17

Correspondence Analysis

17.1 Introduction

Correspondence analysis is an exploratory multivariate technique for simultaneously displaying scores representing the row categories and column categories of a two-way contingency table as the coordinates of points in a low-dimensional (two- or possibly three-dimensional) vector space. The objective is to clarify the relationship between the row and column variates of the table and to discover a low-dimensional explanation for possible deviations from independence of those variates. The methodology has its own nomenclature, and its approach is decidedly geometric, especially for interpreting the resulting graphical displays.

For two-way contingency tables, correspondence analysis is known as *simple* correspondence analysis. For three-way and higher contingency tables, it is known as *multiple* correspondence analysis. Variants of correspondence analysis are *dual* (or *optimal*) *scaling*, *reciprocal averaging*, *perceptual mapping*, and *social space analysis*. In general, correspondence analysis is applicable when the variates are discrete with many categories and, hence, is well-suited for analyzing large contingency tables. It can also be used for continuous variates, such as age, which can be segmented into a finite

A.J. Izenman, *Modern Multivariate Statistical Techniques*, Springer Texts in Statistics,
DOI 10.1007/978-0-387-78189-1_17,

number of ranges, but discretization of a continuous variate usually entails some loss of information.

17.1.1 Example: Shoplifting in The Netherlands

These data[1] were taken from van der Heijden, de Falguerolles, and de Leeuw (1989). It is a three-way contingency table of 33,101 individuals, classified by gender and age, who were suspected of stealing specific goods in The Netherlands in 1978 and 1979. The data were obtained from a survey of about 350 Dutch stores and big retail shops. Cases in which shoplifting consisted of more than a single type of good, or in which more than one person was suspected, were omitted from the study. Age was divided into nine nonoverlapping categories, and shoplifted items were classified into 13 types of goods.

For this example, we arranged the original $2 \times 9 \times 13$ three-way contingency table into a $(2 \times 9) \times 13$ two-way contingency table in which gender has been introduced as separate sets of nine male and nine female rows of ages. The ages were coded by groups: < 12 (1 for boys and 10 for girls), 12–14 (2 and 11), 15–17 (3 and 12), 18–20 (4 and 13), 21–29 (5 and 14), 30–39 (6 and 15), 40–49 (7 and 16), 50–64 (8 and 17), and 65+ (9 and 18). The graphical display from the resulting correspondence analysis is given in Figure 17.1.

We can make the following observations from Figure 17.1. First, points representing males and females are well-separated at each age group, suggesting that their shoplifting profiles are quite different. Second, for both males and females, the age category points are clearly ordered from younger than 12 years old on the left-hand side to older than 65 on the right-hand side, with both sets of points doubling back toward the left after 30 years of age. Third, while there are larger distances between males at the younger age groups than those at older age groups, suggesting that shoplifting behavior changes substantially more for younger than for older males, the distances between female age groups are largest at both the younger and older ages (and, hence, more rapidly changing shoplifting behavior), with smaller distances appearing in the middle age groups (18–49).

The configuration of points in Figure 17.1 also tempts us to identify column points (which types of goods are shoplifted more than average) with nearby row points (age groups), possibly leading to the identification of significant age $\times$ goods interactions. Although interrow distances and intercolumn distances can be compared, row-to-column distances are undefined and, therefore, are essentially meaningless (see, e.g., Greenacre and Hastie,

[1]The contingency table can be downloaded from the book's website.

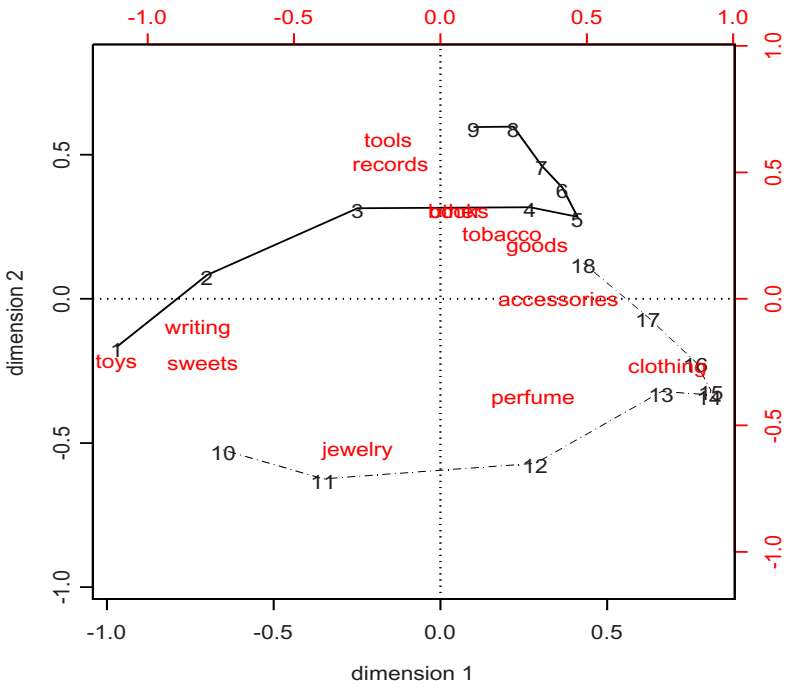

FIGURE 17.1. *Correspondence map for the shoplifting example. The red words are the items shoplifted, the points joined by a solid line represent the progression in male ages (1–9), and the points joined by a dotted line represent the progression in female ages (10–18).*

1987). In other words, row points should not be associated with neighboring column points (and vice versa). Using row percentages obtained from the contingency table, we summarize in Table 17.1 the types of goods most often shoplifted by males and by females at each of the different age groups. In the light of the above comments, it is perhaps instructive for the reader to compare Figure 17.1 with Table 17.1.

17.2 Simple Correspondence Analysis

17.2.1 Two-Way Contingency Tables

Categorical data are count data that are collected in a contingency table $\mathbf{N}$. A *two-way* $(r \times s)$ *contingency table* with r rows (labelled $A_1, A_2, \ldots, A_r$) and s columns (labelled $B_1, B_2, \ldots, B_s$) has rs *cells*. The ijth cell has entry

TABLE 17.1. *Types of goods most often shoplifted by males and by females at each age group, as derived from the two-way contingency table of the example. Superscripts show the percentages of that type of good stolen for that age group and gender. Also listed in parentheses for each age group and gender are those goods that are stolen more than 20% of the time.*

Age	Males	Females
< 12	Toys $^{26.2}$ (writing materials $^{23.5}$)	Writing materials $^{23.8}$
12–14	Writing materials $^{25.1}$	Jewelry $^{26.5}$
15–17	Writing materials $^{14.8}$	Clothing $^{32.3}$ (jewelry $^{20.5}$)
18–20	Clothing $^{22.8}$	Clothing $^{45.4}$
21–29	Clothing $^{27.3}$	Clothing $^{55.8}$
30–39	Clothing $^{25.9}$	Clothing $^{57.2}$
40–49	Clothing $^{21.7}$	Clothing $^{51.7}$
50–64	Hobbies, tools $^{22.6}$	Clothing $^{39.4}$
65+	Provisions, tobacco $^{27.3}$ (hobbies, tools $^{20.9}$)	Provisions, tobacco $^{30.1}$ (clothing $^{24.2}$)

n_{ij}, representing the observed frequency in row category A_i and column category B_j, $i = 1, 2, \ldots, r$, $j = 1, 2, \ldots, s$. The ith *marginal row total* is $n_{i+} = \sum_{j=1}^{s} n_{ij}$, $i = 1, 2, \ldots, r$, and the jth *marginal column total* is $n_{+j} = \sum_{i=1}^{r} n_{ij}$, $j = 1, 2, \ldots, s$. If $n = \sum_{i=1}^{r} \sum_{j=1}^{s} n_{ij}$ individuals are classified by row and column categories, then Table 17.2, which is also called a *correspondence table*, shows the cell frequencies, marginal totals, and total sample size. For interpretation purposes, it is important to distinguish when the n individuals are randomly selected from some very large population or when they actually constitute the entire population of interest.

We denote by π_{ij} the probability that an individual has the properties A_i and B_j, $i = 1, 2, \ldots, r$, $j = 1, 2, \ldots, s$. In the event that the row variable A is independent of the column variable B, we have that $\pi_{ij} = \pi_{i+}\pi_{+j}$, where $\pi_{i+} = \sum_j \pi_{ij}$ and $\pi_{+j} = \sum_i \pi_{ij}$, for all $i = 1, 2, \ldots, r$ and $j = 1, 2, \ldots, s$. We are generally interested in assessing whether A and B are indeed independent variables. Such a question can alternatively be posed in terms of *homogeneity* of the row or column probability distributions; that is, whether all the rows have the same probability distributions across columns, or, equivalently, whether all the columns have the same probability distributions across rows.

17.2.2 Row and Column Dummy Variables

For a two-way $(r \times s)$ contingency table, we are interested in the relationship between the row categories and the column categories. We define two sets of dummy variates, an r-vector $\mathbf{X}_i = (X_{i1}, \cdots, X_{ir})^\tau$ to indi-

TABLE 17.2. *Two-way contingency table, showing observed cell frequencies, row and column marginal totals, and total sample size.*

	Column Variable						
Row Variable	B_1	B_2	$\cdots$	B_j	$\cdots$	B_s	Row Total
A_1	n_{11}	n_{12}	$\cdots$	n_{1j}	$\cdots$	n_{1s}	n_{1+}
A_2	n_{21}	n_{22}	$\cdots$	n_{2j}	$\cdots$	n_{2s}	n_{2+}
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
A_i	n_{i1}	n_{i2}	$\cdots$	n_{ij}	$\cdots$	n_{is}	n_{i+}
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
A_r	n_{r1}	n_{r2}	$\cdots$	n_{rj}	$\cdots$	n_{rs}	n_{r+}
Column total	n_{+1}	n_{+2}	$\cdots$	n_{+j}	$\cdots$	n_{+s}	n

cate which of the n observations fall into the ith row, and an s-vector $\mathbf{Y}_j = (Y_{1j}, \cdots, Y_{sj})^\tau$ to indicate which of the n observations fall into the jth column; that is,

$$X_{ij} = \begin{cases} 1, & \text{if the } j\text{th individual belongs to } A_i \\ 0, & \text{otherwise} \end{cases}$$

$$Y_{ij} = \begin{cases} 1, & \text{if the } i\text{th individual belongs to } B_j \\ 0, & \text{otherwise} \end{cases}$$

$i = 1, 2, \ldots, r$, $j = 1, 2, \ldots, s$. These indicator vectors can be collected into two matrices, an $(r \times n)$-matrix $\mathcal{X} = (x_{ij})$ and an $(s \times n)$-matrix $\mathcal{Y} = (y_{ij})$. Note that even though both $\mathcal{X}$ and $\mathcal{Y}$ are defined by the specific distribution of cell frequencies in the contingency table, it turns out that the summary information will be the same as if we assume, for convenience, that $\mathcal{X}$ and $\mathcal{Y}$ are given by

$$\underset{r\times n}{\mathcal{X}} = \begin{pmatrix} 1 & \cdots & 1 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 1 & \cdots & 1 \end{pmatrix}, \tag{17.1}$$

$$\underset{s\times n}{\mathcal{Y}} = \begin{pmatrix} 1 & \cdots & 1 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 1 & \cdots & 1 \end{pmatrix}, \tag{17.2}$$

respectively.

Matrices derived from $\mathcal{X}$ and $\mathcal{Y}$ reproduce the *observed cell frequencies* and their marginal totals. The $(r \times s)$-matrix $\mathcal{X}\mathcal{Y}^\tau$ reproduces the observed

cell frequencies of the contingency table,

$$\mathcal{X}\mathcal{Y}^\tau = \begin{pmatrix} n_{11} & n_{12} & \dots & n_{1s} \\ n_{21} & n_{22} & \dots & n_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ n_{r1} & n_{r2} & \dots & n_{rs} \end{pmatrix} = \mathbf{N}. \tag{17.3}$$

The $(r \times r)$ matrix $\mathcal{X}\mathcal{X}^\tau$ and the $(s \times s)$ matrix $\mathcal{Y}\mathcal{Y}^\tau$ are both diagonal, $\mathcal{X}\mathcal{X}^\tau$ having as diagonal entries the r marginal row totals and $\mathcal{Y}\mathcal{Y}^\tau$ having as diagonal entries the s marginal column totals,

$$\mathcal{X}\mathcal{X}^\tau = \text{diag}\{n_{1+}, \cdots, n_{r+}\}, \tag{17.4}$$

$$\mathcal{Y}\mathcal{Y}^\tau = \text{diag}\{n_{+1}, \cdots, n_{+s}\}. \tag{17.5}$$

Collecting (17.3), (17.4), and (17.5) together, we can form the $(r+s)\times(r+s)$ block matrix,

$$\begin{pmatrix} \mathcal{X} \\ \mathcal{Y} \end{pmatrix} \begin{pmatrix} \mathcal{X} \\ \mathcal{Y} \end{pmatrix}^\tau = \begin{pmatrix} n\mathbf{D}_r & \mathbf{N} \\ \mathbf{N}^\tau & n\mathbf{D}_c \end{pmatrix}, \tag{17.6}$$

where

$$\mathbf{D}_r = n^{-1}\mathcal{X}\mathcal{X} = \text{diag}\{n_{1+}/n, \dots, n_{r+}/n\}, \tag{17.7}$$

$$\mathbf{D}_c = n^{-1}\mathcal{Y}\mathcal{Y}^\tau = \text{diag}\{n_{+1}/n, \dots, n_{+s}/n\}. \tag{17.8}$$

The matrix (17.6) is known as a *Burt matrix* (Burt, 1950) for a two-way contingency table. It is nonnegative definite and symmetric and is the analogue in the discrete case (after dividing through by n) of the sample covariance matrix of two sets of continuous variates.

17.2.3 Example: Hair Color and Eye Color

This classic two-way contingency table $\mathbf{N}$ with $r = 4$ and $s = 5$ (see Table 17.3) was analyzed by R.A. Fisher (1940) and others. It relates to data on hair color and eye color of a sample of 5,387 schoolchildren from Caithness, Scotland. It is given as a (4×5)-matrix by:

$$\mathbf{N} = \mathcal{X}\mathcal{Y}^\tau = \begin{pmatrix} 326 & 38 & 241 & 110 & 3 \\ 688 & 116 & 584 & 188 & 4 \\ 343 & 84 & 909 & 412 & 26 \\ 98 & 48 & 403 & 681 & 85 \end{pmatrix}.$$

The matrices $\mathcal{X}\mathcal{X}^\tau$ and $\mathcal{Y}\mathcal{Y}^\tau$ are given by:

$$\mathcal{X}\mathcal{X}^\tau = \begin{pmatrix} 718 & 0 & 0 & 0 \\ 0 & 1580 & 0 & 0 \\ 0 & 0 & 1774 & 0 \\ 0 & 0 & 0 & 1315 \end{pmatrix}$$

TABLE 17.3. *Relationship of Hair Color to Eye Color of Scottish Schoolchildren.*

	Hair Color					
Eye Color	Fair	Red	Medium	Dark	Black	Totals
Blue	326	38	241	110	3	718
Light	688	116	584	188	4	1,580
Medium	343	84	909	412	26	1,774
Dark	98	48	403	681	85	1,315
Totals	1,455	286	2,137	1,391	118	5,387

$$\mathcal{Y}\mathcal{Y}^\tau = \begin{pmatrix} 1455 & 0 & 0 & 0 & 0 \\ 0 & 286 & 0 & 0 & 0 \\ 0 & 0 & 2137 & 0 & 0 \\ 0 & 0 & 0 & 1391 & 0 \\ 0 & 0 & 0 & 0 & 118 \end{pmatrix},$$

respectively. The matrices $\mathbf{D}_r$ and $\mathbf{D}_c$ are obtained by dividing both $\mathcal{X}\mathcal{X}^\tau$ and $\mathcal{Y}\mathcal{Y}^\tau$ by $n = 5,387$:

$$\mathbf{D}_r = \begin{pmatrix} 0.1333 & 0 & 0 & 0 \\ 0 & 0.2933 & 0 & 0 \\ 0 & 0 & 0.3293 & 0 \\ 0 & 0 & 0 & 0.2441 \end{pmatrix}$$

$$\mathbf{D}_c = \begin{pmatrix} 0.2701 & 0 & 0 & 0 & 0 \\ 0 & 0.0531 & 0 & 0 & 0 \\ 0 & 0 & 0.3967 & 0 & 0 \\ 0 & 0 & 0 & 0.2582 & 0 \\ 0 & 0 & 0 & 0 & 0.0219 \end{pmatrix}.$$

17.2.4 Profiles, Masses, and Centroids

The $(r \times s)$-matrix

$$\mathbf{P} = n^{-1}\mathbf{N} \tag{17.9}$$

converts the contingency table $\mathbf{N}$ into a *correspondence matrix*. See Table 17.4. If the n individuals constitute a random sample, the entry, $p_{ij} = n_{ij}/n$, in the ith row and jth column of $\mathbf{P}$ can be characterized as either the uniformly minimum variance unbiased (UMVU) estimator or the maximum likelihood (ML) estimator of π_{ij}. For the hair-color/eye-color example,

$$\mathbf{P} = \begin{pmatrix} 0.0605 & 0.0071 & 0.0447 & 0.0204 & 0.0006 \\ 0.1277 & 0.0215 & 0.1084 & 0.0349 & 0.0007 \\ 0.0637 & 0.0156 & 0.1687 & 0.0765 & 0.0048 \\ 0.0182 & 0.0089 & 0.0748 & 0.1264 & 0.0158 \end{pmatrix}.$$

TABLE 17.4. *Correspondence matrix, showing observed cell relative frequencies* $\mathbf{P}$ *(*$p_{ij} = n_{ij}/n$*), row marginal totals* $\mathbf{r}$ *(*$p_{i+} = n_{i+}/n$*), and column marginal totals* $\mathbf{c}^\tau$ *(*$p_{+j} = n_{+j}/n$*)*

	Column Variable						
Row Variable	B_1	B_2	$\cdots$	B_j	$\cdots$	B_s	Row Total
A_1	p_{11}	p_{12}	$\cdots$	p_{1j}	$\cdots$	p_{1s}	p_{1+}
A_2	p_{21}	p_{22}	$\cdots$	p_{2j}	$\cdots$	p_{2s}	p_{2+}
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
A_i	p_{i1}	p_{i2}	$\cdots$	p_{ij}	$\cdots$	p_{is}	p_{i+}
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
A_r	p_{r1}	p_{r2}	$\cdots$	p_{rj}	$\cdots$	p_{rs}	p_{r+}
Column total	p_{+1}	p_{+2}	$\cdots$	p_{+j}	$\cdots$	p_{+s}	1

The row totals and column totals of $\mathbf{P}$ are given by the diagonal elements of $\mathbf{D}_r$ and $\mathbf{D}_c$, respectively.

The $(r \times s)$-matrix $\mathbf{P}_r$ of *row profiles* of $\mathbf{N}$ (or $\mathbf{P}$) consists of the rows of $\mathbf{N}$ divided by their appropriate row totals (e.g., n_{ij}/n_{i+}, which, under random sampling, can be characterized as either the UMVU or ML estimator of π_{ij}/π_{i+}, the conditional probability that an individual has property B_j given that he or she has property A_i), and can be computed as the regression coefficient matrix of $\mathcal{Y}$ on $\mathcal{X}$; that is,

$$\mathbf{P}_r = (\mathcal{X}\mathcal{X}^\tau)^{-1}\mathcal{X}\mathcal{Y}^\tau = \mathbf{D}_r^{-1}\mathbf{P} = \begin{pmatrix} \mathbf{a}_1^\tau \\ \vdots \\ \mathbf{a}_r^\tau \end{pmatrix}, \tag{17.10}$$

where

$$\mathbf{a}_i^\tau = \left(\frac{n_{i1}}{n_{i+}}, \cdots, \frac{n_{is}}{n_{i+}} \right) \tag{17.11}$$

is the ith row profile, $i = 1, 2, \ldots, r$. For the hair-color/eye-color example,

$$\mathbf{P}_r = \begin{pmatrix} 0.4540 & 0.0529 & 0.3357 & 0.1532 & 0.0042 \\ 0.4354 & 0.0734 & 0.3696 & 0.1190 & 0.0025 \\ 0.1933 & 0.0474 & 0.5124 & 0.2322 & 0.0147 \\ 0.0745 & 0.0365 & 0.3065 & 0.5179 & 0.0646 \end{pmatrix}.$$

Similarly, the $(s \times r)$-matrix $\mathbf{P}_c$ of *column profiles* of $\mathbf{N}$ (or $\mathbf{P}$) consists of the columns of $\mathbf{N}$ divided by their appropriate column totals (e.g., n_{ij}/n_{+j}, which, under random sampling, can be characterized as the UMVU or ML estimator of π_{ij}/π_{+j}, the conditional probability that an individual has property A_i given that he or she has property B_j), and computed as the

regression coefficient matrix of $\mathcal{X}$ on $\mathcal{Y}$; that is,

$$\mathbf{P}_c = (\mathcal{Y}\mathcal{Y}^\tau)^{-1}\mathcal{Y}\mathcal{X}^\tau = \mathbf{D}_c^{-1}\mathbf{P}^\tau = \begin{pmatrix} \mathbf{b}_1^\tau \\ \vdots \\ \mathbf{b}_s^\tau \end{pmatrix}, \tag{17.12}$$

where

$$\mathbf{b}_j^\tau = \left(\frac{n_{1j}}{n_{+j}}, \cdots, \frac{n_{rj}}{n_{+j}} \right) \tag{17.13}$$

is the jth column profile, $j = 1, 2, \ldots, s$. For the hair-color/eye-color example,

$$\mathbf{P}_c = \begin{pmatrix} 0.2241 & 0.4729 & 0.2357 & 0.0674 \\ 0.1329 & 0.4056 & 0.2937 & 0.1678 \\ 0.1128 & 0.2733 & 0.4254 & 0.1886 \\ 0.0791 & 0.1352 & 0.2962 & 0.4896 \\ 0.0254 & 0.0339 & 0.2203 & 0.7203 \end{pmatrix}.$$

The row means of the contingency table $\mathbf{N}$ are the row sums of $\mathbf{P}$,

$$\mathbf{P}\mathbf{1}_s = \begin{pmatrix} \bar{X}_1 \\ \vdots \\ \bar{X}_r \end{pmatrix} = \begin{pmatrix} n_{1+}/n \\ \vdots \\ n_{r+}/n \end{pmatrix} = \begin{pmatrix} p_{1+} \\ \vdots \\ p_{r+} \end{pmatrix} = \mathbf{r}, \tag{17.14}$$

and the column means of $\mathbf{N}$ are the column sums of $\mathbf{P}$ (or row sums of $\mathbf{P}^\tau$),

$$\mathbf{P}^\tau\mathbf{1}_r = \begin{pmatrix} \bar{Y}_1 \\ \vdots \\ \bar{Y}_s \end{pmatrix} = \begin{pmatrix} n_{+1}/n \\ \vdots \\ n_{+s}/n \end{pmatrix} = \begin{pmatrix} p_{+1} \\ \vdots \\ p_{+s} \end{pmatrix} = \mathbf{c}, \tag{17.15}$$

where $\mathbf{1}_a$ denotes an a-vector each of whose entries is 1. The vectors $\mathbf{r}$ and $\mathbf{c}$ can be formed from the diagonal elements of $\mathbf{D}_r$ and $\mathbf{D}_c$, respectively; that is, $\mathbf{D}_r = \text{diag}\{\mathbf{r}\}$ and $\mathbf{D}_c = \text{diag}\{\mathbf{c}\}$. For the hair-color/eye-color example,

$$\mathbf{r} = \begin{pmatrix} 0.1333 \\ 0.2933 \\ 0.3293 \\ 0.2441 \end{pmatrix}, \quad \mathbf{c} = \begin{pmatrix} 0.2701 \\ 0.0531 \\ 0.3967 \\ 0.2582 \\ 0.0219 \end{pmatrix}.$$

Powers of these diagonal matrices are given by $\mathbf{D}_r^\alpha = \text{diag}\{\mathbf{r}^\alpha\}$ and $\mathbf{D}_c^\alpha = \text{diag}\{\mathbf{c}^\alpha\}$, where $\mathbf{r}^\alpha$ and $\mathbf{c}^\alpha$ are the column vectors (17.14) and (17.15), respectively, with each entry raised to the αth power. In this chapter, we will be interested in situations where $\alpha = -\frac{1}{2}$ or -1.

The ith element, $p_{i+} = n_{i+}/n$, of the r-vector $\mathbf{r}$ is called the ith *row mass* and, under random sampling, is an estimate of the unconditional

probability, π_{i+}, of belonging to A_i. Similarly, the jth element, $p_{+j} = n_{+j}/n$, of the s-vector $\mathbf{c}$ is called the jth *column mass* and is an estimate of the unconditional probability, π_{+j}, of belonging to B_j. In correspondence analysis, $\mathbf{r}$ is called the *average column profile* and $\mathbf{c}$ is called the *average row profile* of the contingency table. The vector $\mathbf{c}$ is also referred to as the *row centroid* because it can be expressed as the weighted average of the row profiles, namely,

$$\mathbf{c} = \sum_{i=1}^{r} p_{i+}\mathbf{a}_i, \tag{17.16}$$

where the weights are the row masses. Similarly, the vector $\mathbf{r}$ is referred to as the *column centroid* because it can be expressed as the weighted average of the column profiles, namely,

$$\mathbf{r} = \sum_{j=1}^{s} p_{+j}\mathbf{b}_j, \tag{17.17}$$

where the weights are the column masses. It is not difficult to show that the relationship between $\mathbf{r}$ and $\mathbf{c}$ is given by $\mathbf{r} = \mathbf{P}^\tau \mathbf{D}_c^{-1}\mathbf{c}$ and $\mathbf{c} = \mathbf{P}^\tau \mathbf{D}_r^{-1}\mathbf{r}$.

17.2.5 Chi-squared Distances

In correspondence analysis, it is important to be able to visualize distances between different row profiles (i.e., rows of $\mathbf{P}_r$) or between different column profiles (i.e., rows of $\mathbf{P}_c$). To do this, we use the *chi-squared metric* as a measure of distance.

Row Distances

Consider the ith and i'th row profiles, $\mathbf{a}_i$ and $\mathbf{a}_{i'}$, respectively. We will need the fact that $\mathbf{a}_i - \mathbf{a}_{i'}$ is an s-vector whose jth entry is $n_{ij}/n_{i+} - n_{i'j}/n_{i'+}$. The squared distance between $\mathbf{a}_i$ and $\mathbf{a}_{i'}$ is defined as the quadratic form,

$$d^2(\mathbf{a}_i, \mathbf{a}_{i'}) \equiv (\mathbf{a}_i - \mathbf{a}_{i'})^\tau \mathbf{D}_c^{-1}(\mathbf{a}_i - \mathbf{a}_{i'}) \tag{17.18}$$

$$= \sum_{j=1}^{s} \frac{n}{n_{+j}} \left(\frac{n_{ij}}{n_{i+}} - \frac{n_{i'j}}{n_{i'+}} \right)^2. \tag{17.19}$$

We see from (17.19) that the jth column mass, n_{+j}/n, enters the squared distance between row profiles $\mathbf{a}_i$ and $\mathbf{a}_{i'}$ as an inverse element of the jth term in the sum. It follows that those categories having fewer observations contribute more to the inter-row profile distances.

Recall that $\mathbf{c}$ is the row centroid. The $(r \times s)$-matrix of *centered row profiles* $\mathbf{P}_r - \mathbf{1}_r\mathbf{c}^\tau$, where $\mathbf{P}_r = \mathbf{D}_r^{-1}\mathbf{P}$, has ith row $(\mathbf{a}_i - \mathbf{c})^\tau$, with jth

entry $n_{i+}^{-1}(n_{ij} - n_{i+}n_{+j}/n)$, $i = 1, 2, \ldots, r$, $j = 1, 2, \ldots, s$. The squared χ^2-distance between $\mathbf{a}_i$ and $\mathbf{c}$ is, therefore,

$$\begin{aligned} d^2(\mathbf{a}_i, \mathbf{c}) &= (\mathbf{a}_i - \mathbf{c})^\tau \mathbf{D}_c^{-1} (\mathbf{a}_i - \mathbf{c}) \\ &= \frac{1}{n_{i+}} \sum_{j=1}^{s} \frac{n}{n_{i+}n_{+j}} \left(n_{ij} - \frac{n_{i+}n_{+j}}{n} \right)^2. \end{aligned} \tag{17.20}$$

Summing (17.20) over all row profiles yields

$$n \sum_{i=1}^{r} p_{i+} d^2(\mathbf{a}_i, \mathbf{c}) = \sum_{i=1}^{r} \sum_{j=1}^{s} \left(n_{ij} - \frac{n_{i+}n_{+j}}{n} \right)^2 / \left(\frac{n_{i+}n_{+j}}{n} \right), \tag{17.21}$$

which is the Pearson's chi-squared statistic,

$$X^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}, \tag{17.22}$$

where the *observed cell frequency* O_{ij} and the *expected cell frequency* E_{ij} (assuming independence of row and column variates) are given by

$$O_{ij} = n_{ij}, \quad E_{ij} = \frac{n_{i+}n_{+j}}{n}, \tag{17.23}$$

respectively, $i = 1, 2, \ldots, r$, $j = 1, 2, \ldots, s$. Under random sampling, X^2 has approximately (large n) the χ^2 distribution with $(r-1)(s-1)$ degrees of freedom (see, e.g., Rao, 1965, Section 6d.2).

Column Distances

In a similar manner, we define the squared χ^2-distance between the jth and j'th column profiles, $\mathbf{b}_j$ and $\mathbf{b}_{j'}$, respectively, as the quadratic form,

$$d^2(\mathbf{b}_j, \mathbf{b}_{j'}) \equiv (\mathbf{b}_j - \mathbf{b}_{j'})^\tau \mathbf{D}_r^{-1} (\mathbf{b}_j - \mathbf{b}_{j'}) \tag{17.24}$$

$$= \sum_{i=1}^{r} \frac{n}{n_{i+}} \left(\frac{n_{ij}}{n_{+j}} - \frac{n_{ij'}}{n_{+j'}} \right)^2. \tag{17.25}$$

The squared χ^2-distance between the jth column profile and the column centroid is, therefore, given by

$$\begin{aligned} d^2(\mathbf{b}_j, \mathbf{r}) &= (\mathbf{b}_j - \mathbf{r})^\tau \mathbf{D}_r^{-1} (\mathbf{b}_j - \mathbf{r}) \\ &= \frac{1}{n_{+j}} \sum_{i=1}^{r} \frac{n}{n_{i+}n_{+j}} \left(n_{ij} - \frac{n_{i+}n_{+j}}{n} \right)^2. \end{aligned} \tag{17.26}$$

Summing (17.26) over all column profiles yields

$$n \sum_{j=1}^{s} p_{+j} d^2(\mathbf{b}_j, \mathbf{r}) = X^2, \tag{17.27}$$

where X^2 is given by (17.22).

Thus, the weighted average of the squared χ^2-distances of all row profiles to the row centroid (or of all column profiles to the column centroid), where the weights are the row masses (column masses), is the quantity X^2/n. If the row and column variates are independent, then X^2/n will be small, in which case every component of X^2/n — either the $\{p_{i+}d^2(\mathbf{a}_i, \mathbf{c})\}$ or the $\{p_{+j}d^2(\mathbf{b}_j, \mathbf{r})\}$ — will be small. On the other hand, if X^2/n is large, that means that at least one of the $\{p_{i+}d^2(\mathbf{a}_i, \mathbf{c})\}$ or at least one of the $\{p_{+j}d^2(\mathbf{b}_j, \mathbf{r})\}$ will be large. This type of information will be important in determining where independence in the table fails.

For the hair-color/eye-color example, the matrix $\mathbf{E} = (E_{ij})$ of expected cell frequencies is given by:

$$\mathbf{E} = \begin{pmatrix} 193.93 & 38.12 & 284.83 & 185.40 & 15.73 \\ 426.75 & 83.88 & 626.78 & 407.98 & 34.61 \\ 479.15 & 94.18 & 703.74 & 458.07 & 38.86 \\ 355.17 & 69.81 & 521.65 & 339.55 & 28.80 \end{pmatrix}.$$

Compare this matrix with $\mathbf{N} = (O_{ij})$ above. The matrix of values of $(O_{ij} - E_{ij})^2/E_{ij}$ is given by:

$$\begin{pmatrix} 89.95 & 0.00 & 6.74 & 30.66 & 10.30 \\ 159.93 & 12.30 & 2.92 & 118.61 & 27.07 \\ 38.69 & 1.10 & 59.87 & 4.63 & 4.26 \\ 186.22 & 6.82 & 26.99 & 343.36 & 109.63 \end{pmatrix}.$$

The sum of all these values is $X^2 = 1240.05$, which should be compared with 21.03, the tabulated 95th-percentile of the χ^2_{12} distribution. Clearly, independence of row and column variates fails for these data.

17.2.6 Total Inertia and Its Decomposition

We see that using dummy variables for representing a two-way contingency table enables us to view the problem as a special case of canonical variate analysis. The situation is, however, different in that instead of extracting the correlation structure between two sets of stochastic data vectors, we are dealing with the correlation structure of two sets of dummy variables.

Let $\mathbf{x} = (x_{ij})$, where $x_{ij} = X_{ij} - \bar{X}_i$ is either $1 - (n_{i+}/n)$ or $-n_{i+}/n$. Similarly, let $\mathbf{y} = (y_{ij})$, where $y_{ij} = Y_{ij} - \bar{Y}_j$ is either $1 - (n_{+j}/n)$ or $-n_{+j}/n$. Then, the covariance matrices are

$$n^{-1}\mathbf{x}\mathbf{x}^\tau = n^{-1}\mathcal{X}(\mathbf{I}_n - n^{-1}\mathbf{J}_n)\mathcal{X}^\tau = \mathbf{D}_r - \mathbf{r}\mathbf{r}^\tau, \tag{17.28}$$

$$n^{-1}\mathbf{y}\mathbf{y}^\tau = n^{-1}\mathcal{Y}(\mathbf{I}_n - n^{-1}\mathbf{J}_n)\mathcal{Y}^\tau = \mathbf{D}_c - \mathbf{c}\mathbf{c}^\tau, \tag{17.29}$$

where $\mathbf{J}_a = \mathbf{1}_a\mathbf{1}_a^\tau$ is an $(a\times a)$-matrix of 1s. The matrices $\mathbf{x}\mathbf{x}^\tau$ (of rank $r-1$) and $\mathbf{y}\mathbf{y}^\tau$ (of rank $s-1$) are both singular and, hence, their inverses do not exist. We could sidestep this problem by deleting one of the row dummy variables and one of the column dummy variables (see Exercise 17.2), but this would reduce the dimensionality and we would not be able to recover the points from the missing dimensions.

The standard assumption of contingency table analysis is that the row and column totals are considered fixed and the cell frequencies in $\mathbf{N}$ are allowed to vary within those constraints. Accordingly, we center the elements of $\mathbf{N}$ at the values we expect them to have under independence (instead of centering the data $\mathbf{N}$ at the mean). Thus, (17.9) becomes the *relative frequency matrix*,

$$n^{-1}\mathcal{X}(\mathbf{I}_n - n^{-1}\mathbf{J}_n)\mathcal{Y}^\tau = \mathbf{P} - \mathbf{r}\mathbf{c}^\tau = \widetilde{\mathbf{P}}. \tag{17.30}$$

For the hair-color/eye-color example,

$$\widetilde{\mathbf{P}} = \begin{pmatrix} 0.0245 & -0.0000 & -0.0081 & -0.0140 & -0.0024 \\ 0.0485 & 0.0060 & -0.0079 & -0.0408 & -0.0057 \\ -0.0253 & -0.0019 & 0.0381 & -0.0086 & -0.0024 \\ -0.0477 & -0.0040 & -0.0220 & 0.0634 & 0.0104 \end{pmatrix}.$$

The matrix $\widetilde{\mathbf{N}} = n\widetilde{\mathbf{P}}$ is often called the matrix of *residuals* because its ijth entry, $\widetilde{n}_{ij} = O_{ij} - E_{ij}$, shows the difference between the observed cell frequency (O_{ij}) and its expected cell frequency (E_{ij}), assuming independence between row and column variates, $i = 1, 2, \ldots, r$, $j = 1, 2, \ldots, s$ (see (17.23)). Note that because $\widetilde{\mathbf{N}}\mathbf{1}_s = (\mathbf{N} - n\mathbf{r}\mathbf{c}^\tau)\mathbf{1}_s = \mathbf{N}\mathbf{1}_s - n\mathbf{r}\mathbf{c}^\tau\mathbf{1}_s = n\mathbf{r} - n\mathbf{r} = \mathbf{0}$, the rank of $\widetilde{\mathbf{N}}$ (and, hence, of $\widetilde{\mathbf{P}}$) is at most $s-1$.

The $(s\times s)$-matrix $\mathbf{R}$ in (8.76) plays a central role in canonical variate analysis, and it has an obvious analogue in this development. The correspondences between (8.76) and (17.6) are given by

$$\mathbf{\Sigma}_{XX} \leftrightarrow \mathbf{D}_r, \quad \mathbf{\Sigma}_{YY} \leftrightarrow \mathbf{D}_c, \quad \mathbf{\Sigma}_{XY} \leftrightarrow \widetilde{\mathbf{P}}. \tag{17.31}$$

Accordingly, we use (17.7), (17.8), and (17.30) to compute the $(s\times s)$-matrix,

$$\mathbf{R}_0 = \mathbf{D}_c^{-1/2}\widetilde{\mathbf{P}}^\tau\mathbf{D}_r^{-1}\widetilde{\mathbf{P}}\mathbf{D}_c^{-1/2}, \tag{17.32}$$

where $\mathbf{D}_r^{-1} = \text{diag}\{\mathbf{r}^{-1}\}$ and $\mathbf{D}_c^{-1/2} = \text{diag}\{\mathbf{c}^{-1/2}\}$. The entry in the jth row and j'th column of $\mathbf{R}_0$ is given by

$$(n_{+j}n_{+j'})^{-1/2}\sum_{i=1}^{r}\frac{1}{n_{i+}}\left(n_{ij} - \frac{n_{i+}n_{+j}}{n}\right)\left(n_{ij'} - \frac{n_{i+}n_{+j'}}{n}\right) \tag{17.33}$$

and the jth diagonal entry of $\mathbf{R}_0$ is obtained by setting $j = j'$,

$$\frac{1}{n_{+j}}\sum_{i=1}^{r}\frac{1}{n_{i+}}\left(n_{ij} - \frac{n_{i+}n_{+j}}{n}\right)^2. \tag{17.34}$$

For the hair-color/eye-color example,

$$\mathbf{R}_0 = \begin{pmatrix} 0.0881 & 0.0160 & -0.0044 & -0.0798 & -0.0420 \\ 0.0160 & 0.0038 & -0.0001 & -0.0156 & -0.0080 \\ -0.0044 & -0.0001 & 0.0179 & -0.0148 & -0.0099 \\ -0.0798 & -0.0156 & -0.0148 & 0.0923 & 0.0507 \\ -0.0420 & -0.0080 & -0.0099 & 0.0507 & 0.0281 \end{pmatrix}.$$

The trace of $\mathbf{R}_0$, which is also the sum of the eigenvalues of $\mathbf{R}_0$, is

$$\sum_{j=1}^{s} \lambda_j^2 = \text{tr}\{\mathbf{R}_0\} = \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{1}{n_{i+}n_{+j}} \left(n_{ij} - \frac{n_{i+}n_{+j}}{n}\right)^2 = \frac{X^2}{n}, \qquad (17.35)$$

where X^2 is given by (17.22).

If the value of X^2 is very large, as it is in the shoplifting example where $X^2 = 19,949.97$ on $17 \times 12 = 204$ degrees of freedom, the hypothesis of independence of the row and column variates in the contingency table has to be rejected. It then becomes of interest to determine where the deviations from independence occur. Understanding which characteristics of the data are important may be useful for further study.

The quantity X^2/n is referred to as the amount of *total inertia* in the contingency table. The eigenvalues (or *principal inertias*) of $\mathbf{R}_0$ form a decomposition of the total inertia. The accumulated contribution of the first t principal inertias is given by

$$\frac{\lambda_1^2 + \cdots + \lambda_t^2}{\sum_{j=1}^{s} \lambda_j^2}, \qquad (17.36)$$

which is an analogue of the percentage of total variance explained by the first t principal components, where we usually take t to be 2 or 3.

For the hair-color/eye-color example, the eigenvalues of $\mathbf{R}_0$ (and their individual percentages of the total, $\text{tr}(\mathbf{R}_0) = 0.2302$) are 0.1992 (86.6%), 0.0301 (13.1%), 0.0009 (0.4%), 0, and 0. Clearly, the first two eigenvalues account for almost all of the total inertia.

Table 17.5 lists the 12 principal inertias (eigenvalues of $\mathbf{R}_0$) for the shoplifting example. The total inertia is $X^2/n = 19,949.97/33,101 = 0.6027$. We see that the first three eigenvalues account for about 90% of the total inertia, which suggests that almost all of the deviations from independence can be attributed to the first three dimensions. The two-dimensional plot (see Figure 17.1) accounts for about 78% of the total inertia.

17.2.7 Principal Coordinates for Row and Column Profiles

The matrix $\mathbf{R}_0$ in (17.32) can be expressed as

$$\mathbf{R}_0 = \mathbf{M}^\tau \mathbf{M}, \qquad (17.37)$$

TABLE 17.5. *Shoplifting example: Principal inertias (eigenvalues λ_j^2), total inertia, the proportions of total inertia explained by each eigenvalue, and the cumulative proportions.*

Axis	Inertia	Percentage	Cumulative
1	0.3504	58.13	58.13
2	0.1192	19.78	77.91
3	0.0700	11.61	89.52
4	0.0382	6.35	95.86
5	0.0112	1.86	97.72
6	0.0086	1.43	99.14
7	0.0031	0.51	99.66
8	0.0009	0.15	99.81
9	0.0006	0.10	99.91
10	0.0003	0.06	99.97
11	0.0001	0.02	99.99
12	0.0001	0.01	100.00
Total	0.6027		

where the $(r \times s)$-matrix

$$\mathbf{M} = \mathbf{D}_r^{-1/2} \widetilde{\mathbf{P}} \mathbf{D}_c^{-1/2} \tag{17.38}$$

has ijth entry given by the *Pearson residual*,

$$m_{ij} = (n_{i+} n_{+j})^{-1/2} \left(n_{ij} - \frac{n_{i+} n_{+j}}{n} \right), \tag{17.39}$$

$i = 1, 2, \ldots, r$, $j = 1, 2, \ldots, s$. For the hair-color/eye-color example,

$$\mathbf{M} = \begin{pmatrix} 0.1292 & -0.0003 & -0.0354 & -0.0754 & -0.0437 \\ 0.1723 & 0.0478 & -0.0233 & -0.1484 & -0.0709 \\ -0.0847 & -0.0143 & 0.1054 & -0.0293 & -0.02811 \\ -0.1859 & -0.0356 & -0.0708 & 0.2525 & 0.1427 \end{pmatrix}.$$

Thus, from (17.35), the sum of squares of all rs Pearson residuals in the contingency table is the total inertia. Note that because $\text{rank}(\widetilde{\mathbf{P}}) \leq s - 1$, it follows that $\mathbf{M}$ in (17.38) also has rank at most $s - 1$. The singular value decomposition of $\mathbf{M}$ is, therefore, given by

$$\mathbf{M} = \mathbf{U} \mathbf{D}_\lambda \mathbf{V}^\tau, \tag{17.40}$$

where $\mathbf{U}$ is an $(r \times s)$-matrix, $\mathbf{U}^\tau \mathbf{U} = \mathbf{I}_s$, whose columns are the eigenvectors, $\{\mathbf{u}_j\}$, corresponding to the $s - 1$ nonzero eigenvalues of the $(r \times r)$-matrix

$$\mathbf{M}\mathbf{M}^\tau = \mathbf{D}_r^{-1/2} \widetilde{\mathbf{P}} \mathbf{D}_c^{-1} \widetilde{\mathbf{P}}^\tau \mathbf{D}_r^{-1/2} = \mathbf{R}_1, \tag{17.41}$$

$\mathbf{V}$ is an $(s \times s)$-matrix, $\mathbf{V}^\tau\mathbf{V} = \mathbf{I}_s$, whose columns are the eigenvectors, $\{\mathbf{v}_j\}$, corresponding to the eigenvalues of the $(s \times s)$-matrix $\mathbf{M}^\tau\mathbf{M} = \mathbf{R}_0$, and $\mathbf{D}_\lambda = \text{diag}\{\lambda_1, \cdots, \lambda_s\}$ is an $(s \times s)$ diagonal matrix with its principal diagonal having entries the *singular values* (the positive square-roots of the nonzero eigenvalues of either $\mathbf{R}_0$ or $\mathbf{R}_1$).

Combining (17.38) and (17.40), we can write

$$\widetilde{\mathbf{P}} = (\mathbf{D}_r^{1/2}\mathbf{U})\mathbf{D}_\lambda(\mathbf{V}^\tau\mathbf{D}_c^{1/2}) = \mathbf{A}\mathbf{D}_\lambda\mathbf{B}^\tau, \tag{17.42}$$

where

$$\mathbf{A} = \mathbf{D}_r^{1/2}\mathbf{U}, \quad \mathbf{B} = \mathbf{D}_c^{1/2}\mathbf{V}. \tag{17.43}$$

For the hair-color/eye-color example,

$$\mathbf{A} = \begin{pmatrix} -0.1195 & 0.1271 & -0.2917 & -0.1333 & 0 \\ -0.2896 & 0.1496 & 0.3179 & -0.2933 & 0 \\ 0.0248 & -0.4651 & -0.0624 & -0.3293 & 0 \\ 0.3843 & 0.1885 & 0.0362 & -0.2441 & 0 \end{pmatrix}.$$

$$\mathbf{B} = \begin{pmatrix} -0.3292 & 0.2707 & -0.1154 & 0.2741 & 0 \\ -0.0277 & 0.0148 & 0.2138 & 0.0421 & -0.0680 \\ -0.0373 & -0.4764 & -0.0438 & 0.4071 & 0.0259 \\ 0.3406 & 0.1547 & -0.0891 & 0.2186 & -0.2501 \\ 0.0537 & 0.0362 & 0.0345 & 0.0433 & 0.1210 \end{pmatrix}.$$

Note that

$$\mathbf{A}^\tau\mathbf{D}_r^{-1}\mathbf{A} = \mathbf{I}_s, \quad \mathbf{B}^\tau\mathbf{D}_c^{-1}\mathbf{B} = \mathbf{I}_s. \tag{17.44}$$

The expression (17.42) (and (17.44)) is the *generalized singular value decomposition* of $\widetilde{\mathbf{P}}$ in the metrics $\mathbf{D}_r^{-1}$ and $\mathbf{D}_c^{-1}$. The columns of $\mathbf{A}$ and $\mathbf{B}$ are called the *principal axes* of the row and column profiles.

The squared χ^2-distance (in the metric $\mathbf{D}_c^{-1}$) between the $(r \times s)$-matrices of centered row profiles $\mathbf{P}_r - \mathbf{1}_r\mathbf{c}^\tau$ and $\mathbf{B}$ is given by

$$\begin{aligned} \mathbf{G}_P^\tau &= (\mathbf{P}_r - \mathbf{1}_r\mathbf{c}^\tau)\mathbf{D}_c^{-1}\mathbf{B} \\ &= (\mathbf{D}_r^{-1}\widetilde{\mathbf{P}}\mathbf{D}_c^{-1})\mathbf{B} \\ &= \mathbf{D}_r^{-1}(\mathbf{A}\mathbf{D}_\lambda\mathbf{B}^\tau)\mathbf{D}_c^{-1}\mathbf{B} \\ &= \mathbf{D}_r^{-1}\mathbf{A}\mathbf{D}_\lambda, \end{aligned} \tag{17.45}$$

where we have used (17.10), $\mathbf{1}_r = \mathbf{D}_r^{-1}\mathbf{r}$, (17.41), and (17.43). Similarly, we can show that the squared χ^2-distance (in the metric $\mathbf{D}_r^{-1}$) between the $(s \times r)$-matrices of centered column profiles $\mathbf{P}_c - \mathbf{1}_c\mathbf{r}^\tau$ and $\mathbf{A}$ is given by

$$\begin{aligned} \mathbf{H}_P^\tau &= (\mathbf{P}_c - \mathbf{1}_c\mathbf{r}^\tau)\mathbf{D}_r^{-1}\mathbf{B} \\ &= \mathbf{D}_c^{-1}\mathbf{B}\mathbf{D}_\lambda. \end{aligned} \tag{17.46}$$

Substituting (17.42) for the $\mathbf{A}$ and $\mathbf{B}$ in (17.44) and (17.45), respectively, we have that

$$\mathbf{G}_P^\tau = \mathbf{D}_r^{-1/2}\mathbf{U}\mathbf{D}_\lambda, \quad \mathbf{H}_P^\tau = \mathbf{D}_c^{-1/2}\mathbf{V}\mathbf{D}_\lambda. \tag{17.47}$$

For the hair-color/eye-color example,

$$\mathbf{G}_P^\tau = \begin{pmatrix} -0.4003 & 0.1654 & -0.0642 & 0 \\ -0.4407 & 0.0885 & 0.0318 & 0 \\ 0.0336 & -0.2450 & -0.0056 & 0 \\ 0.7027 & 0.1339 & 0.0043 & 0 \end{pmatrix},$$

$$\mathbf{H}_P^\tau = \begin{pmatrix} -0.5440 & 0.1738 & -0.0125 & 0 \\ -0.0233 & 0.0483 & 0.1181 & 0 \\ -0.0420 & -0.2083 & -0.0032 & 0 \\ 0.5887 & 0.1040 & -0.0101 & 0 \\ 1.0944 & 0.2864 & 0.0461 & 0 \end{pmatrix}.$$

The columns of $\mathbf{G}_P^\tau$ and $\mathbf{H}_P^\tau$ are called the *principal coordinates* of the row and column profiles, respectively (hence the subscript P). The matrices $\mathbf{G}_P^\tau$ and $\mathbf{H}_P^\tau$ are related to each other. It can be shown (see Exercise 17.5) that

$$\mathbf{G}_P^\tau = \mathbf{D}_r^{-1}\mathbf{P}\mathbf{H}_P^\tau\mathbf{D}_\lambda^{-1}, \quad \mathbf{H}_P^\tau = \mathbf{D}_c^{-1}\mathbf{P}^\tau\mathbf{G}_P^\tau\mathbf{D}_\lambda^{-1}. \tag{17.48}$$

Similar results can also be obtained directly from the canonical variate analysis developed in Chapter 7 and the correspondences given in (17.31). From (7.61) and (7.62) we compute the $(s \times r)$-matrix $\mathbf{G}_S$ and the $(s \times s)$-matrix $\mathbf{H}_S$, where

$$\mathbf{G}_S = \mathbf{U}^\tau\mathbf{D}_r^{-1/2}, \quad \mathbf{H}_S = \mathbf{V}^\tau\mathbf{D}_c^{-1/2}. \tag{17.49}$$

Note that $\mathbf{G}_S\mathbf{D}_r\mathbf{G}_S^\tau = \mathbf{I}_r$ and $\mathbf{H}_S\mathbf{D}_c\mathbf{H}_S^\tau = \mathbf{I}_s$. The columns of $\mathbf{G}_S^\tau$ and $\mathbf{H}_S^\tau$ in (17.49) are known as the *standard coordinates* of the row and column profiles, respectively (hence the subscript S). Instead of defining the row and column coordinates as (17.49), however, they are generally scaled as in (17.47).

17.2.8 Graphical Displays

In correspondence analysis, one has the choice between analyzing only the row profiles, or analyzing only the column profiles, or analyzing both the row and column profiles together. The graphical displays formed from plotting the row and column coordinates in Table 17.6 are scatterplots that can be of two types:

Symmetric map: Both row and column coordinates are expressed as principal coordinates.

TABLE 17.6. *The t-dimensional formulas for row and column coordinates are the columns of the first t rows of the following matrices, where t is two or three.*

Problem	Row Coordinates	Column Coordinates
Row Profiles	$\mathbf{G}_P = \mathbf{D}_\lambda \mathbf{U}^\tau \mathbf{D}_r^{-1/2}$	$\mathbf{H}_S = \mathbf{V}^\tau \mathbf{D}_c^{-1/2}$
Column Profiles	$\mathbf{G}_S = \mathbf{U}^\tau \mathbf{D}_r^{-1/2}$	$\mathbf{H}_P = \mathbf{D}_\lambda \mathbf{V}^\tau \mathbf{D}_c^{-1/2}$
Both Profiles	$\mathbf{G}_P = \mathbf{D}_\lambda \mathbf{U}^\tau \mathbf{D}_r^{-1/2}$	$\mathbf{H}_P = \mathbf{D}_\lambda \mathbf{V}^\tau \mathbf{D}_c^{-1/2}$

Asymmetric map: The row (or column) coordinates are expressed as principal coordinates while the other is expressed as standard coordinates.

Most users of correspondence analysis prefer to view a symmetric map of both the row and column principal coordinates (17.47) in a two- (or three-) dimensional scatterplot. First, we make a scatterplot of each of the r rows of the first two (or three) columns of $\mathbf{G}_P^\tau$. Then, on the same scatterplot, we overlay a plot of each of the s rows of the first two (or three) columns of $\mathbf{H}_P^\tau$. In Figure 17.2, we have drawn the symmetric correspondence map for the eye-color/hair-color example. If the three-dimensional points are plotted on a dynamic scatterplot, then the display can be rotated in all three dimensions for better viewing. These merged displays provide interpretable views of different features in the data.

There will be $r + s$ points in these scatterplots, which are called *correspondence maps*. For clearer interpretation, different symbols should be used for the row points and column points. It is also useful (unless the plot would look overly cluttered) to identify each point in the plot by a tag showing its corresponding category name. If the row (or column) categories are ordered in some way, such as time-order by year or successive age ranges (as in the shoplifting example), then it is visually helpful to connect those category points in the plot with each other to indicate such order-dependence.

In general, points in the scatterplot that appear "close" to each other tend to correspond to categories that are closely related. More specifically,

- if row points are close, then those rows have similar conditional distributions across columns;
- if column points are close, then those columns have similar conditional distributions across rows;

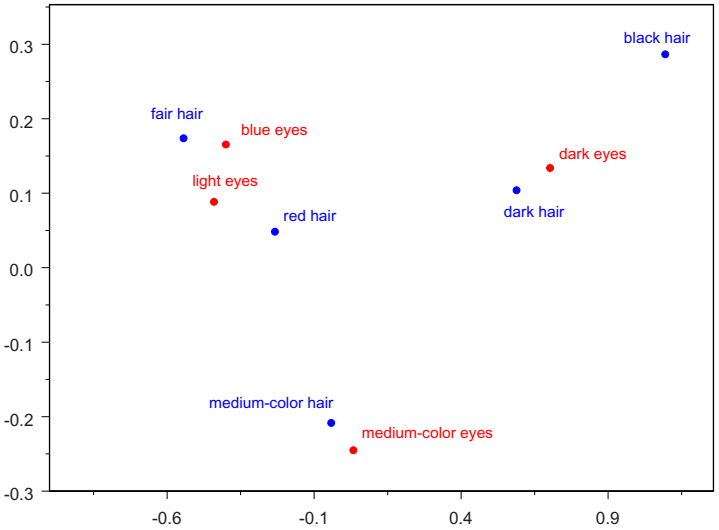

FIGURE 17.2. *Correspondence map for the hair-color/eye-color example. The points exhibit a U-shaped plot with the first principal coordinate (horizontal axis) displaying gradations along the fair-red-medium-dark-black hair scale and the light-blue-medium-dark eyes scale, and the second principal coordinate (vertical axis) displaying a difference between medium-color hair and eyes and the other hair and eye colors.*

- if a row point is close to a column point, then that configuration suggests a particular deviation from independence.

In general, we should not try to compare the positions of row points with the positions of column points and say, for example, that if a particular row point is very close to a particular column point then the corresponding row and column categories are related to each other. (A dissenting view that supports identifying row points with neighboring column points is given by van der Heijden et al, 1989.)

17.3 Square Asymmetric Contingency Tables

An important special case of two-way contingency tables consists of square tables, where $r = s$ and the rows have the same categories as the columns. Examples of square tables include:

- Individuals who are naturally paired, such as husbands and wives or fathers and sons, are classified by occupational or social status.

- Experiments conducted on naturally paired items, such as vision grades of left eye and right eye.
- Two investigators or event judges independently rate each subject in a study using the same Likert-type scale.
- Individuals in a sample are categorized by region of residence at two distinct points in time.
- To study accuracy of a classification rule, the rows give the classes to which the data were assigned by the rule, the columns define the true classes (possibly determined from reference data), the cell entries show how much the classified data and the reference data agree, and the diagonal cells show the numbers of correct classifications.

If a square table $\mathbf{N}$ is symmetric with respect to the r^2 cell frequencies (i.e., $\mathbf{N}^\tau = \mathbf{N}$), then the correspondence map will display coincident pairs of row and column points. In each of the examples listed above, however, the square tables are asymmetric in the sense that $\mathbf{N}^\tau \neq \mathbf{N}$. Unlike rectangular contingency tables, analyzing asymmetric square tables using correspondence analysis has not been very successful. The reason is similar to that for models that try to analyze square tables for symmetry: the data along the principal diagonal tend to have too great an influence on the results.

An innovative way of analyzing square asymmetric tables was proposed by Gower (1977) and Constantine and Gower (1978). Consider a square asymmetric contingency table $\mathbf{N}$ that yields the correspondence table $\mathbf{P}$, also square and asymmetric. Gower showed that $\mathbf{P}$ can be decomposed, prior to analysis, into two orthogonal component tables,

$$\mathbf{P} = \mathbf{M} + \mathbf{Q}, \tag{17.50}$$

where

$$\mathbf{M} = \frac{1}{2}(\mathbf{P} + \mathbf{P}^\tau), \quad \mathbf{Q} = \frac{1}{2}(\mathbf{P} - \mathbf{P}^\tau). \tag{17.51}$$

In (17.51), $\mathbf{M}$ is a symmetric table ($\mathbf{M}^\tau = \mathbf{M}$) and $\mathbf{Q}$ is a skew-symmetric table ($\mathbf{Q}^\tau = -\mathbf{Q}$). Because of the orthogonality of the decomposition (see Exercise 17.4), separate analyses of $\mathbf{M}$ and $\mathbf{Q}$ can be carried out. See van der Heijden et al. (1989). If r is even, the singular vectors of $\mathbf{Q}$ occur in pairs corresponding to pairs of equal singular values (principal inertias). If r is odd, the last singular value of $\mathbf{Q}$ equals zero.

Greenacre (2000) used the decomposition (17.50) to obtain separate correspondence maps of $\mathbf{M}$ and $\mathbf{Q}$. Greenacre showed that these maps could be obtained from a single application of simple correspondence analysis to the $(2r \times 2r)$ block matrix,

$$\mathbf{N}^* = \begin{pmatrix} \mathbf{N} & \mathbf{N}^\tau \\ \mathbf{N}^\tau & \mathbf{N} \end{pmatrix}, \tag{17.52}$$

with correspondence matrix,

$$\mathbf{P}^* = \frac{1}{4}\begin{pmatrix} \mathbf{P} & \mathbf{P}^\tau \\ \mathbf{P}^\tau & \mathbf{P} \end{pmatrix}, \tag{17.53}$$

and row and column totals,

$$\mathbf{w}^* = \frac{1}{2}\begin{pmatrix} \mathbf{w} \\ \mathbf{w} \end{pmatrix}, \tag{17.54}$$

where $\mathbf{w} = (\mathbf{r} + \mathbf{c})/2$. Whereas the usual correspondence analysis is to analyze $\widetilde{\mathbf{P}} = \mathbf{P} - \mathbf{r}\mathbf{c}^\tau$ in the metrics $\mathbf{D}_r^{-1}$ and $\mathbf{D}_c^{-1}$, in this case, we analyze $\mathbf{P} - \mathbf{w}\mathbf{w}^\tau$ in the metrics $\mathbf{D}_w^{-1}$ and $\mathbf{D}_w^{-1}$. Thus, (17.50) becomes $\mathbf{P} - \mathbf{w}\mathbf{w}^\tau = \mathbf{M} - \mathbf{w}\mathbf{w}^\tau + \mathbf{Q}$. We should expect the total inertia attributed to $\mathbf{P} - \mathbf{w}\mathbf{w}^\tau$ to be larger than the usual total inertia (e.g., (17.35)) because $\mathbf{w}\mathbf{w}^\tau$ is not the rank-1 matrix closest to $\mathbf{P}$. The extent of the difference will depend upon how different are $\mathbf{r}$ and $\mathbf{c}$ from each other.

The dimensionality of $\mathbf{N}^*$ is $2r - 1$, of which $r - 1$ dimensions belong to $\mathbf{M}$ and the remaining r dimensions to $\mathbf{Q}$. The correspondence map of $\mathbf{M}$ displays pairs of coincident row and column points (so that it suffices to plot only one set of points). We can, therefore, detect deviations of $\mathbf{N}$ from symmetry by concentrating on the correspondence map of $\mathbf{Q}$.

Thus, there will be two separate correspondence maps for $\mathbf{N}$, one map for the symmetric component $\mathbf{M}$ and the other map for the skew-symmetric component $\mathbf{Q}$. Each map consists of a single set of points. Greenacre recommends that both correspondence maps be scaled equally for comparing the relative sizes of the principal inertias.

17.3.1 Example: Occupational Mobility in England

This 14×14 contingency table (see Table 17.7) of the occupations of a sample of 775 males and their fathers in England was originally studied by Pearson (1904). Figure 17.3 shows the two-dimensional correspondence map of Table 17.7. The total inertia of the contingency table is 1.2974, of which 50.97% is accounted for by the map.

The above decomposition of $\mathbf{P}$ into a symmetric component $\mathbf{M}$ and a skew-symmetric component $\mathbf{Q}$ is accomplished by using (17.52). The resulting total inertia increases by 0.3016 to 1.5990 due to the different type of centering involved. The total symmetric inertia is 1.1484, and the total skew-symmetric inertia is 0.4506. In Table 17.8, we list the 27 principal inertias, of which 13 correspond to the symmetric correspondence analysis and 14 (= 7 pairs) to the skew-symmetric correspondence analysis. Also listed in Table 17.8 are the percentages of the two sets of principal inertias relative to the total symmetric and skew-symmetric inertias. The first pair

TABLE 17.7. *Occupations of fathers and their sons in England (Pearson, 1904). The occupational categories are A army; B art; C teaching, clerical work, civil service; D crafts; E divinity; F agriculture; G landownership; H law; I literature; J commerce; K medicine; L navy; M politics and court; N scholarship and science. Uppercase letters represent occupations of the father and lowercase letters represent occupations of the son. The Pearson chi-squared test for independence gives* $X^2 = 874.9$ *on 169 degrees of freedom, so that an hypothesis of independence is rejected.*

	Sons														
Fathers	*a*	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*	*j*	*k*	*l*	*m*	*n*	Totals
A	28	0	4	0	0	0	1	3	3	0	3	1	5	2	50
B	2	51	1	1	2	0	0	1	2	0	0	0	1	1	62
C	6	5	7	0	9	1	3	6	4	2	1	1	2	7	54
D	0	12	0	6	5	0	0	1	7	1	2	0	0	10	44
E	5	5	2	1	54	0	0	6	9	4	12	3	1	13	115
F	0	2	3	0	3	0	0	1	4	1	4	2	1	5	26
G	17	1	4	0	14	0	6	11	4	1	3	3	17	7	88
H	3	5	6	0	6	0	2	18	13	1	1	1	8	5	69
I	0	1	1	0	4	0	0	1	4	0	2	1	1	4	19
J	12	16	4	1	15	0	0	5	13	11	6	1	7	15	106
K	0	4	2	0	1	0	0	0	3	0	20	0	5	6	41
L	1	3	1	0	0	0	1	0	1	1	1	6	2	1	18
M	5	0	2	0	3	0	1	8	1	2	2	3	23	1	51
N	5	3	0	2	6	0	1	3	1	0	0	1	1	9	32
Totals	84	108	37	11	122	1	15	64	69	24	57	23	74	86	775

of symmetric principal inertias (1 and 2) accounts for 33.85% + 20.20% = 54.05% of the total symmetric inertia, suggesting that higher dimensions contain additional significant information. The first pair of skew-symmetric principal inertias (3 and 4) accounts for 35.15% + 35.15% = 70.30% of the total skew-symmetric inertia (compared with only 9.90% + 9.90% = 19.80% of the total inertia). The symmetric dimensions are, therefore, 1, 2, 5–9, 12, 13, 16, 21, 24, and 27, and the remainder, which occur in pairs, are the skew-symmetric dimensions.

Figure 17.4 shows the correspondence maps of dimensions 1 and 2, and 3 and 4, respectively. The top panel of Figure 17.4 shows the symmetric portion of the table. The points representing the arts (B) and crafts (D) occupations are clearly separated from the other points, but these two points are also not close to each other. One can also argue that these two points account for much of the difference in inertias between the symmetric and skew-symmetric analyses because the variation in points is not that different without points B and D. Points that are close together in this map reflect the fact that there is a lot of movement from father to son

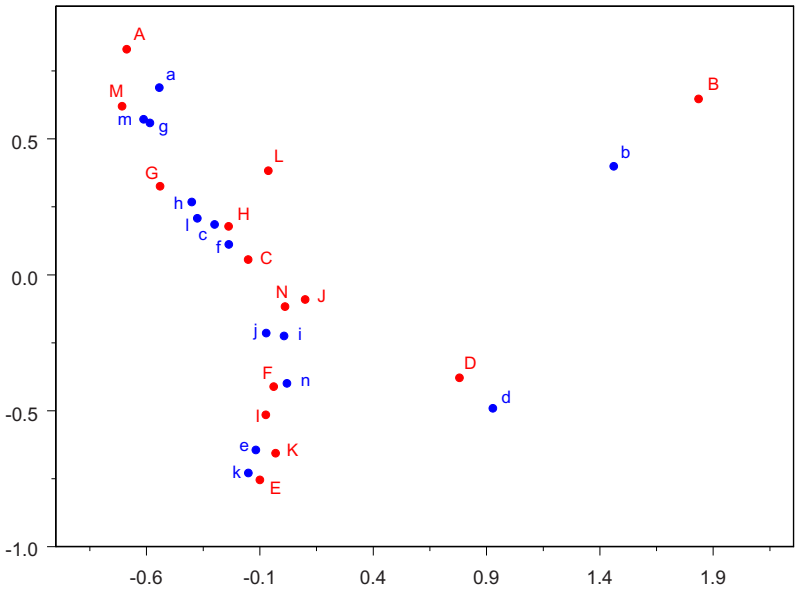

FIGURE 17.3. *Correspondence map for the occupational mobility example. The horizontal axis represents the first principal coordinate and the vertical axis the second principal coordinate. On the left of the map, there is a steady progression in occupations from A to E (and from a to k). The two occupations of B and D (and b and d), representing arts and crafts, stand out from the rest.*

between those occupations, whereas points that are far apart from each other indicate relatively little movement. If we ignore points B and D, there appears to be a progression in the occupations, from the topmost points down through several clusters of points, such as

- army (A), and politics and court (M)
- teaching, clerical work, civil service (C), landownership (G), law (H), and navy (L)
- agriculture (F), literature (I), commerce (J), and scholarship and science (N)
- divinity (E) and medicine (K)

These clusters suggest that occupational mobility from father to son is typically confined to movements within the various clusters only and not between clusters.

TABLE 17.8. *Occupational mobility example: Principal inertias (eigenvalues λ_j^2), total inertia, the percentages and cumulative percentages of total inertia explained by each eigenvalue, and the percentages corresponding to the symmetric (S) and skew-symmetric (SS) correspondence analyses. The total symmetric inertia is 1.1484, and the total skew-symmetric inertia is 0.4506.*

Principal Axis	Principal Inertia	% Inertia	Cumulative	%-S	%-SS
1	0.3887	24.31	24.31	33.85	
2	0.2320	14.51	38.82	20.20	
3	0.1584	9.90	48.72		35.15
4	0.1584	9.90	58.62		35.15
5	0.1439	9.00	67.62	12.53	
6	0.1238	7.74	75.36	10.78	
7	0.0818	5.12	80.48	7.12	
8	0.0707	4.42	84.91	6.16	
9	0.0498	3.12	88.02	4.34	
10	0.0418	2.62	90.64		9.28
11	0.0418	2.62	93.25		9.28
12	0.0229	1.43	94.68	1.99	
13	0.0220	1.38	96.06	1.92	
14	0.0129	0.81	96.87		2.86
15	0.0129	0.81	97.67		2.86
16	0.0104	0.65	98.32	0.91	
17	0.0076	0.47	98.80		1.69
18	0.0076	0.47	99.27		1.69
19	0.0031	0.19	99.46		0.69
20	0.0031	0.19	99.66		0.69
21	0.0017	0.10	99.76	0.15	
22	0.0011	0.07	99.83		0.24
23	0.0011	0.07	99.90		0.24
24	0.0006	0.04	99.94	0.00	
25	0.0004	0.02	99.97		0.00
26	0.0004	0.02	99.99		0.00
27	0.0001	0.01	100.00	0.00	
Total	1.5990				

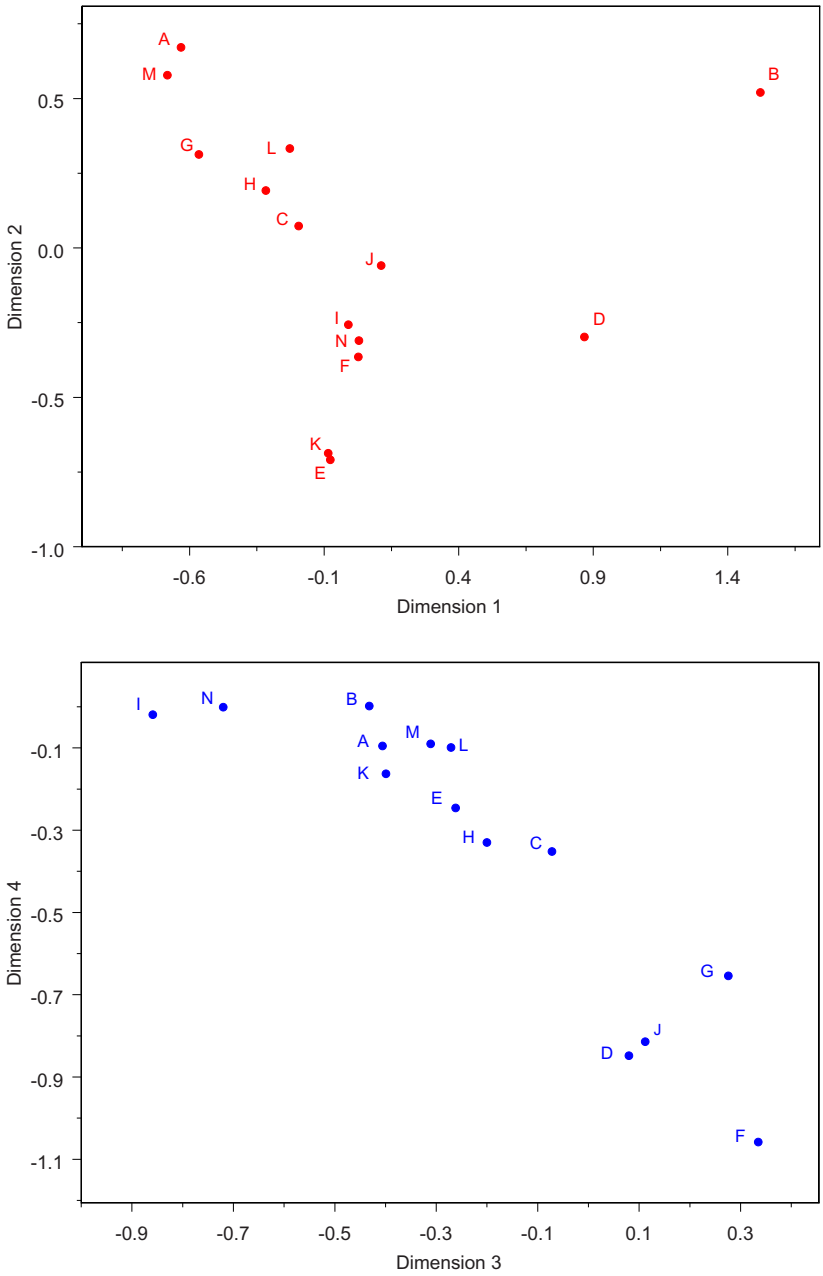

FIGURE 17.4. *Correspondence analysis of the symmetric component (top panel) and skew-symmetric component (bottom panel) for the occupational mobility example.*

The bottom panel of Figure 17.4 shows the deviations from symmetry. Asymmetry between any two points can be envisioned by a triangle constructed with vertices at those two points and the origin; the greater the area of that triangle, the greater the degree of asymmetry between the points. Points that yield triangles with no area (i.e., points on a line through the origin) have no asymmetric relationship. Points that are close to the origin indicate small asymmetries. In this map, there are no points clustered around the origin, suggesting some asymmetry between all occupations. Indeed, all the points in this map lie on one side of a line drawn through the origin, indicating that circular triads are absent in the data. The more drastic asymmetries are those points furthest from the origin, literature (I) and scholarship and science (N) at one extreme and agriculture (F) at the other. The greatest deviation from symmetry is from a father's occupation of literature (I) to a son's occupation in agriculture (F).

17.4 Multiple Correspondence Analysis

Multiple correspondence analysis is intended to be a generalization of simple correspondence analysis, in the sense that it is designed to deal with the graphical representation of contingency tables that have more than two categorical variables. The fact that as currently conceived it is not a true generalization (in the sense that simple correspondence analysis is not a special case) has not, however, detracted from its usefulness. Accordingly, there is much research currently taking place on this topic.

17.4.1 The Multivariate Indicator Matrix

As we did in Section 17.2.2, we can define a dummy (or indicator) variable for each of the Q categorical variables that make up the table. Suppose that the qth variable has J_q categories and that $J = \sum_{q=1}^{Q} J_q$ is the total number of categories over all variables. Suppose further that there are n individuals in the study (who may be some part — a sample — or all of a population). Let $\mathbf{Z} = (Z_{ij})$ be a $(J \times n)$-matrix, where

$$Z_{ij} = \begin{cases} 1, & \text{if the } j\text{th individual belongs to the } i\text{th category} \\ 0, & \text{otherwise,} \end{cases} \tag{17.55}$$

$i = 1, 2, \ldots, J$, $j = 1, 2, \ldots, n$. We assume that there is no row of $\mathbf{Z}$ that contains all 0s. Each column of $\mathbf{Z}$ sums to Q and all Jn entries sum to nQ. The matrix $\mathbf{Z}$ is often called a *multivariate indicator matrix*. One interpretation of the concept of multiple correspondence analysis is that of

carrying out a simple correspondence analysis of the multivariate indicator matrix $\mathbf{Z}$.

We can partition the J rows of $\mathbf{Z}$ into blocks by variable so that

$$\mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 \\ \vdots \\ \mathbf{Z}_Q \end{pmatrix}, \tag{17.56}$$

where $\mathbf{Z}_q$ is a $(J_q \times n)$-matrix corresponding to the qth categorical variable having J_q categories, $q = 1, 2, \ldots, Q$. The following properties of $\mathbf{Z}$ are given in Greenacre (1984). In $\mathbf{Z}_q$, there are $\mathbf{1}_{J_q}^\tau \mathbf{Z}_q \mathbf{1}_n = n$ 1s, $q = 1, 2, \ldots, Q$. Following (17.15), the row masses of $\mathbf{Z}_q$ are defined by the J_q-vector,

$$\mathbf{c}_q^Z \equiv (nQ)^{-1} \mathbf{Z}_q \mathbf{1}_n. \tag{17.57}$$

Because the row masses of $\mathbf{Z}_q$ sum to $\mathbf{1}_{J_q}^\tau \mathbf{c}_q^Z = (nQ)^{-1} n = Q^{-1}$, each of the Q categorical variables has the same total mass. As a result, the row masses over all Q variables sum to 1. The row centroid is a weighted average of the J_q rows of $\mathbf{Z}_q$, where the weights are the row masses,

$$\frac{(\mathbf{c}_q^Z)^\tau \mathbf{Z}_q}{(\mathbf{c}_q^Z)^\tau \mathbf{1}_{J_q}} = \frac{(nQ)^{-1} \mathbf{1}_n^\tau \mathbf{Z}_q^\tau \mathbf{Z}_q}{Q^{-1}} = n^{-1} \mathbf{1}_n^\tau, \tag{17.58}$$

because $\mathbf{Z}_q^\tau \mathbf{Z}_q = \mathbf{I}_n$. Thus, the qth block of J_q row profiles has a row centroid (17.58) that does not depend upon q. Those J_q row profiles are dispersed within a subspace having at most $J_q - 1$ dimensions. All J row profiles are, therefore, dispersed within a subspace having at most $\sum_q (J_q - 1) = J - Q$ dimensions.

17.4.2 The Burt Matrix

A second interpretation of the idea of multiple correspondence analysis is based upon analyzing the $(J \times J)$-matrix

$$\mathbf{B} = \mathbf{Z}\mathbf{Z}^\tau = \begin{pmatrix} \mathbf{Z}_1\mathbf{Z}_1^\tau & \mathbf{Z}_1\mathbf{Z}_2^\tau & \cdots & \mathbf{Z}_1\mathbf{Z}_Q \\ \mathbf{Z}_2\mathbf{Z}_1^\tau & \mathbf{Z}_2\mathbf{Z}_2^\tau & \cdots & \mathbf{Z}_2\mathbf{Z}_Q^\tau \\ \vdots & \vdots & & \vdots \\ \mathbf{Z}_Q\mathbf{Z}_1^\tau & \mathbf{Z}_Q\mathbf{Z}_2^\tau & \cdots & \mathbf{Z}_Q\mathbf{Z}_Q^\tau \end{pmatrix}, \tag{17.59}$$

which is called a *Burt matrix*. See (17.6) for a Burt matrix with $Q = 2$. $\mathbf{B}$ is a symmetric matrix with block structure. The qth diagonal block submatrix, $\mathbf{Z}_q\mathbf{Z}_q^\tau = n\mathbf{D}_q$, say, is a diagonal matrix of the row totals of $\mathbf{Z}_q$ $(q = 1, 2, \ldots, Q)$, where $\mathbf{D}_q$ is the diagonal matrix of row or column masses for the qth variable. The off-diagonal (u, v)-block submatrix, $\mathbf{Z}_u\mathbf{Z}_v^\tau = \mathbf{N}_{uv}$, say, $(u \neq v)$, is a two-way contingency table between the uth variable and

the vth variable ($u, v = 1, 2, \ldots, Q$). Because the total of all entries in each submatrix $\mathbf{Z}_i\mathbf{Z}_j^\tau$ in $\mathbf{B}$ is n, the total of all entries of $\mathbf{B}$ is $b = nQ^2$. The Burt matrix (17.59) is the analogue in the discrete case of the covariance matrix of Q continuous variables.

17.4.3 Equivalence and an Implication

The two primary approaches to multiple correspondence analysis turn out to be equivalent to one another (Greenacre, 1984). From the symmetry of $\mathbf{B}$, a simple correspondence analysis of $\mathbf{B}$ produces the same sets of row and column coordinates, so that one of the two sets can be ignored. Furthermore, the standard coordinates of the rows of $\mathbf{B}$ are identical to the standard coordinates of the rows of $\mathbf{Z}$, and the principal coordinates obtained by analyzing $\mathbf{B}$ are directly related to those obtained by analyzing $\mathbf{Z}$ because the principal inertias of $\mathbf{B}$ are the squares of those of $\mathbf{Z}$.

This equivalence between the two approaches has the following implication. Although the multivariate indicator matrix $\mathbf{Z}$ incorporates information from all Q categorical variables, its multiple correspondence analysis provides no more information than an analysis of all pairs of categorical variables. In other words, multiple correspondence analysis of either $\mathbf{Z}$ or $\mathbf{B}$ offers no insight into three- or higher-way interactions that may be present in the contingency table.

17.4.4 Example: Satisfaction with Housing Conditions

This data set was studied by Madsen (1976) in a study of housing conditions in selected areas of Copenhagen, Denmark. A total of $n = 1,681$ residents living in rented homes built during 1960–1968 were surveyed about their satisfaction (categorized as low (`ls`), medium (`ms`), high (`hs`)), the amount of contact with other residents (low (`lc`), high (`hc`)), and their feeling of influence on apartment management (low (`li`), medium (`mi`), high (`hi`)). The rental units were categorized as tower blocks (`tb`), apartments (`ap`), atrium houses (`ah`), and terraced houses (`th`). The purpose of the study was to assess whether there was any association between degrees of contact, influence, and satisfaction and the type of housing.

The Burt table is given in Table 17.9. The χ^2-statistics for the off-diagonal two-way contingency tables are $X^2_{12} = 16.660$, $X^2_{13} = 39.121$, $X^2_{14} = 60.286$, $X^2_{23} = 17.586$, $X^2_{24} = 106.175$, and $X^2_{34} = 5.140$, where "1" = Housing, "2" = Influence, "3" = Contact, and "4" = Satisfaction. Assuming these two-way tables are independent of each other, we conclude that both housing and influence appear not to be related to either contact or satisfaction. The sum of these χ^2-values is $X^2 = 244.968$.

TABLE 17.9. *Burt table of data on satisfaction with housing conditions in Copenhagen, Denmark (Madsen, 1976). The variables are type of housing (tower blocks:* `tb`*; apartments:* `ap`*; atrium houses:* `ah`*; terraced houses:* `th`*), influence on apartment management (low:* `li`*; medium:* `mi`*; high:* `hi`*), contact with other residents (low:* `lc`*; high:* `hc`*), and satisfaction (low:* `ls`*; medium:* `ms`*; high:* `hs`*). For this table, Q=4,* $J_1 = 4$, $J_2 = 3$, $J_3 = 2$, $J_4 = 3$, $J = 12$, *and* $n = 1681$.

	Housing				Influence			Contact		Satisfaction		
	tb	ap	ah	th	li	mi	hi	lc	hc	ls	ms	hs
tb	400	0	0	0	140	172	88	219	181	99	101	200
ap	0	765	0	0	268	297	200	317	448	271	192	302
ah	0	0	239	0	95	84	60	82	157	64	79	96
th	0	0	0	227	124	106	47	95	182	133	74	70
li	140	268	95	124	627	0	0	234	393	282	170	175
mi	172	297	84	106	0	659	0	279	380	206	189	264
hi	88	200	60	47	0	0	395	200	195	79	87	229
lc	219	317	82	95	234	279	200	713	0	262	178	273
hc	181	448	157	182	393	380	195	0	968	305	268	395
ls	99	271	64	133	282	206	79	262	305	567	0	0
ms	101	192	79	74	170	189	87	178	268	0	446	0
hs	200	302	96	70	175	264	229	273	395	0	0	668

The two-dimensional multiple correspondence map is given in Figure 17.5. The first axis orders from right to left the low, medium, and high categories of the influence and satisfaction variables, whereas the reverse ordering occurs for the contact variable. The second axis separates the high levels from the low levels of influence, contact, and satisfaction, and also separates `th` and `tb` from `ah`, and `ap` is positioned at the center of the map.

Certain points are close to each other and indicate associations. Thus, high influence on management is related to residents being highly satisfied, whereas high contact with other residents produces medium satisfaction. Residents of atrium houses tend to have high contact with other residents and enjoy medium satisfaction, apartment residents have medium influence on management, residents of tower blocks tend to have low contact with other residents, and residents of terraced housing appear to have both low influence and low satisfaction.

17.4.5 A Weighted Least-Squares Approach

There are $Q(Q-1)/2$ distinct two-way contingency tables above the diagonal of $\mathbf{B}$; the tables below the diagonal are transposes of those above. Although we could carry out a simple correspondence analysis for every one of those $Q(Q-1)/2$ tables, such extensive and exhaustive analyses

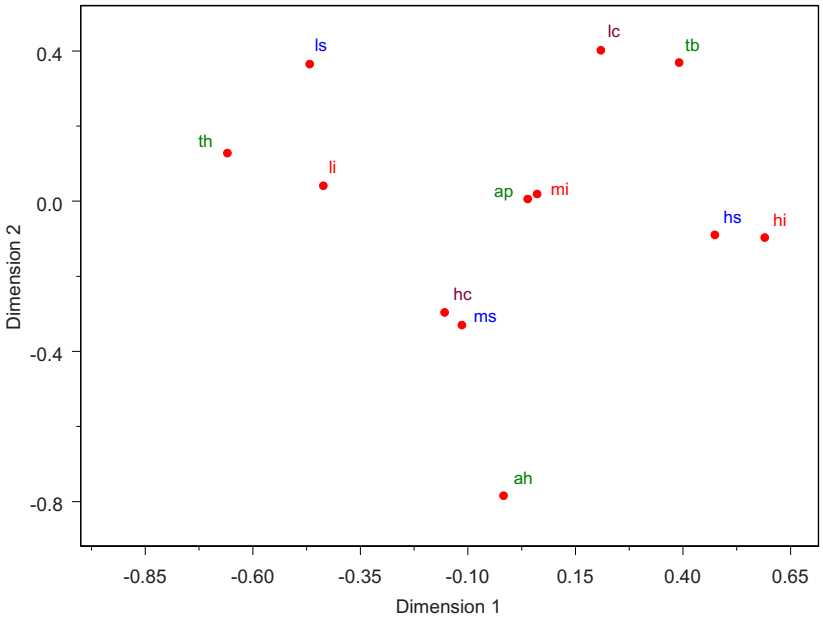

FIGURE 17.5. *Correspondence map for the housing conditions example. The factors in the study were: type of housing (tower blocks,* `tb`*; apartments,* `ap`*; atrium houses,* `ah`*; terraced houses,* `th`*), influence on apartment management (low,* `li`*; medium,* `mi`*; high,* `hi`*), contact with other residents (low,* `lc`*; high,* `hc`*), and satisfaction (low,* `ls`*; medium,* `ms`*; high,* `hs`*).*

would violate the principles of parsimony, efficiency, and dimensionality reduction.

With this in mind, we mention an alternative approach by Greenacre (1988), who proposed a *matrix approximation method* that (*a*) simultaneously fits all the $Q(Q-1)/2$ tables in the upper-triangle of $\mathbf{B}$, and (*b*) reduces to simple correspondence analysis of $\mathbf{N} = \mathbf{N}_{12}$ when $Q = 2$. The idea is to approximate $\mathbf{B}$ by another matrix $\widehat{\mathbf{B}}$, say, having reduced rank that minimized the weighted least-squares criterion

$$n^{-1}\text{tr}\{\mathbf{D}^{-1/2}(\mathbf{B}-\hat{\mathbf{B}})\mathbf{D}^{-1}(\mathbf{B}-\hat{\mathbf{B}})^{\tau}\mathbf{D}^{-1/2}\}, \tag{17.60}$$

where $\mathbf{D} = Q\mathbf{D}_r$ is Q times the diagonal matrix, $\mathbf{D}_r$, of row (or column) masses of $\mathbf{B}$ and is defined so that all its elements sum to 1 (cf. Exercise 17.3). Greenacre suggested the use of an alternating least-squares algorithm as a means of obtaining $\hat{\mathbf{B}}$ but could not guarantee that the minimum of (17.60) would be achieved by that procedure.

17.5 Software Packages

Many of the popular statistical software packages contain simple and multiple correspondence analysis routines. R has the `ca` package; see Charnomordic and Holmes (2001) and the details in Greenacre (2007, Appendix C). MINITAB has a correspondence analysis routine that appears to be matched to the output in Greenacre (1984). There is also a program `CodonW`, written by John Peden and available at `codonw.sourceforge.net`, which provides correspondence analysis of codon and amino acid usage.

Bibliographical Notes

Correspondence analysis was developed by many individuals. Initial work showing the correlation structure of a two-way contingency table appeared during the mid-1930s by H.O. Hirschfield (later Hartley), P. Horst, and others. At the start of the 1940s, R.A. Fisher and L. Guttman constructed scaling theories for contingency tables for biometric and psychometric contexts, respectively. The methodology found its champion, J.-P. Benzecri, in the early 1960s when Benzecri and a group of French statisticians constructed a theory of associations between rows and columns of a two-way contingency table. This was called *analyse des correspondances* in French, which was later loosely translated as "correspondence analysis." Others who have had major impacts on the subject include M.O. Hill, M.J. Greenacre, and L.A. Goodman.

Much of this chapter has benefitted from the treatment of the topic in books and articles by Greenacre; specifically, Greenacre (1981, 1984, 1988, 2000, 2007) and Greenacre and Hastie (1987). An interesting collection of articles on applications of correspondence analysis (and other related topics) is the book edited by Blasius and Greenacre (1998). See also the articles by Gower and Digby (1981) (who provide a general tour of techniques for graphically representing multivariate data), van der Heijden, de Falguerolles, and de Leeuw (1989) (who studied the correspondence analysis of residuals from fitting a log-linear model to a contingency table), and Pack and Jolliffe (1992) (who proposed measures for detecting influential observations in correspondence analysis).

Exercises

17.1 The 4×4 contingency table in Table 17.10 was originally analyzed by Stuart (1953) and has since been studied by many statisticians. It contains frequency data on eye tests, specifically, the right-eye grade and the corresponding left-eye grade in unaided distance vision for 7,477 women,

TABLE 17.10. *Right-eye grade and left-eye grade of 7,477 women with respect to unaided distance vision (Stuart, 1953). The Pearson chi-squared test for independence gives $X^2 = 8,096.877$ on 9 degrees of freedom, so that an hypothesis of independence is rejected.*

	Left-Eye Grade				
Right-Eye Grade	Best	Second	Third	Worst	Totals
Best	1,520	266	124	66	1,976
Second	234	1,512	432	78	2,256
Third	117	362	1,772	205	2,456
Worst	36	82	179	492	789
Totals	1,907	2,222	2,507	841	7,477

aged 30–39, employed in Royal Ordinance factories in Britain, where each eye was graded in one of four categories from best to worst. Carry out a correspondence analysis for this square contingency table and interpret the results.

17.2 Suppose we omit the last row of $\mathcal{X}$ and last row of $\mathcal{Y}$, so that $\mathcal{X}$ has $r-1$ rows and n columns and $\mathcal{Y}$ has $s-1$ rows and n columns. Suppose we center $\mathcal{X}$ and $\mathcal{Y}$ at their means.

(a) Show that

$$(\mathcal{X}_c\mathcal{X}_c^\tau)^{-1} = \text{diag}\left[n_{1+}^{-1}, n_{2+}^{-1}, \ldots, n_{r-1,+}^{-1}\right] + n_{r+}^{-1}\mathbf{J}_{r-1},$$

$$(\mathcal{Y}_c\mathcal{Y}_c^\tau)^{-1} = \text{diag}\left[n_{+1}^{-1}, n_{+2}^{-1}, \ldots, n_{+,s-1}^{-1}\right] + n_{+s}^{-1}\mathbf{J}_{s-1}.$$

(b) Show that the entry in the jth row and ith column of the full-rank regression coefficient matrix, $\hat{\mathbf{\Theta}} = \mathcal{Y}_c\mathcal{X}_c^\tau(\mathcal{X}_c\mathcal{X}_c^\tau)^{-1}$, is

$$\theta_{ji} = \frac{n_{ij}}{n_{i+}} - \frac{n_{rj}}{n_{r+}}, \quad i = 1, 2, \ldots, r-1, \ j = 1, 2, \ldots, s-1,$$

which is just the difference between the ith and rth row proportions for the jth column of the contingency table. Similarly, show that the entry in the ith row and jth column of $\mathcal{X}_c\mathcal{Y}_c^\tau(\mathcal{Y}_c\mathcal{Y}_c^\tau)^{-1}$ is

$$\frac{n_{ij}}{n_{+j}} - \frac{n_{is}}{n_{+s}}, \quad i = 1, 2, \ldots, r-1, \ j = 1, 2, \ldots, s-1.$$

(c) From these two matrices, show that the trace of $\hat{\mathbf{R}}$ is given by

$$\sum_{i=1}^{r}\sum_{j=1}^{s} \frac{1}{n_{i+}n_{+j}}\left(n_{ij} - \frac{n_{i+}n_{rj}}{n_{r+}}\right)\left(n_{ij} - \frac{n_{is}n_{+j}}{n_{+s}}\right),$$

and, under independence of A and B, that $\text{tr}\{\hat{\mathbf{R}}\}$ reduces to X^2 in (17.22).

TABLE 17.11. *Number of children in a family versus yearly income (in units of 1,000 Kroner) for $n = 25263$ Swedish families (Cramér, 1946). The Pearson chi-squared test for independence gives $X^2 = 568.57$ on 12 degrees of freedom, so that an independence hypothesis is rejected.*

Number of Children	Yearly Income (1000s Kroner) 0–1	1–2	2–3	3+	Total
0	2,161	3,577	2,184	1,636	9,558
1	2,755	5,081	2,222	1,052	11,110
2	936	1,753	640	306	3,635
3	225	419	96	38	778
≥ 4	39	98	31	14	182
Total	6,116	10,928	5,173	3,046	25,263

(d) Show that the $s - 1$ eigenvalues of $\hat{\mathbf{R}}$ are identical to the nonzero eigenvalues of $\mathbf{R}_0$ (or $\mathbf{R}_1$).

17.3 (Greenacre, 2000). Another way of deriving the results of simple correspondence analysis is to find an $(r \times s)$-matrix $\hat{\mathbf{P}}$ having reduced-rank $t < \min(r, s)$ that approximates $\mathbf{P}$ by minimizing the weighted least-squares criterion,

$$\text{tr}\{\mathbf{D}_r^{-1/2}(\mathbf{P} - \hat{\mathbf{P}})\mathbf{D}_c^{-1}(\mathbf{P} - \hat{\mathbf{P}})^\tau \mathbf{D}_r^{-1/2}\}.$$

Using the Eckart–Young Theorem, find the matrix $\hat{\mathbf{P}}$ that yields the best reduced-rank approximation of $\mathbf{P}$ in the above sense. Show that the best "rank-1" approximation to $\mathbf{P}$ is the *trivial solution* $\hat{\mathbf{P}} = \mathbf{r}\mathbf{c}^\tau$.

17.4 Let $\mathbf{M} = [m_{ij}]$ and $\mathbf{Q} = [q_{ij}]$ be defined as in (17.51) and let $\mathbf{N} = \mathbf{M} + \mathbf{Q}$. Consider $\text{tr}\{(vec\ \mathbf{N})(vec\ \mathbf{N})^\tau\}$. Show that the cross-product term $\text{tr}\{(vec\ \mathbf{M})(vec\ \mathbf{Q})^\tau\} = 0$, whence, we have the identity,

$$\sum_i \sum_j n_{ij}^2 = \sum_i \sum_j m_{ij}^2 + \sum_i \sum_j q_{ij}^2.$$

17.5 Show that $\mathbf{G}_P^\tau$ and $\mathbf{H}_P^\tau$ are related to each other by proving that $\mathbf{G}_P^\tau = \mathbf{D}_r^{-1}\mathbf{P}\mathbf{H}_P^\tau\mathbf{D}_\lambda^{-1}$ and $\mathbf{H}_P^\tau = \mathbf{D}_c^{-1}\mathbf{P}^\tau\mathbf{G}_P^\tau\mathbf{D}_\lambda^{-1}$.

17.6 The 5×4 contingency table in Table 17.11 is due to Cramér (1946, p. 444); see also Diaconis and Efron (1985). It contains a sample of frequency data from a Swedish census of March 1936 in which 25,263 married couples residing in country districts, who had been married for at most five years, each listed the number of children in their family and their yearly income (in units of 1,000 Kroner). Carry out a correspondence analysis for this table and interpret the results.

17.7 Construct four different contingency tables, each with five rows and three columns, with the restriction that each of the column totals in each table equals 50. Compute the weights in the chi-squared statistic for each table. Compute the inertia for each table and arrange the four tables by increasing inertia. Plot the row profiles for each table as points in a triangular scatterplot. What is the relationship between inertia and these plots?