Multiple Correspondence Analysis

18

Up to now we have analysed the association between two categorical variables or between two sets of categorical variables where the row variables are different from the column variables. In this and the two following chapters we turn our attention to the association *within* one set of variables, where we are interested in how strongly and in which way these variables are interrelated. In this chapter we will concentrate on the two classic ways to approach this problem, called *multiple correspondence analysis*, or MCA for short. One way is to think of MCA as the analysis of the whole data set coded in the form of dummy variables, called the *indicator matrix*, while the other way is to think of it as analysing all two-way cross-tabulations amongst the variables, called the *Burt matrix*. These two ways are very closely connected, but suffer from some deficiencies which we will try to correct in the following chapter, Chapter 19, where several improved versions of MCA are presented.

Contents

A single set of "homogeneous" categorical variables

In this chapter we are concerned with s single set of (more than two) variables, usually in the context of a single phenomenon of interest. For example, the four variables used in Chapter 17, on whether women should work or not, could be such a set of interest, or a set of questions about people's attitudes to science, or a set of categorical variables describing environmental conditions at several terrestrial locations. The point is that the set of variables is "homogeneous" in that the variables are of the same substantive type; that is, there is no mix of attitudinal and demographic variables, for example.

Indicator matrix

As an example let us consider the same set of four variables analysed in Chapter 17. The explanation is simplified by avoiding all the cross-cultural differences seen in previous analyses, using only the data from Germany, but

Exhibit 18.1: *Raw data and the indicator (dummy variable) coding, for the first six respondents out of $N = 3418$.*

Questions				*Qu. 1*				*Qu. 2*				*Qu. 3*				*Qu. 4*			
1	*2*	*3*	*4*	*W*	*w*	*H*	*?*	*W*	*w*	*H*	*?*	*W*	*w*	*H*	*?*	*W*	*w*	*H*	*?*
1	3	2	2	1	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0
2	3	3	2	0	1	0	0	0	0	1	0	0	0	1	0	0	1	0	0
4	3	3	2	0	0	0	1	0	0	1	0	0	0	1	0	0	1	0	0
4	4	4	4	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
4	4	4	4	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
1	3	2	1	1	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0
⋮	⋮	⋮	⋮	⋮				⋮				⋮				⋮			
. . . and so on for 3418 rows																			

including both the West and East German samples, totalling 3418 respondents (three cases with some missing demographic information were omitted from the original samples — see Computational Appendix, page 235). For the moment we are focusing on the four questions about women working, labelled 1 to 4, each of which has four categories of response, labelled as before: *W* (work full-time), *w* (work part-time), *H* (stay at home) and *?* (don't know/non-response). The *indicator matrix* is the 3418×16 matrix which codes all responses as dummy variables, where the 16 columns correspond to the 16 possible response categories. Exhibit 18.1 illustrates this coding for the first six rows: for example, the first respondent has responses 1, 3, 2 and 2 to the four questions, which are then coded as 1 0 0 0 indicating the response 1 (*W*) to question 1, 0 0 1 0 indicating the response 3 (*H*) to question 2, and 0 1 0 0 indicating the response 2 (*w*) to both questions 3 and 4.

MCA definition number 1: CA of the indicator matrix

The most common definition of MCA is that it is simple CA applied to this indicator matrix. This would provide coordinates for all 3418 rows and 16 columns, but it is mainly the positions of the 16 category points that are of interest for the moment, shown in Exhibit 18.2. The first principal axis shows all four non-response categories together, opposing all the substantive responses. In the previous analysis of these questions (see Exhibit 17.4) where the responses were related to demographic variables, the non-response points were not prominent on the first two axes. But here, because we are looking at relationships within the four questions, this is the most important feature: people who do not respond to one question tend to do the same for the others — for example, amongst the first six respondents in Exhibit 18.1 there are already two respondents who have non-responses for all four questions. On the second axis of Exhibit 18.2, we have the line-up of substantive categories from traditional attitudes at the bottom to liberal attitudes on top. Exhibit 18.3 shows the second and third dimensions of the map, which effectively partials out most of the effect of the non-response points, and the positions of the points are now strikingly similar to those in Exhibit 17.4. Notice that the fact that the liberal side of the horizontal dimension is now on the right is of no

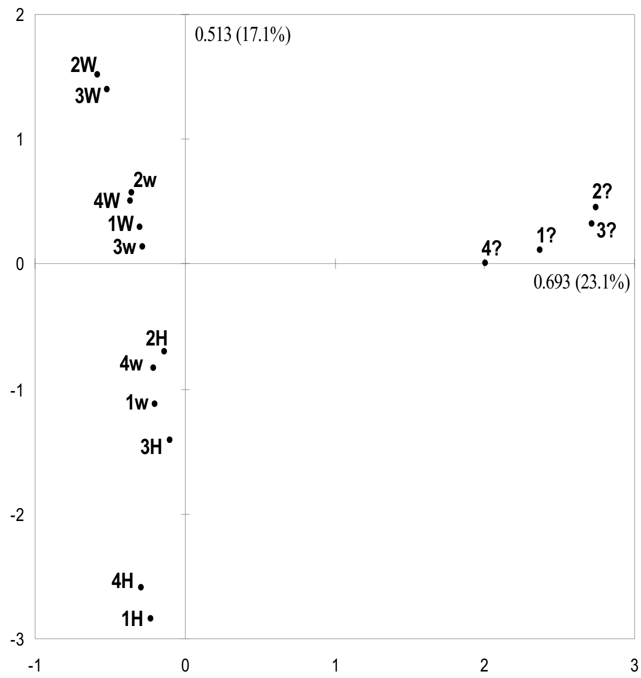


Exhibit 18.2: *MCA map of four questions on women working; total inertia = 3, percentage inertia in map: 40.2%.*

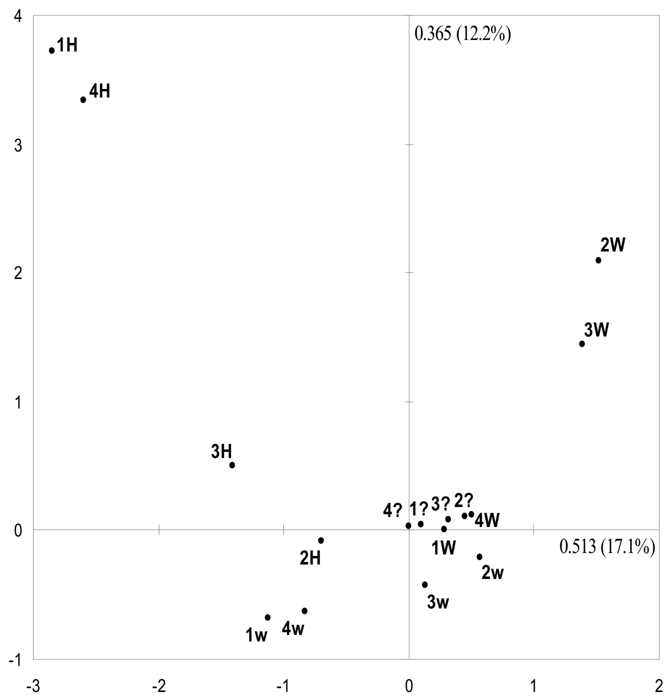


Exhibit 18.3: *MCA map of four questions on women working, showing second and third dimensions; total inertia = 3, percentage inertia in map: 29.3%.*

consequence to the interpretation: in fact, it is always possible to reverse an axis (i.e., multiply all coordinates by -1).

Inertia of indicator matrix

The total inertia of an indicator matrix takes on a particularly simple form, depending only on the number of questions and number of response categories and not on the actual data. Suppose there are Q variables, and each variable q has J_q categories, with J denoting the total number of categories: $J = \sum_q J_q$ (in our example, $Q = 4$, $J_q = 4$, $q = 1, \ldots, Q$, and $J = 16$). The indicator matrix, denoted by $\mathbf{Z}$, with J columns, is composed of a set of subtables $\mathbf{Z}_q$ stacked side by side, one for each variable, and the row margins of each subtable are the same, equal to a column of ones. Thus the result (17.1) in Chapter 17 applies: the total inertia of the indicator matrix is equal to the average of the inertias of the subtables. Each subtable $\mathbf{Z}_q$ has a single one in each row, otherwise zeros, so this is an example of a matrix where all the row profiles lie at the vertices, the most extreme association possible between rows and columns; hence the inertias are 1 on each principal axis of the subtable, and the total inertia of subtable $\mathbf{Z}_q$ is equal to its dimensionality, which is $J_q - 1$. Thus the inertia of $\mathbf{Z}$ is the average of the inertias of its subtables:

$$\text{inertia}(\mathbf{Z}) = \frac{1}{Q}\sum_q \text{inertia}(\mathbf{Z}_q) = \frac{1}{Q}\sum_q (J_q - 1) = \frac{J-Q}{Q} \tag{18.1}$$

Since $J - Q$ is the dimensionality of $\mathbf{Z}$, the average inertia per dimension is $1/Q$. Notice that the first three dimensions that were interpreted in Exhibits 18.2 and 18.3 have principal inertias 0.693, 0.513 and 0.365, all above the average of $1/4 = 0.25$. The value $1/Q$ serves as a threshold for deciding which axes are worth interpreting in MCA (analogous to the threshold of 1 for the eigenvalues in principal component analysis).

Burt matrix

An alternative data structure for MCA is the set of of all two-way cross-tabulations of the set of variables being analysed. The complete set of pairwise cross-tabulations is called the *Burt matrix*, shown in Exhibit 18.4 for the present example. The Burt matrix is a 4×4 block matrix, with 16 subtables. Each of the 12 off-diagonal subtables is a contingency table cross-tabulating the 3418 respondents on a pair of variables. The Burt matrix is symmetric so there are only 6 unique cross-tabulations, which are transposed on either side of the diagonal blocks. The diagonal subtables (by which we mean the tables on the block diagonal) are cross-tabulations of each variable with itself, which is just a diagonal matrix with the marginal frequencies of the variable down the diagonal. For example, the marginal frequencies for question 1 are 2501 *W* responses, 476 *w*s, 79 *H*s and 362 *?*s. The Burt matrix, denoted by $\mathbf{B}$, is simply related to the indicator matrix $\mathbf{Z}$ as follows:

$$\mathbf{B} = \mathbf{Z}^\mathsf{T}\mathbf{Z} \tag{18.2}$$

1W	1w	1H	1?	2W	2w	2H	2?	3W	3w	3H	3?	4W	4w	4H	4?
2501	0	0	0	172	1107	1131	91	355	1710	345	91	1766	538	40	157
0	476	0	0	7	129	335	5	16	261	181	18	128	293	17	38
0	0	79	0	1	6	72	0	1	17	61	0	14	21	38	6
0	0	0	362	1	57	108	196	7	96	55	204	51	45	2	264
172	7	1	1	181	0	0	0	127	48	4	2	165	15	0	1
1107	129	6	57	0	1299	0	0	219	997	61	22	972	239	13	75
1131	335	72	108	0	0	1646	0	24	989	573	60	760	616	84	186
91	5	0	196	0	0	0	292	9	50	4	229	62	27	0	203
355	16	1	7	127	219	24	9	379	0	0	0	360	14	1	4
1710	261	17	96	48	997	989	50	0	2084	0	0	1348	567	23	146
345	181	61	55	4	61	573	4	0	0	642	0	202	286	73	81
91	18	0	204	2	22	60	229	0	0	0	313	49	30	0	234
1766	128	14	51	165	972	760	62	360	1348	202	49	1959	0	0	0
538	293	21	45	15	239	616	27	14	567	286	30	0	897	0	0
40	17	38	2	0	13	84	0	1	23	73	0	0	0	97	0
157	38	6	264	1	75	186	203	4	146	81	234	0	0	0	465

Exhibit 18.4: *Burt matrix of all two-way cross-tabulations of the four variables of the example on attitudes to women working. Down the diagonal are the the cross-tabulations of each variable with itself.*

MCA definition number 2: CA of the Burt matrix

The other "classic" way of defining MCA is the application of CA to the Burt matrix **B**. Since **B** is a symmetric matrix, the row and column solutions are identical, so only one set of points is shown — see Exhibit 18.5. Because of the direct relationship (18.2), it is no surprise that the solutions are related, in fact at first glance Exhibit 18.5 looks identical to Exhibit 18.2, only the scale has changed slightly on the two axes. This is the only difference between the two analyses — the Burt version of MCA gives principal coordinates which are reduced in scale compared to the indicator version, where the reduction is relatively more on the second axis compared to the first.

Comparison of MCA based on indicator and Burt matrices

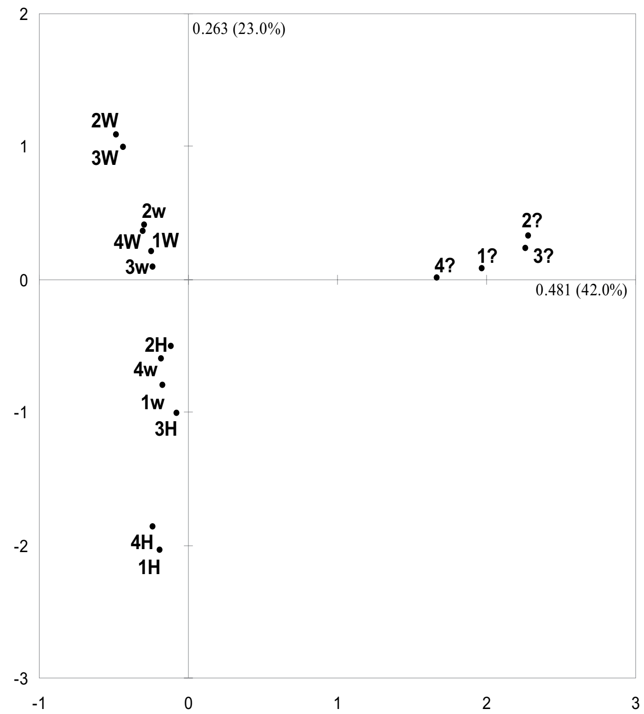
The two ways of defining MCA are related as follows:

- In both analyses the standard coordinates of the category points are identical — this is a direct result of the relationship (18.2).
- Also as a result of (18.2), the principal inertias of the Burt analysis are the squares of those of the indicator matrix.
- Since the principal inertias are less than 1, squaring them makes them smaller in value (and the lower principal inertias relatively smaller still). The principal coordinates are the standard coordinates multiplied by the square roots of the principal inertias, which accounts for the reduction in scale in Exhibit 18.5 compared to Exhibit 18.2.
- The percentages of inertia are thus always going to be higher in the Burt analysis.

Inertia of the Burt matrix

The subtables of the Burt matrix have the same row margins in each set of horizontal tables and the same column margins in each set of vertical tables,

Exhibit 18.5: *MCA map of Burt matrix of four questions on women working, showing first and second dimensions; total inertia = 1.145, percentage inertia in map: 65.0%.*



so the result (17.1) applies exactly: the inertia of $\mathbf{B}$ will be the average of the inertias of the subtables $\mathbf{B}_{qs}$. Exhibit 18.6 shows the 16 individual inertias of the Burt matrix, and their row and column averages. The overall average is equal to the total inertia 1.145 of $\mathbf{B}$. In this table the inertias of the diagonal blocks are exactly 3; in fact their inertias have the same definition (18.1) as the inertias of the subtables of the indicator matrix — they are $J_q \times J_q$ tables of dimensionality $J_q - 1$ with perfect row–column association, and so have maximal inertia equal to the number of dimensions. These high values on the diagonal of Exhibit 18.6 demonstrate why the total inertia of the Burt matrix is so high, which is the cause of the low percentages of inertia on the axes. We return to this topic in the next chapter.

Positioning supplementary variables in the map

Suppose we wish to relate the demographic variables gender, age, etc., to the patterns of association revealed in the MCA maps. There are two ways of doing this, highly related, but one of these has some advantages. The first way is to code these as additional dummy variables and add them as supplementary columns of the indicator matrix. The second way is to cross-tabulate the demographics with the four questions, as we did in the stacked analysis of

QUESTIONS	*Qu. 1*	*Qu. 2*	*Qu. 3*	*Qu. 4*	*Average*
Qu. 1	3.0000	0.3657	0.4262	0.6457	*1.1094*
Qu. 2	0.3657	3.0000	0.8942	0.3477	*1.1519*
Qu. 3	0.4262	0.8942	3.0000	0.4823	*1.2007*
Qu. 4	0.6457	0.3477	0.4823	3.0000	*1.1189*
Average	*1.1094*	*1.1519*	*1.2007*	*1.1189*	*1.1452*

Exhibit 18.6: *Inertias of each of the 16 subtables of the Burt matrix, from their individual CAs.*

Chapter 17, and add these cross-tables as supplementary rows of the indicator matrix or as supplementary rows (or columns) of the Burt matrix. The second strategy is the preferred strategy because it can be used in both forms of MCA as well as in the improved versions that we present in the next chapter. Moreover, it gives the same positions of the supplementary points in both MCA versions and has the same interpretation as the average positions of those cases belonging to the particular demographic category. Exhibit 18.7 shows the positions of five of the demographic variables we used previously, which can be superimposed on the maps of Exhibit 18.2 or 18.5.

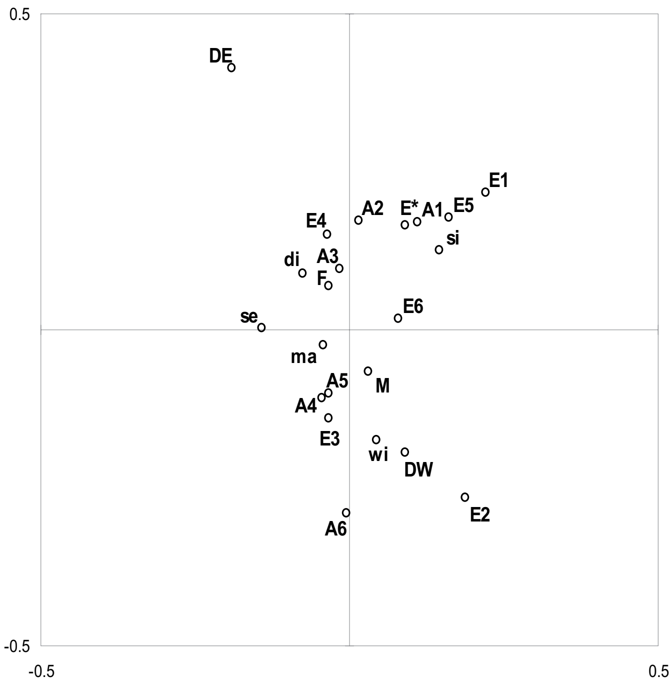


Exhibit 18.7: *Supplementary variables with respect to first two principal axes, to be superimposed on the maps of Exhibits 18.2 or 18.5. These points occupy a small area of the map (note the scale), but will be more spread out in the map of the Burt matrix than that of the indicator matrix.*

Interpretation of supplementary points

Based on the positions of the response categories on the first two dimensions of Exhibit 18.2 (similarly, Exhibit 18.5), the farther a demographic category is to the right, the higher will be the frequency of non-responses. The higher up a category is, the more liberal the attitude, and the lower down it is, the more traditional the attitude. Hence West Germany (DW) has more traditional attitudes and more non-responses than East Germany (DE), a pattern that is mimicked almost identically by the male–female (M–F) contrast but not as much as the difference between the two German regions. The age groups show the same trend as before, from young (A1) at the top (liberal) to old (A6) at the bottom (traditional). The lowest education groups have the highest frequency of non-response and the highest education groups tend to have more liberal attitudes, but so do the lowest education groups E1 and E2. Amongst the marital status groups, single (si) respondents have higher than average non-response and liberal attitudes, opposing separated (se) respondents who have the least non-response, but are otherwise average on the liberal–traditional dimension.

SUMMARY: Multiple Correspondence Analysis

1. MCA is concerned with relationships amongst (or within) a set of variables — usually the variables are homogeneous in the sense that they revolve around one particular issue, and often the response scales are the same.
2. The variables can be recoded as dummy variables in an *indicator matrix*, which has as many rows as cases and as many columns as categories of response. The data in each row are 0s apart from 1s which indicate the particular category of each variable corresponding to the individual case.
3. An alternative coding of such data is as a *Burt matrix*, a square symmetric categories-by-categories matrix formed from all two-way contingency tables of pairs of variables, including on the block diagonal the cross-tabulations of each variable with itself.
4. The two alternative definitions of MCA, applying CA to the indicator matrix or to the Burt matrix, are almost equivalent. Both yield identical standard coordinates for the category points.
5. The difference between the two definitions is in the pricncipal inertias: those of the Burt matrix are the squares of those of the indicator matrix. As a result, the percentages of inertia in the Burt analysis are always more optimistic than those in the indicator analysis.
6. In both approaches, however, the percentages of inertia are artificially low, due to the coding, and underestimate the true quality of the maps as representations of the data.