

As was shown in Chapters 7 and 8, there are several alternative definitions of CA and different ways of thinking about the method. In this book we have stressed Benzécri's geometric approach leading to data visualization. In Chapters 18 and 19 it was clear that the passage from simple two-variable CA to the multivariate form of the analysis is not straightforward, especially if one tries to generalize the geometric interpretation. An alternative approach to the multivariate case, which relies on exactly the same mathematics as MCA, is to see the method as a way of quantifying categorical data, generalizing the optimal scaling ideas of Chapter 7. As before, there are several equivalent ways to think about MCA as a scaling technique, and these different approaches enrich our understanding of the method's properties. The optimal scaling approach to MCA is often referred to in the literature as *homogeneity analysis*.

Contents

<i>Data set 11: Attitudes to science and environment</i>	153
<i>Category quantification as a goal</i>	154
<i>MCA as a principal component analysis of the indicator matrix</i>	154
<i>Maximizing inter-item correlation</i>	155
<i>MCA of scientific attitudes example</i>	155
<i>Individual squared correlations</i>	156
<i>Loss of homogeneity</i>	158
<i>Geometry of loss function in homogeneity analysis</i>	158
<i>Reliability and Cronbach's alpha</i>	160
<i>SUMMARY: Scaling Properties of MCA</i>	160

This data set is taken from the multinational ISSP survey on environment in 1993. We are going to look specifically at $Q = 4$ questions on attitudes towards the role of science. Respondents were asked if they agreed or disagreed with the following statements:

*Data set 11:
Attitudes to
science and
environment*

- A* We believe too often in science, and not enough in feelings and faith.
- B* Overall, modern science does more harm than good.
- C* Any change humans cause in nature, no matter how scientific, is likely to make things worse.
- D* Modern science will solve our environmental problems with little change to our way of life.

There were five possible response categories: 1. strongly agree, 2. somewhat agree, 3. neither agree nor disagree, 4. somewhat disagree, 5. strongly disagree. For simplicity we have used data for the West German sample only and have

omitted cases with missing values on any one of the four questions, leaving us with a sample of $N = 871$ (these data are provided with our **ca** package for R — see the Computational Appendix).

Category quantification as a goal In Chapter 7 CA was defined as the search for quantifications of the categories of the column variable, say, which lead to the greatest possible differentiation, or discrimination, between the categories of the row variable, or vice versa. This is what we would call an “asymmetric” definition because the rows and columns play different roles in the definition and the results reflect this too; for example, the column solution turns out to be in standard coordinates and the row solution in principal coordinates. In Chapter 8 CA was defined “symmetrically” as the search for new scale values which lead to the highest correlation between the row and column variables. Here the rows and columns have an identical role in the definition. These scaling objectives do not include any specific geometric concepts and, in particular, make no mention of a full space in which the data are imagined to lie, which is an important concept in the geometric approach for measuring total inertia and percentages of inertia in lower-dimensional subspaces.

MCA as a principal component analysis of the indicator matrix The asymmetric definition of optimal scaling, when applied to an indicator matrix, resembles closely principal component analysis (PCA). PCA is usually applied to matrices of continuous-scale data, and has close theoretical and computational links to CA — in fact, one could say that CA is a variant of PCA for categorical data. In the PCA of a data set where the rows are cases and the columns variables (m variables, say, x_1, \dots, x_m), coefficients $\alpha_1, \dots, \alpha_m$ (to be estimated) are assigned to the columns, leading to linear combinations for the rows (cases) of the form $\alpha_1 x_1 + \dots + \alpha_m x_m$, called *scores*. The coefficients are then calculated to maximize the variance of the row scores. As before, identification conditions are required to solve the problem, and in PCA this is usually that the sum of squares of the coefficients is 1: $\sum_j \alpha_j^2 = 1$. Applying this idea to the indicator matrix, which consists of zeros and ones only, assigning coefficients $\alpha_1, \dots, \alpha_J$ to the J dummy variables and then calculating linear combinations for the rows, simply means adding up the α coefficients (i.e., scale values) for each case. Then maximizing the variance for each case sounds just like the optimal scaling procedure of Chapter 7 (maximizing discrimination between the rows); in fact this is almost identical except for one aspect, namely the identification conditions. In optimal scaling the identification conditions would be that the weighted variance (inertia) of the coefficients (not the simple sum of squares) be equal to 1: $\sum_j c_j \alpha_j^2 = 1$. Here the c_j 's are the column masses, i.e., the column sums of the indicator matrix divided by the grand total NQ of the indicator matrix — thus each set of c_j 's for one categorical variable adds up to $1/Q$. So with this change in the identification condition, MCA could be called the PCA of categorical data, maximizing the variance across cases. The coefficients are the standard coordinates of the column categories, while the MCA principal coordinate of a case is the average

of that case's scale values, i.e., $1/Q$ times the sum that was called the score before. The first dimension of MCA maximizes the variance (first principal inertia), the second dimension maximizes the variance subject to the scores being uncorrelated with those on the first dimension, and so on.

MCA as a scaling technique, usually called *homogeneity analysis*, is more commonly seen as a generalization of the correlation approach of Chapter 8. In Equation (8.1) on page 63, an alternative way of optimizing correlation between two categorical variables was given, which is easy to generalize to more than two variables. Again we shall use a pragmatic notation for this particular four-variable example but the ideas easily extend to Q variables which have any number of categories (in our example here, $Q = 4$ and the total number of categories is $J = 20$). Suppose that the four variables have (unknown) scale values a_1 to a_5 , b_1 to b_5 , c_1 to c_5 , and d_1 to d_5 . Assigning four of these values a_i , b_j , c_k and d_l to each respondent according to his or her set of responses leads to the quantified responses for the whole sample, which we denote by a , b , c and d (i.e., a denotes all 871 quantified responses to question A, etc.). Each respondent has a score $a_i + b_j + c_k + d_l$ which is the sum of the scale values, so the scores for the whole sample are denoted by $a + b + c + d$. In this context the variables are often referred to as *items* and we talk of the values in a to d as *item scores* and those in $a + b + c + d$ as the *summated scores*. The criterion for finding optimal scale values is thus to maximize the average squared correlation between the item scores and the summated score:

$$\text{average squared correlation} = \frac{1}{4}[\text{cor}^2(a, a + b + c + d) + \text{cor}^2(b, a + b + c + d) + \text{cor}^2(c, a + b + c + d) + \text{cor}^2(d, a + b + c + d)] \quad (20.1)$$

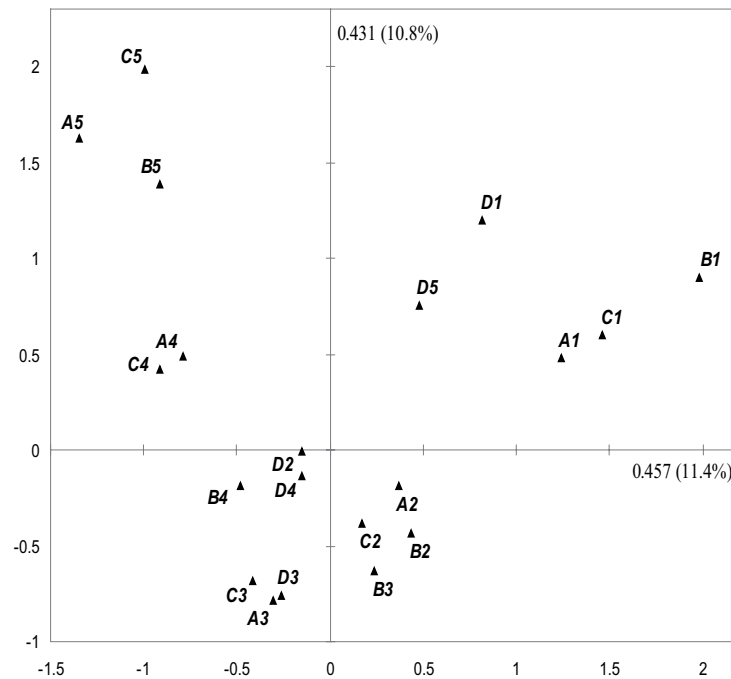
(cf. the two-variable case in (8.1.) on page 63). Again, identification conditions are required, and it is convenient to use the mean 0 and variance 1 conditions on the summated scores: $\text{mean}(a + b + c + d) = 0$, $\text{var}(a + b + c + d) = 1$. The solution to this maximization problem is then given exactly by the standard coordinates of the item categories on the first principal axis of MCA, and the maximized average squared correlation of (20.1) is exactly the first principal inertia (of the indicator matrix version of MCA).

Exhibit 20.1 shows the two-dimensional MCA map based on the indicator matrix, showing again the very low percentages of inertia (the percentages based on adjusted inertias are 44.9% and 34.2% respectively). But in this case the percentages should be ignored, since it is the values of the principal inertias that are of interest *per se*, being average squared correlations. The maximum value of (20.1) is thus 0.457. The second principal inertia, 0.431, is found by looking for a new set of scale values which lead to a set of scores which are uncorrelated with those obtained previously, and which maximize (20.1) — this maximum is the value 0.431. And so it would continue for solutions on subsequent axes, always uncorrelated with the ones already found. Looking

Maximizing
inter-item
correlation

MCA of scientific
attitudes example

Exhibit 20.1:
MCA map
(indicator matrix
version) of science
attitudes, showing
category points in
principal
coordinates. Since
the principal
inertias differ only
slightly (and even
less in their square
roots), the principal
coordinates are
almost the same
contraction of the
standard coordi-
nates on both axes.



at the map in Exhibit 20.1, we see that questions *A*, *B* and *C* follow a very similar pattern, with strong disagreements on the left to strong agreements on the right, in a wedge-shaped horseshoe pattern. Question *D*, however, has a completely different trajectory, with the two poles of the scale very close together. Now the first three questions were all worded negatively towards science whereas question *D* was worded positively, so we would have expected *D5* to lie towards *A1*, *B1* and *C1*, and *D1* on the side of *A5*, *B5* and *C5*. The fact that *D1* and *D5* lie close together inside the horseshoe means that they are both associated with the extremes of the other three questions — the most likely explanation is that some respondents are misinterpreting the change of direction of the fourth question.

Individual squared correlations Knowing the values of the individual squared correlations composing (20.1) will also be interesting information. These can be obtained directly by adding up the individual inertia contributions to the first principal inertia for each question. The results of a MCA usually give these expressed in proportions or permills, so we show these as permills in Exhibit 20.2 as an illustration of how to recover these correlations. Questions *A* to *D* thus contribute proportions 0.279, 0.317, 0.343 and 0.062 of the principal inertia of 0.457. Since 0.457 is the average of the four squared correlations, the squared correlations and thus the correlations are:

CATEGORIES	QUESTIONS				Sum
	A	B	C	D	
1 "strongly agree"	115	174	203	25	518
2 "somewhat agree"	28	21	6	3	57
3 "neither–nor"	12	7	22	9	49
4 "somewhat disagree"	69	41	80	3	194
5 "strongly disagree"	55	74	32	22	182
Sum	279	317	343	62	1000

Exhibit 20.2:
Permill (‰)
contributions to
inertia of first
principal axis of
MCA (indicator
matrix version) of
data on science and
environment.

$$\begin{aligned}
 A: & 0.279 \times 0.457 \times 4 = 0.510 & \text{correlation} &= \sqrt{0.510} = 0.714 \\
 B: & 0.317 \times 0.457 \times 4 = 0.579 & \text{correlation} &= \sqrt{0.579} = 0.761 \\
 C: & 0.343 \times 0.457 \times 4 = 0.627 & \text{correlation} &= \sqrt{0.627} = 0.792 \\
 D: & 0.062 \times 0.457 \times 4 = 0.113 & \text{correlation} &= \sqrt{0.113} = 0.337
 \end{aligned}$$

This calculation shows how much lower the correlation is of question D with the total score. Notice that, although the MCA of the indicator matrix was the worst from the usual CA geometric point of view of χ^2 -distances, total inertia, etc., the principal inertias and the contributions to the principal inertias do have a very interesting interpretation by themselves. In the approach called *homogeneity analysis*, which is theoretically equivalent to the MCA of the indicator matrix but which interprets the method from a scaling viewpoint, the squared correlations 0.510, 0.579, 0.627 and 0.113 are called *discrimination measures*.

In homogeneity analysis the objective function (8.3) (see Chapter 8, page 63) is generalized to many variables. Using the notation above for the present four-variable example, we would calculate the average score $\frac{1}{4}(a_i + b_j + c_k + d_l)$ of the item scores for each respondent and then calculate that respondent's measure of variance within his or her set of quantified responses:

*Loss of
homogeneity*

$$\begin{aligned}
 \text{variance (for one case)} &= \frac{1}{4} \left([a_i - \frac{1}{4}(a_i + b_j + c_k + d_l)]^2 \right. \\
 &\quad + [b_j - \frac{1}{4}(a_i + b_j + c_k + d_l)]^2 \\
 &\quad + [c_k - \frac{1}{4}(a_i + b_j + c_k + d_l)]^2 \\
 &\quad \left. + [d_l - \frac{1}{4}(a_i + b_j + c_k + d_l)]^2 \right) \quad (20.2)
 \end{aligned}$$

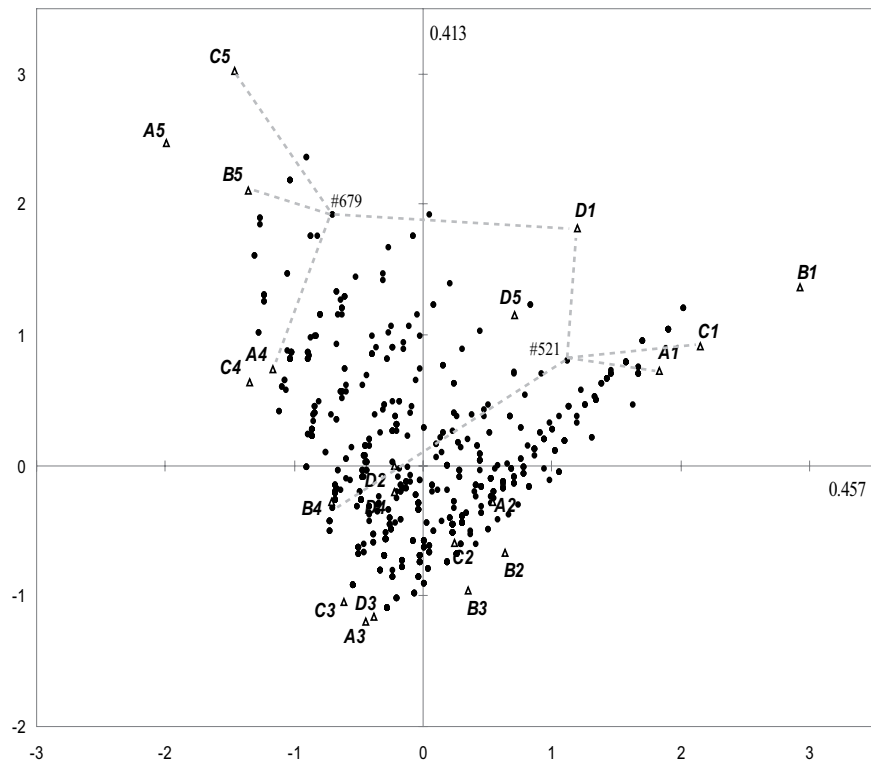
The average of all these values over the N cases is then calculated, called the *loss of homogeneity* and the objective is to minimize this loss. Again the MCA (indicator matrix version) solves this problem and the minimized loss

is 1 minus the first principal inertia; i.e., $1 - 0.457 = 0.543$. Minimizing the loss is equivalent to maximizing the correlation measure defined previously.

*Geometry of
loss function in
homogeneity
analysis*

The objective of minimizing loss has a very attractive geometric interpretation which is closely connected to the row-to-column distance definition of CA discussed in Chapter 7. In fact, the homogeneity loss function is exactly the weighted distance function (7.6) on page 55, applied to the indicator matrix. Exhibit 20.3 shows the asymmetric MCA map of all $N = 871$ respondents (in principal coordinates) and the $J = 20$ category points (in standard coordinates), which means that the respondents lie at the centroids of the categories, where the weights are the relative values in the rows of the indicator matrix. Each respondent has a profile consisting of zeros apart from values of $\frac{1}{4}$ in

Exhibit 20.3:
Asymmetric MCA
map (indicator
matrix version) of
science attitudes,
showing respondents
in principal
coordinates and
categories in
standard
coordinates. Each
respondent is at the
average of the four
categories given as
responses. MCA
minimizes the sum
of squared distances
between category
points and
respondents.



the positions of the four responses; hence each respondent point lies at the ordinary average position of his or her responses. Two respondents, #679 and #521, are labelled in Exhibit 20.3. Respondent #679 chose the categories (A4, B5, C5, D1), disagreeing with the first three questions and agreeing to the fourth — those categories are linked to the respondent point on the left-hand side of the display. This is a strong and consistent position in favour of science.

Respondent #521, however, has a mixed opinion: $(A1, B4, C1, D1)$, strongly agreeing that we believe too much in science and that human interference in nature will make things worse, but at the same time strongly agreeing that science will solve our environmental problems while disagreeing that science does more harm than good. This shows one of the reasons why $D1$ has been pulled to the middle between the two extremes of opinion. Every respondent is at the average of the four categories in his or her set of answers. For any configuration of category points, the respondents could be located at average positions, but the result of Exhibit 20.3 is optimal in the sense that the lines linking the respondents to the category points are the shortest possible (in terms of sum of squared distances). Showing all the links between respondents and their response categories has been called a *star plot*, so the objective of MCA can be seen as obtaining the star plot with the shortest links in the least-squares sense. The number of links between the N respondent points and the corresponding Q category points is NQ , and the value of the loss is actually the average of the squares of the links (for example, in (20.2) where $Q = 4$ the sum of the four squares is divided by 4, and then the average over N is calculated, so that the sum of squared values is divided by $4N$). So the average sum of squared links on the first dimension is $1 - 0.457 = 0.513$, and on the second dimension it is $1 - 0.413 = 0.587$; by Pythagoras' theorem, the average sum of squared links in the two-dimensional map of Exhibit 20.3 is $0.513 + 0.587 = 1.100$.

In the present example of the science and environment data, we saw that the question D is not correlated highly with the others (see page 157). If we were trying to derive an overall measure of attitude towards science in this context, we would say that these results show us that question D has degraded the *reliability* of the total score, and should preferably be removed. In reliability theory, the Q variables, or items, are supposed to be measuring one underlying construct. *Cronbach's alpha* is a standard measure of reliability, defined in general as:

*Reliability and
Cronbach's alpha*

$$\alpha = \frac{Q}{Q-1} \left(1 - \frac{\sum_q s_q^2}{s^2} \right) \quad (20.3)$$

where s_q^2 is the variance of the q -th item score, $q = 1, \dots, Q$ (e.g., variances of a , b , c and d) and s^2 is the variance of the average score (e.g., variance of $\frac{1}{4}(a + b + c + d)$). Applying this definition to the first dimension of the MCA solution, it can be shown that Cronbach's alpha reduces to the following:

$$\alpha = \frac{Q}{Q-1} \left(1 - \frac{1}{Q\lambda_1} \right) \quad (20.4)$$

where λ_1 is the first principal inertia of the indicator matrix. Thus the higher the principal inertia, the higher the reliability. Using $Q = 4$ and $\lambda_1 = 0.4574$ (four significant digits for slightly better accuracy) we obtain:

$$\alpha = \frac{4}{3} \left(1 - \frac{1}{4 \times 0.4574} \right) = 0.605$$

Having seen the behaviour of question D , an option now is to remove it and recompute the solution with the three questions that are highly intercorrelated. The results are not given here, apart from reporting that the first principal inertia of this three-variable MCA is $\lambda_1 = 0.6018$, with an increase in reliability to $\alpha = 0.669$ (use (20.4) with $Q = 3$). As a final remark, it is interesting to notice that the average squared correlation of a set of random variables, with no zero pairwise correlation between them, is equal to $1/Q$, and this corresponds to a Cronbach's alpha of 0. The value $1/Q$ is exactly the threshold used in (19.7) for adjustment of the principal inertias (eigenvalues), and was also the average principal inertia in the MCA of the indicator matrix, mentioned in Chapter 18.

SUMMARY:
Scaling Properties
of MCA

1. Optimal scaling in a two-variable context was defined as the search for scale values for the categories of one variable which lead to the highest separation of groups defined by the other variable. This problem is equivalent to finding scale values for each set of categories which lead to the highest possible correlation between the row and column variables.
2. In a multivariate context, optimal scaling can be generalized as the search for scale values for the categories of all variables so as to optimize a measure of correlation between the variables and their sum (or average). Specifically, the average squared correlation is maximized between the scaled observations for each variable, called *item scores*, and their sum (or average), called simply the *score*.
3. Equivalently, a minimum can be sought for the variance between item scores within each respondent, averaged over the sample. This is the usual definition of *homogeneity analysis*.
4. The scaling approach in general, exemplified by homogeneity analysis, is a better framework for interpreting the results of MCA of an indicator matrix. The principal inertias and their breakdown into contributions are more readily interpreted as squared correlations, rather than quantities with a geometric significance as in simple CA.
5. The first principal inertia in the indicator matrix version of MCA has a monotonic functional relationship with *Cronbach's alpha* measure of reliability: the higher the principal inertia, the higher the reliability.
6. Since the standard coordinates are identical for the MCA of the indicator matrix, the Burt matrix and in the adjusted form, these scaling properties apply to all three versions of MCA.