CHAPTER 2

From Simple to Multiple Correspondence Analysis

Michael Greenacre

CONTENTS

2.1 Introduction 41
2.2 Canonical correlation analysis 43
2.2.1 Two variables 43
2.2.2 Several variables 49
2.2.3 Homogeneity analysis 56
2.3 Geometric approach 58
2.3.1 Chi-square distance scaling 59
2.3.2 Biplot 61
2.3.3 Joint correspondence analysis 65
2.3.4 Adjustment of the inertias in MCA 67
2.4 Supplementary points 70
2.5 Discussion and conclusions 75

2.1 Introduction

Simple correspondence analysis (CA) is primarily applicable to a two-way contingency table, leading to a map that visualizes the association between two categorical variables. Multiple correspondence analysis (MCA) tackles the more general problem of associations among a set of more than two categorical variables. We shall see that the generalization to more than two variables is neither obvious nor well defined. In other areas of multivariate analysis, such as regression and log-linear

modeling, the situation is less complicated: for example, the transition from the regression of a response variable on a single predictor variable to the case of several predictors is quite straightforward. The main problem we face here is that the notion of association between two categorical variables is a complex concept. There are several ways to generalize this concept to more than two variables.

Of the many different ways that exist to define MCA, we shall consider two approaches: first, the definition which is perhaps the easiest to understand, namely that of correlation between sets of variables, known as canonical correlation, and second, the geometric approach, which is directly linked to data visualization and which has many similarities to Pearson-style principal component analysis. In the explanation of each approach, we will consider the case of two variables and then describe possible generalizations to more than two variables.

As an illustration of the theory, we shall use a data set from the International Social Survey Program on environment (ISSP 1993), looking specifically at questions on attitudes toward science. The survey questions that we consider are as follows:

> How much do you agree or disagree with each of these statements?
>
> 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.

Each question has five possible response categories:

1. Agree strongly
2. Agree
3. Neither agree nor disagree
4. Disagree
5. Disagree strongly

To avoid the issue of cross-cultural differences, we use data for the West German sample only (the ISSP surveys still distinguish between former West and East Germany). We shall also show how to relate the

MCA results to the external demographic variables of sex, age, and education, which were also coded as categorical variables as follows:

Sex: male, female
Age (six groups): 16–24, 25–34, 35–44, 45–54, 55–64, 65 and older
Education (six groups): primary incomplete, primary completed, secondary incomplete, secondary completed, tertiary incomplete, tertiary completed

A listwise deletion of respondents with missing data has been performed because we do not want to deal with the further complicating issue of missing data here (see Chapter 8). This reduces the original West German sample by about 14% to leave $n = 871$ respondents with complete data, which form the data set used in this chapter.

In the appendix of this book, most of the numerical results in this chapter are given along with the code in the R language to perform the computations (R Development Core Team 2005).

2.2 Canonical correlation analysis

2.2.1 Two variables

We start by considering just the first two variables, A (concerning belief in science) and B (concerning harm caused by science), both worded unfavorably toward science, so that disagreement indicates a favorable attitude toward science. Because there are only two variables, all 871 responses to these questions can be coded, with no loss of information, in the form of a cross-tabulation, given in Table 2.1. The correlational approach investigates how to measure the association between these two categorical variables. Several measures of association already exist for categorical data, some of which depend on whether the variables are measured on a nominal or ordinal scale, or whether we are trying to predict one variable from the other. In the following, we shall focus our interest on the classical product-moment correlation coefficient applicable to metric data and on the *quantification* of the categories, i.e., how to achieve numerical values for the response categories to calculate a correlation coefficient between the variables. Because the categories are ordered, a simple way out would be to use the existing values 1 to 5, as they are coded in the data file, thereby assuming that there is an equal interval difference between adjacent points on each scale.

Table 2.1 Cross-tabulation of 871 West German respondents with respect to two questions on attitudes to science.

	Overall, modern science does more harm than good					
We believe too often in science, not enough in feelings & faith	B1 agree strongly	B1 agree	B3 neither/ nor	B4 disagree	B5 disagree strongly	SUM
A1-agree strongly	27	28	30	22	12	119
A2-agree	38	74	84	96	30	322
A3-neither/nor	3	48	63	73	17	204
A4-disagree	3	21	23	79	52	178
A5-disagree strongly	0	3	5	11	29	48
SUM	71	174	205	281	140	871

But note that such a choice would be incorrect if one of the variables were nominal, for example "province of residence" or "religious denomination."

There are two ways to calculate the correlation coefficient: one is from the original respondent-level data, which are the 871 pairs of responses to the two questions; the other, more compact, approach is directly from Table 2.1, since this table gives the frequencies of occurrence of all pairs of categories. Suppose that the responses to questions A and B are coded in the indicator matrices $\mathbf{Z}_1$ and $\mathbf{Z}_2$, respectively, whose columns are zero-one dummy variables: that is $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are both 871×5 matrices. Then Table 2.1 is the cross-product $\mathbf{Z}_1^\top \mathbf{Z}_2$ of the two indicator matrices. Furthermore, suppose that the proposed scale values for the categories of the two variables are contained in the vectors $\mathbf{s}_1$ and $\mathbf{s}_2$, so that the individual quantified responses are in the vectors $\mathbf{Z}_1\mathbf{s}_1$ and $\mathbf{Z}_2\mathbf{s}_2$. To simplify the notation greatly, it is convenient to consider the quantified responses as initially mean-centered, $\mathbf{1}^\top\mathbf{Z}_1\mathbf{s}_1 = \mathbf{1}^\top\mathbf{Z}_2\mathbf{s}_2 = 0$, so that the covariance s_{12} between the two variables and their variances s_1^2 and s_2^2 can be written as:

$$s_{12} = (1/n)\mathbf{s}_1^\top\mathbf{Z}_1^\top\mathbf{Z}_2\mathbf{s}_2 = \mathbf{s}_1^\top\mathbf{P}_{12}\mathbf{s}_2$$

$$s_1^2 = (1/n)\mathbf{s}_1^\top\mathbf{Z}_1^\top\mathbf{Z}_1\mathbf{s}_1 = \mathbf{s}_1^\top\mathbf{D}_1\mathbf{s}_1 \quad \text{and} \quad s_2^2 = (1/n)\mathbf{s}_2^\top\mathbf{Z}_2^\top\mathbf{Z}_2\mathbf{s}_2 = \mathbf{s}_2^\top\mathbf{D}_2\mathbf{s}_2$$

where $\mathbf{P}_{12} = (1/n)\mathbf{Z}_1^\top\mathbf{Z}_2$ is called the *correspondence matrix*, containing the relative frequencies, i.e., Table 2.1 divided by its grand total of

n = 871. $\mathbf{D}_1$ and $\mathbf{D}_2$ are diagonal matrices of the marginal relative frequencies, or *masses*, of the two variables. (In Chapter 1 these are denoted by $\mathbf{D}_r$ and $\mathbf{D}_c$ for "rows" and "columns"; in this chapter we use indices 1 and 2, since we are going to extend the concepts to more than two variables.)

Using the above notation the correlation can be written as:

$$r = \frac{s_{12}}{s_1 s_2} = \frac{\mathbf{s}_1^\top \mathbf{P}_{12} \mathbf{s}_2}{\sqrt{\mathbf{s}_1^\top \mathbf{D}_1 \mathbf{s}_1 \mathbf{s}_2^\top \mathbf{D}_2 \mathbf{s}_2}} \tag{2.1}$$

which can be calculated directly from Table 2.1 and its margins. Because this calculation involves some important concepts in CA, we shall go through it in detail, using the values 1 to 5 for the categories of each variable.

- From Table 2.1 we have the values of the marginal relative frequencies (masses) for the categories of the two variables:

$$(1/n)\mathbf{1}^\top \mathbf{Z}_1 = (1/871)[\ 119 \quad 322 \quad 204 \quad 178 \quad 48\ ]$$

$$= [\ 0.137 \quad 0.370 \quad 0.234 \quad 0.204 \quad 0.055\ ]$$

$$(1/n)\mathbf{1}^\top \mathbf{Z}_2 = (1/871)[\ 71 \quad 174 \quad 205 \quad 281 \quad 140\ ]$$

$$= [\ 0.082 \quad 0.200 \quad 0.235 \quad 0.323 \quad 0.161\ ]$$

- Assuming the equal interval scales 1, 2, 3, 4, 5 for the two variables, their averages are

$$(0.137 \times 1) + (0.370 \times 2) + \cdots + (0.055 \times 5) = 2.672$$

$$(0.082 \times 1) + (0.200 \times 2) + \cdots + (0.161 \times 5) = 3.281$$

and the centered vectors $\mathbf{s}_1$ and $\mathbf{s}_2$ are

$$\mathbf{s}_1 = \begin{bmatrix} -1.672 \\ -0.672 \\ +0.328 \\ +1.328 \\ +2.328 \end{bmatrix} \qquad \mathbf{s}_2 = \begin{bmatrix} -2.281 \\ -1.281 \\ -0.281 \\ +0.719 \\ +1.719 \end{bmatrix}$$

- The *correspondence matrix* is the matrix of relative frequencies (we only give some elements of the matrix):

$$\mathbf{P}_{12} = (1/871)\begin{bmatrix} 27 & \cdots & 12 \\ 38 & \cdots & 30 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 29 \end{bmatrix} = \begin{bmatrix} 0.03100 & \cdots & 0.01378 \\ 0.04363 & \cdots & 0.03444 \\ \vdots & \ddots & \vdots \\ 0.00000 & \cdots & 0.03330 \end{bmatrix}$$

  and the diagonal matrices of masses, $\mathbf{D}_1$ and $\mathbf{D}_2$, contain the marginal relative frequencies (masses) computed above.
- Hence, the covariance, variances and correlation are

$$\begin{aligned} s_{12} = \mathbf{s}_1^{\mathsf{T}}\mathbf{P}_{12}\,\mathbf{s}_2 &= (0.03100 \times -1.672 \times -2.281) + \cdots + \\ &\quad (0.03330 \times 2.328 \times 1.719) \\ &= 0.4988 \end{aligned}$$

$$\begin{aligned} s_1^2 = \mathbf{s}_1^{\mathsf{T}}\mathbf{D}_1\mathbf{s}_1 &= 0.137 \times (-1.672)^2 + \cdots + 0.055 \times (2.328)^2 \\ &= 1.233 \end{aligned}$$

$$\begin{aligned} s_2^2 = \mathbf{s}_2^{\mathsf{T}}\mathbf{D}_2\mathbf{s}_2 &= 0.082 \times (-2.281)^2 + \cdots + 0.161 \times (1.719)^2 \\ &= 1.412 \end{aligned}$$

$$r = \frac{s_{12}}{s_1 s_2} = \frac{0.4988}{\sqrt{1.233 \times 1.412}} = 0.3780$$

All of the above calculations clearly depend on the equal-interval scale values in $\mathbf{s}_1$ and $\mathbf{s}_2$ assumed at the start. We now consider these scale values as unknowns to be determined, and we pose the following question: what scale values for $\mathbf{s}_1$ and $\mathbf{s}_2$ will give the highest correlation (Equation 2.1) between the two variables? This is exactly the problem of *canonical correlation* between the five dummy variables in $\mathbf{Z}_1$ and the five dummy variables in $\mathbf{Z}_2$. Because the correlation remains the same if any linear transformations of $\mathbf{s}_1$ and $\mathbf{s}_2$ are made, we need to introduce *identification conditions* that fix the scale of $\mathbf{s}_1$ and $\mathbf{s}_2$ in order to find the optimal solution. The usual identification conditions are that the two variables are standardized, i.e., that the means are zero, as previously: $(1/n)\mathbf{1}^{\mathsf{T}}\mathbf{Z}_1\mathbf{s}_1 = (1/n)\mathbf{1}^{\mathsf{T}}\mathbf{Z}_2\mathbf{s}_2 = 0$ and, furthermore, that the variances are 1: $\mathbf{s}_1^{\mathsf{T}}\mathbf{D}_1\mathbf{s}_1 = \mathbf{s}_2^{\mathsf{T}}\mathbf{D}_2\mathbf{s}_2 = 1$. Under

these conditions, we now show that the optimal solution coincides exactly with the so-called *standard coordinates* of the response categories on the first principal dimension of a simple CA of the original cross-tabulation (see Chapter 1).

Consider the singular-value decomposition (SVD) of the following normalized matrix:

$$\mathbf{D}_1^{-1/2}\mathbf{P}_{12}\mathbf{D}_2^{-1/2} = \mathbf{U}\Sigma\mathbf{V}^\mathsf{T} \quad \text{where} \quad \mathbf{U}^\mathsf{T}\mathbf{U} = \mathbf{V}^\mathsf{T}\mathbf{V} = \mathbf{I} \tag{2.2}$$

where $\mathbf{\Sigma}$ is the diagonal matrix of singular values, and $\mathbf{U}$ and $\mathbf{V}$ are the matrices of left and right singular vectors as columns. Then, writing Equation 2.2 for one pair of left and right vectors, $\mathbf{u}$ and $\mathbf{v}$, corresponding to a singular value σ, we have, after multiplying on the left by $\mathbf{u}^\mathsf{T}$ and on the right by $\mathbf{v}$, and using the orthogonality of the singular vectors:

$$\mathbf{u}^\mathsf{T}\mathbf{D}_1^{-1/2}\mathbf{P}_{12}\mathbf{D}_2^{-1/2}\mathbf{v} = \sigma$$

So if we let $\mathbf{s}_1 = \mathbf{D}_1^{-1/2}\mathbf{u}$ and $\mathbf{s}_2 = \mathbf{D}_2^{-1/2}\mathbf{v}$, then $\mathbf{s}_1^\mathsf{T}\mathbf{P}_{12}\mathbf{s}_2 = \sigma$, which is the formula for the covariance. Furthermore, the identification conditions $\mathbf{s}_1^\mathsf{T}\mathbf{D}_1\mathbf{s}_1 = \mathbf{s}_2^\mathsf{T}\mathbf{D}_2\mathbf{s}_2 = 1$ are satisfied, since the singular vectors have length 1: $\mathbf{u}^\mathsf{T}\mathbf{u} = \mathbf{v}^\mathsf{T}\mathbf{v} = 1$, so it appears that the correlation is given by the singular value σ. However, the centering conditions have not been imposed, and these can be introduced by first centering the matrix to be decomposed as follows, subtracting the product of the row and column margins from each element of the correspondence matrix:

$$\mathbf{D}_1^{-1/2}(\mathbf{P}_{12} - \mathbf{P}_{12}\mathbf{1}\mathbf{1}^\mathsf{T}\mathbf{P}_{12}^\mathsf{T})\mathbf{D}_2^{-1/2} = \mathbf{U}\Sigma\mathbf{V}^\mathsf{T} \tag{2.3}$$

where $\mathbf{P}_{12}\mathbf{1}$ is the (column) vector of row margins of $\mathbf{P}_{12}$, i.e., the row masses (denoted by $\mathbf{r}$ in Chapter 1), and $\mathbf{1}^\mathsf{T}\mathbf{P}_{12}^\mathsf{T}$ is the (row) vector of column margins, the column masses (denoted by $\mathbf{c}^\mathsf{T}$). In the parlance of CA, this is known as "removing the trivial solution" because the uncentered matrix (Equation 2.2) has a trivial maximal solution with a singular value of 1 for $\mathbf{s}_1$ and $\mathbf{s}_2$ equal to $\mathbf{1}$ (thus, $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{\Sigma}$ in Equation 2.2 all have this one extra trivial singular component, which is eliminated by the centering in Equation 2.3).

We thus have the following result: each singular value is a correlation between variables A and B, based on the scale values from the transformed singular vectors $\mathbf{s}_1$ and $\mathbf{s}_2$, and so the maximum correlation is attained for the first (i.e., the largest) singular value of Equation 2.3 or, equivalently, the second largest singular value of the uncentered matrix

(Equation 2.2). The solutions $\mathbf{s}_1$ and $\mathbf{s}_2$ are exactly the vectors of standard coordinates in CA on the first principal axis. The largest singular value σ_1 of Equation 2.3, also called the first canonical correlation, is equal to 0.4106 in our example, compared with the value of 0.3780 obtained with the equal-interval (1 to 5) scales. The scale values are:

$$\mathbf{s}_1^\mathsf{T} = [-1.017 \quad -0.560 \quad -0.248 \quad 1.239 \quad 2.741]$$

$$\mathbf{s}_2^\mathsf{T} = [-1.571 \quad -0.667 \quad -0.606 \quad -0.293 \quad 1.926]$$

These scale values are standardized, but since any linear transformation leaves the correlation unchanged, it is convenient to transform them so that the endpoints also have values 1 and 5, with a range of 4, in order to compare with the previous equal-interval scales. For example, for the first variable, the range of values is $2.741 - (-1.017) = 3.758$, so in order to make the range exactly four units, we should multiply all the values by $4/3.758 = 1.064$, in which case the lowest value is now $-1.017 \times 1.064 = -1.083$. Then the addition of 2.083 to all the values will bring the scale to have lowest and highest values equal to 1 and 5, respectively. This procedure gives the following rescaled values for the two sets of response categories:

$$\text{rescaled row values} = [1 \quad 1.486 \quad 1.818 \quad 3.402 \quad 5]$$

$$\text{rescaled column values} = [1 \quad 2.034 \quad 2.103 \quad 3.132 \quad 5]$$

The scale points do emerge in their expected order in both cases, but it is interesting to study their relative spacings. Compared with the equal-interval values considered previously, these rescaled values show that the categories "disagree" and "disagree strongly" for question A are further spaced out, with relatively small differences between scale values assigned to the categories "strongly agree," "agree," and "neither/nor." For question B the difference between "disagree" and "disagree strongly" is even larger, almost two full units. For both questions the neutral "neither/nor" category is not in the center of the scale but close to the agreement category.

Before moving onto the case of several variables, we remark that, in the above, only one set of scale values has been derived for each variable, corresponding to the first singular value σ_1. Further sets of scale values can be determined in a stepwise manner by maximizing the correlation between another pair of subject scores based on different scale values, say $\tilde{\mathbf{s}}_1$ and $\tilde{\mathbf{s}}_2$, where the subject scores are uncorrelated with those already obtained, i.e., $\tilde{\mathbf{s}}_1^\mathsf{T}\mathbf{D}_1\mathbf{s}_1 = \tilde{\mathbf{s}}_2^\mathsf{T}\mathbf{D}_2\mathbf{s}_2 = 0$. The solution is given by the second set of singular vectors of Equation 2.3, transformed

as before to standard coordinates, corresponding to the second singular value, σ_2, which is the second canonical correlation. For a table of order $I \times J$, this process can be continued to obtain a total of $\min\{I-1, J-1\}$ canonical correlations and associated scale values: in our 5×5 example, four sets of scale values and canonical correlations can be calculated. The canonical correlations are the square roots of the *principal inertias* usually reported on the axes of the map (see Chapter 1 and Section 2.3 below).

2.2.2 Several variables

To make the transition to the case of several variables, notice that the problem is almost identical if we reformulate it as maximizing the correlation between the two variables and their average (or their sum). In general, for two variables z_1 and z_2 with correlation ρ, the correlation between either of them and their average $\frac{1}{2}(z_1 + z_2)$ (or their sum $z_1 + z_2$) is equal to $\sqrt{(1+\rho)/2}$, so that maximizing ρ is equivalent to maximizing the correlation between the variables and their average (or sum). The only real difference is the value of the maximum found: in the latter formulation this will be $\sqrt{(1+\rho)/2}$ and not the value of ρ itself. The average of two categorical variables leads us to consider the matrix of the two indicator matrices $[\mathbf{Z}_1 \quad \mathbf{Z}_2]$, where the average of the two quantifications of the variables, based on $\mathbf{s}_1$ and $\mathbf{s}_2$, respectively, is equal to

$$\frac{1}{2}\left(\mathbf{Z}_1\mathbf{s}_1 + \mathbf{Z}_2\mathbf{s}_2\right) = \frac{1}{2}[\mathbf{Z}_1 \quad \mathbf{Z}_2]\begin{bmatrix}\mathbf{s}_1\\ \mathbf{s}_2\end{bmatrix}$$

Consider now what happens when we apply the standard CA algorithm to the superindicator matrix $\mathbf{Z} = [\mathbf{Z}_1 \; \mathbf{Z}_2]$. Because $\mathbf{Z}$ has total sum $2n$, with each of the n rows summing to a constant 2 and column sums equal to the marginal frequencies of each variable, the correspondence matrix is $[1/(2n)]\mathbf{Z}$, the row mass matrix is $(1/n)\mathbf{I}$, and the column mass matrix is $\mathbf{D} = \frac{1}{2}\text{diag}(\mathbf{D}_1,\mathbf{D}_2)$, where $\text{diag}(\mathbf{D}_1,\mathbf{D}_2)$ is the diagonal matrix formed by the two diagonal matrices $\mathbf{D}_1$ and $\mathbf{D}_2$ defined. Hence, the SVD to compute the CA solution of $\mathbf{Z}$ is (in its uncentered form, see Equation 2.2):

$$\sqrt{n}\,\frac{\mathbf{Z}}{2n}\mathbf{D}^{-1/2} = \mathbf{U}\Gamma\mathbf{V}^{\mathsf{T}} \quad \text{where} \quad \mathbf{U}^{\mathsf{T}}\mathbf{U} = \mathbf{V}^{\mathsf{T}}\mathbf{V} = \mathbf{I}$$

which—in one of its symmetric eigenvalue formulations—can be written as:

$$\left(\sqrt{n}\,\frac{\mathbf{Z}}{2n}\mathbf{D}^{-1/2}\right)^{\mathsf{T}}\left(\sqrt{n}\,\frac{\mathbf{Z}}{2n}\mathbf{D}^{-1/2}\right)=\frac{1}{4n}\mathbf{D}^{-1/2}\mathbf{Z}^{\mathsf{T}}\mathbf{Z}\mathbf{D}^{-1/2}=\mathbf{V}\boldsymbol{\Gamma}^{2}\mathbf{V}^{\mathsf{T}}$$

$$\text{where } \mathbf{V}^{\mathsf{T}}\mathbf{V}=\mathbf{I}$$

that is,

$$\frac{1}{4n}\mathbf{D}^{-1/2}\mathbf{C}\mathbf{D}^{-1/2}=\mathbf{V}\boldsymbol{\Gamma}^{2}\mathbf{V}^{\mathsf{T}}=\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{\mathsf{T}} \tag{2.4}$$

where $\mathbf{C}=\mathbf{Z}^{\mathsf{T}}\mathbf{Z}$ and $\boldsymbol{\Lambda}=\boldsymbol{\Gamma}^{2}$. The matrix $\mathbf{C}$, called the *Burt matrix*, is an important data structure in MCA: it is the matrix of all two-way cross-tabulations of the categorical variables, which in the present case of two categorical variables can be written as:

$$\mathbf{C}=\begin{bmatrix}\mathbf{Z}_1^{\mathsf{T}}\mathbf{Z}_1 & \mathbf{Z}_1^{\mathsf{T}}\mathbf{Z}_2\\ \mathbf{Z}_2^{\mathsf{T}}\mathbf{Z}_1 & \mathbf{Z}_2^{\mathsf{T}}\mathbf{Z}_2\end{bmatrix}=n\begin{bmatrix}\mathbf{D}_1 & \mathbf{P}_{12}\\ \mathbf{P}_{21} & \mathbf{D}_2\end{bmatrix}$$

We use the notation $\boldsymbol{\Lambda}=\boldsymbol{\Gamma}^{2}$, that is, the squares γ^2 of the singular values, or principal inertias of $\mathbf{Z}$ that appear on the diagonal of $\boldsymbol{\Gamma}^2$, are denoted by λ on the diagonal of $\boldsymbol{\Lambda}$. Writing Equation 2.4 for a single eigenvector $\mathbf{v}$, partitioned into two subvectors $\mathbf{v}_1$ and $\mathbf{v}_2$ (one corresponding to the rows of the original table, the other to the columns) and multiplying as before on the left by $\mathbf{v}^{\mathsf{T}}$ and on the right by $\mathbf{v}$, defining $\mathbf{s}=\mathbf{D}^{-1/2}\mathbf{v}$ similarly partitioned into $\mathbf{s}_1$ and $\mathbf{s}_2$, we obtain the eigenequation:

$$\frac{1}{4}\begin{bmatrix}\mathbf{s}_1\\ \mathbf{s}_2\end{bmatrix}^{\mathsf{T}}\begin{bmatrix}\mathbf{D}_1 & \mathbf{P}_{12}\\ \mathbf{P}_{21} & \mathbf{D}_2\end{bmatrix}\begin{bmatrix}\mathbf{s}_1\\ \mathbf{s}_2\end{bmatrix}=\gamma^2=\lambda$$

that is,

$$^1/_4(\mathbf{s}_1^{\mathsf{T}}\mathbf{D}_1\mathbf{s}_1+\mathbf{s}_1^{\mathsf{T}}\mathbf{P}_{12}\mathbf{s}_2+\mathbf{s}_2^{\mathsf{T}}\mathbf{P}_{21}\mathbf{s}_1+\mathbf{s}_2^{\mathsf{T}}\mathbf{D}_2\mathbf{s}_2)=\gamma^2=\lambda \tag{2.5}$$

The maximum value of Equation 2.5, given by the largest nontrivial eigenvalue $\lambda_1=\gamma_1^2$, coincides with the solution of simple CA of the single two-way table with correspondence matrix $\mathbf{P}_{12}$, except that its maximum is now equal to $^1/_4(1+\sigma_1+\sigma_1+1)=^1/_2(1+\sigma_1)$, where σ_1 is the maximized canonical correlation in simple CA. According to our previous remarks, $^1/_2(1+\sigma_1)$ is exactly the square of the correlation between either of the

two (quantified) categorical variables and their average (or sum). Hence $\mathbf{s}_1$ and $\mathbf{s}_2$ derived above are identical to the $\mathbf{s}_1$ and $\mathbf{s}_2$ of simple CA, and what we have derived are the standard coordinates of the columns of the indicator matrix $\mathbf{Z}$. Notice that the eigenvalue λ_1 above is also the singular value of $\mathbf{C}$ because $\mathbf{C}$ is symmetric: in the language of geometric CA (see Chapter 1 and Section 2.3 below), λ_1 is the square root of the principal inertia of the Burt matrix $\mathbf{C}$.

The alert reader will have noticed that in Equation 2.4 the identification condition on $\mathbf{s}$ implied by the standardization of $\mathbf{v}$ in the SVD is that its weighted sum of squares is equal to 1, that is, $\frac{1}{2}(\mathbf{s}_1^\top \mathbf{D}_1\mathbf{s}_1 + \mathbf{s}_2^\top \mathbf{D}_2\mathbf{s}_2) = \mathbf{s}^\top\mathbf{D}\mathbf{s} = 1$, and not that the subvectors $\mathbf{s}_1$ and $\mathbf{s}_2$ are individually normalized to be 1. It can be shown, however, that if $\mathbf{s}_1$ and $\mathbf{s}_2$ constitute a solution corresponding to an eigenvalue $\frac{1}{2}(1 + \sigma)$, then $\mathbf{s}_1$ and $-\mathbf{s}_2$ constitute another solution corresponding to the eigenvalue $\frac{1}{2}(1 - \sigma)$ (see, for example, Greenacre 1984: section 5.1). The orthogonality of these eigenvectors, $\mathbf{s}_1^\top\mathbf{D}_1\mathbf{s}_1 - \mathbf{s}_2^\top\mathbf{D}_2\mathbf{s}_2 = 0$, together with the overall normalization constraint, imply the individual normalizations $\mathbf{s}_1^\top\mathbf{D}_1\mathbf{s}_1 = \mathbf{s}_2^\top\mathbf{D}_2\mathbf{s}_2 = 1$. As far as the individual centering constraints are concerned, these are automatic, since each set of dummy variables (columns of $\mathbf{Z}_1$ and $\mathbf{Z}_2$) has the same sum, equal to $\mathbf{1}$, the vector of ones, so that each set has the same centroid, equal to the overall centroid $(1/n)\mathbf{1}$.

The scene is now set for one possible generalization of CA to the multivariable case, where there are Q categorical variables, coded in indicator matrices $\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_Q$. The problem can be defined as finding a set of scale values $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_Q$ for the variables so that an overall measure of correlation is maximized. To generalize the two-variable case, the measure of choice is the sum of squared correlations of the individual scores $\mathbf{Z}_1\mathbf{s}_1, \mathbf{Z}_2\mathbf{s}_2, \ldots, \mathbf{Z}_Q\mathbf{s}_Q$ with the summated score $\mathbf{Z}\mathbf{s}$, where $\mathbf{Z}$ and $\mathbf{s}$ are the concatenations of the $\mathbf{Z}_q$'s and $\mathbf{s}_q$'s, respectively. We specify an overall identification constraint $\mathbf{s}^\top\mathbf{D}\mathbf{s} = 1$, where $\mathbf{D} = (1/Q)\mathrm{diag}(\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_Q)$. This overall constraint does not imply that individual variances $\mathbf{s}_q^\top\mathbf{D}_q\mathbf{s}_q$ will be 1 in the final solution—in contrast to the case $Q = 2$ described in the previous paragraph.

Again there are two ways to achieve the solution, one way by performing a CA of the superindicator matrix $\mathbf{Z} = [\mathbf{Z}_1\ \mathbf{Z}_2, \ldots, \mathbf{Z}_Q]$, alternatively a CA of the Burt matrix $\mathbf{C}$, which is now a block matrix with Q blocks row-wise and columnwise. We denote the number of categories for the qth categorical variable by J_q and let $J = \sum_q J_q$ be the total number of categories. Then $\mathbf{Z}$ is of order $n \times J$ and $\mathbf{C}$ is of order $J \times J$. Since $\mathbf{Z}$ has total sum nQ, with row sums equal to a constant Q and column sums equal to the marginal frequencies of each variable, the correspondence matrix is $(1/Qn)\mathbf{Z}$, the row mass matrix

is $(1/n)\mathbf{I}$, and the column mass matrix is $\mathbf{D}$. Hence, the SVD to compute the CA solution of $\mathbf{Z}$ is (in its uncentered form, see Equation 2.2):

$$\sqrt{n}\,\frac{\mathbf{Z}}{Qn}\mathbf{D}^{-1/2} = \mathbf{U}\Gamma\mathbf{V}^{\mathsf{T}} \quad \text{where} \quad \mathbf{U}^{\mathsf{T}}\mathbf{U} = \mathbf{V}^{\mathsf{T}}\mathbf{V} = \mathbf{I} \tag{2.6}$$

To eliminate the trivial solution, the matrix to be decomposed is (see Equation 2.3):

$$\sqrt{n}\left(\frac{\mathbf{Z}}{Qn} - \frac{1}{n}\mathbf{1}\mathbf{1}^{\mathsf{T}}\mathbf{D}\right)\mathbf{D}^{-1/2}$$

where $(1/n)\mathbf{1}$ is the vector of row masses and $\mathbf{1}^{\mathsf{T}}\mathbf{D}$ is the vector of column masses of the indicator matrix (denoted by $\mathbf{c}^{\mathsf{T}}$ in simple CA). The SVD for the CA of the Burt matrix $\mathbf{C}$ (uncentered) is

$$\mathbf{D}^{-1/2}\,\frac{\mathbf{C}}{Q^2n}\,\mathbf{D}^{-1/2} = \mathbf{V}\Gamma^2\mathbf{V}^{\mathsf{T}} = \mathbf{V}\Lambda\mathbf{V}^{\mathsf{T}} \quad \text{where} \quad \mathbf{V}^{\mathsf{T}}\mathbf{V} = \mathbf{I} \tag{2.7}$$

and $\mathbf{C} = \mathbf{Z}^{\mathsf{T}}\mathbf{Z}$. Once again, the centered form of the matrix on the left-hand side of Equation 2.7 removes the trivial solution in the form of the expected relative frequencies:

$$\mathbf{D}^{-1/2}\left(\frac{\mathbf{C}}{Q^2n} - \mathbf{D}\mathbf{1}\mathbf{1}^{\mathsf{T}}\mathbf{D}\right)\mathbf{D}^{-1/2}$$

The right-hand singular vectors, which give us the scale values for the Q variables, are identical in the two problems. The maximum value of the average squared correlation is given by the square of the first singular value in the (centered) analysis of $\mathbf{Z}$, that is, the first singular value in the (centered) analysis of $\mathbf{C}$. Notice that the singular values λ in the analysis of $\mathbf{C}$ are also eigenvalues, since the matrix being decomposed is positive definite symmetric. The standard coordinates $\mathbf{x}$ that provide the scale values, partitioned into $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_Q$ for the Q variables, are given by the usual transformation of the singular vectors:

$$\mathbf{x} = \mathbf{D}^{-1/2}\mathbf{v}$$

where $\mathbf{v}$ is the first right-hand singular vector, that is, the first column of $\mathbf{V}$. Only the principal coordinates are slightly different in the two problems, since the singular values differ.

We now apply the above theory to the four variables described in Section 2.2.1. In terms of our notation: $Q = 4$, $J_q = 5$ for all q, $J = 20$, $\mathbf{Z}$

is 871×20, and $\mathbf{C}$ is 20×20. Some rows of the original data matrix and the indicator matrix are given in Table A.1 and Table A.2 of the computational appendix at the end of this book, which details all of the computing steps to obtain the solution in this example. The full Burt matrix is reproduced in Table 2.2. In Table 2.3 we reproduce the standard coordinates for the first and second optimal solutions, along with their corresponding correlation measures. In addition, the squared correlation of each quantified variable with the total score is given, showing that the correlation measure is equal to their average. The first and optimal set of scale values, with an average squared correlation of 0.457, is monotonically increasing for questions A, B, and C, but question D has a quite different pattern, with the extreme poles opposing the intermediate categories. This is evidence that there is a possible problem with the responses to question D, which was worded in the reverse sense compared with the other questions. The second set of scale values captures an axis of "polarization," where all four questions have the pattern of the extreme categories opposing the intermediate ones, and here question D fits in more with the others. This interpretation is supported by the squared correlations, which show a low value for question D in the first solution. MCA thus effectively acts as an item analysis, and this result shows us that question D has degraded the reliability of the total score based on the second optimal solution and should preferably be removed.

To clarify the link between MCA and reliability theory, consider the Q variables as items measuring an underlying construct. Using the average squared correlation of 0.457, that is, 0.676 in the square root, as a measure of the reliability is an overestimate because even for random data we would find positive correlation between items and their sum (in fact, the average squared correlation between Q uncorrelated items and their sum is equal to $1/Q$). Cronbach's alpha is a measure of reliability that compensates for this and is classically defined as:

$$\alpha = \left(\frac{Q}{Q-1} \right) \left(1 - \frac{\sum_q s_q^2}{s^2} \right) \tag{2.8}$$

where s_q^2 is the variance of the qth item score, and s^2 is the variance of the summated score. In MCA the sum of item score variances ($\sum_q s_q^2$) is equal to $\mathbf{a}^{\mathsf{T}}\mathbf{D}\mathbf{a}$, which from the identification conditions described above is a fixed value, equal to Q. The variance of the summated score (s^2) is equal to Q^2 times the variance of the average score $\mathbf{z}$, that is, Q^2

Table 2.2 Data on attitudes to science and the environment, showing the complete Burt matrix of all pairwise cross-tables of the four variables.

	Variable A					Variable B					Variable C					Variable D				
	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	C1	C2	C3	C4	C5	D1	D2	D3	D4	D5
A1	119	0	0	0	0	27	28	30	22	12	49	40	18	7	5	15	25	17	34	28
A2	0	322	0	0	0	38	74	84	96	30	67	142	60	41	12	22	102	76	68	54
A3	0	0	204	0	0	3	48	63	73	17	18	75	70	34	7	10	44	68	58	24
A4	0	0	0	178	0	3	21	23	79	52	16	50	40	56	16	9	52	28	54	35
A5	0	0	0	0	48	0	3	5	11	29	2	9	9	16	12	4	9	13	12	10
B1	27	38	3	3	0	71	0	0	0	0	43	19	4	3	2	9	17	10	10	25
B2	28	74	48	21	3	0	174	0	0	0	36	88	34	15	1	16	51	42	45	20
B3	30	84	63	23	5	0	0	205	0	0	37	90	57	19	2	10	53	63	51	28
B4	22	96	73	79	11	0	0	0	281	0	27	88	75	74	17	6	66	70	92	47
B5	12	30	17	52	29	0	0	0	0	140	9	31	27	43	30	19	45	17	28	31
C1	49	67	18	16	2	43	36	37	27	9	152	0	0	0	0	25	24	15	38	50
C2	40	142	75	50	9	19	88	90	88	31	0	316	0	0	0	15	97	67	89	48
C3	18	60	70	40	9	4	34	57	75	27	0	0	197	0	0	5	51	83	41	17
C4	7	41	34	56	16	3	15	19	74	43	0	0	0	154	0	6	44	30	51	23
C5	5	12	7	16	12	2	1	2	17	30	0	0	0	0	52	9	16	7	7	13
D1	15	22	10	9	4	9	16	10	6	19	25	15	5	6	9	60	0	0	0	0
D2	25	102	44	52	9	17	51	53	66	45	24	97	51	44	16	0	232	0	0	0
D3	17	76	68	28	13	10	42	63	70	17	15	67	83	30	7	0	0	202	0	0
D4	34	68	58	54	12	10	45	51	92	28	38	89	41	51	7	0	0	0	226	0
D5	28	54	24	35	10	25	20	28	47	31	50	48	17	23	13	0	0	0	0	151

Note: Table 2.1 is the A × B block of this matrix.

Table 2.3 Results of CA of 871 × 20 indicator matrix **Z** (see Table A.2 in the appendix) or, equivalently, of Burt matrix **C** in Table 2.2, showing standard coordinates (scale values) for the four variables on the first two dimensions of the solution (F1 and F2).

	F1	F2
A1	–1.837	0.727
A2	–0.546	–0.284
A3	0.447	–1.199
A4	1.166	0.737
A5	1.995	2.470
sq. corr.	0.510	0.382
B1	–2.924	1.370
B2	–0.642	–0.667
B3	–0.346	–0.964
B4	0.714	–0.280
B5	1.354	2.108
sq. corr	0.579	0.517
C1	–2.158	0.909
C2	–0.247	–0.592
C3	0.619	–1.044
C4	1.349	0.635
C5	1.468	3.017
sq. corr.	0.627	0.488
D1	–1.204	1.822
D2	0.221	–0.007
D3	0.385	–1.159
D4	0.222	–0.211
D5	–0.708	1.152
sq. corr.	0.113	0.337
rho	0.457	0.431
Cronbach's alpha	0.605	0.560

Note: *sq. corr.* is the squared correlation of the quantified variable with the total score; *rho* is the corresponding singular value of **C**, i.e., the squared singular value (or principal inertia) of **Z**, which is the arithmetic average of the four corresponding squared correlations; *Cronbach's alpha* is the measure of reliability discussed in Section 2.2.2.

times the λ that we are maximizing. Hence, we can write the maximum value of Equation 2.8 as:

$$\alpha = \left(\frac{Q}{Q-1}\right)\left(1-\frac{Q}{Q^2\lambda}\right) = \left(\frac{Q}{Q-1}\right)\left(1-\frac{1}{Q\lambda}\right) \tag{2.9}$$

so that maximum λ (the first singular value of **C** in Equation 2.7, which is also an eigenvalue as we have said previously) corresponds to maximum reliability. Hence, the maximum value of Cronbach's alpha for the first two solutions is, respectively (see Table 2.3):

$$\alpha_1 = \frac{4}{3}\left(1-\frac{1}{4\times 0.457}\right) = 0.605 \quad \text{and} \quad \alpha_2 = \frac{4}{3}\left(1-\frac{1}{4\times 0.431}\right) = 0.560$$

If question D is removed, as would be suggested by its low item correlation with the total, a recomputation of the solution gives a much higher value, 0.602, of the maximum average squared correlation, and an increase in Cronbach's alpha to 0.669. (We do not report the complete results here.)

Table 2.4 shows all the squared intercorrelations as well as the variances and covariances of the four quantified questions, according to the first optimal solution. This table also demonstrates empirically that the optimal λ can be computed either as (a) the variance of the total score, or (b) the average of the four squared correlations of the respective questions with the total, or (c) the average of all the elements of the full variance–covariance matrix between the four questions.

2.2.3 Homogeneity analysis

An alternative but equivalent definition of the correlational definition of MCA is based on Guttman's criterion of "internal consistency" (see, for example, Nishisato 1994). The idea is to look for scale values $\mathbf{s}_1$, $\mathbf{s}_2,\ldots,\mathbf{s}_Q$ that give individual scores $\mathbf{Z}_1\mathbf{s}_1$, $\mathbf{Z}_2\mathbf{s}_2,\ldots,\mathbf{Z}_Q\mathbf{s}_Q$ that are as close to one another (i.e., *homogeneous*) as possible. Lack of closeness can be measured by the sum of squares of the differences of each individual's Q scale values from the corresponding mean score in the vector $(1/Q)(\mathbf{Z}_1\mathbf{s}_1+ \mathbf{Z}_2\mathbf{s}_2 + \cdots + \mathbf{Z}_Q\mathbf{s}_Q) = (1/Q)\mathbf{Z}\mathbf{s}$, which we shall again denote by $\mathbf{z}$. The overall objective is thus to

Table 2.4 Squared intercorrelations as well as variances and covariances of the four quantified questions according to the first optimal solution.

	A	B	C	D	Squared correlation with total
A	1.1151	*0.1396*	*0.1270*	*0.0059*	**0.5100**
B	0.4440	1.2666	*0.1868*	*0.0059*	**0.5793**
C	0.4406	0.5697	1.3716	*0.0480*	**0.6273**
D	0.0403	0.0369	0.1274	0.2467	**0.1129**
Ave. covariance	0.5100	0.5793	0.6273	0.1129	**0.4574**

Note: The squared correlations between the four variables A, B, C, and D are quantified by their scale values on the first dimension (upper right-hand-side triangle of table, in italics) as well as their squared correlations with the total score (right-hand column; cf. *sq.corr.* in column F1 of Table 2.3). The variances (diagonal of table) and covariances (lower left-hand-side triangle of table) are also quantified, with the average covariance of each variable with itself and the others shown in the last row in boldface (e.g., 0.5793 = (0.4440 + 1.2666 + 0.5697 + 0.0369)/4). Note that these average covariances are identical to the squared correlations with the total. Hence, the variance of the average score (the quantity maximized by MCA, underlined) is both (a) the average of the four squared correlations of the question scores with the total score and (b) the average of the four average covariances; in other words, it is the average of the full 4×4 variance–covariance matrix. Note further the sum of the variances of the four variables, 1.1151 + 1.2666 + 1.3716 + 0.2467 = 4, which is the identification condition on the scale values. (To calculate variances and covariances, divide by $n = 871$, not $n - 1$.)

minimize, in this case, the following function of $\mathbf{s}$, which is the average of the Q squared differences for each individual, averaged in turn over all n individuals:

$$\frac{1}{nQ}\left[(\mathbf{Z}_1\mathbf{s}_1-\mathbf{z})^{\mathsf{T}}(\mathbf{Z}_1\mathbf{s}_1-\mathbf{z})+(\mathbf{Z}_2\mathbf{s}_2-\mathbf{z})^{\mathsf{T}}(\mathbf{Z}_2\mathbf{s}_2-\mathbf{z})+\cdots+(\mathbf{Z}_Q\mathbf{s}_Q-\mathbf{z})^{\mathsf{T}}(\mathbf{Z}_Q\mathbf{s}_Q-\mathbf{z})\right] \tag{2.10}$$

This approach is known as *homogeneity analysis* (Gifi 1990), and the objective function (Equation 2.10) is called the loss function. Here "loss" refers to loss of homogeneity, since perfect homogeneity would be when all the differences $\mathbf{Z}_q\mathbf{s}_q - \mathbf{z}$ are zero. Once more an identification condition on $\mathbf{s}$ is required, otherwise the trivial solution when all elements of $\mathbf{s}$ are constant will be found, giving a loss of zero. With

the same quadratic constraint $\mathbf{s}^T\mathbf{D}\mathbf{s} = 1$ as previously, it can be shown that minimum loss is achieved by the same optimal scale values described above, and the value of the minimum is equal to 1 minus the value of the corresponding largest eigenvalue of the superindicator matrix $\mathbf{Z}$. In our example, the successively maximized eigenvalues of 0.457 and 0.431 (see Table 2.4) correspond to minimum losses of 0.543 and 0.569, respectively.

2.3 Geometric approach

The geometric approach to CA, introduced in Chapter 1, turns out to be slightly more problematic to generalize to the multivariable case. A lot of the controversy about CA stems from this difficulty, and here we shall clarify the issues involved in MCA as a graphical method as well as propose a specific version of MCA that acceptably addresses these problems. We shall approach the geometry from both the chi-square distance scaling perspective and the biplot perspective.

Figure 2.1 shows the usual CA map of the contingency table in Table 2.1. The map is established using the theory described in Chapter 1 and Section 2.2.1, namely the SVD of the matrix of standardized residuals, followed by the calculation of the principal coordinates to represent the points in a map. The principal coordinates are the standard coordinates multiplied by the respective singular values (see Chapter 1). Since standard coordinates have unit normalization, principal coordinates are normalized to have (weighted) sum of squares equal to the respective squared singular value of the associated solution.

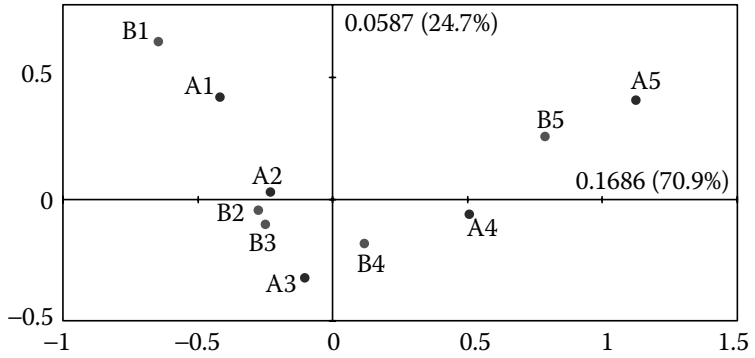

Figure 2.1 Symmetric CA map of Table 2.1. Percentage of inertia displayed in the map is 95.6%. Categories for questions A and B are from 1 (agree strongly) to 5 (disagree strongly).

The squared singular value σ^2 is called the *principal inertia*, corresponding to a principal axis, or dimension (see Chapter 1). Figure 2.1 shows the points represented by their principal coordinates calculated for the first two principal axes.

2.3.1 Chi-square distance scaling

Simple CA is justified mainly by its use of the chi-square (χ^2) distance as a measure of dissimilarity between row profiles and between column profiles of a two-way table. In Figure 2.1, where both rows and columns are displayed in principal coordinates, distances between row points optimally approximate the χ^2 distances between row profiles, and the distances between column points optimally approximate the χ^2 distances between column profiles. Recall from Chapter 1 that the squared χ^2 distance between the row profiles, for example, has this form:

$$d^2(i,i') = \sum_{j=1}^{J} \left(\frac{p_{ij}}{r_i} - \frac{p_{i'j}}{r_{i'}} \right)^2 \Big/ c_j \qquad (2.11)$$

so that every jth squared difference between the profile elements is weighted inversely by the column margin c_j.

MCA is the application of CA to either the superindicator matrix $\mathbf{Z}$ or the Burt matrix $\mathbf{C}$. While the χ^2 distance makes sense for a two-way contingency table, it has less justification when applied to the rows and to the columns of the superindicator matrix or the Burt matrix. As an illustration of this problem, consider the same four-variable example on attitudes to science in the environmental context. As shown in Equation 2.11, the χ^2 distances between rows and between columns are calculated between their profiles: in the case of distances between row profiles, the column masses of the correspondence matrix are used inversely as weights in the calculation of distance. The row profiles of the superindicator matrix $\mathbf{Z} = [\mathbf{Z}_1 \quad \mathbf{Z}_2 \quad \mathbf{Z}_3 \quad \mathbf{Z}_4]$ are vectors with elements equal to zero apart from four values of 1/4 in the positions of the categories selected by the corresponding case. When calculating the distance between two rows, differences between coincident zero values and coincident values of 1/4 are zero, thus making no contribution to the distance measure, and so it is only differences between noncoincident categories that count in the distance function. These nonzero squared differences (each equal to 1/16 in this case), arising from disagreements between respondents, are then weighted

by the inverses of the corresponding column masses, proportional to the respective categories' marginal frequencies, and added up to give the χ^2 distances. For the rows, this distance measure appears fairly reasonable, and the weighting is in accordance with the χ^2 concept that the contribution of categories with low frequency needs to be boosted because their variance is inherently lower. However, Gower (Chapter 3, this volume) prefers an unweighted version of this distance measure.

The situation for the column profiles of $\mathbf{Z}$, however, is quite different and difficult, if not impossible, to justify. Here we make the distinction between calculating (a) distances between two categories of the same variable and (b) distances between two categories of different variables. Let us denote the relative frequency of the jth column category by c_j (that is, for a particular variable, the quantities c_j sum to 1). As shown by Greenacre (1989), the squared χ^2 distances between two column categories of a superindicator matrix are, in the two cases:

1. $1/c_j + 1/c_{j'}$ between categories j and j' of the same variable q
2. $1/c_j + 1/c_{j'} - 2p_{jj'}/(c_j c_{j'})$ between categories j and j' of different variables q and q'

where $p_{jj'}$ is the relative frequency of occurrence of categories j and j' (in fact, the above formulas are the same, since the frequency of co-occurrence of categories j and j' of the same variable is zero). The former "within-variable" distance makes little sense, since it depends only on the marginal frequencies, no matter what the relationship with the other variables is. The latter "between-variable" distance has at least a slight justification in that the distance decreases as association between categories j and j' increases, but again the dominant role played by the marginal frequencies is hard to defend.

The situation improves if we consider intercategory distances calculated on the Burt matrix rather than the indicator matrix. Because the Burt matrix is symmetric, it makes no difference whether we calculate the χ^2 distances between rows or between columns. The squared distance between categories can be described verbally as follows:

1. Between categories j and j' of the same variable q: This within-variable squared distance is the average of the $(Q - 1)$ squared χ^2 distances between categories j and j' calculated in the cross-tabulations of variable q with all the other variables $q' \neq q$, but also including an unnecessary term from the cross-tabulation of q with itself. (This term involves the distance

between two unit profiles in a submatrix on the diagonal of **C** and is thus a large component of the overall distance, tending to inflate the distance.)

2. Between categories j and j' of different variables q and q': This between-variable squared distance is an average of $(Q - 2)$ squared χ^2 distances between profiles of categories j and j' across variables not equal to q or q', but including two additional terms that can also be considered unnecessary. (These measure distances between a profile and a unit profile again on the diagonal of **C**, again tending to inflate the between-category distance.)

In spite of the above theoretical difficulties to justify the full-space chi-square geometry, MCA as regularly applied — that is, the CA of **Z** or of **C** — successfully recovers interesting patterns of association between the variables. It seems that the low-dimensional projections of the points are more valid than their full-dimensional counterparts, which is a paradox from the multidimensional scaling viewpoint. Another worrying aspect is the inflation of the total inertias of **Z** and of **C**, which leads to all percentages of inertia on the principal axes being artificially low. This inflation can also be understood by considering the calculation of total inertia for the Burt matrix **C** and the high contributions made by the diagonal matrices on its block diagonal. It is clear that MCA of a two-variable data set will not give the same results as a CA; the standard coordinates will be the same, but the principal inertias (and hence the principal coordinates) and their percentages of inertia will be different. In Section 2.3.3 we define another variant of MCA, called joint correspondence analysis, that resolves all these issues to a certain extent. We shall also show that a simple adjustment of the scale in the MCA solution dramatically improves the fit from a multidimensional scaling viewpoint.

2.3.2 Biplot

The biplot is concerned with data reconstruction in a joint map of the rows and columns, rather than distance reconstruction. In the simple case of a two-way table, we can think of reconstructing different variants of the table depending on the way we think of the table: either as a set of rows, or a set of columns, or just a two-way table of entries where rows and columns are symmetric entities (see Chapter 3). As an illustration of this approach we consider a two-way table as a set of rows. For example, Table 2.5a shows the row profiles of the two-way

Table 2.5 (a) Row profiles of Table 2.1, including average row profile. (b) Approximate row profiles estimated from biplot of Figure 2.2 (the average profile is always represented exactly by the origin of the map).

(a) Original profiles

	B1	B2	B3	B4	B5	Sum
A1	0.227	0.235	0.252	0.185	0.101	1
A2	0.118	0.230	0.261	0.298	0.093	1
A3	0.115	0.235	0.309	0.358	0.083	1
A4	0.017	0.118	0.129	0.444	0.292	1
A5	0.000	0.063	0.104	0.229	0.604	1
Average	0.075	0.176	0.211	0.303	0.235	1

(b) Estimated profiles

	B1	B2	B3	B4	B5	Sum
A1	0.226	0.239	0.253	0.181	0.102	1
A2	0.117	0.229	0.265	0.294	0.094	1
A3	0.024	0.226	0.283	0.393	0.074	1
A4	0.002	0.135	0.169	0.387	0.307	1
A5	0.026	0.034	0.034	0.329	0.578	1
Average	0.075	0.176	0.211	0.303	0.235	1

Note: The difference between the two tables is the error of biplot approximation, measured as 100 − 95.6% = 4.4% of the total inertia of the table.

table in Table 2.1, that is, conditional on each response category of question A, the proportions of respondents falling into the response categories of question B. The biplot can be thought of as a way to reconstruct these row profiles in a map. Greenacre and Hastie (1987) and Greenacre (1993a) show how the asymmetric map of CA, with row points in principal coordinates and column points in standard coordinates, is a biplot of these profiles. The direction vector defined by each column point, called a *biplot axis*, can be calibrated in profile units, and the approximate value of the profile can be read off the map by simply projecting the row points onto the column axis (Figure 2.2). The success of the reconstruction of the data from the biplot in this way is measured by the percentage of inertia explained by the map: in this case it is 95.6%, so the reconstruction has an error of only 4.4%. Table 2.5b reports the estimated values from the biplot of Figure 2.2, testifying to the high accuracy of the data reconstruction.

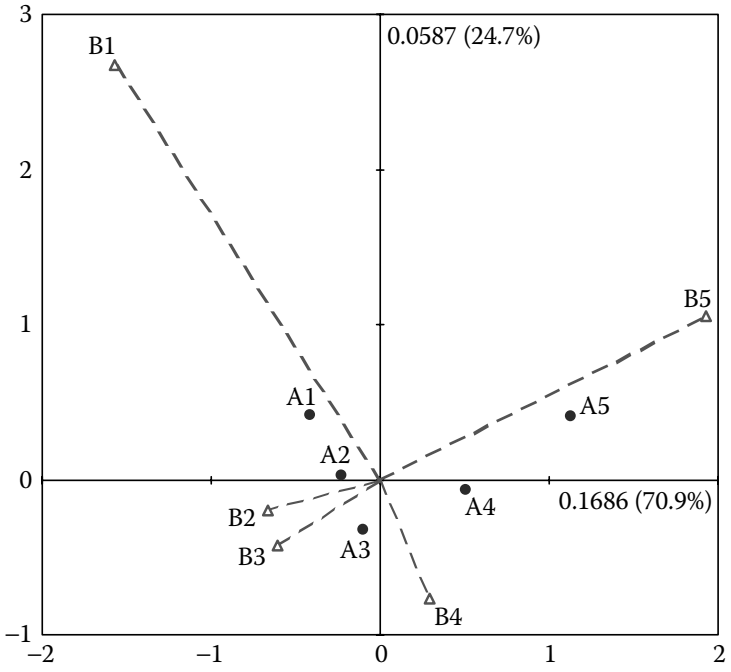

Figure 2.2 Asymmetric CA map of Table 2.1 showing biplot axes. The categories of the row points A1 to A5 can be projected onto each biplot axis to read off approximations of corresponding profile values. The accuracy of this approximation is as good as the percentage of inertia displayed, which is 95.6%; hence it is excellent.

Interpreting Figure 2.2 we can see, for example, a direction opposing the categories B2 and B3 pointing bottom left and category B5 top right. If we project A1, A2, and A3 onto this diagonal "dimension," it is clear that they project more or less at the same position, showing that their profile values on B2, B3, and B5 are similar, with profile values on B2 and B3 above average and those on B5 below average (the origin of the biplot always represents the average profile values exactly). This deduction from the map can be confirmed in Table 2.5a, and it is only for categories A4 and A5 that there are distinct changes along this direction, increasing in percentage response to B5 and decreasing on B2 and B3. Likewise, with respect to the other diagonal "dimension" from top left to bottom right, which opposes B1 and B4, we see that A3, A4, and A5 project at the same positions and thus are estimated to have similar profile values on B1 and B4. This can be mostly verified in Table 2.5a, the only exception being the profile of A5 on B4, which has an observed frequency much lower than the corresponding values for A3 and A4. This error of approximation would be part of the 4.4% unexplained inertia.

Thinking about the joint map in this way sheds light on the problematic aspects of the CA of the indicator matrix **Z** or the Burt matrix **C**. In the case of **Z**, it is futile to expect a good approximation of a matrix of zeros and ones in a two-dimensional map of points. Another measure of quality is needed; for example, one could deduce from a joint map the most likely set of responses of each case (row) and then count how many of these are correct predictions (see Gower 1993; Greenacre 1994). The situation is similar for the Burt matrix **C**: any attempt to approximate the diagonal matrices down the diagonal of the Burt matrix is clearly in conflict with the approximation of the more interesting and relevant contingency tables in the rest of the matrix. In both cases the percentages of inertia will be artificially low because of the structurally high-dimensional nature of the matrices being analyzed.

Figure 2.3 shows the CA of the Burt matrix of Table 2.3, which represents only 35.1% of the total inertia; yet its interpretation is clear: we can see the same pattern of association for questions A and B already seen in Figure 2.1, along with a similar pattern of association with question C. But the response categories for question D are not

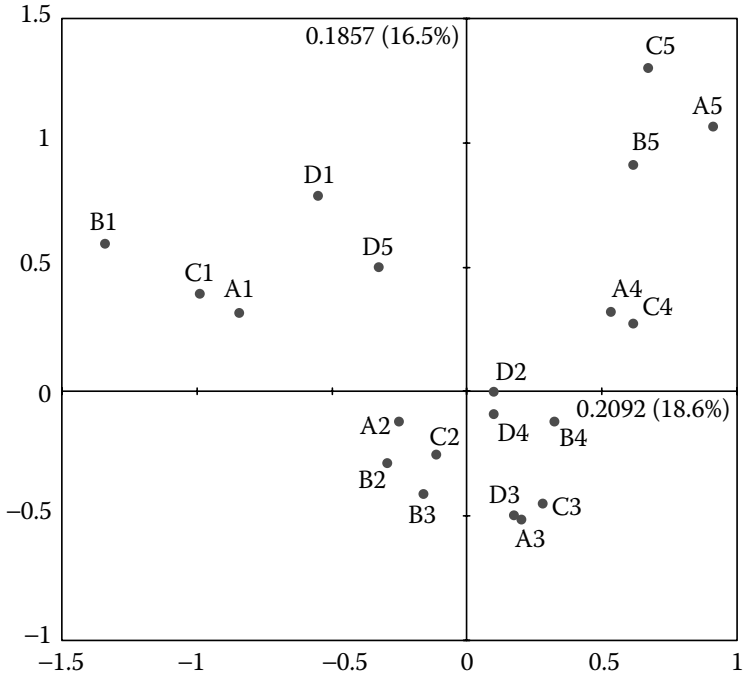

Figure 2.3 Multiple correspondence analysis map of the Burt matrix of Table 2.3. The percentage of explained inertia is 35.1%.

at all in line with the other three. The categories D1 and D5 of strong agreement and strong disagreement lie within the arch formed by the other questions, quite close together even though they are at opposite ends of the scale. This shows clearly the incompatibility of this question with the others.

Notice how different the scale of Figure 2.3 is compared with Figure 2.1, and how the points have been pulled outward in the analysis of the Burt matrix. Most of the problem of low percentages of inertia is due to this scale change, and this can be rectified by a simple scale adjustment of the solution. This is best explained after an account of joint correspondence analysis.

2.3.3 Joint correspondence analysis

As we have just seen, when applying CA to the Burt matrix, the diagonal submatrices on the "diagonal" of the block matrix **C** inflate both the chi-square distances between profiles and the total inertia by artificial amounts. In an attempt to generalize simple CA more naturally to more than two categorical variables, joint correspondence analysis (JCA) accounts for the variation in the "off-diagonal" tables of **C** only, ignoring the matrices on the block diagonal. Hence, in the two-variable case ($Q = 2$) when there is only one off-diagonal table, JCA is identical in all respects to simple CA (which is not the case for MCA of **Z** or **C**, which give different principal inertias).

The solution can no longer be obtained by a single application of the SVD, and various algorithms have been proposed by Greenacre (1988), Boik (1996), and Tateneni and Browne (2000). For example, Greenacre (1988) describes an alternating least-squares algorithm that treats the matrices on the block diagonal as missing values. The algorithm proceeds in the following steps:

1. Perform MCA by applying CA to the Burt matrix **C**, and choose the dimensionality S^* of the solution (e.g., $S^* = 2$ is the most typical).
2. Optionally perform an adjustment of the solution along each of the S^* dimensions to improve the approximation to the off-diagonal block matrices (see Section 2.3.4 below).
3. From the resulting map, reconstruct the values in the diagonal blocks of **C** in the same way as the data were reconstructed in the biplot, using the reconstruction formula (see appendix, Equation A.7). Replace the original values in the diagonal blocks by these estimates, calling this the modified Burt matrix $\mathbf{C}^*$.

4. Perform another CA on the resultant matrix **C*** with modified block diagonal.
5. Repeat steps 3 and 4, that is, substitute the diagonal blocks from the reconstructed values in the new solution, performing again the CA to obtain another solution, and so on, until the process converges. Convergence can be measured by the maximum absolute difference between the values substituted into the diagonal blocks at the present iteration and their corresponding values substituted during the previous iteration.

Figure 2.4 shows the results of a two-dimensional JCA applied to the Burt matrix of Table 2.3, where we have intentionally left the scale exactly as in Figure 2.3. Comparing these two figures we can see the high degree of similarity in the pattern of the response categories, but mainly a change in the scale, with the JCA map being reduced in scale on both axes, but especially on the second. Most of the properties of simple CA carry over to JCA, most importantly the reconstruction of profiles with respect to biplot axes (Greenacre 1993a: chapter 16).

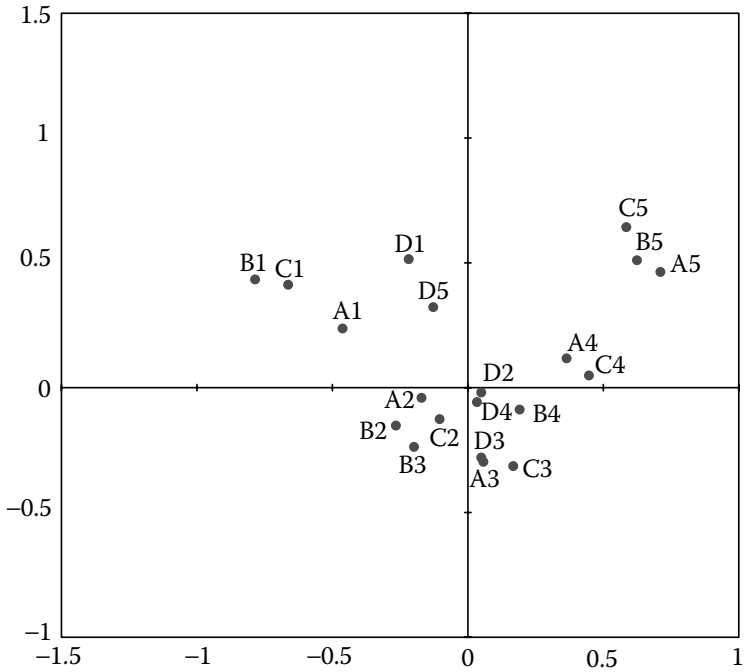

Figure 2.4 Joint correspondence analysis map of the Burt matrix of Table 2.3. The percentage of explained inertia is 85.7%.

Compared with regular CA and MCA, there are two aspects to remember in computing the percentage of inertia explained by the map. First, the percentage has to be calculated for the two dimensions of the solution together, not separately, since the dimensions in JCA are not nested. Second, in the final solution (the CA of the modified Burt matrix at the final iteration), the usual way of calculating the proportion of explained inertia involves the ratio between the sum of the first two principal inertias and the total inertia, but both the numerator and denominator of this sum include an amount due to the modified diagonal blocks, which are fitted exactly by the solution. This amount, which can be calculated in various ways (see the computational appendix, Section A.5), needs to be discounted from both the numerator and denominator to obtain the percentage of (off-diagonal) inertia explained. In this example, the percentage of inertia accounted for by the JCA map is 85.7%, much higher than the 35.1% explained in the MCA map based on the Burt matrix. The value of 85.7% appropriately measures the success of approximating the off-diagonal blocks relative to the total inertia of these blocks only, unaffected by the diagonal blocks. This would be the quality of the map considered as an MCA biplot as well: that is, express all six off-diagonal blocks as profiles (rows or columns profiles, in upper or lower triangle of the Burt matrix), then the quality of reconstructing these profiles as described in Section 2.3.2 would be 85.7%, and the error of reconstruction, or residual, would be 14.3%.

2.3.4 Adjustment of the inertias in MCA

Since the main difference between MCA and JCA in Figure 2.3 and Figure 2.4 is change in scale, it is possible to remedy partially the percentage-of-inertia problem in a regular MCA by a compromise between the MCA solution and the JCA objective by using simple scale readjustments of the MCA solution. In this approach the total inertia is measured (as in JCA) by the average inertia of all off-diagonal blocks of **C**, calculated either directly from the tables themselves or by adjusting the total inertia of **C** by removing the fixed contributions of the diagonal blocks as follows:

$$\text{average off-diagonal inertia} = \frac{Q}{Q-1}\left(\text{inertia}(\mathbf{C}) - \frac{J-Q}{Q^2}\right) \qquad (2.12)$$

Parts of inertia are then calculated from the principal inertias λ_s^2 of $\mathbf{C}$ (or from the principal inertias λ_s of $\mathbf{Z}$) as follows: for each $\lambda_s \geq 1/Q$, calculate the adjusted inertias:

$$\lambda_s^{\text{adj}} = \left(\frac{Q}{Q-1}\right)^2 \left(\lambda_s - \frac{1}{Q}\right)^2 \tag{2.13}$$

and then express these as percentages of Equation 2.12. Although these percentages underestimate those of a JCA, they dramatically improve the results of an MCA and are recommended in all applications of MCA. A further property of the adjusted principal inertias λ_s^{adj} is that they are identical to the principal inertias σ_s^2 of simple CA in the case of two categorical variables, where $Q = 2$: $\lambda_s^{\text{adj}} = 4(\lambda_s - 1/2)^2$, since we have shown earlier in Section 2.2.2 the relationship $\lambda_s = 1/2(1 + \sigma_s)$.

In our example the total inertia of $\mathbf{C}$ is equal to 1.1138, and the first seven principal inertias are such that $\lambda_s \geq 1/Q$, that is, $\lambda_s^2 \geq 1/Q^2 = 1/16$. The average off-diagonal inertia is equal to 0.17024, as shown in Table 2.6 along with the different possibilities for inertias and percentages of inertia. Thus what appears to be a percentage explained in two dimensions of 22.2% (= 11.4 + 10.8) in the analysis of the indicator matrix $\mathbf{Z}$, or 35.1% (= 18.6 + 16.5) in the analysis of the Burt matrix C, is shown to have a lower bound of 79.1% (= 44.9 + 34.2) when the principal inertias are adjusted. Compared with the adjusted MCA solution, the JCA solution for this example (Figure 2.4) gives an additional benefit of 6.4 percentage points in the explained inertia, with a percentage explained of 85.7%. We stress again that in JCA the solutions are not nested, so the percentages are reported for the whole solution (in this case, a two-dimensional one), not for individual dimensions.

We propose the adjusted solution as the one to be routinely reported; not only does it considerably improve the measure of fit, but it also removes the inconsistency about which of the two matrices to analyze, indicator or Burt. The adjusted solution is given in Figure 2.5 and has the same standard coordinates as Figure 2.3, but it uses the adjusted principal inertias to calculate the principal coordinates, leading to the improved quality of display. Again we have left the scale identical to Figure 2.3 and Figure 2.4 for purposes of comparison.

Benzécri (1979) has proposed the same adjusted inertias (Equation 2.13), but expresses them as percentages of their own sum over the dimensions s for which $\lambda_s \geq 1/Q$ (see an example of this in Chapter 5). This approach goes to the opposite extreme of giving an overly optimistic expression of explained inertia, since it explains 100% in the space of the dimensions for which $\lambda_s \geq 1/Q$ (there are six dimensions

Table 2.6 Eigenvalues (principal inertias) of indicator matrix **Z** and Burt matrix **C**, their percentages of inertia, the adjusted inertias, and their lower-bound estimates of the percentages of explained inertia for the off-diagonal tables of the Burt matrix.

Dimension	Eigenvalue of **Z**	Percentage Explained	Eigenvalue of **C**	Percentage Explained	Adjusted Eigenvalue	Percentage Explained
1	0.4574	11.4	0.2092	18.6	0.07646	44.9
2	0.4310	10.8	0.1857	16.5	0.05822	34.2
3	0.3219	8.0	0.1036	9.2	0.00920	5.4
4	0.3065	7.7	0.0939	8.3	0.00567	3.3
5	0.2757	6.9	0.0760	6.7	0.00117	0.7
6	0.2519	6.3	0.0635	5.6	0.00001	0.0

Note: The average off-diagonal inertia on which these latter percentages are based is equal to $\frac{4}{3}(1.12768 - \frac{16}{16}) = 0.17024$, where 1.12768 is the total inertia of the Burt matrix (i.e., the average of the inertias of all its submatrices, including those on the block diagonal). Note that the sum of explained inertia over these six dimensions must be less than 100% in the case of the adjusted eigenvalues because the percentages are lower-bound estimates and, in any case, the six-dimensional MCA solution does not fully explain all pairwise cross-tabulations.

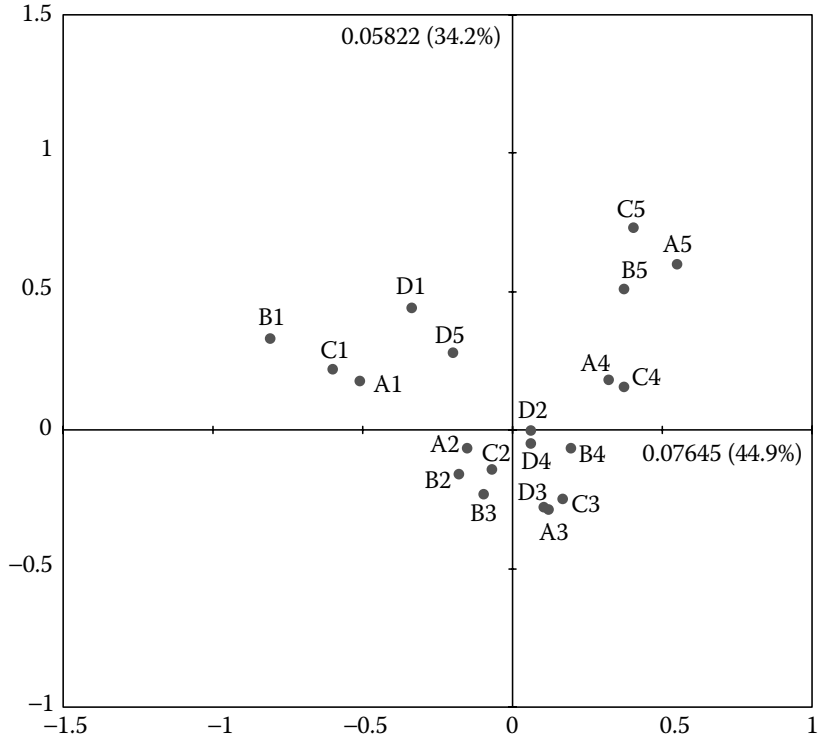

Figure 2.5 Multiple correspondence map of Table 2.3 with adjustment of the principal inertias (and thus the scale of the principal coordinates) along each dimension. The percentage of explained inertia is at least 79.1%. Note the change in scale compared with Figure 2.3.

in our example, as seen in Table 2.6) when in fact the data are not exactly reconstructed in the map.

2.4 Supplementary points

Up to now we have described three different ways to perform MCA:

Variant 1: CA of the indicator matrix
Variant 2: CA of the Burt matrix
Variant 3: An adjusted version of variants 1 or 2 that unifies and rectifies the scaling issue, giving a single solution irrespective of which matrix is considered, with highly improved measures of explained inertia

In all these variations, the standard coordinates of category points remain the same; only the principal inertias change. Furthermore, we have introduced an alternative method, JCA, that has a different solution from the above variants and that is analogous to least-squares

factor analysis in that it concentrates on between-variable associations only. In all of these cases it is possible to display supplementary points in the usual way to enrich the interpretation (see Chapter 5, where this aspect is discussed in depth for the MCA case). Here we define a way of displaying supplementary points that does not depend on the variant of the method used. In our illustration, we will consider three supplementary demographic variables: gender, age and education (full category descriptions are given in Section 2.1).

To motivate our approach we consider the case of the indicator matrix, where supplementary categories can be thought of as either row or column points. For example, male and female categories can be added as two supplementary-column dummy variables or as two supplementary rows containing the frequencies for males and for females across the response categories. These two alternatives are equivalent up to scale factors, as we now explain. To position a supplementary column category using the so-called transition, or barycentric, relationship between rows and columns (see, for example, Greenacre 1984 and Chapter 1 of this volume), we have to consider all the respondent points (rows) in standard coordinate positions in the map. Then any column category, active or supplementary, is situated (in principal coordinates) at the average of the respondents who fall into that category. Alternatively, to position a supplementary row, for example "male," we have to consider all the active column categories in standard coordinate positions; then the "male" row point will be at the weighted average of column points, using the profile of "male" across the active columns. Remember that the "male" frequency profile sums to 1 across the Q questions, so its position is an average of averages; for each question, the group "males" has an average position according to male frequencies across the categories of that particular question, and the final position of "male" is a simple average of these averages. We can show that the position of a supplementary point as a row is the same as the supplementary column dummy, but it is shrunk on each dimension by the corresponding singular value, that is, by the same scale factor that links principal to standard coordinates on each dimension (see Greenacre 1984: chapter 5.1 for a proof of this result). Thus a simple way to unify the representation of supplementary points in all situations would be to think of supplementary categories always as the averages of the principal coordinate positions of respondents, in which case both approaches will give exactly the same results.

Our proposal is thus the following: using the principal coordinates of respondent points, calculate average positions for supplementary categories, for example, the average position for male points, female points. Since it is only for the indicator matrix (variant 1 listed above) that we

(automatically) have respondent points in the CA, we need to make precise what we mean by respondent points in the case of the Burt matrix and the adjusted analysis (variants 2 and 3 above, respectively). In these last cases, respondent points are displayed, at least theoretically, as supplementary points, that is, as averages of the column categories, in standard coordinates, according to their respective response patterns. Because in all three variants of MCA these standard coordinates are identical, respondent points will have exactly the same positions in all three cases. Thus when we average their positions according to a supplementary variable, showing for example average male and average female points, the results will also be identical. But notice that, to obtain these average positions, we do not actually have to calculate all the original respondent points. The calculations can be made much more efficiently, thanks to transition relationships, by simply adding cross-tabulations as supplementary rows or columns. The following summarizes the calculations in each case, assuming that a supplementary variable is coded in indicator form as $\mathbf{Z}_s$, so that $\mathbf{Z}_s^\top\mathbf{Z}$ denotes the concatenated set of cross-tabulations of the supplementary variable with the Q active variables:

1. In the case of the indicator matrix $\mathbf{Z}$, we would already have the principal coordinates of the respondents, so we can either make the calculation of averages directly or add as supplementary row points the cross-tabulations $\mathbf{Z}_s^\top\mathbf{Z}$ of the supplementary variable with the active variables.
2. In the case of the Burt matrix $\mathbf{C}$, we do not need to calculate the positions of respondents (if required for other reasons, this would be done by adding $\mathbf{Z}$ as supplementary rows to the Burt matrix $\mathbf{C}$). Instead, we can simply add the cross-tabulations $\mathbf{Z}_s^\top\mathbf{Z}$ as supplementary rows (or $\mathbf{Z}^\top\mathbf{Z}_s$ as supplementary columns), which leads to the same positions for the supplementary categories as in variant 1.
3. In the case of the adjusted analysis, we do as in variant 2, since it is only *a posteriori* that we adjust the eigenvalues, and this adjustment affects only the positions of the principal coordinates of the active category points, not the supplementary categories that, we repeat, are defined as averages of the respondent points;
4. In the case of JCA, it is again a simple addition of supplementary rows or columns, as in variants 2 and 3, to the modified Burt matrix $\mathbf{C}^*$ at the final iteration of the algorithm.

Table 2.7 shows the cross-tabulations $\mathbf{Z}_s^\top\mathbf{Z}$, and Figure 2.6 shows the positions of the supplementary points in the adjusted MCA (variant 3,

Table 2.7 Supplementary rows added to the indicator matrix **Z**, Burt matrix **C**, or modified Burt matrix **C*** to represent the supplementary variable sex, age, and education. These are the cross-tabulations of the three supplementary variables with four active variables A, B, C, and D.

Supplementary Variables	Variable A					Variable B					Variable C					Variable D				
	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	C1	C2	C3	C4	C5	D1	D2	D3	D4	D5
Sex																				
Male	43	144	120	92	28	25	75	104	146	77	58	157	105	82	25	25	136	92	101	73
Female	76	178	84	86	20	46	99	101	135	63	94	159	92	72	27	35	96	110	125	78
Age [a]																				
age1	9	34	17	25	6	5	21	16	33	16	13	32	22	19	5	7	26	19	30	9
age2	33	68	45	51	13	15	30	52	76	37	35	60	57	43	15	10	46	48	56	50
age3	18	63	40	26	11	15	27	31	62	23	24	61	35	25	13	12	39	42	35	30
age4	19	45	47	27	8	13	29	34	41	29	24	57	33	26	6	10	39	38	38	21
age5	18	43	33	25	5	10	26	35	37	16	21	45	26	24	8	9	48	23	28	16
age6	22	69	22	24	5	13	41	37	32	19	35	61	24	17	5	12	34	32	39	25
Education [b]																				
edu1	7	15	7	8	1	6	10	11	7	4	7	8	11	12	0	5	9	12	8	4
edu2	59	155	84	68	12	34	93	95	112	44	79	156	73	54	16	28	110	90	95	55
edu3	29	84	65	54	10	19	47	55	82	39	34	90	68	36	14	11	69	58	63	41
edu4	11	27	18	26	12	6	12	18	37	21	18	24	18	28	6	7	13	23	29	22
edu5	5	20	11	8	5	4	5	11	16	13	6	14	14	9	6	4	12	7	14	12
edu6	8	21	19	14	8	2	7	15	27	19	8	24	13	15	10	5	19	12	17	17

[a] age1: 16–24 years; age2: 25–34 years; age3: 35–44 years; age4: 45–54 years; age5: 55–64 years; age6: 65 years and older.

[b] edu1: no or some primary education; edu2: primary education completed; edu3: some secondary education; edu4: secondary education completed; edu5: some tertiary education; edu6: tertiary education completed.

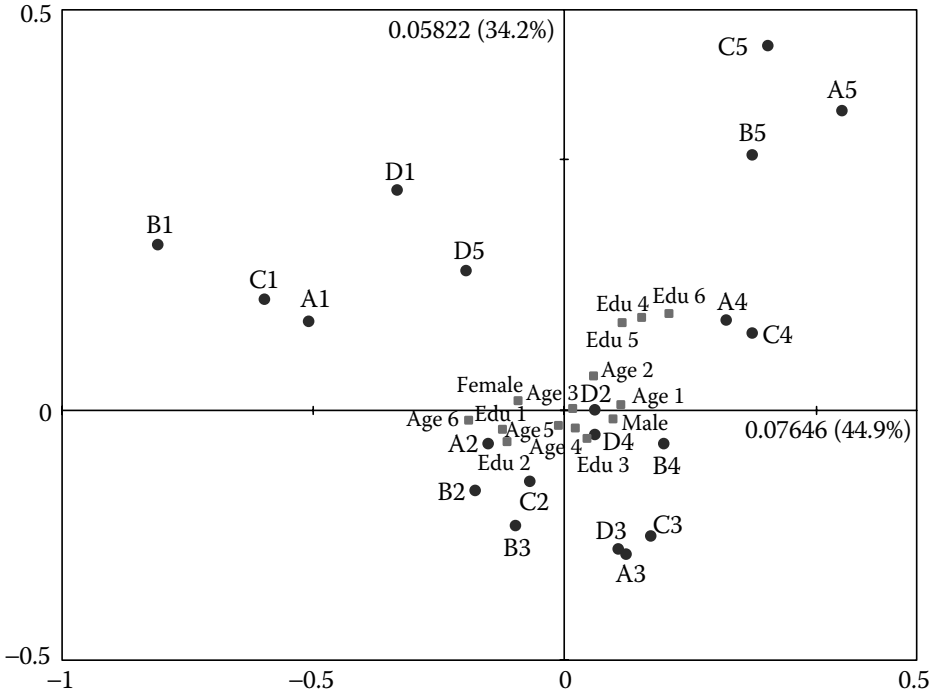

Figure 2.6 Adjusted MCA solution, showing positions of supplementary categories of sex, age, and education.

that is, the supplementary points are superimposed on Figure 2.5). Here we can see that the age groups and education groups show a horizontal trend, with younger respondents on the right moving over to older respondents on the left, and lower education groups on the left moving over to higher-education groups on the right. In addition, we can see that the three upper education groups (from secondary education completed upward) separate out at the right toward the strong disagreement poles of the questions, indicating that they are particularly strongly in favor of science. We also find the average male point on the right-hand side and average female point on the left-hand side. Remember that these supplementary variables have been added separately to the map and not in combination, that is, the fact that male is on the right and higher-education groups are on the right does not imply that it is only higher-educated males that are more favorable toward science. To see the positions of higher-educated females, for example, interactive coding of the demographic variables would have to be performed. This can be done, for example, by coding the six education groups for males and for females separately, giving 12 combinations of gender and education, each represented as a separate supplementary point.

2.5 Discussion and conclusions

We have shown that extending CA of two variables to the case of several variables is not a simple issue, especially in the geometric case. As explained in Chapter 1, simple CA is typically applied to situations where two different types of variables are cross-tabulated, for example, country of residence and a response to a survey question. MCA is applied to one set of variables, preferably all with the same response scales, that revolve around a particular issue, and where we are interested in the association patterns among the variables. Putting this in another way, in simple CA we are interested in associations between two variables or between two sets of variables, while in MCA we are interested in associations within a set of variables.

Although the geometric concepts of simple CA do not carry over easily to the multiple-variable case, adjustment of the principal inertias and alternative methods, such as JCA, partially rectify the situation. Because MCA has attractive properties of optimality of scale values (thanks to achieving maximum intercorrelation and thus maximum reliability in terms of Cronbach's alpha), the compromise offered by the adjusted MCA solution is the most sensible one and the one that we recommend. The adjustment, described in Section 2.3.4, is easy to calculate and simply changes the scale on each dimension of the map to best approximate the two tables of association between pairs of variables, leaving everything else in the solution intact. Thanks to this adjustment we obtain estimates of the explained inertias that are much closer to the true values than the pessimistic values obtained in MCA. For this reason we propose the adjusted MCA solution as the one to be used as a matter of course in all (graphical) MCA applications. The adjusted solution also has the nice property that it is identical to simple CA of a cross-tabulation in the case of two variables. JCA also perfectly reproduces simple CA in the two-variable case, since it is also focused exclusively on the single off-diagonal cross-tabulation. JCA has the additional advantage in the multivariable case of optimizing the fit to all of the off-diagonal cross-tabulations of interest.

Categorical supplementary points can be added to any of the variants of MCA, as well as JCA, as averages of respondents that fall into the corresponding categories. This amounts to simply adding the cross-tabulations of the supplementary variables with the active variables as supplementary rows or columns. The JCA and adjusted MCA solutions have the advantage that the active points are reduced in scale compared with the solutions for the indicator and Burt matrices, thus

leading to a greater relative dispersion of the supplementary points in the joint map of the active and supplementary points.

We have not dealt with the important topic of imposing constraints on the MCA solution, such as linear or order constraints. Chapter 4 gives a comprehensive and up-to-date treatment of these issues.

Software note

The analyses of this chapter were performed using XLSTAT (www.xlstat.com), distributed by Addinsoft, and the R functions for CA, MCA, and JCA, written by Oleg Nenadić. Both implementations, described in the appendix of this book, include the adjustment of inertias described in this chapter.