds, the profiles in the matrix will be the same. Often one is interested in differences between profiles.

In the next section a method is discussed which facilitates the study of these differences.

3.0 Correspondence analysis

For the analysis of contingency tables, loglinear analysis is already a very popular technique in the English speaking countries. In the last few years there is a growing interest in correspondence analysis, which has been the most important data analytic technique in France for many years. The basic works are those of the group around Benzécri (1973, 1980). In the English speaking world the growing interest is apparent from works written by De Leeuw (1973), Nishisato (1980), Gifi (1981) and Greenacre (1984). Apart from these books, the number of articles and contributions at conferences is growing rapidly.

Strangely enough, correspondence analysis was already known in the English literature for a long time be it under several other names. Nishisato (1980) gives a full survey of the history of correspondence analysis. Greenacre (1984) accentuates that the various approaches have a different rationale and interpretation. He discusses this for the approaches 'reciprocal averaging', 'dual (or optimal) scaling', 'canonical correlation analysis', and 'simultaneous linear regressions'. Van der Heijden (1984) and Van der Heijden & De Leeuw (1985) discuss relations between loglinear analysis and correspondence analysis. The recent flourishing in the use of correspondence analysis as a data analytic technique is probably due to the heavy emphasis put on the geometrical aspect of the method. On the other hand, canonical correlation analysis of categorical data (Kendall & Stuart, 1973, p. 588-598), which is proved by De Leeuw (1971) to be formally identical to correspondence analysis, emphasizes the quantification aspect.

Here we will treat correspondence analysis briefly, emphasizing both geometrical and quantification aspects. A more elaborate description can be found in Van der Heijden & De Leeuw (1985). For details and proofs we refer to Gifi (1981) or Greenacre (1984), and the references mentioned there. First we discuss classical correspondence analysis, and secondly a generalisation of correspondence analysis, proposed by Escofier (1983), and applied in Van der Heijden (1984) and Van der Heijden & De Leeuw (1985).

3.1 Classical correspondence analysis

Correspondence analysis is a technique with which it is possible to construct a multi-dimensional representation of the dependence between the row and column variable of a two-way contingency table. This representation can be constructed using scores found for row and column categories as coordinates for category points. These scores can be normalized in such a way that distances between row points or between column points in Euclidean space are equal to chi-square distances. This property implies that the use of correspondence analysis can be recommended in those cases in which the chi-squared distance is a meaningful measure for the difference between row or column entities of the matrix at hand (D. Sikkel, pers. comm., 1985). One field of application is in the analysis of contingency tables.

Consider a two-way contingency table F with elements f_{ij}, having $\underline{I}$ rows $(i=1,..,i,.,i',.,I)$ and $\underline{J}$ columns $(j=1,..,j,.,j',.,J)$. The chi-square distances are computed on the profiles of the corresponding rows or columns, where for instance the profile of row i is the row of values f_{ij}/f_{i+}. So $\Sigma_j f_{ij}/f_{i+} = 1$. The chi-square distance between rows i and i' is defined as

$$(5)\quad \delta^2(i,i') = \Sigma_j \frac{(f_{ij}/f_{i+} - f_{i'j}/f_{i'+})^2}{f_{+j}/n} .$$

Formula (5) shows that $\delta^2(i,i')$ is a measure of the difference between the profiles of row i and i': when i and i' have the same profile, $\delta^2(i,i') = 0$.

Correspondence analysis proceeds as follows: let F be the matrix to be analysed; D_r and D_c diagonal matrices with marginal row frequencies f_{i+} and column frequencies f_{+j} respectively; $E = D_r tt' D_c/n$,, where $n = f_{++}$ and t is a vector of ones, the length of which depends on the context. Elements of E have the following form

(6) $e_{ij} = f_{i+}f_{+j}/n$,

and thus E is the matrix with expected frequencies computed under independence model (1). Subsequently the singular value decomposition of the matrix $D_r^{-\frac{1}{2}}(F-E)D_c^{-\frac{1}{2}}$ is computed. Elements of this matrix have value $(1/n^{\frac{1}{2}})((f_{ij}-e_{ij})/e_{ij}^{\frac{1}{2}})$, which are standardized residuals scaled by $(1/n)^{\frac{1}{2}}$. These residuals are decomposed with (7):

(7) $D_r^{-\frac{1}{2}}(F-E)D_c^{-\frac{1}{2}} = U\Lambda V'$,

where $U'U = I$, $V'V = I$, and Λ is a diagonal matrix with singular values λ_α in descending order; α is the index for dimension. The dimensionality of the solution is equal to min $(\underline{I}-1, \underline{J}-1)$. For the remaining dimensions $\lambda_\alpha = 0$.
U and V contain scores corresponding with the row and column categories. The scores for rows and columns are normalized as follows:

(8a) $R = D_r^{-\frac{1}{2}}\, U\, n^{\frac{1}{2}}$

(8b) $C = D_c^{-\frac{1}{2}}\, V\, n^{\frac{1}{2}}$.

So $R'D_rR = nI$ and $C'D_cC = nI$. Furthermore $t'D_rR = 0$ and $t'D_cC = 0$: for each dimension row scores and column scores have a weighted variance of 1 and a weighted average of 0.
One can make a simultaneous representation of row and column points in three ways (Gifi, 1981, p. 134-151):
a) by using scores R and $\tilde{C}=C\Lambda$ as coordinates, so that the Euclidean distances between column points are equal to chi-square distances. The weighted variance of the coordinates of the column points equals the eigenvalue λ_α^2 for each dimension.
b) By using $\tilde{R}=R\Lambda$ and C as coordinates so that the analogous result holds for the row points.
c) By using $R\Lambda^{\frac{1}{2}}$ and $C\Lambda^{\frac{1}{2}}$, so that a symmetric representation of row and column points is chosen.
Row scores can be derived from column scores (and column scores from row scores) with the so-called 'transition formulas':

(9a) $R = D_r^{-1}FC\Lambda^{-1}$,

(9b) $C = D_c^{-1}F'R\Lambda^{-1}$.

Bringing Λ from the right to the left side of (9a) and (9b), it can be seen that in the above mentioned simultaneous representation b) the row scores R are in the weighted average of the column scores C, and in a) the column scores C in the weighted average of the row scores R. This property is called the baricentric principle. In these averages the weighting is done by the column and row profiles. The transition formulas define the rationale for the 'reciprocal averaging' approach, since, apart from the multiplicative constant Λ, the row points are in the weighted average of the column points, while at the same time the column points are in the weighted average of the row points.
The so-called reconstitution formula (Benzécri et al., 1973, 1980; Greenacre, 1984, p.93) can be found by substituting (8) in (7):

(10a) $D_r^{-1}(F-E)D_c^{-1}n = R\Lambda C'$,

so that

(10b) $F = E + D_rR\Lambda C'D_cn^{-1} = n^{-1}D_r(tt' + R\Lambda C')D_c$.

Elements of $R\Lambda C'$ are equal to $(f_{ij}-e_{ij})/e_{ij}$. Formulas (10a) and (10b) show that correspondence analysis decomposes the departure from independence in a matrix. This decomposition has the following relation with the well-known Pearson goodness-of-fit $\underline{X}^2$ statistic, which is defined in (2):

(11) trace $\Lambda^2 = \underline{X}^2/n$,

where trace Λ^2, the sum of the eigenvalues, is called the total 'inertia'. Thus correspondence analysis decomposes the $\underline{X}^2$-value of a matrix (Kendall & Stuart, 1971, p. 588-594). The importance of dimension α can be evaluated by the ratio of the inertia of dimension α and the total inertia $\lambda_\alpha^2/\Sigma_\alpha\lambda_\alpha^2$. This quantity can be interpreted as the proportion 'explained' inertia for dimension α, or the proportion of $\underline{X}^2$ that is decomposed in dimension α. Thus the technique tries to picture the most important respects of the dependence of the row and column variables in the first few dimensions.

Clouds of points can be interpreted using chi-square distances: when two row points (or column points) are near each other, their profiles are similar. When profiles differ considerably, the distance between the points is large. The profiles of the marginal row and column frequencies of F, i.e. the profiles with values f_{i+}/n and f_{+j}/n, are projected into the origin. When the distance of a category point to the origin is small, the profile of this category point does not differ much from the mean profile. The distance of row i and column j can be interpreted with the transition formulas; roughly one can say that i and j will be near each other when $f_{ij} >> e_{ij}$, and that i and j are far apart when $f_{ij} << e_{ij}$.

An important aid for interpretating a solution is the property that the sum of the weighted squared distances of the row points (or column points) to the origin, is equal to λ_α^2 for dimension α:

$$(12)\quad \lambda_\alpha^2 = \Sigma_i (f_{i+}/n) r_{i\alpha}^2 \lambda_\alpha^2 = \Sigma_j (f_{+j}/n) c_{j\alpha}^2 \lambda_\alpha^2 .$$

Using (12) one can evaluate the relative contribution of row i to dimension α with the ratio $((f_{i+}/n) r_{i\alpha}^2 \lambda_\alpha^2)/\lambda_\alpha^2$. The same holds for column point j, when one uses the last term of (12). Using Pythagoras' equation, it is also possible to compute for row i on dimension α the ratio of the squared projected distance and squared total distance to the origin. With this ratio it is possible to evaluate how good the total chi-square distance of row i to the origin is represented on dimension α.

In the introduction of this section it was stated that correspondence analysis is formally identical to canonical correlation analysis of contingency tables. From this follows the special relation between λ_1 and the Pearson product-moment correlation coefficient: the correlation between the row and column variable is, under all possible rescalings of the row and column categories, maximal and equal to λ_1, when as quantification for the categories of both variables the scores for the first dimension is taken. λ_2 is equal to the maximal correlation of the quantified variables, where the quantification is restricted to be orthogonal to the quantification for the first dimension, etc. (Kendall & Stuart, 1973, p. 588-594). Correspondence analysis thus finds the maximal canonical correlations between the quantified row and column variable.

Practice with ordinary contingency tables shows that correspondence analysis is a suitable method to gain insight into the relation between the variables of a contingency table when the number of categories is large.

3.2 A generalisation of correspondence analysis

In this section we will briefly describe a generalization of correspondence analysis proposed by Escofier (1983). A more elaborate description can be found in Escofier (1983), Van der Heijden (1984) and Van der Heijden & De Leeuw (1985).

Escofier generalizes correspondence analysis by computing the singular value decomposition of the matrix $S_r^{-\frac{1}{2}}(G_1-G_2)S_c^{-\frac{1}{2}}$ instead of the matrix $D_r^{-\frac{1}{2}}(F-E)D_c^{-\frac{1}{2}}$ (see formula (7)), to find row and column scores R and C, and singular values Λ. Here G_1 and G_2 are matrices of the same size, and S_r and S_c are diagonal matrices with weights for row and column categories. In contrast to classical correspondence analysis, S_r, S_c, G_1 and G_2 are not necessarily related in the way that D_r, D_c and E are to the matrix F.

In the above mentioned references it is indicated that this generalization is difficult to interpret in its most general form. It is advised to use this generalization only in cases that G_1 and G_2 have identical marginal frequencies, which are also taken as diagonal elements of S_r and S_c. Thus the generalization simplifies to the situation that in formula (3) only for E a matrix different from the independence model is taken. When we denote this matrix as G, formula (7), (8), (10) and (12) remain unchanged (apart from replacing E by G). Thus,

$$(13)\quad D_r^{-\frac{1}{2}}(F-G)D_c^{-\frac{1}{2}} = U\Lambda V',$$

etcetera. It should be noted that formula (11) does not hold anymore, since elements of the left term of (13) are not standardized residuals, but are equal to $(f_{ij}-g_{ij})/e_{ij}^{\frac{1}{2}}$, where e_{ij} has the form specified in (2). Formula (9) becomes

$$(14a)\quad R = D_r^{-1}(F-G)C\Lambda^{-1}$$

$$(14b)\quad C = D_c^{-1}(F-G)'R\Lambda^{-1}.$$

Using the appropriate normalization, the Euclidean distances between the points are still equal to chi-square distances. A row point represents the difference between the profiles of the row in F and G. Interpretation of solutions remain basically the same (see Van der Heijden & De Leeuw, 1985).
The reconstitution formula can be found by substituting (8) into (13):

$$(15)\ F = G + D_r R \Lambda C' D_c n^{-1},$$

which is the same as the first two terms of (10b), apart from the fact that E is replaced by G.
This generalization can be used for the analysis of residuals of various sorts of models. In Van der Heijden (1984) and Van der Heijden & De Leeuw (1985) it was used to decompose the difference between two loglinear models. Other possible examples of applications are to take for G expected frequencies following specific models for social mobility tables, confusion matrices, and import-export tables. In this paper we compare observed frequencies with the quasi-independence model.
A further comment has to be made about the choice of the weights in D_r and D_c. When the margins of F and G are equal, it is usually advisable to take these margins as weights in D_r and D_c (see the references mentioned above). In this way the row points are in the weighted average of the column points, and the other way around. Furthermore, a point represents the difference between the profiles of the corresponding category in F and G. A point of a different nature is that by choosing margins as weights for D_r and D_c, it is possible to find the generalized solutions using programs for classical correspondence analysis by taking (F-G+E) as input matrix.
In the context of the quasi-independence model it may be useful to take other weights for D_r and D_c. One possibility is to take as weights the margins of the cells to which the model $m_{ij}=a_i b_j$ (see equation (4)) is fitted. Of course, this makes no difference in case of structural zeros, when $f_{ij}=m_{ij}=0$. However, when there are cells for which $f_{ij}=m_{ij}>0$, this weighting is more "fair", since we are only interested in the cells to which expected values are fitted. Another possible choice for the weights, as suggested by D. Sikkel (pers. comm.), is to take elements a_i as row weights, and b_j as column weights. This has two advantages: first, elements of the left term of (13) become equal to standardized residuals $(f_{ij}-g_{ij})/g_{ij}^{\frac{1}{2}}$ scaled by $(1/n^{\frac{1}{2}})$. (When $g_{ij}=0$, the residuals will be zero). Therefore, (11) still holds, and the solution shows us the decomposition of the appropriate chi-squared statistic. Secondly, this weighting is also "fair". However, in spite of these advantages, we haven chosen for the usual weighting, since in this application we want to study the differences between profiles in F and G. Using the alternative weightings, this property is lost (compare equation (14)). However, in most cases these three alternatives will not lead to substantially different solutions.

4.0 Correspondence analysis of transition matrices

The rationale for applying correspondence analysis to transition matrices is the following. First, we saw in section 2 that profiles are an important concept in the analysis of transition matrices. It is often interesting to study whether and how two row profiles differ. This can be done using correspondence analysis, since, with the appropriate normalization, Euclidean distances between the rows are equal to chi-square distances. These distances can be interpreted as a measure of the difference between rows. Secondly, one of the arguments in section 2 for not recommending the study of significant cells was that one might loose sight of the relationship that may exist between significant cells. Correspondence analysis can be used to find this relationship. A third reason to apply correspondence analysis is that it can be used to find maximal canonical correlations between the row and column variable. This will be especially interesting in the case of social mobility tables, since there the number of time points T equals 2, and the canonical correlation can be interpreted as the correlation between for instance the occupations of a father and son. Finally, in ethology there is a tradition of factor analysis of standardized residuals of transition matrices. However, this approach can be critized for various reasons. Correspondence analysis, which also uses standardized residuals, circumvents many of these criticisms. We will come back to factor analysis of transition matrices in section 4.2.
We have the following reasons for using Escofier's generalisation of correspondence analysis. First, we need it to compare the observed transition frequencies with other models than the independence model.

Often the diagonal elements of transition matrices cause a degeneration of classical correspondence analysis solutions. These elements are often either extremely high (compared with the diagonal elements for the independence model) or zero. In the sequel we will show examples of this. Here we will use the generalization to compare the observed frequencies with the quasi-independence model. In this model, for some cells the expected frequencies are equal to the observed frequencies, so that for these cells there is no difference to be reconstituted (see formula (15)). We think this is an elegant way to get rid of the usual dominating influence of the diagonal elements in classical correspondence analysis. A second reason to use a generalization of correspondence analysis is the following: single cells sometimes cause the degeneration of a correspondence analysis solution, in the sense that the solution is uninteresting. This is likely to happen when cell f_{ij} is an outlier in the sense we discussed in section 2. In section 3 we pointed out that the categories of the row and column variable get a quantification in one or more dimensions (depending on the smallest number of rows and columns). A single cell can sometimes produce a dichotomy in the quantifications: for instance row category i and column category j have an extreme quantification, whereas the other categories have about the same non-extreme quantification. This problem is solved by defining these cells as structural zeros (Bishop, Fienberg & Holland, 1975, p.177-202) before computing expected frequencies, and decomposing the difference between thus defined quasi-independence model and the observed frequencies.
We continue with two examples of transition matrices: one from ethology, and one from social mobility research. We have chosen for examples with large numbers of categories, since in these situations the advantages of correspondence analysis are most clearly visible. All correspondence analysis plots are made using the symmetric normalization.

4.1 An example from ethology

In ethology sequences are frequently encountered and they are analysed in various ways. A recent survey is Van Hooff (1982). Sometimes transition matrices are analysed with factor analysis on correlation matrices constructed from a transition matrix. This procedure is particularly useful when transition matrices are large, and when it is assumed that the manifest behavior states are triggered by a

smaller number of latent motives, drives, or comparable concepts. In this context factor analysis is used to find these concepts. First we will provide a more detailed description of this approach, then we discuss some criticisms. Finally it is shown how correspondence analysis circumvents these criticisms.
Wiepkema (1961) was the first to apply factor analysis to transition matrices, in order to study the reproductive behavior of the bitterling. He transformed the transition matrix to a matrix with elements $f_{ij}/\hat{m}_{ij}$, where the expected values m_{ij} have the form as in formula (1). Subsequently Spearman rank correlations were computed between the rows, and between the columns, yielding two correlation matrices. On these matrices factor analysis was performed. A varimax rotation was used to facilitate interpretation. In some later applications this approach was modified by computing elements $(f_{ij}-\hat{m}_{ij})/\hat{m}_{ij}^{\frac{1}{2}}$ - standardized residuals - instead of $f_{ij}/\hat{m}_{ij}$. Balthazar (1972) compared results using not only the Spearman rank correlation, but also Kendall's rank correlation and Pearson's product-moment correlation.
The following criticisms were raised against this procedure. First, factor analysis of the correlation matrix for the rows can produce results different from the factor analysis for the columns. These differences are due to the asymmetry of the matrix. Differences can be found, for instance for the number of extracted factors, for the interpretation of the factors, and for the values of the component loadings. These differences lead to different conclusions on the number and types of latent drives, between which it is difficult to choose. A second criticism is raised by Balthazar (1972), who showed examples in which correlations sometimes do not reflect the observed associations between the data: it is possible that a correlation is around zero or negative while two kinds of behavior trigger each other. He stressed that correlations should be meaningful in the context in which they are used.
We propose to analyse transition matrices with correspondence analysis rather then with factor analysis. It seems that the two criticisms can be circumvented using correspondence analysis: first correspondence analysis produces the same number of factors for the rows and for the columns. Corresponding states can have a different position on these factors, which is the result from asymmetries in the data matrix. Secondly, the chi-square distance seems a meaningful association

measure for the similarity between two rows or two columns, as was pointed out in section 4.0. Clearly, in this context the chisquare distance seems to be a more meaningful measure then the correlation coefficient. Another advantage is that it is not necessary anymore to choose between $f_{ij}/\hat{m}_{ij}$ versus standardized residuals, and between the three sorts of correlation coefficients. Furthermore, in section 3 it was stressed that classical correspondence analysis can be described as a method with which a matrix with standardized residuals is decomposed. In the ethological tradition sometimes correlations are computed on these standardized residuals, followed by the factor analysis decomposition of the correlation matrix. From this it is clear that correspondence analysis stays much 'closer' to the original data then factor analysis does. A last advantage is that correspondence analysis seems more malleable when certain cells should not influence the final solution, because it is possible to choose between a large variety of quasi-independence models.

We will now discuss the analysis of a transition matrix, that we have taken from Slater & Ollason (1972). This matrix has the special property that there are off-diagonal structural zeros. We will show how to deal with such a situation. The behavior of isolated male zebra finches was recorded for several birds. As an example Slater & Ollason provided the transition matrix for bird 27 (table 1). The behavior states are stretching (STRTCH), locomotion (LOCOM), singing (SING), preening (PREEN), scratching (SCRTCH), bill-wiping (BWIP), sand taking (SAND), stereotypes (STYP), drinking (DRINK), feeding (FEED), gaps in behavior (GAP), ruffling (RUFFL), wing shaking (WSHK) and cuttle fish bone taking (FBONE). Precise descriptions of the

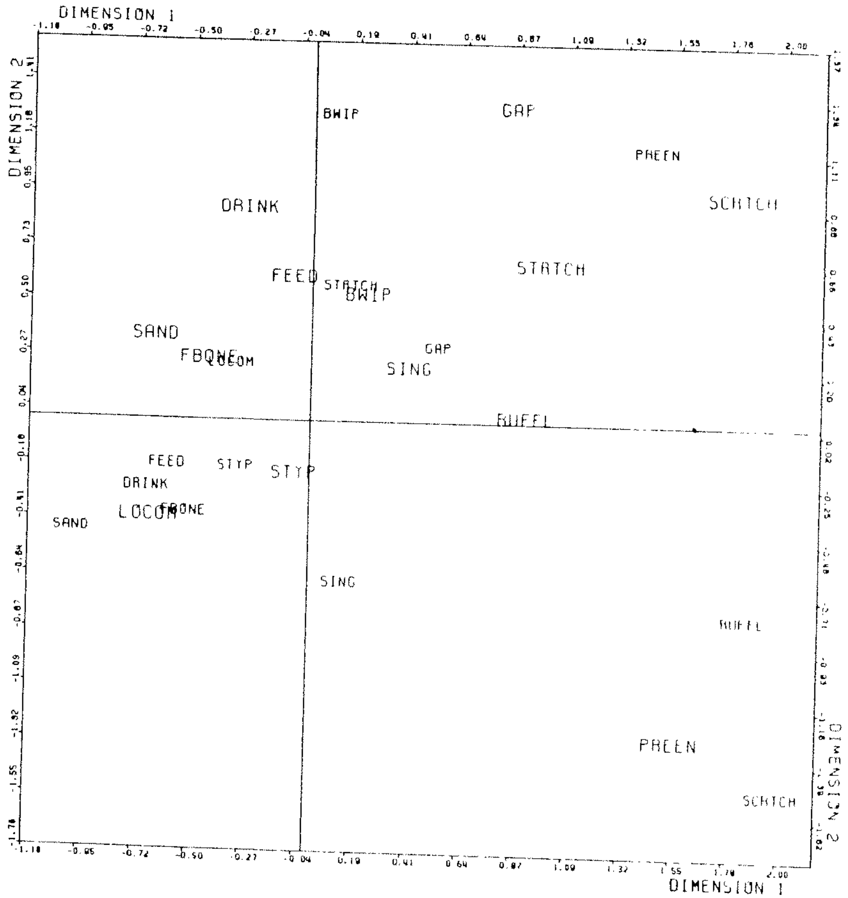
Table 1: The matrix of transitions shown by bird 27.
Source: Slater & Ollason (1972).

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.
1. Stretching	-	3	3	3	0	2	0	0	1	0	1	1	0	0
2. Locomotion	5	-	129	8	3	16	2	23	82	42	7	5	0	32
3. Singing	2	121	-	21	2	3	0	4	1	3	2	11	0	4
4. Preening	1	19	18	-	9	0	0	1	2	1	2	15	8	2
5. Scratching	0	2	4	7	-	1	0	0	0	0	1	1	2	0
6. Bill wiping	2	25	2	6	0	-	0	0	6	2	0	2	0	0
7. Sand taking	0	2	0	0	0	0	-	0	-	-	0	0	0	-
8. Stereotyping	0	18	2	0	1	1	0	-	3	3	1	2	0	2
9. Drinking	3	69	0	3	1	16	-	1	-	-	2	0	0	-
10. Feeding	1	39	1	5	0	2	-	1	-	-	3	1	0	-
11. Gaps in behavior	0	7	0	5	0	3	0	2	0	1	-	1	0	0
12. Ruffling	0	16	12	8	2	1	0	0	0	0	0	-	0	0
13. Wing shaking	0	1	0	9	0	0	0	0	0	0	0	0	-	0
14. Cuttle fish bone taking	0	33	3	2	0	0	-	1	-	-	0	0	0	-
	14	355	174	77	18	45	2	33	95	52	19	39	10	40

behaviors can be found in Slater & Ollason (1972). Only transitions to other states were recorded: a state cannot be followed by itself. Therefore the diagonal frequencies are zero. Furthermore, the matrix contains off-diagonal structural zeroes, since the four states sand taking, drinking, feeding and cuttle fish bone taking cannot follow each other: these activities take place in different locations, and are therefore interspersed with locomotion. We have omitted the category wing shaking from all analyses, since it has a very low marginal frequency, and it is followed and follows preening almost exclusively. In the plots we made, it was placed at an extreme and therefore clouded the relationship between the other states.

Since we do not want the structural zeros to influence the solution, we use generalized correspondence analysis, where the differences between the observed frequencies and expected frequencies according to a quasi-independence model are decomposed. In this example, for all structural zeros $f_{ij} = \hat{m}_{ij} = 0$, and for the other cells $m_{ij} = a_i b_j$. The departure from the quasiindependence model is significant: $X^2 = 557$ with 119 degrees of freedom. The resulting correspondence analysis solution is displayed in figure 1. On the right hand side of the plot the grooming behaviors can be found, while on the left hand side the feeding and drinking behaviors are shown near locomotion, for (t) and (t+1). So, two clusters of behaviors are distinguished, accounting for 47% of the departure from quasiindependence. On the one hand, bird 27 performs different grooming behaviors which follow each other, on the other hand it performs a lot of nurturing behaviors with locomotion in between. It is not possible to interpret the second dimension in terms of clusters of behaviors, because corresponding points are on opposite sides from the origin. Therefore we should interprete this dimension in terms of asymmetries in the data (since if the data matrix would have been symmetric, corresponding points would have the same coordinates). The important asymmetries can be found by comparing the row and column contributions for corresponding points to the first and second dimension. On dimension 1 large differences can be found for e.g. ruffling (contribution to first dimension on time is t: .07; time is t+1: .29) and locomotion (time t: .32; t+1: .08), on dimension 2 for singing (t: .03; t+1: .22), bill wiping (t: .04; t+1: .21) and ruffling (t: .00; t+1: .07). A further inspection of the data teaches us e.g. that ruffling is followed (relatively) much more by locomotion then the other way around; that preening is followed much more by bill wiping than the other way around. It is not clear to us how these peculiarities should be interpreted. We do not discuss this further, because our main purpose was to show how one can deal with structural zeroes. For substantive discussions on zebra finches we refer to Slater & Ollason (1972). In the next example we discuss the situation where some frequencies are extremely high.

Figure 1: Slater & Ollason, generalized correspondence analysis, structural zeroes excluded; $\underline{X}^2 = 557$, df is 119; large label denotes t, small label t+1.
$\lambda_1 = .492\ (.465)$; $\lambda_2 = .361\ (.250)$; $\lambda_3 = .255\ (.125)$



4.2 An example from social mobility

Social mobility tables are most often analyzed using extensions of loglinear models: a simple model such as the independence model does not fit, while the saturated model uses too many parameters to describe the departure from independence. Haberman (1974) mentions two special aspects of social mobility tables. First, inheritance of status requires special attention - which implies special attention to the diagonal of the table. Secondly, in case of transitions to a state different from the original state, the stepsize is important: small steps are more likely than large steps. Models should account for the fact that the states are ordered in this respect social mobility tables differ from other transition matrices. To account for these properties, very many models have been proposed, e.g. by Goodman (1979), Haberman (1974), Bishop, Fienberg & Holland (1975), and Duncan (1979). In these models parameters are fitted to account for concepts as 'occupational inheritance', 'occupational immobility and/or persistence', 'overall upward or downward mobility', 'occupational mobility inertia', 'occupational mobility barriers' and 'perfect mobility'.

In examples matrices are usually highly aggregated before models are fitted to them. Only recently has attention been given to the question in which situation this results in a loss of information (Breiger, 1981; Goodman, 1981). We will show that from a data analytic point of view such an aggregation is not necessary. Furthermore we will illustrate how correspondence analysis can be used to trace the structure in social mobility tables. Earlier this was done by Klatzky and Hodge (1971), using classical correspondence analysis. We will

TABLE 2: Social mobility from father's occupation to son's first full-time occupation

FATHER'S OCCUPATION ↴	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.
1. Professional, self-employed	25	107	20	11	3	27	8	8	8	8	5	12	11	4	7	2	1
2. Professional, salaried	8	395	64	42	9	116	40	34	59	15	40	61	75	20	30	4	16
3. Managers	14	317	116	89	6	144	68	40	56	31	33	82	89	16	60	5	7
4. Sales, other	7	120	34	45	2	73	27	13	25	7	11	31	33	5	21	1	4
5. Proprietors	19	187	69	52	33	112	82	34	50	36	35	69	103	24	63	11	14
6. Clerks	4	203	41	26	4	145	42	23	49	19	34	74	95	28	42	3	13
7. Sales, retail	5	77	20	20	2	57	47	13	23	8	12	48	46	11	19	3	4
8. Crafts, manufacturing	7	208	49	33	2	174	52	151	67	29	59	104	262	81	75	8	16
9. Crafts, other	6	215	54	38	5	172	54	65	195	53	57	195	175	47	110	6	34
10. Crafts, construction	6	132	29	20	6	131	51	54	71	170	43	142	158	39	130	4	40
11. Service	8	122	47	24	3	170	49	56	65	36	126	130	174	51	110	6	37
12. Operatives, other	.5	142	33	18	6	184	71	55	102	49	81	391	239	67	149	11	58
13. Operatives, manufacturing	9	160	37	28	1	188	75	108	93	36	89	171	529	118	123	10	36
14. Laborers, manufacturing	2	33	5	5	0	40	13	26	22	13	25	55	97	92	38	1	26
15. Laborers, other	4	54	11	8	6	86	24	37	42	30	53	101	126	47	142	6	45
16. Farmers	13	252	58	34	10	188	94	86	145	121	102	323	399	150	259	457	981
17. Farm laborers	2	39	8	3	3	39	23	27	42	21	40	96	114	46	83	18	376

also do this using generalized correspondence analysis. The data from our example were taken from Featherman & Hauser (1975), and were also analysed by Breiger (1981) (see table 2). Breiger aggregated the categories on different theoretical grounds and tested which level of aggregation is permissible. After this he fitted several models to his aggregated data.

First we analyzed these data with classical correspondence analysis. The first singular value, which can be interpreted as the first canonical correlation between the occupation of the father and the son, is .512. However, this correlation is for the greatest part based on the difference between farmers and farm laborers versus the other occupations. After ommitting these two categories, the analysis on the 15x15 matrix resulted in a correlation of .356.

Let us presume we are interested in the cases that sons do not have the same occupations as their fathers. In this case an interesting analysis is to compare the observed frequencies with the quasi-independence model as defined in (4). In the context of social mobility this model is called the 'quasi-perfect mobility model' with which is meant that, apart from occupation inheritence, the occupation of father and son are not related. The result of this correspondence analysis is shown in figure 2. The solution is dominated by the outlier frequency for farmer (father) (contribution on first dimension .693) and farm laborer (son) (.738). To have a better view of the relations between the other off-diagonal cells, we repeat this analysis by defining the quasi-independence model in such a way that, not only for the diagonal elements, but also for this outlier $f_{ij}=\hat{m}_{ij}$.

The correspondence analysis solution can be found in figure 3. Only the first dimension is shown since 68% of the total inertia is decomposed in this dimension. The eigenvalues of the second and third dimension are not clearly separated. So, apart from the transition from farmer to farm laborer, there seems to be a one-dimensional structure for the cases that fathers and sons do not have the same occupation. From this dimension it can be seen how a certain occupation for fathers is followed by different occupations for the sons, and vice versa. Note that the categories of fathers and sons are not ordered in the same way. For fathers this order is self-employed professional, manager, other sales, proprietor, clerk etcetera, for sons the order is other sales, salaried professional, manager, self-employed professional and so on. If the solution were perfectly one-dimensional, this could be interpreted as that for instance fathers who are self-employed professional, have sons who become (corrected for marginal frequency) other sales, salaried professional, manager etc, while sons who are other sales have mostly fathers are self-employed professional, then manager, proprietor, clerk and so on.

<u>Figure 2</u>: Breiger, generalized correspondence analysis, diagonal elements excluded; $\underline{X}^2$ = 4785, df is 239.
large labels denote occupations of fathers.
λ_1 = .367 (.658); λ_2 = .210 (.215); λ_3 = .081 (.032)

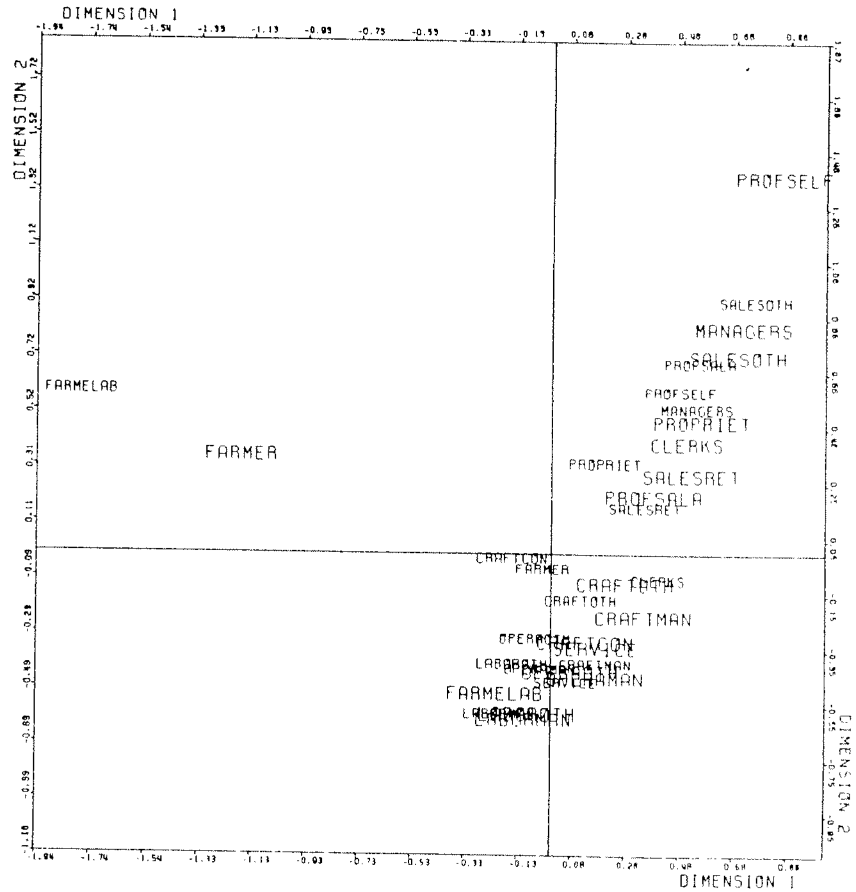
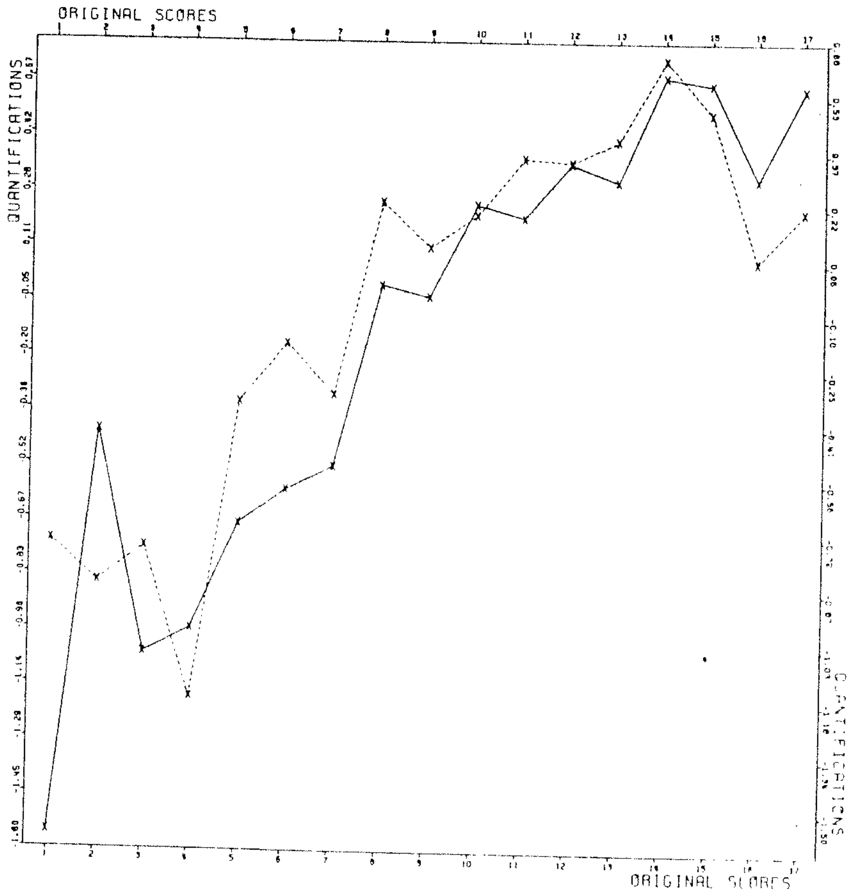


Figure 3: Breiger, generalized correspondence analysis, diagonal elements and cel (FARMER;FARMLAB) excluded; $\underline{X}^2$ = 2058, df is 238.
$\lambda_1 = .248$ (.680); $\lambda_2 = .083$ (.077); $\lambda_3 = .078$ (.068)
Original category numbers vs. quantifications for first dimension.
Father-line is solid.



5.0 Conclusions and discussion

It is shown that correspondence analysis is a suitable method for the analysis of transition matrices. This has the following reasons: in transition matrices the profile concept is an important one, while in correspondence analysis distances between these profiles are projected on low-dimensional spaces; correspondence analysis can help to find

relations between significant cells; maximal canonical correlations between the row and column variable are provided; correspondence analysis is an alternative for the use of factor analysis of transition matrices in ethology. Furthermore, correspondence analysis provides a clear view of the important asymmetries in the transition matrix.

A generalization of correspondence analysis can be used to decompose the departure from the quasi-independence model. In this way problems with diagonal cells and outlier cells can be circumvented. Correspondence analysis is probably also a suitable method to study the departure from models other than the quasi-independence model. Furthermore, the study of the departure from the quasi-independence model will probably also be a good way to tackle similar problems with diagonal cells and outlier cells in other matrices, such as confusion matrices, import-export tables, and ordinary contingency tables. These two areas need further research.

References

Balthazar, J. (1972). Validité de l'application de l'analyse factorielle á l'étude causale d'une sequence comportementale avec un essai d'analyse du comportement agonistique chez Tilapia macrochir (Boulenger 1912). Annales de la Société Royale Zoologique de Belgique, 102, 3-34.

Benzecri, 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.

Breiger R.L. (1981). The social class structure of occupational mobility. Amer. Journ. of Soc., 87, 578-611.

Budescu, D.V. (1984). Tests of lagged dominance in sequential dyadic interaction. Psychological Bulletin, 96, 402-414.

Duncan, O.D. (1979). How destination depends on origin in the occupational mobility table. Amer. Journ of Soc., 84, 793-803.

Escofier B. (1983). Analyse de la difference entre deux mesures sur le produit de deux mêmes ensembles. Cahiers de l'analyse des données, 3, 325-329.

Featherman, D.L. & Hauser, R.M. (1975). Design for a replicate study of social mobility in the United States. Pp. 219-252 in Land, K.C., & Spilerman, S. (eds.) Social indicator Models. New York: Russell Sage.

Feller, W. (1968). An introduction to probability theory and its applications. (Vol. 1, 3rd ed.). New York: Wiley.

Gifi A. (1981). Non-linear multivariate analysis. Leiden: R.U.L./F.S.W., Department of Data Theory. (2nd edition, in press. Leiden: DSWO-Press.

Gokhale, D.V., & Kullback, S. (1979). The information in contingency tables. New York: Dekker.

Goodman, L.A. (1979). Multiplicative models for the analysis of occupational mobility tables and other kinds of cross-classifications. Amer. Journ. of Soc., 84, 804-819.

Goodman, L.A. (1981). Criteria for determining whether certain categories in a cross-classification table should be combined, with special reference to occupational categories in an occupational mobility table. American Journal of Sociology, 87, 612-650.

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

Haberman, S.J. (1974) The analysis of frequency data. Chicago: Univ. of Chicago Press.

Heijden, P.G.M. van der (1984). Relaties tussen loglineaire analyse en correspondentieanalyse. Leiden: R.U.L./F.S.W., Department of Data Theory.

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

Hooff, J.A.R.A.M. (1982). Categories and sequences of behavior: methods of description and analysis. In: Scherer, K.R. & Ekman, P. (eds.) Handbook of methods in nonverbal behavior research, Cambridge: Cambridge Univ. Press.

Kendall, D.G. & Stuart, A. (1973). The advanced theory of statistics Vol. 2, 3rd. ed. London: Griffin.

Klatzky, S.R. & Hodge, W.J. (1971). A canonical correlation analysis of occupational mobility. Journal of the American Statistical Association, 66, 16-22.

Leeuw, J. de (1973). Canonical analysis of categorical data. Dissertation, University of Leiden, Leiden. (1985, Leiden: DSWO-Press).

Leeuw, J. de (1971). Canonical analysis of contingency tables.(RB-002-71). Leiden: F.S.W./R.U.L., Department of Datatheorie.

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

Slater, P.J.B., & Ollason, J.C. (1972). The temporal patterning of behaviour in isolated male zebra finches: transition analysis. Behaviour, 42, 248-269.

Snijders, T. (1975) A test for randomness in behaviour. Statistica Neerlandica, 29, 39-48.

Wiepkema, P.R. (1961) An ethological analysis of the reproductive behaviour of the bitterling (Rhodeus amarus Bloch). Archives Néerlandaises Zoologie, 14, 103-199.

KM 19(1985)
pag 73 - 87

PATH ANALYSIS FOR MIXED QUALITATIVE AND QUANTITATIVE VARIABLES

by

Abby Z. Israëls*)

Abstract

In this report path analysis models are considered for mixed qualitative/quantitative variables. Only endogenous variables that are dependent in all its relations are supposed to be quantitative, but this restriction can easily be dropped. Qualitative variables are handled using a dummy-variable for each category. Parameters are estimated by ordinary least squares. The method allows decomposition of total effects in direct and indirect effects, which makes interpretation easier.

*) Netherlands Central Bureau of Statistics, Department for Statistical Methods, P.O. Box 959, 2270 AZ VOORBURG. Tel. 070-694341. The views expressed in this paper are those of the author and do not necessarily reflect the policies of the Netherlands Central Bureau of Statistics. The author wishes to thank an anonymous reviewer for several valuable suggestions.