Chapter 1

Dual Scaling

Shizuhiko Nishisato

1.1. Why Dual Scaling?

Introductory and intermediate courses in statistics are almost exclusively based on the following assumptions: (a) the data are continuous, (b) they are a random sample from a population, and (c) the population distribution is normal. In the social sciences, it is very rare that our data satisfy these assumptions. Even if we manage to use a random sampling scheme, the data may not be continuous but qualitative, and the assumption of the normal distribution then becomes irrelevant. What can we do with our data, then? Dual scaling will offer an answer to this question as a reasonable alternative.

More important, however, the traditional statistical analysis is mostly what we call *linear analysis*, which is a natural fate of using continuous variables, for which such traditional statistical procedures as analysis of variance, regression analysis, principal component analysis, and factor analysis were developed. In traditional principal component analysis, for example, we can look into such a linear phenomenon as "blood pressure increases as one gets older" while failing to capture a nonlinear phenomenon such as "migraines occur more frequently when blood pressure is very low or very high." When we look at possible forms of relations between two variables, we realize that most relations are nonlinear and that it is not advantageous to restrict our attention only to the linear relation. Dual scaling captures linear and nonlinear relations among variables, without modeling the forms of relations for analysis.

Dual scaling is also referred to as "optimal scaling" (Bock, 1960) because all forms of relations among variables are captured through optimally spacing categories of variables. The main purpose of data analysis lies in delineating relations among variables, linear or nonlinear, or, more generally, in extracting as much information in data as possible. We will find that dual scaling is an optimal method to extract a maximal amount of information from multivariate categorical data. We will see later that dual scaling can be applied effectively to many kinds of psychological data such as observation data, teacher evaluation forms, attitude/aptitude data, clinical data, and all types of questionnaire data. This chapter contains a minimal package of information about all aspects of dual scaling.

1.2. Historical Background

1.2.1. Mathematical Foundations in Early Days

Two major contributions to the area from the past are (a) algebraic *eigenvalue theory*, pioneered

AUTHOR'S NOTE: This work was supported by the Natural Sciences and Engineering Research Council of Canada. The paper was written while the author was a Visiting Professor at the School of Business Administration, Kwansei Gakuin University, Nishinomiya, Japan.

by mathematicians (e.g., Euler, Cauchy, Jacobi, Cayley, and Sylvester) in the 18th century, and (b) the theory of *singular value decomposition* (SVD) by Beltrami (1873), Jordan (1874), and Schmidt (1907).

The eigenvalue decomposition (EVD) was for orthogonal decomposition of a square matrix, put into practice as principal component analysis (Hotelling, 1933; Pearson, 1901). SVD was for the joint orthogonal decomposition of row structure and column structure of any rectangular matrix and reappeared much later in metric multidimensional scaling as the *Eckart-Young decomposition* (Eckart & Young, 1936). Both EVD and SVD are based on the idea of principal hyperspace, that is, space described in terms of principal axes.

1.2.2. Pioneers in the 20th Century

With these precursors, Richardson and Kuder (1933) presented the idea of what Horst (1935) called the *method of reciprocal averages* (MRA) for the analysis of multiple-choice data. Hirschfeld (1935) provided a formulation for weighting rows and columns of a two-way table in such a way that the regression of rows on columns and that of columns on rows could be simultaneously linear, which Lingoes (1964) later called *simultaneous linear regressions*. Fisher (1940) considered discriminant analysis of data in a contingency table, in which he, too, suggested the algorithm of MRA. Most important contributions in the early days were by Guttman (1941) for his detailed formulation for the scaling of multiple-choice data and Maung (1941) for elaborating Fisher's scoring method for contingency tables. Guttman (1946) further extended his approach of internal consistency to rank-order and paired-comparison data. Thus, solid foundations were laid by 1946.

1.2.3. Period of Rediscoveries and Further Developments

We can list Mosier (1946), Fisher (1948), Johnson (1950), Hayashi (1950, 1952), Bartlett (1951), Williams (1952), Bock (1956, 1960), Lancaster (1958), Lord (1958), Torgerson (1958), and many other contributors. Among others, there were four major groups of researchers: the Hayashi school in Japan since 1950, the Benzécri school in France since the early 1960s, the Leiden group in the Netherlands since the late 1960s, and the Toronto group in Canada since the late 1960s.

Because of its special appeal to researchers in various countries and different disciplines, the method has acquired many aliases, mostly through rediscoveries of essentially the same technique—among others, the method of reciprocal averages (Horst, 1935; Richardson & Kuder, 1933), simultaneous linear regressions (Hirschfeld, 1935; Lingoes, 1964), appropriate scoring and additive scoring (Fisher, 1948), principal component analysis of categorical data (Torgerson, 1958), optimal scaling (Bock, 1960), correspondence analysis (Benzécri, 1969; Escofier-Cordier, 1969), biplot (Gabriel, 1971), canonical analysis of categorical data (de Leeuw, 1973), reciprocal averaging (Hill, 1973), basic structure content scaling (Jackson & Helmes, 1979), dual scaling (Nishisato, 1980), homogeneity analysis (Gifi, 1980), centroid scaling (Noma, 1982), multivariate descriptive statistical analysis (Lebart, Morineau, & Warwick, 1984), nonlinear multivariate analysis (Gifi, 1990), and nonlinear biplot (Gower & Hand, 1996). Because all of these are based on singular value decomposition of categorical data, they are either mathematically identical or not much different from one another.

1.2.4. Dual Scaling

The name *dual scaling* (DS) was coined by Nishisato (1980) as a result of the discussion at the symposium on optimal scaling during the 1976 annual meeting of the Psychometric Society in Murray Hill, New Jersey (see Nishisato & Nishisato, 1994a). With the general endorsement among the participants, he adopted it in the title of his 1980 book. Franke (1985) states that he "uses Nishisato's term for its generality and lack of ambiguity" (p. 63).

Under the name *dual scaling*, Nishisato has extended its applicability to a wider variety of categorical data, including both incidence data and dominance data. This aspect of DS is reflected in Meulman's (1998) statement that "dual scaling is a comprehensive framework for multidimensional analysis of categorical data" (p. 289). For those interested in the history of quantification theory, see de Leeuw (1973), Benzécri (1982), Nishisato (1980), Greenacre (1984), Gifi (1990), Greenacre and Blasius (1994), and van Meter, Schiltz, Cibois, and Mounier (1994).

1.3. An Intuitive Introduction to Dual Scaling

1.3.1. Is Likert Scoring Appropriate?

Suppose subjects were asked two multiple-choice questions.[1]

Q1: What do you think of taking sleeping pills? (1) *strongly disagree*, (2) *disagree*, (3) *indifferent*, (4) *agree*, (5) *strongly agree*

Q2: Do you sleep well every night? (1) *never*, (2) *rarely*, (3) *sometimes*, (4) *often*, (5) *always*

The data are in Table 1.1. Likert scores are often used for ordered sets of categories (Likert, 1932). Suppose we assign −2, −1, 0, 1, 2 to the five ordered categories of each set in the above example. Our question here is if these Likert scores are appropriate. There is a simple way to examine it.

First, we calculate the mean of each category, using Likert scores. For example, the mean of category *never* is $[15 \times (-2) + 5 \times (-1) + 6 \times 0 + 0 \times 1 + 1 \times 2]/27 = -1.2$. Likewise, we calculate the means of row categories and those of column categories, which are summarized in Table 1.2. We now plot those averages against the original scores (−2, −1, 0, 1, 2), as seen in Figure 1.1. The two lines are relatively close to a straight line, which indicates that the original scores are "pretty good." Suppose we use, instead of those subjective category weights, the weights derived by DS and calculate the weighted category means and plot these against the DS weights. We then obtain Figure 1.2.

Notice that the two lines are now merged into a single straight line. This is "mathematically optimal," as seen later. We will also see shortly that the slope of the line in Figure 1.2 is equal to the maximal "nontrivial" singular value for this data set.

But how do we arrive at the DS weights? It is simple: Once we obtain the mean category scores as in Figure 1.1, replace the original scores (e.g., −2, −1, etc.) with the corresponding mean scores, and then calculate the new mean category scores in the same way as before and plot the new category scores against the first mean scores, replace the old mean scores with the new mean scores, and calculate new mean category scores and plot them. This is a convergent process (Nishisato, 1980, pp. 60–62, 65–68). Horst (1935) called the above process the *method of reciprocal averages* (MRA), used by Richardson and Kuder (1933), also suggested by Fisher (1940), and fully illustrated by Mosier (1946). MRA is one of the algorithms for DS.

1. With permission from Nishisato (1980).

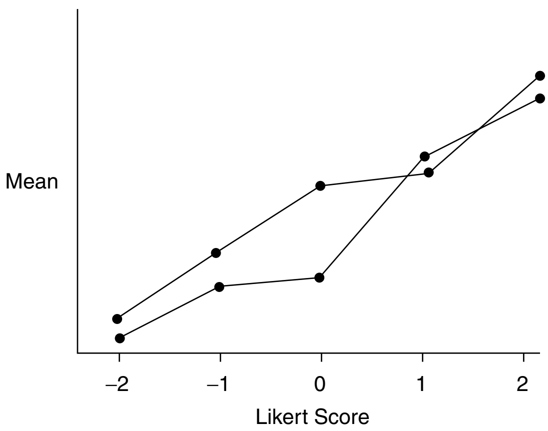
Table 1.1 Sleeping and Sleeping Pills

	Never	*Rarely*	*Sometimes*	*Often*	*Always*	*Sum*	*Score*
Strongly against	15	8	3	2	0	28	−2
Against	5	17	4	0	2	28	−1
Neutral	6	13	4	3	2	28	0
For	0	7	7	5	9	28	1
Strongly for	1	2	6	3	16	28	2
Sum	27	47	24	13	29	140	
Score	−2	−1	0	1	2		

Table 1.2 Likert Scores and Weighted Means

Score	*Mean*	*Score*	*Mean*
−2	−1.2	−2	−1.3
−1	−0.5	−1	−0.8
0	0.4	0	−0.6
1	0.5	1	0.6
2	1.3	2	1.1

Figure 1.1 Likert Scores



1.3.2. The Method of Reciprocal Averages (MRA)

Let us illustrate the process of MRA.[2] Suppose three teachers (White, Green, and Brown) were rated on their teaching performance by students (see Table 1.3).

2. With permission from Nishisato and Nishisato (1994a).

Figure 1.2 Dual-Scaling Optimal Weights

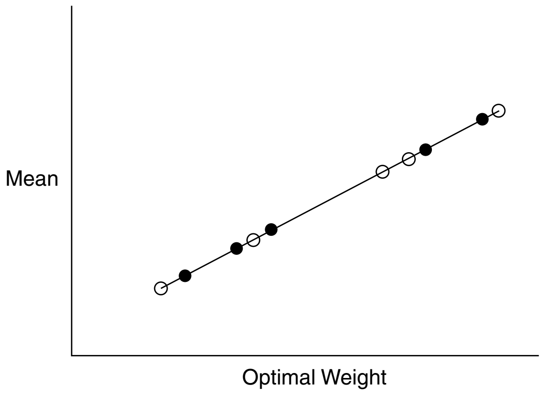


Table 1.3 Evaluating Teachers

Teacher	*Good*	*Ave*	*Poor*	*Total*
White	1	3	6	10
Green	3	5	2	10
Brown	6	3	0	9
Total	10	11	8	29

The MRA is carried out in the following way:

Step 1: The MRA starts with assigning arbitrary weights to columns (or rows, if preferred). Although such values are arbitrary, one must avoid identical weights for all columns (or rows), including zero. It is always a good strategy to use "reasonable" values. As an example, consider the following:

$$x_1(\text{good}) = 1, \quad x_2(\text{average}) = 0, \quad x_3(\text{poor}) = -1. \tag{1}$$

Step 2: Calculate the weighted averages of the rows:

$$y_1(\text{White}) = \frac{1 \times x_1 + 3 \times x_2 + 6 \times x_3}{10} = \frac{1 \times 1 + 3 \times 0 + 6 \times (-1)}{10} = -0.5, \tag{2}$$

$$y_2(\text{Green}) = \frac{3 \times 1 + 5 \times 0 + 2 \times (-1)}{10} = 0.1000, \tag{3}$$

$$y_3(\text{Brown}) = \frac{6 \times 1 + 3 \times 0 + 0 \times (-1)}{9} = 0.6667. \tag{4}$$

Step 3: Calculate the mean responses weighted by y_1, y_2, y_3:

$$M = \frac{10y_1 + 10y_2 + 9y_3}{29} = \frac{10 \times (-0.5) + 10 \times 0.1 + 9 \times 0.6667}{29} = 0.0690. \tag{5}$$

Step 4: Subtract M from each of y_1, y_2, y_3, and adjusted values should be indicated again by y_1, y_2, y_3, respectively:

$$y_1 = -0.5000 - 0.0690 = -0.5690, \tag{6}$$

$$y_2 = 0.1000 - 0.0690 = 0.0310, \tag{7}$$

$$y_3 = 0.6667 - 0.0690 = 0.5977. \tag{8}$$

Step 5: Divide y_1, y_2, y_3 by the largest absolute value of y_1, y_2, y_3, say, g_y. At this stage, $g_y = 0.5977$. Adjusted values should again be indicated by y_1, y_2, y_3:

$$y_1 = \frac{-0.5690}{0.5977} = 0.9519, \quad y_2 = \frac{0.0310}{0.5977} = 0.0519, \quad y_3 = \frac{0.5977}{0.5977} = 1.0000. \tag{9}$$

Step 6: Using these new values as weights, calculate the averages of the columns:

$$x_1 = \frac{1y_1 + 3y_2 + 6y_3}{10} = \frac{1 \times (-0.9519) + 3 \times 0.0519 + 6 \times 1.0}{10} = 0.5204, \tag{10}$$

$$x_2 = \frac{3 \times (-0.9519) + 5 \times 0.0519 + 3 \times 1.0000}{11} = 0.0367, \tag{11}$$

$$x_3 = \frac{6 \times (-0.9519) + 2 \times 0.0519 + 0 \times 1.0000}{8} = -0.7010. \tag{12}$$

Step 7: Calculate the mean responses weighted by x_1, x_2, x_3:

$$N = \frac{10 \times 0.5204 + 11 \times 0.0367 + 8 \times (-0.7010)}{29} = 0. \tag{13}$$

Step 8: Subtract N from each of x_1, x_2, x_3.

Table 1.4 Iterative Results

	Iter2 y	*Iter2 x*	*Iter3 y*	*Iter3 x*	*Iter4 y*	*Iter4 x*	*Iter5 y*	*Iter5 x*
1	−0.9954	0.7321	−0.9993	0.7321	−0.9996	0.7311	−0.9996	0.7311
2	0.0954	0.0617	0.0993	0.0625	0.0996	0.0625	0.0996	0.0625
3	1.0000	−1.0000	1.0000	−1.0000	1.0000	−1.0000	1.0000	−1.0000
g	0.5124	0.7227	0.5086	0.7246	0.5083	0.7248	0.5083	0.7248

Step 9: Divide each element of x_1, x_2, x_3 by the largest absolute value of the three numbers, say, g_x. Because -0.7010 has the largest absolute value, $g_x = 0.7010$. Adjusted values are indicated again by x_1, x_2, x_3:

$$x_1 = \frac{0.5204}{0.7010} = 0.7424,$$

$$x_2 = \frac{0.0367}{0.7010} = 0.0524,$$

$$x_3 = \frac{-0.7010}{0.7010} = -1.0000. \qquad (14)$$

Reciprocate the above averaging processes (Steps 2 through 9) until all the six values are stabilized. Iteration 5 provides the identical set of numbers as Iteration 4 (see Table 1.4). Therefore, the process has converged to the optimal solution in four iterations. Notice that the largest absolute values at each iteration, g_y and g_x, also converge to two constants, 0.5083 and 0.7248. Nishisato (1988) showed that the eigenvalue, ρ^2, is equal to the product, $g_y g_x = 0.5083 \times 0.7248 = 0.3648$, and the singular value, ρ, is the geometric mean,

$$\rho = \text{singular value} = \sqrt{g_y g_x}$$
$$= \sqrt{0.5083 \times 0.7248} = 0.6070. \qquad (15)$$

If we start with the cross-product symmetric table, instead of the raw data (the present example), the process will converge to one constant of g, which is the *eigenvalue*, and its positive square root is the *singular value* (Nishisato, 1980). See Nishisato (1994, p. 89) for why the final value of g is the eigenvalue.

Step 10: In the DUAL3 for windows (Nishisato & Nishisato, 1994b), the unit of weights is chosen in such a way that the sum of squares of weighted responses is equal to the number of responses. In this case, the constant multipliers for adjusting the unit of y (say, c_r) and x (c_c) are given by

$$c_r = \sqrt{\frac{29}{10y_2^1 + 10y_2^2 + 9y_3^2}} = 1.2325,$$

$$c_c = \sqrt{\frac{29}{10x_2^1 + 11x_2^2 + 8x_3^2}} = 1.4718. \qquad (16)$$

Table 1.5 Two Types of Optimal Weights

	Normed y	*Normed x*	*Projected y*	*Projected x*
1	−1.2320	1.0760	−0.7478	0.6531
2	0.1228	0.0920	0.0745	0.0559
3	1.2325	−1.4718	0.7481	−0.8933

The final weights are obtained by multiplying y_1, y_2, y_3 by c_r and x_1, x_2, x_3 by c_c. These weights are called *normed weights*. The normed weights, multiplied by the singular value—that is, ρy_i and ρx_j—are called *projected weights*, which reflect the relative importance of categories. The distinction between these two types of weights will be discussed later. In the meantime, let us remember that normed weights and projected weights are what Greenacre (1984) calls *standard coordinates* and *principal coordinates*, respectively, and that projected weights are the important ones because they reflect relative importance of the particular solution (component, dimension). The final results are in Table 1.5. These weights thus obtained are scaled in such a way that (a) the sum of responses weighted by y is zero, and the sum of responses weighted by x is zero; (b) the sum of squares of responses weighted by y is the total number of responses, and the same for x. Once the first solution is obtained, calculate the residual frequencies, and apply the MRA to the residual table to obtain the second solution. This process will be discussed later.

1.4. Two Types of Categorical Data

Nishisato (1993) classified categorical data into two distinct groups, *incidence data* (e.g., contingency tables, multiple-choice data, sorting data) and *dominance data* (e.g., rank-order, paired-comparison data).

1.4.1. Incidence Data

Elements of data are 1 (presence), 0 (absence), or frequencies, as we see in contingency tables, multiple-choice data, and sorting data. DS of incidence data is characterized by (a) the use of the "chi-square metric" (Greenacre, 1984; Lebart et al., 1984; Nishisato & Clavel, 2003), (b) a lower rank approximation to input data, (c) "a trivial solution" (Gifi, 1990; Greenacre, 1984; Guttman, 1941; Nishisato, 1980, 1994), and (d) more than one dimension needed to describe the data (Nishisato, 2002, 2003). This last point is true even when all variables are perfectly correlated to one another. *Correspondence analysis* and *multiple correspondence analysis* were originally developed in France specifically for incidence data for the contingency table and multiple-choice data, respectively.

1.4.2. Dominance Data

Elements of data are greater than, equal to, or smaller than, as we see in rank-order data and paired-comparison data. Because the information is typically given in the form of inequality relations, without any specific amount of the discrepancy between the two attributes or stimuli indicated, it is not possible to approximate the value of the data directly as is done with the incidence data. Instead, the objective here is to derive new measurements for objects in such a way that the ranking of the measurements best approximates the corresponding ranking of the original dominance data. DS of dominance data is characterized by (a) the use of the Euclidean metric (Nishisato, 2002), (b) a lower rank approximation to the *ranks* of the data (Nishisato, 1994, 1996), (c) no trivial solution (Greenacre & Torres-Lacomba, 1999; Guttman, 1946; Nishisato, 1978; van de Velden, 2000), and (d) one dimension to describe the data when all variables are perfectly correlated to one another (Nishisato, 1994, 1996).

1.4.3. Scope of Dual Scaling

DS is applicable not only to the incidence data but also to the dominance data. The DUAL3 for Windows (Nishisato & Nishisato, 1994b), a computer program package for DS, handles both types of categorical data. Recently, Greenacre and Torres-Lacomba (1999) and van de Velden (2000) reformulated correspondence analysis for dominance data, which were not much different from Nishisato's (1978) earlier study. After all, they are all based on singular-value decomposition.

1.5. Scaling of Incidence Data

1.5.1. Contingency Tables

Contingency tables are often used to summarize data. For example, a small survey on the popularity of five movies, collected from three age groups, can be summarized into a 5 × 3 table of the number of people in each cell. Similarly, we often see a large number of tabulation tables on voting behavior, typically on two categorical variables (e.g., age and education). These are contingency tables.

1.5.1.1. Some Basics

Consider an n-by-m contingency table with typical element f_{ij}. DS first eliminates from this table the frequencies expected when rows and columns are statistically independent, that is, $f_{i.}f_{.j}/f_t$, where f_t is the total frequency in the table. This is called a trivial solution. Then, the residual table, consisting of typical elements for row i and column j, say,

$$f_{ij} - \frac{f_{i.}f_{.j}}{f_t} = f_{ij} - h_{ij}, \tag{17}$$

is decomposed into independent components, called solutions. Let $\min(n, m)$ be the smaller value of n and m. Then the n-by-m residual table can be exhaustively explained by at most $[\min(n, m) - 1]$ solutions. In other words, the total number of nontrivial solutions, that is, proper solutions $T(\text{sol})$, is given by

$$T(\text{sol}) = \min(n, m) - 1. \tag{18}$$

The variance of solution k is called the eigenvalue, ρ_k^2, which is a measure of information conveyed by solution k. The total information contained in the residual matrix, $T(\text{inf})$, is the sum of the $[\min(n, m) - 1]$ eigenvalues, which is equal to

$$T(\text{inf}) = \sum_{k=1}^{p} \rho_k^2 = \frac{\chi^2}{f_t}, \text{ where}$$
$$\chi^2 = \sum_{i}^{n} \sum_{j}^{m} \frac{(f_{ij} - h_{ij})^2}{h_{ij}}, \tag{19}$$

and h_{ij} is the frequency expected when the ith row and the jth column are statistically independent. The percentage of the total information explained by solution k is indicated by δ_k and is given by

$$\delta_k = \frac{100\rho_k^2}{T(\text{inf})}. \tag{20}$$

1.5.1.2. Example: Biting Habits of Laboratory Animals

The biting habits of four laboratory animals were investigated. The following data were obtained from Sheskin's (1997) book.[3] Because this is a small example, let us list the main output from the program DUAL3 (Nishisato & Nishisato, 1994b) (Table 1.7).

Because this data set is a 4×3 table, $T(\text{sol}) = 2$, and the analysis shows that δ_1 and δ_2 are 94.2% and 5.8%, respectively. The order-0 approximation is the trivial solution. The trivial solution is removed from the data, and the residual table is analyzed into components. The order-1 approximation is what one can predict from the trivial solution and Solution 1:

$$f^*_{ij(1)} = \frac{f_{i.}f_{.j}}{f_t}[1 + \rho_1 y_{i1} x_{j1}]. \quad (21)$$

Because the value of δ_1 is 94.2% (the contribution of Solution 1), this approximation to the input data is very good, and the residual table does not contain much more information to be analyzed. In the current example, the order-2 approximation perfectly reproduces the input data:

$$f^*_{ij(2)} = \frac{f_{i.}f_{.j}}{f_t}[1 + \rho_1 y_{i1} x_{j1} + \rho_2 y_{i2} x_{j2}]. \quad (22)$$

See also the residual table (Table 1.7), which shows no more information left to be analyzed. Notice that it is not clear what relations between the animals and biting habits are from the input table, but see the graph based on DS: The two-dimensional graph (Figure 1.3) shows, among other things, that (a) guinea pigs are flagrant biters, (b) mice are between flagrant biters and mild biters, (c) mild biters and nonbiters are relatively closely located, (d) gerbils are nonbiters, and (e) hamsters are between mild biters and nonbiters. The graph is much easier to understand than the original table.

Table 1.6 Sheskin's Data on Biting Habits of Laboratory Animals

Animals	*Not a Biter*	*Mild Biter*	*Flagrant Biter*
Mice	20	16	24
Gerbils	30	10	10
Hamsters	50	30	10
Guinea pigs	19	11	50

1.5.2. Multiple-Choice Data

Multiple-choice data are ubiquitous in psychological research, particularly in personality, social, and clinical research. We should question, however, how arbitrarily such data are typically analyzed. When response options are ordered (e.g., never, sometimes, often, always), researchers often use the integer scores 1, 2, 3, and 4 for these ordered categories and analyze the data. This practice of using the so-called Likert scores is by no means effective in retrieving information in data. We will see this problem very shortly. In contrast, dual scaling can analyze such multiple-choice data in a very effective way in terms of information retrieval. We will see an example of dual-scaling analysis shortly.

1.5.2.1. Some Basics

Consider n multiple-choice items, with item j having m_j options. Consider further that each of N subjects is asked to choose one option per item. Let m be the total number of options of n items. For DS, multiple-choice data are expressed in the form of (1,0) response patterns (see the example in 1.5.2.2) and also have a trivial solution. The aforementioned statistics of multiple-choice data are as follows:

$$T(\text{sol}) = m - n \text{ or } N - 1, \text{ whichever is smaller.} \quad (23)$$

$$T(\text{inf}) = \sum_{k=1}^{m-n} \rho_k^2 = \frac{\sum_{j=1}^{n} m_j}{n} - 1 = \overline{m} - 1. \quad (24)$$

The definition of δ_k is the same as the contingency table, but in practice we will modify it as we discuss later. Option weights are determined, as Lord (1958) proved, to yield scores with a maximal value of the generalized Kuder-Richardson internal consistency reliability, or Cronbach's α (Cronbach, 1951), which can be inferred from the following relations (Nishisato, 1980):

$$\alpha = 1 - \frac{1-\rho^2}{(n-1)\rho^2} = \frac{n}{n-1}\left(\frac{\sum r_{jt}^2 - 1}{\sum r_{jt}^2}\right) \text{ since}$$
$$\rho^2 = \frac{\sum_j^n r_{jt}^2}{n}, \quad (25)$$

where r_{jt}^2 is the square of correlation between item j and the total score. It is known (Nishisato, 1980, 1994) that the average information in multiple-choice data, that is—$T(\text{inf})/T(\text{sol})$—is $1/n$ and that α becomes negative when ρ^2 is smaller than the average information. Therefore, Nishisato (1980, 1994) suggests

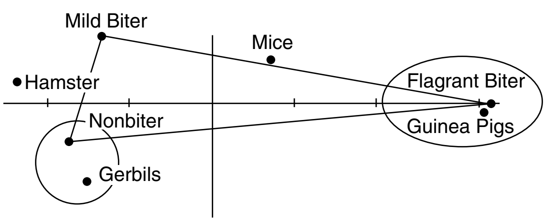
3. Reprinted with permission from Sheskin (1997).

Table 1.7 Approximation to Input

ORDER 0 APPROXIMATION			*RESIDUAL MATRIX*			
25.5	14.4	20.1	−6.5	1.6	3.9	
21.3	12.0	16.8	8.8	−2.0	−6.8	
38.3	21.5	30.2	11.8	8.5	−20.2	
34.0	19.1	26.9	−15.0	−8.1	23.1	
ORDER 1 APPROXIMATION			*RESIDUAL MATRIX*			*SOLUTION 1*
22.7	13.1	24.2	−2.7	2.9	−0.2	Eigenvalue = 0.20
26.1	14.2	9.7	3.9	−4.2	0.3	Singular value = 0.45
52.1	27.8	10.2	−2.1	2.2	−0.2	Delta = 94.2%
18.1	12.0	49.9	0.9	−1.0	0.1	CumDelta = 94.2
ORDER 2 APPROXIMATION			*RESIDUAL MATRIX*			*SOLUTION 2*
20.0	16.0	24.0	0.0	0.0	0.0	Eigenvalue = 0.01
30.0	10.0	10.0	0.0	0.0	0.0	Singular value = 0.11
50.0	30.0	10.0	0.0	0.0	0.0	Delta = 5.8%
19.0	11.0	50.0	0.0	0.0	0.0	CumDelta = 100.0

	PROJECTED WEIGHTS			*PROJECTED WEIGHTS*	
	Sol-1	*Sol-2*		*Sol-1*	*Sol-2*
Mice	0.14	0.12	Not a biter	−0.34	−0.10
Gerbils	−0.30	−0.21	Mild biter	−0.27	0.19
Hamsters	−0.47	0.06	Flagrant biter	0.63	−0.01
Guinea pigs	0.61	−0.03			

Figure 1.3 Biting Habits of Four Animals



stopping the extraction of solutions as soon as ρ^2 becomes smaller than $1/n$. Accordingly, we redefine the statistic δ_k as the percentage of ρ_k^2 over the sum of ρ_j^2 greater than $1/n$.

1.5.2.2. Example: Blood Pressure, Migraines, and Age

As mentioned earlier, Torgerson (1958) called DS "principal component analysis of categorical data." Because principal component analysis (PCA) is a method to find a linear combination of continuous variables (PCA) and that of categorical variables (DS), it would be interesting to look at differences between them. The following example is adopted from Nishisato (2000):

1. How would you rate your blood pressure? (Low, Medium, High): coded 1, 2, 3
2. Do you get migraines? (Rarely, Sometimes, Often): 1, 2, 3 (as above)
3. What is your age group? (20–34, 35–49, 50–65): 1, 2, 3
4. How would you rate your daily level of anxiety? (Low, Medium, High): 1, 2, 3
5. How would you rate your weight? (Light, Medium, Heavy): 1, 2, 3
6. What about your height? (Short, Medium, Tall): 1, 2, 3

Suppose we use the traditional Likert scores for PCA—that is, 1, 2, 3 as scores for the three categories of each question. DS uses response patterns of 1s and 0s. See the two data sets from 15 subjects in Table 1.8 and the product-moment correlation matrix for PCA in Table 1.9. Examine the correlation between blood pressure (BP) and age (Age) ($r = 0.66$) and that between BP and migraines (Mig) ($r = -0.06$) using the data in the contingency table format (Table 1.10).

Notice a linear relation between BP and Age and a nonlinear relation between BP and Mig. It seems that the nonlinear relation between BP and Mig is much clearer than the linear relation between BP and Age:

Table 1.8 Likert Scores for PCA and Response Patterns for DS

	PCA						DS					
Subject	*Bpr Q1*	*Mig Q2*	*Age Q3*	*Anx Q4*	*Wgt Q5*	*Hgt Q6*	*Bpr 123*	*Mig 123*	*Age 123*	*Anx 123*	*Wgt 123*	*Hgt 123*
1	1	3	3	3	1	1	100	001	001	001	100	100
2	1	3	1	3	2	3	100	001	100	001	010	001
3	3	3	3	3	1	3	001	001	001	001	100	001
4	3	3	3	3	1	1	001	001	001	001	100	100
5	2	1	2	2	3	2	010	100	010	010	001	010
6	2	1	2	3	3	1	010	100	010	001	001	100
7	2	2	2	1	1	3	010	010	010	100	100	001
8	1	3	1	3	1	3	100	001	100	001	100	001
9	2	2	2	1	1	2	010	010	010	100	100	010
10	1	3	2	2	1	3	100	001	010	010	100	001
11	2	1	1	3	2	2	010	100	100	001	010	010
12	2	2	3	3	2	2	010	010	001	001	010	010
13	3	3	3	3	3	1	001	001	001	001	001	100
14	1	3	1	2	1	1	100	001	100	010	100	100
15	3	3	3	3	1	2	001	001	001	001	100	010

Table 1.9 Product-Moment Correlation Based on Likert Scores

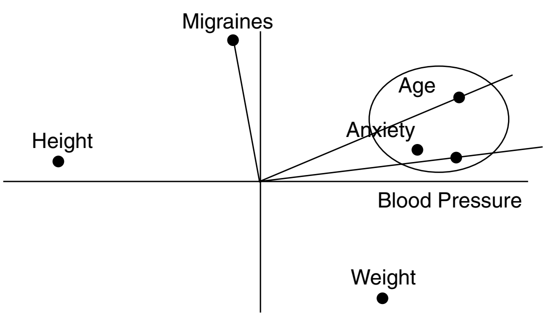
	BP	*Mig*	*Age*	*Anx*	*Wgt*	*Hgt*
Blood pressure (BP)	1.00					
Migraine (Mig)	−.06	1.00				
Age (Age)	.66	.23	1.00			
Anxiety (Anx)	.18	.21	.22	1.00		
Weight (Wgt)	.17	−.58	−.02	.26	1.00	
Height (Hgt)	−.21	.10	−.30	−.23	−.31	1.00

Table 1.10 Relation of Blood Pressures to Age and Migraines

	Age			*Migraine*		
	20–34	*35–49*	*50–65*	*Rarely*	*Sometimes*	*Often*
High BP	0	0	4	0	0	4
Mid BP	1	4	1	3	3	0
Low BP	3	1	1	0	0	5

“If you have frequent migraines, your blood pressure is either high or low.” The first two principal components of Likert scores are plotted in Figure 1.4. Notice that it captures only linear relations. The data for DS are expressed in terms of chosen response patterns, and the units of analysis are response options, not items as in the case of PCA. PCA is a method to determine the most informative weighted combinations of items, whereas DS looks for the most informative weighted combinations of categories of items. This means that DS yields an inter-item correlation matrix for each solution, rather than one for the entire data set as in PCA.

Figure 1.4 Two Solutions From Principal Component Analysis



The current data yield four solutions associated with positive values of reliability coefficient α (see Table 1.11).

The adjusted delta is the one redefined in terms of solutions associated with positive values of reliability α. CumDelta and CumAdjDelta are cumulative values of delta and adjusted delta, respectively. For the limited space, we will look at only the first two solutions and their projected option weights (see Table 1.12). Notice that the weights for options of BP and Mig for Solution 1 are weighted in such a way that the nonlinear relation is captured. Study the weights to convince yourself. Using these weights, inter-item

Table 1.11 Four Solutions

	Solution 1	*Solution 2*	*Solution 3*	*Solution 4*
Eigenvalue	0.54	0.37	0.36	0.31
Singular value	0.74	0.61	0.59	0.55
Delta	27	19	17	15
CumDelta	27	46	63	79
Adjusted delta	34	24	22	20
CumAdjDelta	34	58	80	100

Table 1.12 Projected Option Weight of Two Solutions

	Solution 1	*Solution 2*
Blood Pressure		
Low	−0.71	0.82
Medium	1.17	−0.19
High	−0.86	−0.74
Anxiety		
Low	1.55	1.21
Medium	0.12	0.31
High	−0.35	−0.33
Migraine		
Rarely	1.04	−1.08
Sometimes	1.31	0.70
Often	−0.78	0.12
Weight		
Light	−0.27	0.46
Medium	0.32	0.01
Heavy	0.50	−1.40
Age		
20–34	0.37	0.56
35–49	1.03	0.22
50–65	−0.61	−0.56
Height		
Short	−0.56	−0.63
Medium	0.83	−0.35
Tall	−0.27	0.98

correlation matrices are obtained for the two DS solutions (see Table 1.13).

BP and Mig are now correlated at 0.99 in Solution 1. This was attained by assigning similar weights to high BP, low BP, and frequent migraines, which are very different from the weights given to medium BP, rare migraines, and occasional migraines. The same correlation for Solution 2 is 0.06. Characteristics of the first two DS solutions can be obtained by putting options of similar weights together (see Table 1.14). "Nonlinear combinations" of response categories are involved in each solution. In DS, linear correlation is maximized by transforming categories linearly or nonlinearly, depending on the data, whereas PCA filters out all nonlinear relations in the process of analysis, which is why it is called linear analysis. The first two DS solutions are plotted in Figure 1.5. Unlike PCA solutions, three categories of a single variable are not forced to be on a single line but usually form a triangle, the area of which is monotonically related to the contribution of the variable to these dimensions. PCA can never reveal a strong relation between BP and Mig, but this relation is the most dominant one in DS. In DS, high and low BP are associated with frequent migraines, but the second dimension identifies a different association between low and high BP—the former with young, skinny, and tall subjects and the latter with old, heavy, and short subjects.

1.5.3. Sorting Data

Sorting data are not as popular as contingency tables and multiple-choice data, but in some areas, such as cognitive psychology, we often see references to sorting data. So, in this section, we will learn how sorting data are collected and optimally analyzed by dual scaling.

1.5.3.1. Some Basics

Sorting data are collected in the following way. Consider the first object to be a member of the first pile and assign 1 to it; go down the list, and each time you find an object similar to the first object, assign 1 to it. When you finish identifying all the objects with 1, go to the next object that has not been chosen so far and give it 2; go down the list and identify all the objects that are similar to the object with number 2. In this way, you classify all objects on the list into piles. Takane (1980) demonstrated that DS can be used to analyze sorting data by transposing the data or exchanging the roles of subjects and item options in multiple-choice data with objects and subject piles in sorting data, respectively. With this understanding, $T(\text{sol})$ and $T(\text{inf})$ are the same as those of multiple-choice data.

1.5.3.2. Example: Sorting 19 Countries Into Similar Groups

The data in Table 1.15 were collected from Nishisato's class in 1990. The last two columns of the table indicate the optimal (projected) weights of the countries on the first two solutions. Note that prior to DS analysis, the data are first transformed to (1, 0) response patterns, as was the case of multiple-choice data. One of the outcomes is the inter-subject correlation matrix, just like the inter-item correlation matrix in multiple-choice data. Table 1.16 shows the

Table 1.13 Correlation Matrices From Two DS Solutions

	Solution 1						*Solution 2*					
	BP	*Mig*	*Age*	*Anx*	*Wgt*	*Hgt*	*BP*	*Mig*	*Age*	*Anx*	*Wgt*	*Hgt*
BP	1.0						1.0					
Mig	.99	1.0					.06	1.0				
Age	.60	.58	1.0				.59	−.31	1.0			
Anx	.47	.52	.67	1.0			.07	.35	.35	1.0		
Wgt	.43	.39	.08	−.33	1.0		.28	.62	−.01	.19	1.0	
Hgt	.56	.57	.13	.19	.20	1.0	.31	.29	.32	.17	.38	1.0

Table 1.14 Characteristics of Two DS Solutions

Solution 1		*Solution 2*	
One End	*The Other End*	*One End*	*The Other End*
Low BP	Medium BP	High BP	Low BP
High BP	Rare migraine	Rare migraine	Occasional migraine
Frequent migraine	Middle age	Old	Young
Old age group	Low anxiety	Heavy	Tall
High anxiety	Medium height	Short	
Short			

Figure 1.5 First Two Dual-Scaling Solutions

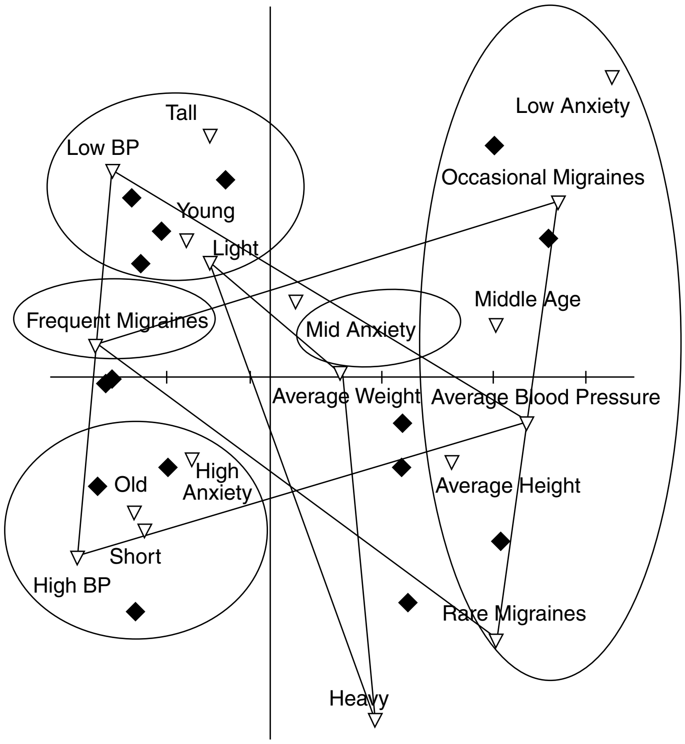


Table 1.15 Sorting of 19 Countries by Five Subjects

Country	*S1*	*S2*	*S3*	*S4*	*S5*	*Solution 1*	*Solution 2*
Britain	1	1	1	1	1	−0.50	−0.69
Canada	5	2	2	2	1	1.06	−0.81
China	2	3	3	3	2	1.53	0.52
Denmark	1	1	1	1	3	−0.73	−0.71
Ethiopia	3	5	5	4	4	−1.00	2.15
Finland	1	4	1	1	3	−0.81	−0.71
France	1	1	1	1	5	−0.73	−0.71
Germany	1	4	1	5	8	−0.50	−0.60
India	4	3	4	3	6	1.02	0.81
Italy	1	4	5	5	7	−0.93	−0.17
Japan	2	3	6	2	8	1.21	−0.01
New Zealand	4	1	6	1	1	0.24	−0.31
Nigeria	3	5	4	4	4	−0.76	2.34
Norway	1	4	1	1	3	−0.81	−0.71
Singapore	4	3	6	3	8	1.12	0.24
Spain	1	5	5	1	7	−0.92	0.34
Switzerland	1	4	1	5	5	−0.85	−0.71
Thailand	4	3	6	3	6	1.20	0.46
United States	5	2	2	2	8	1.17	−0.73

Table 1.16 Inter-Subject Correlation for Two DS Solutions

	Solution 1					*Solution 2*				
Subject 1	1.00					1.00				
Subject 1	0.90	1.00				0.63	1.00			
Subject 3	0.93	0.82	1.00			0.60	0.90	1.00		
Subject 4	0.88	0.99	0.81	1.00		0.98	0.67	0.63	1.00	
Subject 5	0.77	0.87	0.75	0.85	1.00	0.90	0.87	0.82	0.90	1.00

inter-subject correlation matrices associated with the two solutions. In both solutions, the correlation between subjects is relatively high. Figure 1.6 shows only the configuration of 18 of the 19 countries (France is missing because it occupies the same point as Denmark) captured by the first two solutions. The graph clearly shows geographical similarities of the countries.

One commonly observed characteristic of sorting data is that there are often too many dominant solutions to interpret. It must be a reflection of the freedom that the subjects can enjoy in terms of the number of piles and the sizes of piles that are completely in the hands of the subjects. The δ values of the first eight solutions are 19%, 18%, 16%, 11%, 9%, 7%, 6%, and 5%, an unusually gradual drop in percentage from solution to solution. This poses in practice a problem of how many solutions to extract and interpret.

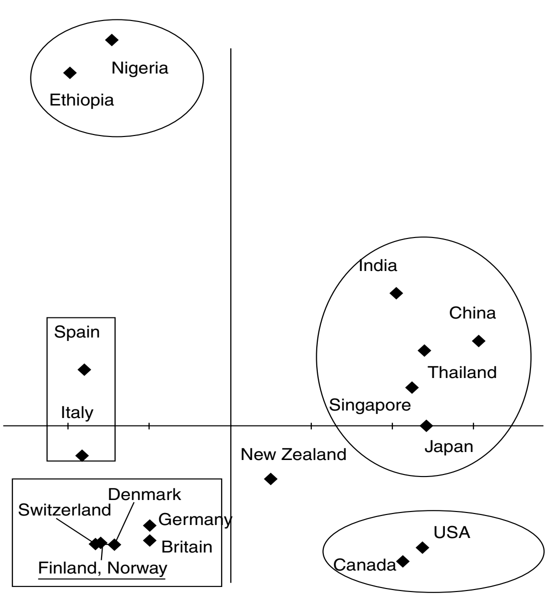
1.6. Scaling of Dominance Data

We will discuss only rank-order data and paired-comparison data. As for DS of successive categories data, see Nishisato (1980, 1986,1994), Nishisato and Sheu (1980), and Odondi (1997).

1.6.1. Rank-Order Data

Ranking is a very popular task in psychological research. For instance, we ask people to rank a number of candidates for a committee and choose the winner in terms of the average ranks of the candidates. Although this popular method for processing ranking data looks reasonable, it is far from even being good and is rather misleading. Why? We will see why such averaged ranks should not be used to evaluate candidates or

Figure 1.6 Sorting of 19 Countries



voters, which becomes obvious once we analyze the same ranking data with dual scaling.

1.6.1.1. Some Basics

Suppose that each of N subjects ranks all of n objects, according to the order of preference, with 1 being the first choice and n being the last choice. Assuming that the number of subjects is greater than that of the objects, the total number of solutions and the total information from the data are given by the following:

$$T(\text{sol}) = n - 1 \text{ and } T(\text{inf}) = \frac{n+1}{3(n-1)}. \qquad (26)$$

When dominance data are subjected to DS, the original rank-order data are first converted to a dominance table. Let us indicate by R_{ij} the rank given to object j by subject i. Then, assuming that each subject ranks n objects, the corresponding dominance number, e_{ij}, is given by the formula

$$e_{ij} = n + 1 - 2R_{ij}, \qquad (27)$$

where e_{ij} indicates the number of times subject i ranked object j before other objects minus the number of times the subject ranked it after other objects. So it indicates relative popularity of each object within each subject. The sum of dominance numbers for each subject is always zero, and the dominance number is bounded between $-(n-1)$ and $(n-1)$. Because dominance numbers are *ipsative* (i.e., each row sum is a constant), we must modify the process of MRA by redefining each row marginal to be $n(n-1)$ and that of column $N(n-1)$. The total number of responses in the dominance table is $Nn(n-1)$. These numbers are based on the fact that each element in the dominance table is the result of $(n-1)$ comparisons between each object and the remaining $(n-1)$ objects (Nishisato, 1978). Using these redefined marginals, we may use MRA for analysis.

The ipsative property of dominance numbers has another implication for quantification: There is no centering constraint on weights for subjects. Thus, the weights for subjects can be all positive or negative. This aspect of quantification of dominance data is very different from that of incidence data, in which both weights for subjects and those for stimuli are centered within each set.

1.6.1.2. Example: Ranking of Municipal Services

Table 1.17 contains ranking of 10 municipal services by 31 students, collected from Nishisato's class in 1982, together with the dominance table. If there were no individual differences, the reasonable scale values or satisfaction values of the 10 government services would be given by the average dominance numbers of the services over subjects. However, in DS, we assume that individual differences are worthwhile variates. The scale values of the services are calculated as averages differentially weighted by subjects' weights. Its main task is to determine appropriate weights for subjects, appropriate in the sense that the variance of the weighted means be a maximum. Individual differences are responsible for multidimensional data structure. $T(\text{sol})$ is 9, and the δ values are in Table 1.18. Considering a relatively sharp drop from Solution 2 to Solution 3, one may decide to look at two solutions, as is done here.

For dominance data, there exists a strict rule for plotting (Nishisato, 1996), namely, plot-normed weights of subjects and projected weights of objects. Then, in the total space, we obtain a configuration such that each subject ranks the closest object first, second closest second, and so on for all subjects and objects—that is, a solution to the Coombs problem of multidimensional unfolding (Coombs, 1964).

Figure 1.7 (p. 18) shows a plot of the first two solutions. A large number of subjects are furthest from postal service, which indicates that postal service is the least satisfactory. This is partly due to the fact that

Table 1.17 Ranking of 10 Government Services in Toronto and Dominance Table

	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
1	1	7	9	10	2	6	3	8	5	4	9	−3	−7	−9	7	−1	5	−5	1	3
2	6	10	9	5	3	1	7	2	4	8	−1	−9	−7	1	5	9	−3	7	3	−5
3	9	8	4	3	5	6	10	2	1	7	−7	−5	3	5	1	−1	−9	7	9	−3
4	2	10	5	6	3	1	4	8	7	9	7	−9	1	−1	5	9	3	−5	−3	−7
5	2	10	6	7	4	1	5	3	9	8	7	−9	−1	−3	3	9	1	5	−7	−5
6	1	3	5	6	7	8	2	4	10	9	9	5	1	−1	−3	−5	7	3	−9	−7
7	7	10	1	6	5	3	8	4	2	9	−3	−9	9	−1	1	5	−5	3	7	−7
8	2	10	6	7	4	1	5	3	9	8	7	−9	−1	−3	3	9	1	5	−7	−5
9	2	10	5	8	4	1	6	3	7	9	7	−9	1	−5	3	9	−1	5	−3	−7
10	2	10	5	9	8	7	4	1	3	6	7	−9	1	−7	−5	−3	3	9	5	−1
11	9	10	7	6	5	1	4	2	3	8	−7	−9	−3	−1	1	9	3	7	5	−5
12	6	10	7	4	2	1	3	9	8	5	−1	−9	−3	3	7	9	5	−7	−5	1
13	1	10	3	9	6	4	5	2	7	8	9	−9	5	−7	−1	3	1	7	−3	−5
14	8	6	5	3	10	7	9	2	1	4	−5	−1	1	5	−9	−3	−7	7	9	3
15	8	10	9	6	4	1	3	2	5	7	−5	−9	−7	−1	3	9	5	7	1	−3
16	3	5	10	4	6	9	8	2	1	7	5	1	−9	3	−1	−7	−5	7	9	−3
17	1	10	8	9	3	5	2	6	7	4	9	−9	−5	−7	5	1	7	−1	−3	3
18	5	4	9	3	10	8	7	2	1	6	1	3	−7	5	−9	−5	−3	7	9	−1
19	2	10	6	7	8	1	5	4	3	9	7	−9	−1	−3	−5	9	1	3	5	−7
20	1	4	2	10	9	7	6	3	5	8	9	3	7	−9	−7	−3	−1	5	1	−5
21	2	10	5	7	3	1	4	6	8	9	7	−9	1	−3	5	9	3	−1	−5	−7
22	6	3	9	4	10	8	7	2	1	5	−1	5	−7	3	−9	−5	−3	7	9	1
23	6	9	10	4	8	7	5	2	1	3	−1	−7	−9	3	−5	−3	1	7	9	5
24	5	2	1	9	10	4	8	6	3	7	1	7	9	−7	−9	3	−5	−1	5	−3
25	2	10	6	7	9	1	3	4	5	8	7	−9	−1	−3	−7	9	5	3	1	−5
26	7	10	9	5	2	6	3	1	4	8	−3	−9	−7	1	7	−1	5	9	3	−5
27	8	7	10	3	5	9	4	2	1	6	−5	−3	−9	5	1	−7	3	7	9	−1
28	3	8	6	7	5	10	9	2	4	1	5	−5	−1	−3	1	−9	−7	7	3	9
29	2	10	7	9	4	1	5	3	6	8	7	−9	−3	−7	3	9	1	5	−1	−5
30	2	10	9	1	4	7	5	3	6	8	7	−9	−7	9	3	−3	1	5	−1	−5
31	4	10	9	7	5	1	3	2	6	8	3	−9	−7	−3	1	9	5	7	−1	−5

Table 1.18 Nine Solutions and Their Contributions

	Solution								
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
Delta	37.9	22.4	13.4	10.6	4.9	4.2	2.7	2.2	1.9
CumDelta	37.9	60.2	73.6	84.2	89.0	93.2	95.9	98.1	100.0

the data were collected shortly after a major postal strike. There are groups who prefer theaters first and restaurants second, or vice versa, suggesting that those who go to theaters must go to good restaurants near the theaters. The most dominant group considers public libraries most satisfactory. One important message of this graphical analysis is that it is very difficult, if not impossible, to interpret the configuration of only services. When we plot subjects and see they are all scattered in the space, the configuration of the services suddenly becomes meaningful because they provide us with how they view those services in terms of satisfaction.

One can calculate the distance from each subject (normed) to each service (projected) in the two-dimensional graph and see if indeed the ranking of distances between each subject and each of the 10 services is close to the ranking in the input data. The ranking thus derived from the first two solutions is called rank-2 approximation to the input ranking. The DUAL3 (Nishisato & Nishisato, 1994b) provides these distances and approximated ranks. The distances between each of the first five subjects and the 10 services and the rank-2 and rank-8 approximations to input ranks are in Tables 1.19 and 1.20. The rank-9 approximation perfectly reproduces the

Table 1.19 Rank 2: Distances and Ranks of Distances

	Service									
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Distances										
Subject 1	0.20	2.05	0.66	1.32	0.26	0.13	0.29	1.29	1.81	1.29
Subject 2	2.05	5.42	3.65	2.79	2.43	1.81	2.33	1.14	2.09	3.64
Subject 3	3.03	3.48	3.40	1.76	3.08	3.35	2.94	1.08	0.94	2.49
Subject 4	1.33	4.82	2.48	3.56	1.57	0.95	1.63	2.97	4.10	3.60
Subject 5	1.31	5.26	2.81	3.40	1.65	0.87	1.66	2.29	3.55	3.72
Ranks of distances										
Subject 1	2	10	5	8	3	1	4	7	9	6
Subject 2	3	10	9	7	6	2	5	1	4	8
Subject 3	6	10	9	3	7	8	5	2	1	4
Subject 4	2	10	5	7	3	1	4	6	9	8
Subject 5	2	10	6	7	3	1	4	5	8	9

Table 1.20 Rank 8: Distances and Ranks of Distances

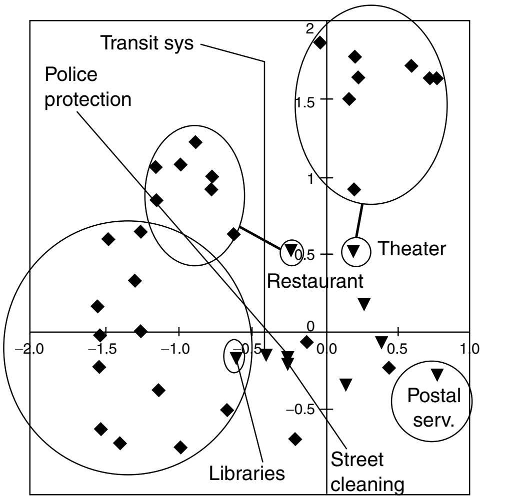
	Service									
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Distances										
Subject 1	13.90	16.75	17.42	17.76	14.20	16.15	14.63	16.91	15.67	15.02
Subject 2	5.79	8.16	6.69	4.99	4.40	4.03	5.49	4.36	4.49	6.78
Subject 3	11.29	11.02	9.04	8.48	9.36	9.98	11.58	8.07	7.73	10.18
Subject 4	4.99	8.49	6.52	6.75	5.24	4.32	6.05	7.30	7.44	7.71
Subject 5	2.70	6.79	4.09	4.59	3.52	2.66	3.36	3.43	5.45	5.42
Ranks of distances										
Subject 1	1	7	9	10	2	6	3	8	5	4
Subject 2	7	10	8	5	3	1	6	2	4	9
Subject 3	9	8	4	3	5	6	10	2	1	7
Subject 4	2	10	5	6	3	1	4	7	8	9
Subject 5	2	10	6	7	5	1	3	4	9	8

Table 1.21 Average Squared Rank Discrepancies

	Rank k										
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*Solution 1*	*Solution 2*
Subject 1	8.8	7.8	9.0	4.6	4.2	1.4	1.6	0.0	0.0	0.65	−0.51
Subject 2	6.2	2.8	1.4	0.2	0.4	0.4	0.2	0.4	0.0	1.15	1.08
Subject 3	19.6	8.0	8.0	1.2	1.2	0.0	0.0	0.0	0.0	−0.16	1.51
Subject 4	1.4	1.0	1.2	1.6	1.6	1.6	0.6	0.2	0.0	1.39	−0.73
Subject 5	1.2	0.8	1.4	1.4	1.4	1.0	0.8	0.6	0.0	1.54	−0.21

input ranks. It is useful to look at average squared rank discrepancies between these approximated ranks and the original ranks (see Table 1.21). Notice that the rank-9 approximation reproduced the input ranks, thus showing no discrepancies. Table 1.21 also lists normed weights for those five subjects, which should be all equal to 1.00 if no individual differences were involved.

Figure 1.7 Ten Government Services



1.6.2. Paired-Comparison Data

The method of paired comparison (see Bock & Jones, 1968) has been one of the pillars in the history of psychological scaling. For a unidimensional preference scale to be constructed from paired-comparison data, we must avoid intransitive judgments (e.g., A is preferred to B, B to C, and C to A), and we must consider individual differences as random fluctuations of judgments. But in real data, we see many intransitive judgments and substantial individual differences. For us to analyze such paired-comparison data, therefore, we must consider a multidimensional scale and treat individual differences as legitimate variates for analysis. This mode of more realistic analysis than the traditional method of paired comparisons is what dual scaling offers. There is no need to worry about unidimensionality, for dual scaling yields as many dimensions as data dictate. We will see how paired-comparison data can be effectively analyzed by dual scaling.

1.6.2.1. Some Basics

For n objects, create all $n(n-1)/2$ possible pairs, present each pair to N subjects, and ask which object in the pair they like better. Collected in this way, such paired-comparison data have mathematically the same structure as the N-by-n rank-order data: $T(\text{sol})$ and $T(\text{inf})$ are identical to those of rank-order data. The only difference is that in rank order, one must arrange all objects in a single order, whereas in paired comparisons, one can anticipate so-called intransitive choices (e.g., A is preferred to B, B is preferred to C, and C is preferred to A). For subject i and pair (X_j, X_k), Nishisato (1978) defined a response variable as follows:

$$ {}_i f_{jk} = \begin{cases} 1 & \text{if } X_j > X_k \\ 0 & \text{if } X_j = X_k \\ -1 & \text{if } X_j < X_k \end{cases}. \tag{28} $$

The subjects-by-objects dominance table can be obtained by transforming ${}_i f_{jk}$ to e_{ij} by the following formula:

$$ e_{ij} = \sum_{\substack{k=1 \\ k \neq j}}^{n} {}_i f_{jk}. \tag{29} $$

Recall that the dominance numbers were easily obtained for rank-order data by a simpler formula than this. The meaning is the same; that is, e_{ij} is the number of times subject i preferred X_j to X_k minus the number of times subject i preferred other objects to X_j.

1.6.2.2. Wiggins's Christmas Party Plans

As a course assignment, Ian Wiggins, now a successful consultant in Toronto, collected paired-comparison data[4] from 14 researchers at a research institute on his eight Christmas party plans:

1. A potluck at someone's home in the evening
2. A potluck in the group room
3. A pub/restaurant crawl after work
4. A reasonably priced lunch in an area restaurant
5. Keep to one's self
6. An evening banquet at a restaurant
7. A potluck at someone's home after work
8. A ritzy lunch at a good restaurant (tablecloths)

Table 1.22 contains data in the form of subjects (14) by pairs (28 pairs), with elements being 1 if the subject prefers the first plan to the second one and 2 if the second plan is preferred to the first ("2" will be later changed to "−1" for analysis). Dominance numbers are in Table 1.23. As is the case with rank-order data, each element of the 14×8 dominance table is based on seven comparisons. Or, more generally, for the $N \times n$ dominance table, each element is based on $(n-1)$ comparisons. Therefore, the marginal frequency of responses for each row is $n(n-1)$ and that of each column is $N(n-1)$.

4. Data used with permission from Ian Wiggins.

Table 1.22 Wiggins's Christmas Party Plans Data

j	*1111111*	*222222*	*33333*	*4444*	*555*	*66*	*7*
k	*2345678*	*345678*	*45678*	*5678*	*678*	*78*	*8*
1	1121121	222222	21121	1121	121	21	2
2	2221212	121212	21112	1112	222	12	2
3	1111121	111121	11121	1121	222	21	1
4	2121112	111112	21222	1112	222	22	2
5	2221212	221222	21212	1111	222	12	2
6	1111111	221222	21222	1111	222	22	1
7	1111121	121121	21121	1121	222	22	1
8	1111121	121221	21221	1221	221	21	1
9	1221121	221122	11121	1121	222	22	1
10	1211222	221222	11111	1222	222	11	2
11	1211111	222222	11111	1111	222	22	2
12	2222122	121111	21111	1111	111	22	1
13	1211212	222222	11111	1212	222	11	2
14	2222121	211111	11111	2121	121	21	1

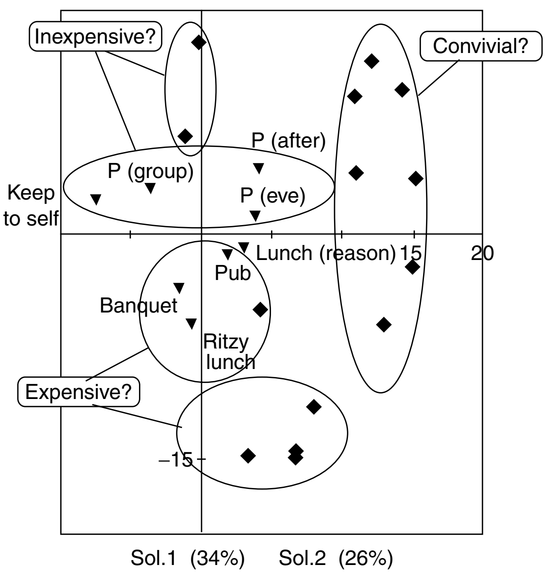
Table 1.23 Dominance Table

j	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
1	3	−7	1	5	−1	−3	5	−3
2	−3	1	−1	5	−7	1	−5	7
3	5	3	1	−1	−7	−3	7	−5
4	1	5	−5	3	−7	−3	−1	7
5	−3	−3	1	7	−7	3	−3	5
6	7	−5	−3	5	−7	−1	3	1
7	5	1	−1	3	−7	−5	7	−3
8	5	−1	−3	1	−5	3	7	−7
9	1	−3	5	3	−7	−5	7	−1
10	−1	−5	7	−3	−7	5	1	3
11	5	−7	7	3	−5	−3	−1	1
12	−5	5	3	7	1	−7	−1	−3
13	1	−7	7	−1	−5	5	−3	3
14	−3	5	7	−1	1	−5	3	−7

Table 1.24 Contributions of Seven Solutions to Total Information

	Solution						
	1	*2*	*3*	*4*	*5*	*6*	*7*
Delta	34	26	16	13	7	3	1
CumDelta	34	60	76	89	96	99	100

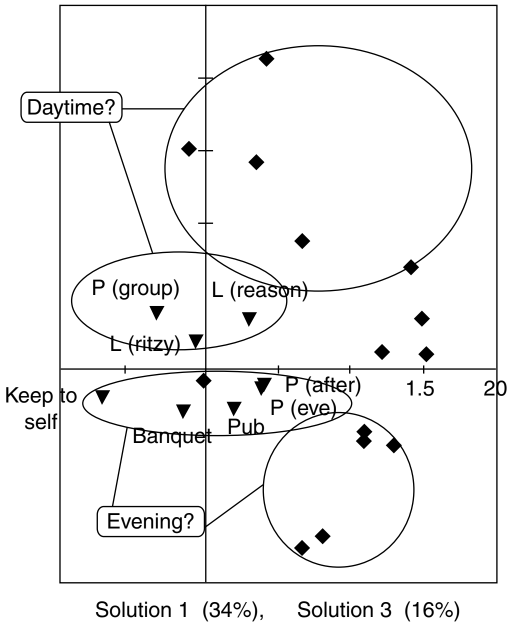
Figure 1.8 Solutions 1 and 2



From the dominance table, it is clear that Plan 5 is not very popular because the corresponding elements from 14 subjects are mostly negative. If we calculate the mean dominance numbers of the eight columns, they may provide good unidimensional estimates of preference values of the party plans, provided that individual differences are negligible. In DS, we weight subjects differentially in such a way that the variance of the eight weighted averages be a maximum. For the present data set, $T(\text{sol})$ is 7, and the corresponding δ values are in Table 1.24. Although weights are not listed here, Solution 4 is dominated only by one variable, that is, "pub/restaurant crawl." In contrast, the first three solutions present a variety of preference patterns. Therefore, let us look at the first three solutions. Figures 1.8 and 1.9 show the following: Dimension 1 divides party plans into the convivial side and the "Keep to one's self" side, Dimension 2 separates plans into expensive and nonexpensive, and Dimension 3 divides party plans into daytime parties and evening parties. Note that the weights of subjects on Solution 1 are mostly positive, but that those on Solutions 2 and 3 are much more evenly distributed than those on Solution 1. This is a reflection of the property of dominance data that the weights for subjects are not centered, due to the row-ipsative nature of dominance data, and are free to vary.

That subjects are scattered in the three-dimensional space means that different subjects prefer different party plans. As noted earlier, each subject in total space ranks the closest plan first. The graphs offer an interesting way to look at individual differences in judgment: DS can accommodate any patterns or combinations of different aspects of the party, such as daytime-inexpensive, daytime-expensive, evening-inexpensive, and evening-expensive.

Figure 1.9 Solutions 1 and 3



1.7. Forced Classification for Multiple-Choice Data

We have seen dual scaling of multiple-choice data, and it was noted that dual scaling maximizes the average of all possible inter-item correlation coefficients. There are occasions, however, when we are not interested in all the items but only one item. For instance, if we collect children's background medical and psychological information in addition to whether or not they have allergy problems, we would be interested in finding which of the medical and psychological variables may be related to the allergy problems. In this case, we are no longer interested in scaling data to maximize the average inter-variable correlation, but our interest now lies in the scaling method that maximizes the correlation between the allergy variable and the other variables. This task is carried out by the procedure called *forced classification*.

Nishisato (1984) proposed a simple procedure to carry out the above task, which is nothing but discriminant analysis with categorical data. It is based on two principles: principle of internal consistency (PIC) and principle of equivalent partitioning (PEP). Let us denote the data of n multiple-choice questions from N subjects as

$$\mathbf{F} = [\mathbf{F}_1, \mathbf{F}_2, \ldots, \mathbf{F}_j, \ldots, \mathbf{F}_n], \tag{30}$$

where $\mathbf{F}_j$ is an N-by-m_j matrix, in which the row i consists of subject i's response to item j, with 1 being the choice and 0s the nonchoices out of m_j options. Each subject chooses only one option per item. Suppose that we repeat $\mathbf{F}_j$ k times in the data matrix. As k increases, the response patterns in $\mathbf{F}_j$ become more dominant in the data set, and eventually we will see that the response patterns in the repeated $\mathbf{F}_j$ determine the first solution (PIC). Instead of repeating $\mathbf{F}_j$ k times, it is known that the same dual-scaling results can be obtained from analysis of the following matrix (PEP):

$$[\mathbf{F}_1, \mathbf{F}_2, \ldots, k\mathbf{F}_j, \ldots, \mathbf{F}_n]. \tag{31}$$

This matrix is obtained from the original matrix by replacing each 1 in $\mathbf{F}_j$ with a k. Thus, the computation involved here is ordinary DS with an altered data matrix by multiplying the chosen submatrix by a large enough scalar k. Possible applications of this procedure are, for instance, the following:

1. to identify personality traits that are closely related to the school dropout,
2. to find out if academic performance is influenced by some environmental factors (school buildings, computers, etc.),
3. to see if the high blood pressure is related to the regions where people live,
4. to collect questions related to anxiety for the construction of an anxiety scale
5. to eliminate age effects, if any, from consumer data on purchase patterns of cosmetics after finding significant age effects.

Due to the limited space for this chapter, a numerical example of forced classification is not given here. Please refer to Nishisato and Gaul (1990) for its applications to marketing research and to Nishisato and Baba (1999) for the latest development.

1.8. Mathematics of Dual Scaling

1.8.1. Structure of Data

Given a two-way table of data with typical element f_{ij}, singular-value decomposition can be described as bilinear decomposition:

$$f_{ij} = \frac{f_{i.}f_{.j}}{f_{..}}[1 + \rho_1 y_{i1} x_{j1} + \rho_2 y_{i2} x_{j2} + \cdots + \rho_k y_{ik} x_{jk}], \tag{32}$$

where ρ_k is the kth largest singular value, y_{ik} is the ith element of singular vector y_k for the rows, and x_{jk} is the jth element of singular vector x_k for the columns of the table. These singular vectors can be viewed as weight vectors for the rows and the columns. The first term inside the bracket—that is, the element 1—is called a trivial solution associated with the case in which the rows and the columns are statistically independent. Another well-known expression of the singular-value decomposition is what Benzécri et al. (1973) call *transition formulas* and Nishisato (1980) refers to as *dual relations:*

$$y_{ik} = \frac{1}{\rho_k}\frac{\Sigma f_{ij}x_{jk}}{f_{i.}};\ x_{jk} = \frac{1}{\rho_k}\frac{\Sigma f_{ij}y_{ik}}{f_{.j}}. \qquad (33)$$

These weights, y_{ik}, x_{jk}, are called *normed weights* (Nishisato, 1980) or *standard coordinates* (Greenacre, 1984). If we multiply the formulas by ρ_k, the resultant weights are called *projected weights* (Nishisato, 1980) or *principal coordinates* (Greenacre, 1984). The projected weights are

$$\rho_k y_{ik} = \sum_{j=1}^{n}\frac{f_{ij}x_{jk}}{f_{i.}},\ \rho_k x_{jk} = \sum_{i=1}^{m}\frac{f_{ij}y_{ik}}{f_{.j}}. \qquad (34)$$

The above sets of formulas hold for any data matrix (f_{ij}).

To arrive at these formulas, one can define the task in many ways, which is probably one of the reasons why so many researchers have discovered the method independently and coined their own names. For example, one may state the problem in any of the following ways:

- Determine x_{jk} and y_{ik} in such a way that the data weighted by x_{jk} and the data weighted by y_{ik} attain the maximal product-moment correlation.
- Determine x_{jk} to make the between-row sum of squares, relative to the total sum of squares, be a maximum; determine y_{ik} so as to make the between-column sum of squares to the total sum of squares be a maximum.
- Determine those two sets of weights to make the regression of the rows on the columns and the regression of the columns on the rows be simultaneously linear.
- Determine those two sets of weights in such a way to make the sum of the squared differences between f_{ij} and $\frac{f_{i.}f_{.j}}{f_{..}}x\rho y_{ik}x_{jk}$ be a minimum.

All of these lead to the identical solution set (ρ_k, y_{ik}, x_{jk}). For detailed mathematical derivations, see Benzécri (1973), Nishisato (1980, 1994), Greenacre (1984), and Gifi (1990).

1.8.2. Row Space and Column Space Are Different

We are interested in the relations between rows and columns of a two-way table, for example, relations between subjects and chosen objects. Unfortunately, the space for row variables and the space for column variables are different, the discrepancy of which is related to the cosine of the singular values. In other words, when singular values are relatively large, the discrepancy between the row space and the column space is comparatively small. When we want to put both row and column variables in the same space, we must plot normed weights of rows (or columns) and projected weights of columns (or rows). Then, both sets of weights span the same space. We often talk about symmetric scaling to indicate that both projected row and projected column weights are plotted, in which case care must be exercised in judging their distances because of the discrepancy of the two spaces. Or, rather, symmetric scaling may be justified only when singular values are close to 1. Nonsymmetric scaling of one set of weights to be projected to the other set is the mathematically correct one, but we must often deal with a rather nasty problem of a large difference between the spread of normed weights and that of projected weights, the latter being often too much smaller than the former, making comparisons between them difficult. See Nishisato and Clavel (2003) for a discussion on the discrepant spaces and the calculation of distances between points in two different spaces.

1.8.3. Chi-Square Metric and Data Types

One of the difficult problems in quantifying incidence data lies in its use of the chi-square metric, which is necessitated by the sheer characteristics of the data. When Point A has one observation and Point B nine observations, the midpoint between them is 9 units away from A and one unit away from B. This is an example of a chi-square metric, which is a reciprocal function of the number of observations. In the above example, the distance between A and the midpoint times 1 (observation) is equal to the distance between the midpoint and B times 9. Thus, the point with more observations has a stronger pull than the point with fewer observations.

In contrast, each cell in the dominance table is represented by a constant number of observations (i.e., $n - 1$). Therefore, the chi-square metric is reduced to the Euclidean metric, where the midpoint between A and B is located halfway between A and B. It should be remembered, however, that the way in which DS handles dominance data is to treat dominance numbers as cardinal numbers, rather than ordinal. At the present moment, we have not developed an ordinal way of handling dominance numbers. This is one problem for future research. Another point of caution is that both *chi-square metric* and *Euclidean metric* are defined for the Euclidean space.

1.9. Linear Analysis and Dual Scaling

In the principal coordinate system, each continuous variable is expressed as a straight line (axis), whereas categories of each variable in DS no longer lie on a straight line. In consequence, when data are in multidimensional space, the contribution or information of each variable in PCA is expressed by the length of its vector, which increases as the dimensionality increases, whereas the contribution of each variable in DS increases as the dimensionality increases in a distinctively different way from PCA. The DS contribution of each variable to the given space is not expressed by the length of any vector but by the area or volume formed by connecting the points of those categories of the variable.

If an item has three categories, the information of the variable in the given dimension is the area of a triangle obtained by connecting the three category points in the space. The area of the triangle monotonically increases as the dimensionality of the space for the data increases. If a variable has four categories, the information of the variable in three-dimensional or higher dimensional space is given by the volume of the form created by connecting four-category points. If the variable has n categories, the information of the variable in $n-1$ or higher dimensional space is given by the volume of the form created by connecting n points.

Thus, by stretching our imagination to the continuous variable, where the number of categories is considered very large but finite, we can conclude that the information of the variable in the given space must be expressed by the volume of a shape and not by the length of a vector. This conjecture can be reinforced by the fact that many key statistics associated with dual scaling are related to the number of categories of variables. Some of the examples are given below.

The total number of dimensions required to accommodate a variable with m_j categories is

$$N_j = m_j - 1. \tag{35}$$

The total number of dimensions needed for n variables is

$$N_T = \sum_{j=1}^{n}(m_j - 1) = \sum_{j=1}^{n} m_j - n = m - n. \tag{36}$$

The total amount of information in the data—that is, the sum of the squared singular values, excluding 1—is given by

$$\sum_{k=1}^{K} \rho_j^2 = \frac{\sum_{j=1}^{n} m_j}{n} - 1 = \overline{m} - 1. \tag{37}$$

Therefore, as the number of categories of each variable increases, so does the total information in the data set. The information of variable j with m_j categories is given by

$$\sum_{k=1}^{m-n} r_{jt(k)}^2 = m_j - 1. \tag{38}$$

These are all related to the number of categories of each variable. Thus, we can imagine what will happen as m_j increases to infinity or, in practice, to the number of observations (subjects) N. An inevitable conclusion, then, seems to be that the total information in the data set is much more than the sum of the lengths of vectors of the variables in multidimensional space: It is the sum of the volumes of hyperspheres associated with categories of individual variables.

The above conclusion (Nishisato, 2002) suggests how little information popular linear analyses such as PCA and factor analysis capture. Traditionally, the total information is defined by the sum of the eigenvalues associated with a linear model. But we have just observed that it seems inappropriate unless we are totally confined within the context of a linear model. In a more general context, in which we consider both linear and nonlinear relations among variables, DS offers the sum of the eigenvalues as a reasonable statistic of the total information in the data. As the brain wave analyzer filters a particular wave such as alpha, most statistical procedures—particularly PCA, factor analysis, other correlational methods, and multidimensional scaling—play the role of a linear filter and filter out most of the information from the data, that is, a nonlinear portion of the data. In this context, dual scaling should be reevaluated and highlighted as a means for analyzing both linear and nonlinear information in the data, particularly in the behavioral sciences, where it seems that nonlinear relations are more abundant than linear relations.

REFERENCES

Bartlett, M. S. (1951). The goodness of fit of a single hypothetical discriminant function in the case of several groups. *Annals of Eugenics, 16*, 199–214.

Beltrami, E. (1873). Sulle funzioni bilineari (On the bilinear functions). *Giornale di Mathematiche, 11*, 98–106.

Benzécri, J. P. (1969). Statistical analysis as a tool to make patterns emerge from data. In S. Watanabe (Ed.), *Methodologies of pattern recognition* (pp. 35–74). New York: Academic Press.

Benzécri, J. P., et al. (1973). *L'analyse des données: II. L'analyse des correspondances* (The analysis of data: Vol. 2. The analysis of correspondences). Paris: Dunod.

Benzécri, J. P. (1982). *Histoire et préhistoire de l'analyse des données* (History and prehistory of the analysis of data). Paris: Dunod.

Bock, R. D. (1956). The selection of judges for preference testing. *Psychometrika, 21*, 349–366.

Bock, R. D. (1960). Methods and applications of optimal scaling (University of North Carolina Psychometric Laboratory Research Memorandum, No. 25). Chapel Hill: University of North Carolina.

Bock, R. D., & Jones, L. V. (1968). *Measurement and prediction of judgment and choice*. San Francisco: Holden-Day.

Coombs, C. H. (1964). *A theory of data*. New York: John Wiley.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297–334.

de Leeuw, J. (1973). *Canonical analysis of categorical data*. Leiden, The Netherlands: DSWO Press.

Eckart, C., & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika, 16*, 211–218.

Escofier-Cordier, B. (1969). L'analyse factorielle des correspondances (Factor analysis of correspondences). Paris: Bureau Univesitaire de Recherche operationelle, Cahiers, Série Recherche, Université de Paris.

Fisher, R. A. (1940). The precision of discriminant functions. *Annals of Eugenics, 10*, 422–429.

Fisher, R. A. (1948). *Statistical methods for research workers*. London: Oliver and Boyd.

Franke, G. R. (1985). Evaluating measures through data quantification: Applying dual scaling to an advertising copytest. *Journal of Business Research, 13*, 61–69.

Gabriel, K. R. (1971). The biplot graphical display of matrices with applications to principal component analysis. *Biometrics, 58*, 453–467.

Gifi, A. (1980). *Nonlinear multivariate analysis*. Unpublished manuscript.

Gifi, A. (1990). *Nonlinear multivariate analysis*. New York: John Wiley.

Gower, J. C., & Hand, D. J. (1996). *Biplots*. London: Chapman & Hall.

Greenacre, M. J. (1984). *Theory and applications of correspondence analysis*. London: Academic Press.

Greenacre, M. J., & Blasius, J. (Eds.). (1994). *Correspondence analysis in the social sciences*. London: Academic Press.

Greenacre, M. J., & Torres-Lacomba, A. (1999). *A note on the dual scaling of dominance data and its relationship to correspondence analysis* (Working Paper Ref. 430). Barcelona, Spain: Departament d'Economia i Empresa, Universidat Pompeu Fabra.

Guttman, L. (1941). The quantification of a class of attributes: A theory and method of scale construction. In the Committee on Social Adjustment (Ed.), *The prediction of personal adjustment* (pp. 319–348). New York: Social Science Research Council.

Guttman, L. (1946). An approach for quantifying paired comparisons and rank order. *Annals of Mathematical Statistics, 17*, 144–163.

Hayashi, C. (1950). On the quantification of qualitative data from the mathematico-statistical point of view. *Annals of the Institute of Statistical Mathematics, 2*, 35–47.

Hayashi, C. (1952). On the prediction of phenomena from qualitative data and the quantification of qualitative data from the mathematico-statistical point of view. *Annals of the Institute of Statistical Mathematics, 3*, 69–98.

Hill, M. O. (1973). Reciprocal averaging: An eigenvector method of ordination. *Journal of Ecology, 61*, 237–249.

Hirschfeld, H. O. (1935). A connection between correlation and contingency. *Cambridge Philosophical Society Proceedings, 31*, 520–524.

Horst, P. (1935). Measuring complex attitudes. *Journal of Social Psychology, 6*, 369–374.

Hotelling, H. (1933). Analysis of complex of statistical variables into principal components. *Journal of Educational Psychology, 24*, 417–441, 498–520.

Jackson, D. N., & Helmes, E. (1979). Basic structure content scaling. *Applied Psychological Measurement, 3*, 313–325.

Johnson, P. O. (1950). The quantification of qualitative data in discriminant analysis. *Journal of the American Statistical Association, 45*, 65–76.

Jordan, C. (1874). Mémoire sur les formes binlinieres (Note on bilinear forms). *Journal de Mathématiques Pures et Appliquées, deuxiéme Série, 19*, 35–54.

Lancaster, H. O. (1958). The structure of bivariate distribution. *Annals of Mathematical Statistics, 29*, 719–736.

Lebart, L., Morineau, A., & Warwick, K. M. (1984). *Multivariate descriptive statistical analysis*. New York: John Wiley.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychlogy, 140*, 44–53.

Lingoes, J. C. (1964). Simultaneous linear regression: An IBM 7090 program for analyzing metric/nonmetric or linear/nonlinear data. *Behavioral Science, 9*, 87–88.

Lord, F. M. (1958). Some relations between Guttman's principal components of scale analysis and other psychometric theory. *Psychometrika, 23*, 291–296.

Maung, K. (1941). Measurement of association in contingency tables with special reference to the pigmentation of hair and eye colours of Scottish children. *Annals of Eugenics, 11*, 189–223.

Meulman, J. J. (1998). Review of W. J. Krzanowski and F. H. C. Marriott, "Multivariate analysis: Part I. Distributions, ordinations, and inference." *Journal of Classification, 15*, 297–298.

Mosier, C. I. (1946). Machine methods in scaling by reciprocal averages. In *Proceedings, research forum* (pp. 35–39). Endicath, NY: International Business Corporation.

Nishisato, S. (1978). Optimal scaling of paired comparison and rank order data: An alternative to Guttman's formulation. *Psychometrika, 43*, 263–271.

Nishisato, S. (1980). *Analysis of categorical data: Dual scaling and its applications*. Toronto: University of Toronto Press.

Nishisato, S. (1984). Forced classification: A simple application of quantification technique. *Psychometrika, 49*, 25–36.

Nishisato, S. (1986). Multidimensional analysis of successive categories. In J. de Leeuw, W. Heiser, J. Meulman, & F. Critchley (Eds.), *Multidimensional data analysis*. Leiden, The Netherlands: DSWO Press.

Nishisato, S. (1988, June). *Effects of coding on dual scaling*. Paper presented at the annual meeting of the Psychometric Society, University of California, Los Angeles.

Nishisato, S. (1993). On quantifying different types of categorical data. *Psychometrika, 58*, 617–629.

Nishisato, S. (1994). *Elements of dual scaling*. Hillsdale, NJ: Lawrence Erlbaum.

Nishisato, S. (1996). Gleaning in the field of dual scaling. *Psychometrika, 61*, 559–599.

Nishisato, S. (2000). Data analysis and information: Beyond the current practice of data analysis. In R. Decker & W. Gaul (Eds.), *Classification and information processing at the turn of the millennium* (pp. 40–51). Heidelberg: Springer-Verlag.

Nishisato, S. (2002). Differences in data structure between continuous and categorical variables as viewed from dual scaling perspectives, and a suggestion for a unified mode of analysis. *Japanese Journal of Sensory Evaluation, 6*, 89–94 (in Japanese).

Nishisato, S. (2003). Geometric perspectives of dual scaling for assessment of information in data. In H. Yanai, A. Okada, K. Shigemasu, Y. Kano, & J. J. Meulman (Eds.), *New developments in psychometrics* (pp. 453–463). Tokyo: Springer-Verlag.

Nishisato, S., & Baba, Y. (1999). On contingency, projection and forced classification of dual scaling. *Behaviormetrika, 26*, 207–219.

Nishisato, S., & Clavel, J. G. (2003). A note on between-set distances in dual scaling and correspondence analysis. *Behaviormetrika, 30*(1), 87–98.

Nishisato, S., & Gaul, W. (1990). An approach to marketing data analysis: The forced classification procedure of dual scaling. *Journal of Marketing Research, 27*, 354–360.

Nishisato, S., & Nishisato, I. (1994a). *Dual scaling in a nutshell*. Toronto: MicroStats.

Nishisato, S., & Nishisato, I. (1994b). *The DUAL3 for Windows*. Toronto: MicroStats.

Nishisato, S., & Sheu, W. J. (1980). Piecewise method of reciprocal averages for dual scaling of multiple-choice data. *Psychometrika, 45*, 467–478.

Noma, E. (1982). The simultaneous scaling of cited and citing articles in a common space. *Scientometrics, 4*, 205–231.

Odondi, M. J. (1997). *Multidimensional analysis of successive categories (rating) data by dual scaling*. Unpublished doctoral dissertation, University of Toronto.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazines and Journal of Science, Series 6, 2*, 559–572.

Richardson, M., & Kuder, G. F. (1933). Making a rating scale that measures. *Personnel Journal, 12*, 36–40.

Schmidt, E. (1907). Zür Theorie der linearen und nichtlinearen Integralgleichungen. Erster Teil. Entwickelung willkürlicher Functionen nach Systemaen vorgeschriebener (On the theory of linear and nonlinear integral equations. Part one. Development of arbitrary functions according to prescribed systems). *Mathematische Annalen, 63*, 433–476.

Sheskin, D. J. (1997). *Handbook of parametric and nonparametric procedures*. Boca Raton, FL: CRC Press.

Takane, Y. (1980). Analysis of categorizing behavior. *Behaviormetrika, 8*, 75–86.

Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: John Wiley.

van de Velden, M. (2000). Dual scaling and correspondence analysis of rank order data. In R. D. H. Heijmans, D. S. G. Pollock, & A. Satorra (Eds.), *Innovations in multivariate statistical analysis* (pp. 87–99). Dordrecht, The Netherlands: Kluwer Academic.

van Meter, K. M., Schiltz, M., Cibois, P., & Mounier, L. (1994). Correspondence analysis: A history and French sociological perspectives. In M. Greenacre & J. Blasius (Eds.), *Correspondence analysis in the social sciences* (pp. 128–137). London: Academic Press.

Williams, E. J. (1952). Use of scores for the analysis of association in contingency tables. *Biometrika, 39*, 274–289.