~~To appear in:~~ Did appear in:
Diday (Ed.), Data Analysis & Informatics IV.
Amsterdam: North Holland. 1986

TRANSITION MATRICES, MODEL FITTING AND CORRESPONDENCE ANALYSIS

Peter G.M. van der Heijden
Dept. of Research Methods and Psychometrics
Hoolgracht 15
2312 KM Leiden
The Netherlands

Transition matrices are frequently analysed using the independence and quasi-independence model. In the context of transition matrices this last model is used to eliminate the influence of diagonal elements. Usually the independence models do not fit. In this paper we propose to analyze the residuals from these models with (generalized) correspondence analysis. It is also discussed how to deal with structural zeros in the context of correspondence analysis.

1.0 Introduction. Longitudinal categorical data are frequently summarized in two-way transition matrices with elements f(ij), where f(ij) is the number of individuals in state i at time t (1,..,t,..,T) and in state j at time t+1. For each individual the number of transitions equals T-1. When k is the number of individuals, the total frequency os the transition matrix equals $f(++)=k(T-1)$, where '+' denotes summing over the corresponding index.
In this paper we propose to analyse two-way transition matrices with correspondence analysis, and a generalization of correspondence analysis. We will show that this has several advantages for applications in ethology and social mobility respectively. 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.

2. Usual ways to analyze transition matrices. A starting point in the analysis of sequences of observations (also called 'sequential analysis') is often to compare the observed transition frequencies with the transition frequencies expected under the assumption that the independence model would hold, 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 $m(ij) = m(i+)m(+j)/m(++)$. The difference between the observed frequencies and expected frequencies can be tested with Pearson's chi-square statistic $X^2 = \Sigma\Sigma(f(ij)-\hat{m}(ij))^2/\hat{m}(ij)$. 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 independence. This last approach is often used in ethology. Although the testing of all cells individually is often done in applications, it should not be recommended because the tests are not independent, and one may loose sight of the structure which may exist between significant cells.
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 an 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 from the chosen sampling strategy: for example when time sampling is used, the observed behavior is recorded after each fixed time-interval; when only the transitions to other states are notified, the diagonal is zero by design (structural zeros). In this case the quasi-independence model can be written as

(1) $m(ij) = f(ij)$ when $i=j$; $m(ij) = a(i)b(j)$ when $i \neq j$,

where a(i) and b(j) can be estimated iteratively (cf. [2]). The quasi-independence model can also be used when there are off-diagonal structural zeros. When the quasi-independence model does not fit, less restricted models can be fitted (see section 4.2). Of course, it is possible to test individual cells for departure from model (1), but generally this is not done.
In the sequel transition probabilities, which add row-wise up to one, will be called 'profiles'. Profiles are important since, when the state at time t is known, the state's profile specifies the probabilities that some other state will follow at time t+1. Often one is interested in differences between profiles. Correspondence analysis facilitates the study of these differences.

3.0 Correspondence analysis. We will treat correspondence analysis briefly here, emphasizing geometrical and quantification aspects. For details and proofs we refer to [1], [5] and [6]. First we discuss classical correspondence analysis, and secondly a generalization of correspondence analysis, proposed in [4].
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 variables of a two-way contingency table: scores found for the categories can be used as coordinates for the respective row and column 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. The solution can be found as follows: let F be the matrix to be analyzed; D_r and D_c diagonal matrices with marginal frequencies f(i+) and f(+j); E the matrix with expected frequencies computed under the independence model. First 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 are equal to standardized residuals, scaled by $(1/n^{\frac{1}{2}})$. These residuals are decomposed with (2):

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

where U'U = I, V'V = I, and Λ is a diagonal matrix with singular values $\lambda(\alpha)$ in descending order; α is the index for dimension. The dimensionality of the solution is min (I-1, J-1). The scores for rows and columns are normalized as in (3):

$$(3a)\ R = D_r^{-\frac{1}{2}}Un^{\frac{1}{2}};\quad (3b)\ C = D_c^{-\frac{1}{2}}Vn^{\frac{1}{2}}.$$

The so-called reconstitution formula is found by substituting (3) in (2):

$$(4)\quad F = E + D_r R\Lambda C' D_c n^{-1}.$$

Formula (4) shows that the departure from independence is decomposed. This decomposition has the following relation with the X^2-statistic: trace $\Lambda^2 = X^2/n$. The importance of dimension α can be evaluated by the ratio of the inertia of dimension α and the total inertia: $\lambda(\alpha)^2/\Sigma\lambda(\alpha)^2$. This quantity can be interpreted as the proportion of X^2 that is decomposed in dimension α.
Clouds of points can be interpreted using the chi-square distances: when two rows (or columns) are near each other, their profiles are similar. Row i and column j will be near each other when f(ij)>>e(ij). An important aid in the interpretation can be found in the property that the sum of the weighted squared distances of the row points (or column points) to the origin, is equal to $\lambda(\alpha)^2$ for dimension α. Using this one can evaluate the relative contribution of row i to dimension α. Correspondence analysis is formally identical to canonical correlation analysis of contingency tables. $\lambda(1)$ and the product-moment correlation are related: the correlation between the row and the column variable is, under all possible rescalings of the categories, maximal and equal to $\lambda(1)$, when as quantification for the categories the scores for the first dimension are taken. $\lambda(2)$ is the second maximal correlation of the (orthogonally quantified) variables, etc. Correspondence analysis thus finds the maximal canonical correlations between the quantified row and column variables.
3.2 A generalization of correspondence analysis. In this section we will describe a generalization of correspondence analysis proposed in [4] (see also [7]). In [4], Escofier generalizes correspondence analysis by computing the singular value decomposition of the matrix $S_r^{-2}(G_1-G_2)S_c^{-2}$ to find scores R and C, and singular values Λ. Here G_1 and G_2 matrices of the same size, and S_r and S_c are diagonal

matrices with weights for row and column categories. S_r, S_c, G_1 and G_2 and not necessarily related in the way that D_r, D_c and E are to the matrix F. This generalization is difficult to interpret in its most general form. It is advized to use this generalization only in cases that G_1 and G_2 have identical marginal frequencies, and to take these as diagonal elements of S_r and S_c. Thus the generalization simplifies to the situation that for E a matrix different from the independence model is taken. When we denote this matrix as G, formulas (2), (3) and (4) remain unchanged (apart from replacing E by G). 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 remains basically the same.

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 the profile concept is an important one in the analysis of transition matrices. The difference between two rows of transition probabilities can be studied using correspondence analysis since, with the appropriate normalization, Euclidean distances between the rows are equal to chi-square distances. 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 these cells. Correspondence analysis can be used to find this structure. A third reason to apply correspondence analysis is to find the maximal canonical correlations between the row and column variable. Finally, in ethology there is a tradition of factor analysis of transition matrices. However, this approach can be criticized for various reasons. Correspondence analysis circumvents many of these criticisms.
We need Escofier's generalization of correspondence analysis to compare observed transition frequencies with expected frequencies following the quasi-independence model (1). Thus for the diagonal cells there is no difference to be reconstituted. We think this is an elegant way to get rid of the usual dominating influence of these elements in classical solutions.
4.1 An example from ethology. A special procedure sometimes used in the analysis of transition matrices is factor analysis performed on correlation matrices constructed from a transition matrix. This procedure is thought to be useful when the transition matrix is large, and when it is assumed that the manifest behavior states are triggered by a smaller number of latent motives, drives, or a comparable concept. References can be found in [8]. In this procedure the transition matrix is transformed to a matrix with elements $f(ij)/\hat{m}(ij)$, where $m(ij)=m(i+)m(+j)/m(++)$. Subsequently Spearman rank corelations are computed between rows, and columns, yielding two correlation matrices. On these matrices factor analysis is performed. A varimax rotation is used to make interpretation easier. In some later applications this approach is modified by computing standardized residuals $(f(ij)-\hat{m}(ij))/\hat{m}(ij)^{\frac{1}{2}}$. The following criticisms were raised against this procedure. Firstly, factor analysis of the correlation matrix for the rows and for the columns yields different results, due to the asymmetry of the matrix. A second criticism is that correlations do sometimes not reflect the observed associations in the data: it can occur that a correlation is around zero or negative while two behaviors trigger each other.
We propose to analyze transition matrices with correspondence analysis rather than with factor analysis. It seems that the two criticisms can be circumvented using correspondence analysis: the same number of factors for the rows and for the columns is produced, and the chi-square distance seems a more meaningfull measure for the similarity between two rows or two columns. Furthermore, correspondence analysis is a method ith which standardized residuals can be decomposed: it stays 'closer' to the original data than factor analysis does.
We will now discuss the analysis of a transition matrix, that we have taken from [9]. 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. The transition matrix for bird 27 is shown in table 1. Only transitions to other states were recorded: the diagonal cells are structural zeros. Furthermore, the matrix contains off-diagonal

TABLE 1: Transitions shown by bird 27

time t		time t+1												
stretching	(STR)	-	3	3	3	0	2	0	0	1	0	1	1	0
locomotion	(LOCO)	5	-	129	8	3	16	2	23	82	42	7	5	32
singing	(SNG)	2	121	-	21	2	3	0	4	1	3	2	11	4
preening	(PR)	1	19	18	-	9	0	0	1	2	1	2	15	2
scratching	(SCR)	0	2	4	7	-	1	0	0	0	0	1	1	0
bill-wiping	(BW)	2	25	2	6	0	-	0	0	6	2	0	2	0
sand taking	(SND)	0	2	0	0	0	0	-	0	-	-	0	0	-
stereotyping	(STYP)	0	18	2	0	1	1	0	-	3	3	1	2	2
drinking	(DR)	3	69	0	3	1	16	-	1	-	-	2	0	-
feeding	(FD)	1	39	1	5	0	2	-	1	-	-	3	1	-
gaps in behavior	(GAP)	0	7	0	5	0	3	0	2	0	1	-	1	0
ruffling	(RF)	0	16	12	8	2	1	0	0	0	0	0	-	0
cuttle fish bone taking	(CFB)	0	33	3	2	0	0	-	1	-	-	0	0	-
		14	354	174	68	18	45	2	33	95	52	19	39	40

NOTE: '-' denotes structural zero

structural zeros, since the four states SND, DR, FD and CFB cannot follow each other: these activities take place in different locations, and are therefore interspersed with LOCO. The frequencies differ significantly from the independence model: X^2=898, df is 144. We firstly performed an analysis to investigate whether there is any structure in the departure from independence (figure 1). The first dimension is dominated by the structural zeros: since for the diagonal elements f(ij)<<$\hat{m}$(ij), it follows that for almost all states the points for t and t+1 are placed at different sides of the origin. The off-diagonal structural zeros cause LOCO(t) to be on the left with CFB, DR, FD and SND (t+1). On the right the reverse situation is found. The second dimension is somehow connected with grooming: here we find points for SCR and RF following PR. We are not satisfied with this solution: the first dimension is trivial because it reflects the design.

A soltion that is not influenced by the structural zeros can be found using the generalization of correspondence analysis, where the difference between observed frequencies and expected frequencies following the quasi-independence model is decomposed. In this example, for all structural zeros f(ij)=$\hat{m}$(ij)=0, for the other cells m(ij)=a(i)b(j). The departure from the quasi-independence model is also significant: X^2=557, df is 119. The resulting solution can be found in figure 2. On the right the grooming behaviors can be found, on the left the nurturing behaviors are near LOCO (t) and LOCO (t+1). The second dimension is showing

Figure 1: Correspondence analysis of table 1
λ_1 = .78 (.65), λ_2 = .38 (.15); time t is large

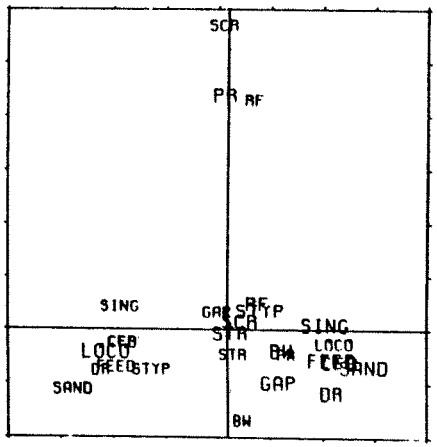


Figure 2: Structural zeros excluded
λ_1 = .49 (.47), λ_2 = .36 (.25)

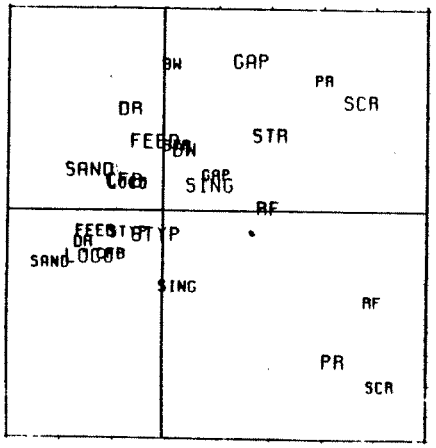


TABLE 2: Social mobility from father's occupation to son's first full-time occupation

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

asymmetries in the matrix. Space limitations withhold us from treating the interpretation more thoroughly.

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, and the saturated model uses to many parameters to describe the departure from independence. Two special properties of social mobility tables should be mentioned. Firstly, inheritance of status requires special attention, which implies special attention to the diagonal of the table. Secondly, in case of transition 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 in which parameters are fitted to describe these properties.

In examples the matrices are usually highly aggrgated before models are fitted to them. We will show that such an aggregation is not necessary from a data-analytic point of view. We will illustrate how correspondence analysis can be used to trace the structure in social mobility tables. The data from our example stem from [3] (see table 2).

Firstly we analyzed this data with classical correspondence analysis. The first two dimensions are plotted in figure 3, and 'explain' 72.5% of the value of X^2. on the left we see occupations 16 and 17, from down-middle to top-right the other occupations are smoothly ordered. 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, we can see that this correlation is for the greatest part based on the difference between 16 and 17 versus the other occupations. After ommiting these two categories, the analysis on the 15x15 matrix resulted in a correlation of .356.

When we are interested in the cases that sons do not have the same occupations as their fathers, we should compare the observed frequencies with the quasi-independence model as defined in (1). The result of this correspondence analysis is not shown. The solution is dominated by the cell for father 16 (contribution on first dimension .693) and 17 (.738). To have a better view on 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 cell f(ij)=$\hat{m}$(ij). The correspondence analysis solution can be found in figure 4. In this solution only the first dimension seems to be important: 68% of the total inertia is decomposed in this dimension. So, apart from the transition from 16 to 17, there seems to be a one-dimensional structure for the cases that fathers and sons do not have the same occupation.

Figure 3: Correspondence analysis of table 2
λ_1 = .51 (.53), λ_2 = .31 (.20); father is large

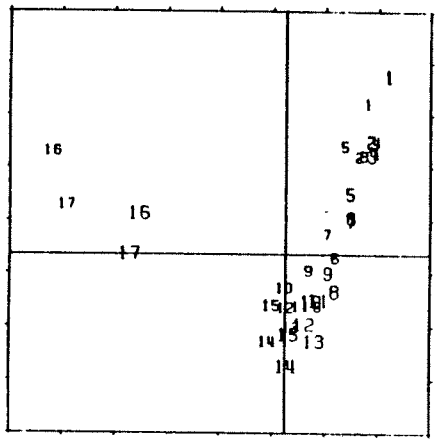
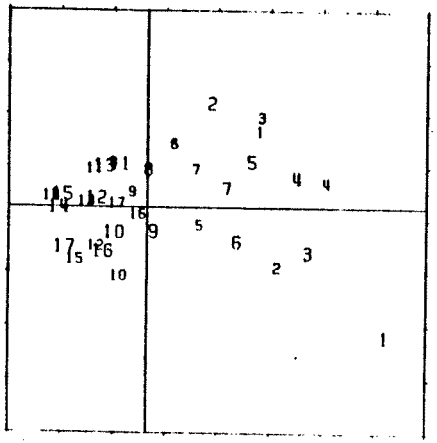


Figure 4: Diagonal cells and (16,17) excluded
λ_1 = .25 (.68), λ_2 = .08 (.08)



5.0 Conclusions and discussion. It is shown that (generalized) correspondence analysis is a suitable method for the analysis of transition matrices, especially when the number of categories is large. The reasons for this are outlined in section 4.0, and illustrated in 4.1 and 4.2. Correspondence analysis is probably also a suitable method to study the departure from models different from 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 in for instance confusion matrices and import-export tables.

References

[1] Benzécri, J.P. et al.. Analyse des données (2 vols.) (Dunod: Paris, 1973)
[2] Bishop, Y.M.M., Fienberg, S.E., & Holland, P.W.. Discrete multivariate analysis: theory and practice (MIT-press: Cambridge, 1975)
[3] Breiger, R.L.. The social class structure of occupational mobility, Amer. Journ. of Soc., 87, (1981) 578-611.
[4] Escofier, B. Analyse de la difference entre deux mésures sur le produit de deux mêmes ensembles, Cahiers de l'analyse des données, 3, (1983) 325-329.
[5] Gifi, A. Non-linear multivariate analysis (R.U.L./F.S.W., Department of Data Theory: Leiden, 1981).
[6] Greenacre, M.J. Theory and applications of correspondence analysis. (Academic Press: London, 1984).
[7] Heijden, P.G.M. van der, & Leeuw, J. de. Correspondence analysis used complementary to loglinear analysis. Psychometrika (in press).
[8] Hooff, J.A.R.A.M. van, Categories and sequences of behavior: methods of description and analysis, in: Scherer, K.R. & Ekman, P. (eds.) Handbook of methods in non-verbal behavior research (Cambridge Univ. Press: Cambridge, 1982).
[9] Slater, P.J.B., & Ollason, J.C. The temporal patterning of behavior in isolated male zebra finches: transition analysis, Behaviour, 42, (1972) 248-269.