406

MULTIPLE CORRESPONDENCE ANALYSIS

Donna L. Hoffman
School of Management
University of Texas at Dallas
Richardson, TX 75083-0688

Jan de Leeuw
Department of Mathematics
UCLA
Los Angeles, CA 90024-1555

Ramesh Arjunji
School of Management
University of Texas at Dallas
Richardson, TX 75083-0688

January 1993

INTRODUCTION

Multiple correspondence analysis (MCA) is a popular tool in marketing research (Green, Krieger, and Carroll 1987; Hoffman and Batra 1991; Hoffman and Franke 1986; Kaciak and Louviere 1990; Valette-Florence and Rapacchi 1991). In this chapter, we introduce MCA as a nonlinear multivariate analysis method which integrates ideas from both classical multivariate analysis and multidimensional scaling (MVA and MDS, from now on, respectively).

In MDS models, a geometrical representation of the brands, say, is derived from information about the *dissimilarities* among these brands. The brands are represented in a metric map in such a way that dissimilar brands are relatively far apart, while similar brands are relatively close. In most MDS approaches, the map is a low-dimensional Euclidean space (Cooper 1983; Green 1975). Compare this with cluster analysis, which can be interpreted as an MDS technique. However, the metric map in which we display the brands is now a tree or some other combinatorial structure (Aldenderfer and Blashfield 1984). Nevertheless, it is clear that MDS has a very strong geometrical orientation and that the key notion is one of *distance*.

In these respects, MDS is quite different from MVA, at least in its usual formulations. The traditional MVA approaches proceed either by constructing *linear* combinations of variables with optimality properties defined in terms of *correlation coefficients*, or by specifying structural models for correlated variables, which are usually assumed to be *multinormally distributed*. In these classical formulations, the geometrical notions play a relatively minor role, with the emphasis shifted to linear algebra in the form of matrix calculus. Distance in low-dimensional Euclidean space, the key concept in MDS, is replaced by the vector inner product in high-dimensional space. Nevertheless, it is critical to realize that the basic mathematical structure used in most forms of MVA and MDS is identical. It is none other than the familiar Euclidean space with the inner product defining the angle, and the accompanying Pythagorean distance measure.

We draw this comparison because the basic similarity between MVA and MDS can be impressively exploited. Gifi (1981; 1990), for example, organizes the better known MVA methods into a system which takes *multiple correspondence analysis* (MCA) as the basic technique from which all others are derived as special cases. MCA, as defined there, and as we shall develop it in this chapter (see also Hoffman and de Leeuw 1992), is an MDS method for categorical variables in which the focus is on the distance among the points in a low-dimensional map.

In MCA, our concern is primarily with the following: 1) What are the similarities and differences among the brands, say, with respect to the various variables describing them?; 2) What are the similarities and differences among the variables with respect to the brands?; 3) What is the interrelationship among the brands and the variables?; and 4) Can these relationships be represented graphically in a joint low-dimensional space?

As we shall develop it, MCA is an MDS method that answers these questions in terms of the notion of closeness. This means that between brands, two brands are close together if they share similar variables, and between each variable category, two categories are close if they occur in the same brands to the same degree; it also implies that a brand is close to a variable category if the brand falls into that category.

We organize our chapter as follows. First, we offer some philosophy, which serves to fix ideas. We then present the theory underlying this philosophy, discussing in turn the homogeneity loss function and computational aspects of our approach. Next, we focus on the geometry of the loss function, which emphasizes interpretation through the links between MCA and some of the better known MVA techniques. An empirical example is then presented which illustrates the primary geometric features of MCA. The chapter concludes with a discussion of the issues involved in representing brands and variable categories in the same map.

Some Philosophy by Way of History

The French literature (see, for example, Benzécri et. al. 1973) discusses MCA in the context of metric MDS suitable for frequency matrices, contingency tables or cross-tables. Others formulate MCA as factorial analysis of qualitative data using scale analysis (Bock 1960; Guttman 1941; Nishisato 1980) or principal component analysis (Burt 1950; de Leeuw 1973; Greenacre 1984; Hayashi 1950) perspectives.

Key papers in the history of simple correspondence analysis (CA) are Pearson (1904) and Hirschfeld (1935). The history is complicated and somewhat confused because the early papers dealt with theoretical questions and not with actual data analysis. The first paper applying CA in actual data analysis was Fisher (1940).

For multiple correspondence analysis, the situation is somewhat simpler. The technique was introduced, from the start, as a data analysis method, and the first paper was undoubtedly the one by Guttman (1941). Nevertheless, there are some predecessors. In the very early days of psychometrics, Edgerton and Kolbe (1936), Horst, (1936), Richardson and Kuder (1933), and Wilks (1938) derived principal components as a form of regression in which the predictor is missing.

MCA is the analysis of *interdependence* among a set of categorical variables, as distinct from the analysis of dependence (with pre-defined sets of dependent and independent variables). The approach we present in this chapter is particularly intuitive and should appeal to marketing researchers, borrowing, as it does, concepts and terminology from discriminant analysis and analysis of variance. We emphasize construction of not only an aesthetically pleasing map, but also one that is easy to interpret, and hence, managerially relevant.

THEORY

The data are m categorical variables on n objects, with the j^{th} variable taking on k_j different values, its categories. Consider the example in Table 1, with m=7, n=31, $k_1=4$, $k_2=4$, $k_3=4$, $k_4=5$, $k_5=4$, $k_6=4$, and $k_7=4$. Here, the objects are 31 Swedish industries which groups of customers rated in 1990 on seven variables tapping different aspects of satisfaction.

The data we present are aggregated over hundreds of customer responses in each service or product category to provide an average view of the levels of each variable in each industry. As shown in Table 1, except for the *Price Increase Tolerance* variable, which was scaled using five categories, all variables were scaled using four categories.

Price measures the perceived price level relative to quality, ranging from very unreasonable to very reasonable. *Quality* measures the perceived quality level relative to price ranging from very low quality to very high quality. *Repurchase Intention* evaluates how likely, from very unlikely to very likely, it is that the next time the customer purchases in the category, the purchase will be the same manufacturer or brand again. *Price Increase Tolerance* assesses the tolerable price increase given quality before the customer is likely to switch, ranging from very likely to switch to very unlikely to switch. *Satisfaction* measures the customer's overall satisfaction with the product or service from very dissatisfied to very satisfied. *Expectation of quality* measures the customer's prior expectations of the quality of the product or service from very low to very high. Finally, *Ideal* measures how close the product or service comes to the ideal in the category, from very far away to very close.

Sets of variables are hypothesized to measure different aspects relating to customer satisfaction. Thus, quality and price represent performance aspects of satisfaction. Satisfaction, expectations, and closeness to the ideal indicate the satisfaction construct and price increase tolerance and repurchase tap aspects of customer loyalty. Fornell (1992) provides a fuller discussion of these data in the context of his Customer Satisfaction Barometer project[1].

Insert Table 1 about here

We code the variables using indicator matrices, which allow for easy expression in matrix notation. An indicator matrix is a binary matrix (exactly one element equal to one in each row) which indicates the category that an industry is in for a particular variable. Thus, if variable j has k_j categories, the indicator matrix $\mathbf{G}_j$ for this variable is $n \times k_j$. The rows of $\mathbf{G}_j$

[1]We thank Claes Fornell, Professor of Business Administration and Director of the Office for Customer Satisfaction Research, School of Business Administration, the University of Michigan, for graciously providing us with these data.

add up to one. The seven indicator matrices for our example appear in Table 2. Notice that the variable Expectation of Quality is missing for the industry FOOD, as indicated by the zeros for every category of that variable. The indicator matrix for this variable is thus *incomplete*.

Insert Table 2 about here

The purpose of multiple correspondence analysis is to construct a joint map of the industries and satisfaction variable categories in such a way that an industry will be relatively close to a category it is in, and relatively far from the categories it is not in. By the triangle inequality, this implies that industries mostly occurring in the same categories tend to be close, while categories sharing mostly the same industries tend to be close, as well.

Imagine for a moment an arbitrary map that we could construct from Table 2 by connecting industries with categories of the variables they are in. We might locate industries randomly in the map, while variable categories could be positioned at the centroid or average of all the industries in that category. We could draw lines that show the connections between the industries and the variable categories. Such a map would contain the same information as the data matrix in Table 2, but would be rather unappealing to the eye and difficult to interpret. There would be many lines in the map and they would cross every which way. Our rule for constructing such a map thus means that the map would give the impression that industries are as far from the categories they occur in as they are from the categories they do not occur in. We would deem such an arbitrary representation of the data highly unsatisfactory.

Now suppose we think of this map as a multivariable representation, i.e. as a joint map of the industries and the variable categories, in two-dimensional Euclidean space. How can we improve upon the map? The map will be much more useful to us if industries are close to the categories of the variables that they occur in. This is, in words, the basic premise of multiple correspondence analysis. We desire a map of the data in low-dimensional Euclidean space such that the points connected by a line are relatively close together (and the points not connected by lines are relatively far apart). By the triangle inequality this implies that industries with similar profiles (i.e. industries that are often in the same categories) will be close, and categories containing roughly the same industries will be close, as well. The resultant map will capture the essence of our original idea for a map, but in a way that yields easier and better interpretation. We now formalize these ideas by defining a suitable loss function to be minimized.

The Homogeneity Loss Function

The concept of *homogeneity* serves as the basis for our theoretical development of multiple correspondence analysis. We use homogeneity in a data theoretical sense as being closely related to the concept of data reduction. That is, homogeneity refers to the extent to which different variables measure the same characteristic or characteristics (Gifi 1981; 1990). Homogeneity thus specifies a type of similarity. In order to measure homogeneity, we need a measure for the difference or the similarity of the variables. There are different measures of homogeneity and different approaches to find maps with some distances smaller than others. The particular choice of loss function defines the former and the specific algorithm employed determines the latter. Note that when the variables measure more than one property or characteristic, we may wish to proceed in order to find another, orthogonal, solution. This is in keeping with the principle of data reduction which advocates that a small number of dimensions should be used to explain a maximum amount of information contained in the data.

The extent to which a particular representation **X** of the industries and particular representations $\mathbf{Y}_j$ of the categories, satisfy the axioms of multiple correspondence analysis is quantified by the *loss of homogeneity*, a least squares loss function:

$$\sigma(\mathbf{X};\mathbf{Y}_1,\ldots,\mathbf{Y}_m) = \Sigma_j \, SSQ(\mathbf{X}-\mathbf{G}_j\mathbf{Y}_j) \tag{1}$$

where SSQ(.) is shorthand for the sum of squares of the elements of a matrix or vector. The loss function in (1), giving the sum of squares of the distances between industries and the variable categories they occur in, measures departure from perfect fit. In words, Loss =

$Dist^2$(Airline 1, unreasonable) + $Dist^2$(Airline 2, unreasonable) + ... + $Dist^2$(TV Station 2, very far away) + $Dist^2$(TV Station 3, very far away). A total of n*m = 31*7 = 217 squared distances are summed and these squared distances correspond exactly to the lines connecting industries and variable categories in our original messy map. Quite simply, multiple correspondence analysis produces the map with the smallest possible loss.

There are two sets of unknowns in the MCA problem: the $n \times p$ matrix of industry coordinates $\mathbf{X}$ and the m matrices of variable category coordinates $\mathbf{Y}_j$, each of order $k_j \times p$, where p is the number of dimensions. We could say that multiple correspondence analysis is a method that minimizes (1) over $\mathbf{X}$ and the $\mathbf{Y}_j$, but this would not be sufficient. In the first place, we can set $\mathbf{X}=\mathbf{0}$ and $\mathbf{Y}_j = \mathbf{0}$, for all j. This gives loss equal to zero (i.e. the map is a single point) and consequently, perfect homogeneity. In fact, more generally, taking all elements of $\mathbf{X}$ equal to a constant c, and taking all elements of $\mathbf{Y}_j$ equal to c as well, gives loss zero. These trivial solutions are excluded by imposing suitable normalizations. As we discuss subsequently, it is these normalizations which define a particular coordinate scaling. For now, we choose to minimize over all $\mathbf{Y}_j$ and all *normalized* $\mathbf{X}$, which means we require $\mathbf{X}'\mathbf{u} = \mathbf{0}$ and $\mathbf{X}'\mathbf{X}=n\mathbf{I}$, where $\mathbf{u}$ is a unit vector (all elements equal to one) and $\mathbf{I}$ is the identity matrix. A matrix $\mathbf{X}$ satisfying these restrictions is said to be *normalized*. We now define multiple correspondence analysis more precisely as minimization of the loss function (1) over all $\mathbf{Y}_j$ and over all normalized $\mathbf{X}$ (or equivalently over all normalized $\mathbf{Y}_j$ and all $\mathbf{X}$).

Computational Aspects: Reciprocal Averaging

Our algorithm, which is exceedingly simple, uses alternating least squares (or, equivalently, reciprocal averaging (Hirschfeld 1935; Horst 1935)). This means that we start with an arbitrarily normalized $\mathbf{X}$ and then compute the *optimal* $\mathbf{Y}_j$,

$$\mathbf{Y}_j = \mathbf{D}_j^{-1}\mathbf{G}_j'\mathbf{X} \qquad (2)$$

with $\mathbf{G}_j$ defined as above, and $\mathbf{D}_j = \mathbf{G}_j'\mathbf{G}_j$ the $k_j \times k_j$ diagonal matrix containing the univariate marginals of variable j[2]. In words, the optimal coordinate for a variable category is the average or *centroid* of the (optimal) coordinates of the industries in that category. Using these new $\mathbf{Y}_j$ we now compute new *optimal* **X**,

$$\mathbf{X} = m^{-1}\Sigma_j \mathbf{G}_j \mathbf{Y}_j \tag{3}$$

In words, the optimal coordinate for an industry is the centroid of the (optimal) coordinates of the categories containing that industry. We now normalize **X** by Gram-Schmidt orthogonalization and go back to (2) until convergence. The algorithm is implemented in the SPSS-X program CATEGORIES (SPSS Inc. 1989). Equations (2) and (3) make clear the centroid principle.

GEOMETRIC ASPECTS OF MCA

To highlight interpretation we focus on the geometry of the loss function. Our development proceeds from initially assuming that the industry coordinates, **X**, are normalized and the satisfaction variable category coordinates, $\mathbf{Y}_j$, are free. Loss function (1) has a natural lower bound, because obviously $\sigma(\mathbf{X};\mathbf{Y}_1,\ldots,\mathbf{Y}_m) \geq 0$, but no natural upper bound. The $\mathbf{Y}_j$ can be arbitrarily far from **X**, and thus σ is really unbounded. This means that we have no standard to compare loss with, and no factor to normalize it with. In order to remedy this, we define the loss function

$$\sigma(\mathbf{X};*,\ldots,*) = \min_{\mathbf{Y}} \sigma(\mathbf{X};\mathbf{Y}_1,\ldots,\mathbf{Y}_m) \tag{4}$$

which is the minimum of the homogeneity loss function (1) over the variable category coordinates. It is now easy to derive an upper bound for (4), because $\sigma(\mathbf{X};*,\ldots,*) \leq \sigma(\mathbf{X};0,\ldots,0) = m\,\mathrm{tr}(\mathbf{X}'\mathbf{X}) = mnp$.

[2] If some of the categories are empty, then $\mathbf{D}_j^{-1}$ becomes $\mathbf{D}_j^{+}$, where + denotes the Moore-Penrose inverse.

MCA as Discriminant Analysis and ANOVA

We can restate the MCA problem in discriminant analysis or ANOVA terms. Suppose we knew the industry coordinates, **X**. Each variable defines a partitioning of these industries. This means we can compute the total variance of **X**, which is just the sum of the between and within group (variable) variance. In matrix notation, this is simply $\mathbf{T} = \mathbf{B} + \mathbf{W}$. We now wish to scale the industries, that is, solve for optimal **X**, in such a way that **W** will be as small as possible while keeping **T** equal to a constant; **I**, for example.

This leads to a trivial solution: all industries in the first category of the variable are a single point, all industries in the second category are another point, and so on. The location of the points in the map is arbitrary, although $\mathbf{W} = \mathbf{0}$ and $\mathbf{B} = \mathbf{T} = \mathbf{I}$. But in MCA we have more than one variable, so a trivial solution for one variable will not be a trivial solution for another. This leads us to seek a compromise solution to the problem. For a given **X**, let us define $\mathbf{T}_*$, $\mathbf{W}_*$, and $\mathbf{B}_*$, which are averages over variables. Clearly, for all variables, **T**, the total variance of **X**, is the same, and now we search for the smallest $\mathbf{W}_*$ with $\mathbf{T} = \mathbf{T}_* = \mathbf{I}$. This defines MCA.

Substituting the optimal category coordinates from (2) gives

$$\sigma(\mathbf{X};*,\ldots,*) = \Sigma_j \operatorname{tr} \mathbf{X}'(\mathbf{I} - \mathbf{P}_j)\mathbf{X}, \qquad (5)$$
$$= mn(p - \operatorname{tr}\mathbf{X}'\mathbf{P}_*\mathbf{X}/n),$$

with $\mathbf{P}_j = \mathbf{G}_j\mathbf{D}_j^{-1}\mathbf{G}_j'$ an $n \times n$ orthogonal projector matrix and $\mathbf{P}_*$ equal to the average of the $\mathbf{P}_j$. Now, from equation (5), $\mathbf{X}'\mathbf{P}_j\mathbf{X}$ is the *variance between categories* of **X** for variable j, and $\mathbf{X}'(\mathbf{I}-\mathbf{P}_j)\mathbf{X}$ is the *variance within categories*. Then $\mathbf{X}'\mathbf{X}$, which we fix at $n\mathbf{I}$, is the *total variance.*

Thus, MCA maximizes the average between-category variance, while keeping the total variance fixed (or, equivalently, minimizes the average within-category variance). Consequently, the main difference between discriminant analysis and MCA is that with the former we have one categorical variable and **X** must be of the form **ZA**, with **Z** known and weights **A** unknown. In MCA, the number of variables m is greater than one and **X** is completely unknown (or $\mathbf{Z} = \mathbf{I}$).

MCA as a Dual Eigenproblem

This development shows that multiple correspondence analysis of $\mathbf{X}$ normalized and the $\mathbf{Y}_j$ free solves the eigenvalue problem:

(6) $$\mathbf{P}_{\bullet}\mathbf{X} = \mathbf{X}\Lambda.$$

where $\lambda_1(\mathbf{P}_{\bullet}) \geq \ldots \geq \lambda_p(\mathbf{P}_{\bullet})$ are the p largest nontrivial eigenvalues of the average projector $\mathbf{P}_{\bullet}$. We use "nontrivial" because $\mathbf{P}_{\bullet}$ always has a largest trivial eigenvalue $\lambda_0(\mathbf{P}_{\bullet}) = 1$, corresponding with the trivial eigenvector $\mathbf{u} = \mathbf{1}$. All other eigenvectors can consequently be chosen such that $\mathbf{X}'\mathbf{u} = \mathbf{0}$. The optimal coordinates for the industries in p dimensions are given by the p eigenvectors, with corresponding eigenvalues, of $\mathbf{P}_{\bullet}$. The corresponding optimal coordinates for the categories of variable j are then $\mathbf{Y}_j = \mathbf{D}_j^{-1}\mathbf{G}_j'\mathbf{X}$, as in equation (2) and follow from the centroid principle.

MCA as Categorical PCA

MCA is also identical to a form of principal component analysis of categorical data outlined by Guttman (1941) and Burt (1950). To show this clearly, suppose that we normalize $\mathbf{Y}$ by $\Sigma_j\mathbf{Y}_j'\mathbf{D}_j\mathbf{Y}_j = n\mathbf{I}$ and $\Sigma_j\mathbf{Y}_j'\mathbf{D}_j\mathbf{u} = 0$ with the condition that $\mathbf{X}$ is free. Thus, we alternatively define the MCA problem as minimizing (1) over free $\mathbf{X}$ and normalized $\mathbf{Y}_j$. As a first step, we examine the loss function:

(7) $$\sigma(*;\mathbf{Y}_1,\ldots,\mathbf{Y}_m) = \min_{\mathbf{X}} \sigma(\mathbf{X};\mathbf{Y}_1,\ldots,\mathbf{Y}_m)$$

Then we find, after substituting the optimal industry coordinates from (3), that MCA amounts to solving the eigenvalue problem:

(8) $$\mathbf{CY} = m\mathbf{DY\Lambda},$$

with $\mathbf{C} = \mathbf{G'G}$ the $\Sigma k_j \times \Sigma k_j$ "Burt matrix," so-called by the French; it contains the bivariate marginals (cross-tables), where $\mathbf{G} = [\mathbf{G}_1|...|\mathbf{G}_m]$ is the super-indicator matrix. The Burt matrix has a block structure, wherein each off-diagonal submatrix $\mathbf{G}_{jl} = \mathbf{G}_j'\mathbf{G}_l$, $j \neq l$, is the cross-table of variables j and l containing the bivariate marginals across the n industries. Each diagonal submatrix $\mathbf{D}_j = \mathbf{G}_j'\mathbf{G}_j$ is the $k_j \times k_j$ diagonal matrix with the univariate marginals of variable j. Then $\mathbf{D}=\text{diag}(\mathbf{C})$ is the diagonal super-matrix of univariate marginals. The optimal coordinates for the variable categories in p dimensions are given by the p eigenvectors with corresponding eigenvalues of $\mathbf{C}$. The corresponding optimal industry coordinates are then $\mathbf{X} = m^{-1}\Sigma_j\mathbf{G}_j\mathbf{Y}_j$, as in equation (3), and follow from the centroid principle.

The dual eigenproblems in equations (6) and (8) amount to a singular value decomposition (SVD) of the matrix $m^{-1/2}\mathbf{GD}^{-1/2}$. It can be shown (van Rijckevorsel and de Leeuw 1988) that both problems have the same eigenvalues; the eigenvectors of interest are then the left and right singular vectors of $m^{-1/2}\mathbf{GD}^{-1/2}$. Moreover, the SVD of this matrix also solves the problem of minimizing the loss function (1) over all normalized $\mathbf{X}$ and all normalized $\mathbf{Y}$. Computing this SVD effectively performs a correspondence analysis on $\mathbf{G}$.

Correspondence Analysis as a Special Case of MCA

Our approach allows us to *specialize* MCA to the situation in which there are just two variables; it then becomes identical to simple or two-way correspondence analysis (Benzecri, et.al. 1973; Greenacre 1984; Lebart, Morineau, and Warwick 1984). See also Carroll, Green, and Schaffer (1986). We have already shown above that, in general, MCA can be formulated as a type of categorical PCA.

Now, suppose we have only two variables, i.e. m=2. Then, $\mathbf{G} = [\mathbf{G}_1|\mathbf{G}_2]$ and $\mathbf{G}_1'\mathbf{G}_2 = \mathbf{F}$, the contingency table for these two variables. We write the univariate marginals for variables 1 and 2 along the diagonals of $\mathbf{D}_1$ and $\mathbf{D}_2$, respectively. For two variables, the Burt matrix $\mathbf{C}$ now has the very special form:

$$\mathbf{C} = \begin{bmatrix} \mathbf{D}_1 & \mathbf{F} \\ \mathbf{F}' & \mathbf{D}_2 \end{bmatrix}$$

Since two-way correspondence analysis is given by the SVD of $\mathbf{D}_1^{-1/2}\mathbf{F}\mathbf{D}_2^{-1/2}$ (cf. Hoffman and Franke 1986, equation (11)), it is immediately seen that the SVD of $\mathbf{D}^{-1/2}\mathbf{C}\mathbf{D}^{-1/2}$, with $\mathbf{D}$ containing $\mathbf{D}_1$ followed by $\mathbf{D}_2$ on the diagonal, equal to the SVD of

$$\begin{bmatrix} \mathbf{I} & \mathbf{D}_1^{-1/2}\mathbf{F}\mathbf{D}_2^{-1/2} \\ \mathbf{D}_1^{-1/2}\mathbf{F}\mathbf{D}_2^{-1/2} & \mathbf{I} \end{bmatrix}$$

gives the same solution.

MCA as an MDS Method

We can also link MCA to multidimensional scaling through the notion of distance. Suppose we were to perform a multidimensional unfolding on **G**, the super-indicator matrix. The MDS solution for unfolding requires a representation where the distance between an industry point and a variable category it occurs in is always smaller than the distance from that industry to a "non-chosen" variable category point. The relation with MCA is obvious. MCA plots a category point in the center of gravity of the industry points for those industries which "choose" that category, with the consequence that, overall, industry points will be closer to the chosen variable categories than to the non-chosen variable categories.

Interpretation of the industry points is guided by the fact that we solve for **X** (with unit total sum of squared distances, $\mathbf{X}'\mathbf{X}$) such that the within-category squared Euclidean distances, $\mathbf{X}'(\mathbf{I} - \mathbf{P}_j)\mathbf{X}$, are as small as possible (or, equivalently, that the between-category squared Euclidean distances, $\mathbf{X}'\mathbf{P}_j\mathbf{X}$, are as large as possible). The primary difference between MCA and MDS is that the MCA solution is obtained at the expense of stronger normalization conditions and a metric interpretation of the data. That is, MCA approaches perfect fit, i.e.

distance $d_{ij}(X,Y)$ between industry point i and variable category point j equals zero, if g_{ij} in the indicator matrix, equals one. This is a stricter requirement than in MDS, which requires that if $g_{ij} = 0$ and $g_{il} = 1$ then $d_{ij}(X,Y) \geq d_{il}(X,Y)$. However, MDS methods for unfolding make weaker assumptions, but also tend to produce degenerate solutions[3].

Our MDS interpretation repeats the multivariate analysis development, which is in terms of variance, and reformulates it in terms of distance. The variance interpretation has appeal for those researchers familiar with discriminant analysis, analysis of variance, and principal component analysis, while the distance interpretation should appeal to those more comfortable with multidimensional unfolding. Both interpretations, however, are correct.

Handling Missing Data

A distinct advantage of nonlinear multivariate analysis in general, and multiple correspondence analysis in particular, is the treatment of missing data. Because all variables are treated as categorical, missing data may always be considered as a separate and bona fide category: the "missing category." Thus, there is no need to impute, estimate, or otherwise insert substitute information for the missing datum. This also means there is no need to discard observations with missing data.

In the context of multiple correspondence analysis, there are three general approaches to handle missing data. In the first place, we may leave the indicator matrix incomplete and proceed with analysis. Gifi (1990) refers to this as *missing data passive*. For our example, this means that we do not alter the four zeros in Table 2 for the variable Expectation of Quality for industry FOOD. If we choose this option for missing data, the missing values are not quantified.

A second treatment for missing data completes the indicator matrix with a single additional column for each variable with missing data. This is called *missing data single* (Gifi 1990). Missing data single has the effect of treating missing data as if they are in a category

[3]But see DeSarbo and Hoffman (1987) for a multidimensional unfolding solution for G, incorporating reparameterization, which avoids the degeneracy problem.

by themselves. If we were to choose this option for our example, we would add a single column to the indicator matrix in Table 2 and enter a single 1 for the industry FOOD, to reflect that it had a missing datum for one of the variables. If further industries had missing data for that variable, they too would have 1's entered in that additional single column. This option therefore treats industries as if they are in the same category with respect to what generated the missing data. Put another way, the process assumed to underlie the occurrence of missing data is assumed to be homogeneous (across industries) for each variable. Such a treatment produces a scaling for the missing data that is the average industry score for industries with missing data on a variable.

The third option requires the indicator matrix to be completed by adding to G_j as many additional columns as there are missing data for the j^{th} variable. Gifi (1990) refers to this as *missing data multiple*. In this case, as many extra columns as there are industries with missing data on the j^{th} variable are added to G^j to complete it, and each such column has only a single 1. The treatment of missing data with the missing data multiple option assumes that whatever process generated the missing data differs for different industries. Each missing value receives the scaling of the associated industry.

Not surprisingly, if there are not too much missing data and they are randomly distributed over row objects and variable categories, the differences among the three treatments will be minor, with highly similar interpretation of the resulting solutions. However, if missing data seem to collect at some objects or variables, the results can be highly different. Treating missing data as a category in its own right allows special study of those objects with missing values.

EXAMPLE

In this section, we offer a detailed example based on the multivariate indicator matrix of customer satisfaction variables presented in Table 2. The reader is referred to Anderson and Sullivan (1993) for an insightful substantive treatment of these data. We use this example to illustrate the most important geometrical aspects of MCA maps, but remind the reader that the rules apply in general to objects (rows) and variable categories (columns) of the scaled multivariate data matrix.

Customer Satisfaction Across 31 Swedish Industries

Remember that the purpose of MCA is to produce a map with loss as small as possible and where the distances between industries and the variable categories they occur in are as small as possible. Our original idea for an "arbitrary" map discussed earlier actually represents the initial (i.e. iteration 0) MCA solution, based on $\mathbf{Y}_j$ estimated according to the centroid principle and **X** arbitrarily normalized. Thus, this solution is already a half-step in the right direction. The loss for this solution is .9650.

The initial solution is highly unsatisfactory, however, as the map (not shown in the interests of space) is very cluttered. Since loss (i.e. fit) is simply the sum of squares of the line lengths in the plot, the optimal solution, in keeping with the principles of MCA, is the one where the distances connecting points are minimized. After ten iterations, the solution is much more satisfactory as the lines connecting industries to their categories are as short as possible, and the fit is improved considerably (loss=.2354). Although both solutions represent the super-indicator matrix **G**, the final solution in two-dimensional Euclidean space is the more appealing.

The MCA of the data in Table 2 produces dominant eigenvalues of .60 and .38. Since the singular values from an MCA are canonical correlations, we interpret the eigenvalues (squared singular values) as squared canonical correlation coefficients. Industries and variable categories are represented as points in a joint low-dimensional map. This joint map appears in Figure 1. Industries are represented by open circles and satisfaction variable categories by closed circles. In this Figure, we leave the variable categories unlabeled for ease of presentation.

Insert Figure 1 about here

MCA requires industries corresponding to a certain category of a variable to have a position in the map in the direction of the associated category; other industry points will have a position in the opposite direction. Stated differently, industries are relatively close to categories they are in and relatively far from categories they are not in. Further, industries mostly occurring in the same category tend to be close to each other and categories sharing mostly the same industries tend to be close, as well[4]. Interpretation of category points is guided by the centroid principle: category coordinates are the center of gravity, or centroid, of industry coordinates occurring in that category.

The variable category points are plotted in Figure 2. In this Figure, we have omitted the industries for ease of interpretation. Clear regions of customer satisfaction are revealed in this plot. We can see that in the upper right quadrant are the variable categories associated with the most extreme levels of dissatisfaction, price increase intolerance, unreasonable price, negative repurchase intentions, low quality, low expectations of that quality, and far distance from the ideal product or service in the category. Thus, industries in this area of the map are associated with these categories. Looking back at Figure 1, we can see that this includes the TV stations, telecommunications businesses with business clients, rail, police and the post office packet service with public clients. Customers of these services and products are decidedly unhappy! Perhaps not surprisingly, these services represent state-owned monopolies.

Insert Figure 2

To a large extent, the analysis separates the extremely disenchanted (the upper right portion of the figure) from the rest of the customers. We can examine Figure 2 further and identify the region of extreme satisfaction slightly below and to the left of the origin. Industries here include the airlines, mainframe computers, pharmacies, automobiles, and food. Moving up in the figure we identify another region of satisfaction, with the corresponding industries of charter travel, both public and business banking, shipping, groceries, gasoline, public life insurance and mail order. In the lower center region of the map, we find those

[4]This will only be true approximately in reduced dimensionalities.

variable categories indicating dissatisfaction. The industries associated with these judgments include public and business insurance, furniture, business post office packets, clothing, public and business post office letters, department stores, and public telecommunications.

Thus, the multiple correspondence analysis of Table 2 has clearly identified four distinct regions relating to customer satisfaction. Notice also that the analysis has revealed distinctly nonlinear patterns of satisfaction. That is, variable categories are not linear with the dimensions of the space. This illustrates yet another advantage of nonlinear multivariate analysis. By treating all variables as categorical, we may discover patterns in the data that would be hidden by conventional linear multivariate analysis.

The "object plot" in Figure 3, simply the plot of industry points only, makes it easy to see relationships among the industries. The distance between two industry points is related to the homogeneity of their profiles, or more generally, their response-patterns[5]. Industries with identical patterns are plotted as identical points. This is illustrated in Figure 3 for TV Stations 1 and 2 in the upper right, which as may be verified in Table 2, are identical in profile. Industries which are very similar include the group in the upper left, comprised of mail order, public life insurance, groceries, gasoline, shipping, and public banking. If a category applies uniquely to only one industry, then the industry point and this category point will coincide. The same is true when a category applies uniquely to a group of industries with identical response patterns. For example, the price increase tolerance category "very likely to switch" applies uniquely to the three TV stations, and their similar profiles mean these points are very close to each other in the two-dimensional map.

Insert Figure 3 about here

Along the same lines, the position of "very likely to switch" indicates that a category point with low marginal frequency will be plotted farther towards the periphery of the map, while a category with high marginal frequency will be plotted nearer to the origin of the map ("very reasonable" and "reasonable" price). As a corollary, industries with response patterns

[5]Note that the reverse will not necessarily be true. Two industry points that are close together in a map of the first two-dimensions may be far apart in higher dimensionalities.

similar to the "average" response pattern will be plotted more towards the origin (mainframe computers, airline 1), while industries with "unique" patterns (for example, post office letter service for businesses and TV Stations 1,2, and 3) appear in the periphery. These statements, however, are only precisely true when considering all dimensions, and not necessarily the map for the first two dimensions only.

Star Plots

The star plots displayed in Figures 4, 5, and 6 for the variables quality, satisfaction and distance to ideal, respectively, illustrate a number of important properties of MCA. Each star plot maps a particular variable's categories with all the industry points and shows loss for each variable. Relative loss in the two-dimensional solution is the sum of the squared distances between industry points in a cluster and their average, the category point[6]. We have drawn lines in the star plots to illustrate this. The star plots also illustrate the centroid principle: that the optimal category coordinates are centroids of industry points in those categories.

Insert Figures 4, 5, and 6

Since category points are the average of the industry points that share the category, for each variable, categories of that variable divide the industry points into clusters, and the category points are the means of the clusters. For example, Figure 4 depicts clearly the four different clusters of product or service quality, Figure 5 reveals the corresponding levels of satisfaction with these industries, and Figure 6 displays the groups of industries classified according to how far away customers viewed them in relation to the ideal in that industry.

A variable *discriminates* better to the extent that its category points are further apart. The discrimination measures, shown in Table 3, are quantified as the squared correlations between the industry coordinates $\mathbf{X}$ and the optimally transformed variables, $\mathbf{G}_j\mathbf{Y}_j$, and are interpreted as squared factor loadings. Quality (Figure 4), Satisfaction (Figure 5) and Ideal (Figure 6) discriminate best among industries on both dimensions.

[6]In the perfect solution (loss equal zero), all car points will coincide with their category points, but there must be at least as many categories as cars for this to happen.

Insert Table 3 about here

To better understand the discrimination measures, we begin by noting that the covariances between the industry coordinates and the optimally transformed variables are simply $\mathbf{X'G}_j\mathbf{Y}_j/n$, but, using equation (2), the optimally transformed variables $\mathbf{G}_j\mathbf{Y}_j$ can be expressed as $\mathbf{G}_j\mathbf{D}_j^{-1}\mathbf{G}_j'\mathbf{X}$ and this equals $\mathbf{P}_j\mathbf{X}$, where $\mathbf{P}_j$ is the orthogonal projector matrix from equation (5). Thus, the covariances can be written as $\mathbf{X'P}_j\mathbf{X}/n$ and these are the covariances of the optimally transformed variables, as well, since $\mathbf{Y}_j'\mathbf{D}_j\mathbf{Y}_j/n = \mathbf{X'P}_j\mathbf{X}/n$. Of course, our normalization fixes $\mathbf{X'X}/n$ at $\mathbf{I}$. Thus, for corresponding dimensions, the squared correlations between the industry coordinates and the variables are written on the diagonals of $\mathbf{X'P}_j\mathbf{X}/n$. It follows directly that the average over variables of the discrimination measures are the diagonals of $\mathbf{X'P}_\star\mathbf{X}/n$ which is equal to Λ. For each dimension, MCA thus maximizes the sum total of the discrimination measures.

Horseshoes

It turns out that the map in Figure 1 exhibits what is called the "horseshoe" effect. It has a quadratic structure, in the sense that industries are on or close to a second degree polynomial in two-dimensions. That is, the second dimension is a quadratic function of the first dimension, in this case contrasting satisfied and very dissatisfied groups of customers with those very satisfied and dissatisfied. Because industries are on a horseshoe, and category points are close to the industries occurring in them, the category points will tend to be on a horseshoe as well. In this case, the star plots will tend to be (pieces of) horseshoes, and the stars will be elongated along the structure.

In order to explain the idea of horseshoes, we use a technique described by Gifi (1990) as *gauging*. In gauging, we use a mathematically defined data structure (that is, a model which we call a gauge), apply the technique to the model, and observe how the known properties of the model are reproduced by the technique. In our case, the technique is MCA and there are various gauges we can select which lead to horseshoes.

The first gauge that is relevant is the multivariate normal distribution. If data are a sample from a multivariate normal, classified into a discrete number of categories, then an MCA of these data (on a large enough sample), will show components which are linear, components which are quadratic, components which are cubic, and so on. The horseshoe corresponds with the case in which the first component (the one corresponding with the largest eigenvalue) is linear and the second component is quadratic (the one corresponding with the second largest eigenvalue). Multinormal data do not necessarily give horseshoes: it is also possible that both the first and second components are linear with the original category values. An example of this is given in Gifi (1990, pp.382-384).

Another famous gauge leading to horseshoes is the Guttman scale (Guttman 1950). If individuals and categories can be jointly ordered in such a way that the indicator matrices have a banded or parallelogram structure (each individual only gives positive responses to a number of adjacent categories in the order), then again the first two dimensions will form a horseshoe. This generalizes to various item-response models such as the Rasch model (Rasch 1966), the unidimensional unfolding model (Coombs 1964) unimodal structures used in ecology (Ter Braak 1986), and so on. All such gauges will give horseshoes when MCA is applied.

It is somewhat of a problem to decide if a horseshoe is desirable or undesirable. In one sense, we should be happy with one, because it shows a strong underlying order structure in our data. In effect, the horseshoe in multiple correspondence analysis is equivalent to the general factor in principal component analysis. In this example, that underlying variable is satisfaction. The first dimension dominates and the industries are ordered according to the ordering in the data. On the other hand, we could also be unhappy, because the horseshoe uses two linear dimensions to present this ordering. Basically, we use two dimensions to present a one-dimensional structure, and any higher-dimensional information remains hidden. For ways of dealing with this problem, we refer the reader to Bekker and de Leeuw (1988, pp. 29-30).

In some cases, horseshoe type structures can be decomposed, basically by collecting all linear transformations in one solution (which then corresponds to the ordinary linear PCA solutions), by collecting all quadratic components in another solution, and so on. In other cases, we require the transformations to be linear (or monotonic), which forces the MCA solutions away from the horseshoe. This could be implemented, for example, in programs such as PRINCALS (SPSS Inc. 1989) or PRINQUAL (SAS® Institute Inc. 1988).

DISCUSSION

For MCA maps to be useful in marketing, rules for representation and interpretation must be explicit and unambiguous. Consider our example. There we had a data matrix indicating which categories of various variables measuring satisfaction a series of Swedish industries falls into. We observed from Table 1, for example, that each industry fell into one of the following categories for the price variable: very unreasonable, unreasonable, reasonable, and very reasonable. As we saw in Table 2, the data matrix had a 1 whenever an industry fell into its category of the variable and a 0 otherwise.

The joint map produced from the MCA in Figure 1 has points for each industry and for each variable category. A dimensional interpretation (as is typically done in factor analysis) is problematic because we must identify constructs which can simultaneously describe both industries and variables, and it is somewhat difficult to find constructs which can convincingly do that job. On the other hand, interpretations in terms of closeness of the within-set distances, i.e. between each industry and each variable category, are quite natural and compelling. This illustrates one of the problems in graphically representing rectangular categorical data matrices: how to construct an interpretable joint map of the row and column points. The fundamental issue concerns the appropriate way to represent both the objects corresponding with the rows and variables corresponding with the columns of the matrix in the same map.

This problem is more important than ever, as the three major statistical packages have MCA programs in which the choice of scaling of row and column coordinates is left largely to the user (BMDP 1988; SAS® Institute Inc. 1988; SPSS Inc. 1989). In addition, the variety of commercially available PC-based programs offer numerous options but little guidance to the user (BMDP 1988; Greenacre 1986; Nishisato and Nishisato 1986; SAS® Institute Inc. 1988; Smith, 1988; see also Hoffman 1991 for a review).

Recently, this debate has raged in the marketing literature. Carroll, Green and Schaffer (1986; 1987; 1989) and Greenacre (1989) provide an interesting and rather heated discussion of this seemingly innocuous topic of scaling row and column points. It seems to us that the points of view of CGS on the one side and Greenacre on the other reflect to some extent the bias that each of these researchers brings to marketing.

Greenacre was trained in the "French" school, which appears to correspond nicely with the fact that he takes simple correspondence analysis (CA) to be a more fundamental and satisfactory technique than MCA. It also means that he tends to emphasize the so-called "chi-square distance" interpretation of the within-set distances[7]. CGS have their starting point in multidimensional scaling and unfolding theory, which naturally leads them to emphasize between-set distance relations.

Although our "psychological" interpretation of the debate may be interesting, it does little to resolve any practical or theoretical problems. Marketing researchers still wish to know what is "best", or at least what they should do in any particular situation. Our geometrical approach to MCA leads directly to a set of unambiguous rules for representation and interpretation of MCA maps.

MCA as a "Model"

Let us consider distance models for a moment. In multidimensional unfolding, we start explicitly with a model formulated in terms of fitting between-set distances to data. It is possible to formulate MCA as a particular, although somewhat peculiar, approximate solution

[7]The "chi-square" distance between two row points, say, is equal to the weighted sum of squared differences between row "profile" values, with weights equal to the inverse of the relative frequencies of the columns. A similar definition holds for column points. These within-set distances are denoted *chi-square* because if the data are a contingency table, then the numerator creates squared differences between conditional row probabilities (the profiles), while the denominator weights the squared differences by inverse relative column marginals; thus, as Novak and Hoffman (1990) show, distances can be interpreted in terms of a) standardized residuals (components of chi-square), b) $O - E^{I}/E^{I}$, observed minus expected counts under the log-linear model of independence as a proportion of expected counts under independence, and c) "profiles" (conditional probabilities).

to the unfolding problem (Heiser, 1981), but in general, MCA is not thought of in this way. Rather, many marketing researchers think of MCA "merely" as a technique for graphically representing a data matrix. Unfortunately, it is precisely this lack of explicitness that can lead to problems with interpretation.

Selecting a specific model implies, at the same time, choosing a framework for interpretation of the results. The results are interpreted within the model and are related to its assumptions. The interpretations use the terminology in which the model is formulated. MDS models, for instance, are formulated in terms of (dis)similarity and distance, and thus MDS results are also interpreted naturally in these terms. Factor analysis is formulated in terms of unobserved variables called factors, and consequently we can interpret a factor analysis solution as soon as we can interpret the factors. This imposed framework can be a very positive aspect of the analysis: interpretation is unambiguous, key concepts are known to many researchers in the field, and communication is made easy. But it can also be a great hindrance to communication with other fields, as the history of factor analysis amply shows. A model, like any cognitive map of the world, is a filter of reality. In many cases, such a filter is needed in order to proceed; in other cases, some of the more important, stable, and interesting aspects of the data may be filtered out or distorted by the imbedding in a rigid framework.

In the case of MCA, it seems natural for marketing researchers, perhaps owing to the popularity of multidimensional unfolding, to concentrate on simple geometrical aspects of the MCA map (e.g. interpoint distances), and observe what aspects of the data matrix they are trying to represent. This means, of course, that we look at MCA as if it is, in some devious way, still trying to fit a model to the interpoint distances. It merely does not make its loss function explicit, and thus it is inferior (at least in this sense) to unfolding techniques.

In our framework, distances corresponding to 1's in the indicator matrix must be small, but this requirement alone is not sufficient to produce a map, since the trivial solution satisfies it. Hence, we need a normalization. A natural normalization would be to examine all the distances and simply minimize the between-set distances (i.e. the sum of squares) keeping all other distances fixed. However, this always leads to a one-dimensional solution. Thus, we require something stricter, and so impose dimension orthogonality and normalization constraints. Which way we choose to normalize (i.e. normalize the objects **X** and leave the variables $\mathbf{Y}_j$ free, or the reverse) is immaterial geometrically, since the problem is formulated in a joint space. However, the choice affects the interpretation. Therefore, substantive considerations will almost always guide the researcher's choice of normalization.

Choosing the "Best" Normalization

Case I. The first approach is to normalize the set of object coordinates **X** and leave the variable category coordinates $\mathbf{Y}_j$ free. Normalizing **X** means that the $\mathbf{Y}_j$ are found by the centroid principle. In this case, the optimal scaling of a variable category (equation 2) satisfies $\mathbf{Y}_j'\mathbf{D}_j\mathbf{Y}_j = \mathbf{X}'\mathbf{P}_j\mathbf{X}$ and thus $\mathbf{Y}'\mathbf{D}\mathbf{Y} = m\mathbf{X}'\mathbf{P}_*\mathbf{X} = m\Lambda$. Quite simply, in words, a variable category coordinate is the centroid of the coordinates of the objects in that category. Gifi (1981; 1990) calls this the *first centroid principle.*

Case II. The second approach is to normalize the set of variables and leave the objects free. Normalizing the $\mathbf{Y}_j$ means that **X** is found by the centroid principle. Here, the optimal scaling of the objects (equation 3) satisfies $\mathbf{X}'\mathbf{X} = m^{-1}\mathbf{Y}'\mathbf{D}\mathbf{Y}\Lambda = \Lambda$. In words, the optimal coordinate of an object is the centroid of the coordinates of the variable categories the object occurs in. This is the *second centroid principle* (Gifi 1981; 1990).

Case III. Finally, we may choose to normalize *both* the objects and the variable categories. This is usually referred to as the "French scaling." This option treats rows and columns symmetrically and drops the centroid principle. Within-set relations are interpretable as chi-square distances, but no between-set interpretation is possible.

Cases I and II, the centroid principles, *define* graphical representation and interpretation of the MCA map. We adopt Case I as convention, but note that which case the researcher chooses is completely arbitrary, *from a geometrical standpoint,* as we can switch from one to the other without changing anything essential. MCA is an elegant multivariate method because these two normalizations can be translated into each other, through the transition formula.

The rationale for choosing the centroid principle to guide normalization, as opposed to Case III or any other normalization that one could devise, lies in the inherent asymmetry of multivariate data. All applications of MCA, and consequently all interpretations, are inherently asymmetric as multivariate data are, by definition, row or column conditional. In other words, we treat rows and columns differently since each represents distinct entities we wish to characterize graphically. Thus, we define our data matrix as row or column conditional and proceed from there.

Row conditionality implies that we primarily wish to emphasize rows and scale them such that in the map, row points are closer together to the extent that rows are more similar with respect to the variables making up the columns. This suggests that it is logical to think of ordering objects by variables. Columns, i.e. variables, are the center of gravity of the rows. Practically speaking, choosing the Case I normalization, implied by row conditional data, means that objects will be equally spread in all directions in the map, with category points indicating the averages of subgroups of objects. In other words, objects are sorted into their respective categories of a variable. If our concern is primarily with the objects, as it would be when objects are brands, for example, then objects are normalized and the centroid interpretation applies with respect to the variable categories as weighted averages of the brands in that category. This leads, as in our industry data example, to a joint map for the industries and variable categories and a set of star plots for each separate variable.

Column conditionality implies that we primarily wish to represent the columns as points in a map and scale them such that columns close together are more similar with respect to the objects in the matrix. In situations where the objects represent individuals, for example in the Q-technique, then the Case II normalization of MCA orders variables by these individuals. In this case, variables are normalized and the individuals are free. Then, we obtain a single map for all the variable categories and a set of star plots for each individual with categories of all variables in the plots.

These arguments make clear why Case III, with symmetric treatment of rows and columns, is the least defensible normalization, both from the geometric and substantive points of view. It will almost always be the case that primary focus is on either the objects or the variables, but not both equally. *The aims of the investigation guides the researcher's choice of normalization.*

Greenacre (1989) prefers to normalize according to Case III (symmetrically scaling both sets of points in "principal" coordinates) and emphasize the within-set "chi-square" distances at the expense of any between-set interpretation. Carroll, Green and Schaffer (1989) recommend a variant of Case III, which despite their arguments to the contrary, does *not* allow for between-set interpretations. Quite simply, this is because the correspondence analysis of the two-way contingency table and the MCA of the super-indicator matrix (what CGS call the "pseudo-contingency table") are well known to give equivalent parameter estimates. As the researcher loses the centroid principle, fundamental to MCA and a critical aspect of interpretability, in the symmetric Case III normalization or its variant, we do not recommend it.

SUMMARY

In this chapter, we introduced multiple correspondence analysis as a nonlinear multivariate analysis method which integrates ideas from both classical multivariate analysis and multidimensional scaling. We formulated MCA as a graphical method which seeks to connect brands, say, with all the variable categories they are in and uses a least squares loss function as the rule to do this. Our approach emphasized the geometrical aspects of multiple correspondence analysis. Considering MCA in this light leads directly to a set of unambiguous rules for representation and interpretation of MCA maps.

As we saw, interpretation of the joint map stems not from terms of "chi-square distance" or "profiles," but rather, follows from *le principe barycentrique*, the centroid principle, which says that brands close together are similar to each other. In keeping with this view, we developed simple correspondence analysis as a special case of MCA with the number of variables equal to two, rather than as a method to approximate within-set chi-square distance.

We also showed how MCA is related to a number of familiar MVA techniques, including analysis of variance, discriminant analysis, principal component analysis and multidimensional scaling. Our detailed example illustrated the most important geometrical properties of MCA maps. We hope this chapter has demonstrated how MCA, a powerful multivariate methodology, may suit many and varied applications in marketing research.

REFERENCES

Aldenderfer, M.S. and R.K. Blashfield (1984), *Cluster Analysis*, Beverly Hills, CA: Sage.

Anderson, Eugene W. and Mary W. Sullivan (1993), "The Antecedents and Consequences of Customer Satisfaction for Firms," *Marketing Science*, forthcoming.

Bekker, P. and Jan de Leeuw (1988), "Relations between Variants of Non-linear Principal Component Analysis," In J.L.A. van Rijckevorsel and J. de Leeuw (Eds.), *Component and Correspondence Analysis: Dimension Reduction by Functional Approximation*, pp. 1-31. New York: John Wiley & Sons.

Benzecri, J.-P., et. al. (1973), *Analyse des Donnees* (2 vols.) Paris: Dunod.

Bock, R. Darrell (1960), "Methods and Applications of Optimal Scaling," Laboratory Report No. 25, L.L. Thurstone Psychometric Laboratory, University of North Carolina, Chapel Hill.

BMDP Statistical Software, Inc. (1988), CA: Correspondence Analysis. Technical Report No. 87. Los Angeles, CA.

Burt, Cyril (1950), "The Factorial Analysis of Qualitative Data," *British Journal of Psychology (Statistical Section)*, 3 (November), 166-185.

Carroll, J. Douglas, Paul E. Green, and Catherine M. Schaffer (1986), "Interpoint distance Comparisons in Correspondence Analysis," *Journal of Marketing Research*, 23 (August), 271-80.

Carroll, J. Douglas, Paul E. Green, and Catherine M. Schaffer (1987), "Comparing Interpoint Distances in Correspondence Analysis: A Clarification," *Journal of Marketing Research*, 24 (November), 445-50.

Carroll, J. Douglas, Paul E. Green, and Catherine M. Schaffer (1989), "Reply to Greenacre's Commentary on the Carroll-Green-Schaffer Scaling of Two-Way Correspondence Analysis Solutions," *Journal of Marketing Research*, 26 (August), 366-8.

Coombs, Clyde (1964), *A Theory of Data*, New York: Wiley.

Cooper, Lee (1983), "A Review of Multidimensional Scaling in Marketing Research," *Applied Psychological Measurement*, 7 (Fall), 427-50.

de Leeuw, Jan (1973), *Canonical Analysis of Categorical Data*. Unpublished doctoral dissertation, Psychological Institute, University of Leiden, The Netherlands. Reissued DSWO-Press, Leiden, 1984.

DeSarbo, Wayne S. and Donna L. Hoffman (1987), "Constructing MDS Joint Spaces from Binary Choice Data: A New Multidimensional Unfolding Model for Marketing Research," *Journal of Marketing Research*, 24 (February), 40-54.

Edgerton, H.A. and L.E. Kolbe (1936), "The Method of minimum Variation for the Combination of Criteria," *Psychometrika*, 1, 183-187.

Fisher, Ronald A. (1940), "The Precision of Discriminant Functions," *Annals of Eugenics*, 10, 422-9.

Fornell, Claes (1992), "A National Customer Satisfaction Barometer: The Swedish Experience," *Journal of Marketing*, 56 (January), 1-18.

Gifi, Albert (1981), *Non-linear Multivariate Analysis*. Leiden: Department of Data Theory.

Gifi, Albert (1990), *Nonlinear Multivariate Analysis*. New York: John Wiley & Sons.

Green, Paul E. (1975), "Marketing Applications of MDS: Assessment and Outlook," *Journal of Marketing*, 39 (January), 24-31.

Green, Paul E., Abba M. Krieger and J. Douglas Carroll (1987), "Multidimensional Scaling: A Complementary Approach," *Journal of Advertising Research*, October/November, 21-27.

Greenacre, Michael J. (1984), *Theory and Applications of Correspondence Analysis*. London: Academic Press.

Greenacre, Michael J. (1986), "SIMCA: A program to perform simple correspondence analysis. *American Statistician*, 51, 230-231.

Greenacre, Michael J. (1989), "The Carroll-Green Schaffer Scaling in Correspondence Analysis: A Theoretical and Empirical Appraisal," *Journal of Marketing Research*, 26 (August), 358-65.

Guttman, Louis (1941), "The Quantification of a Class of Attributes: A Theory and Method of Scale Construction." In *The Prediction of Personal Adjustment*, Paul Horst, ed. New York: Social Science Research Council, 319-348.

Guttman, Louis (1950), "The Basis for Scalogram Analysis," in S.A. Stouffer, et. al. (Eds.), *Measurement and Prediction*. Princeton: Princeton University Press.

Hayashi, Chikio (1950), "On the Quantification of Qualitative Data from the Mathematico-Statistical Point of View," *Annals of the Institute of Statistical Mathematics*, 2 (1), 35-47.

Heiser, Willem J. (1981), *Unfolding Analysis of Proximity Data*. Department of Data Theory, University of Leiden, The Netherlands.

Hirschfeld, H.O. (1935), "A Connection between Correlation and Contingency," *Proceedings of the Cambridge Philosophical Society*, 31 (October), 520-524.

Hoffman, Donna L. (1991), "Review of Four Correspondence Analysis Programs for the IBM PC," *American Statistician*, 45 (4), November, 305-311.

Hoffman, Donna L. and Rajeev Batra (1991), "Viewer Response to Programs: Dimensionality and Concurrent Behavior," *Journal of Advertising Research*, August/September, 46-56.

Hoffman, Donna L. and George R. Franke (1986), "Correspondence Analysis: Graphical Representation of Categorical Data in Marketing Research," *Journal of Marketing Research*. 23 (August), 213-27.

Hoffman, Donna L. and Jan de Leeuw (1992), "Interpreting Multiple Correspondence Analysis as a Multidimensional Scaling Method," *Marketing Letters*, 3(3), 259-272.

Horst, Paul (1935), "Measuring Complex Attitudes," *Journal of Social Psychology*, 6 (3), 369-374.

Horst, Paul (1936), "Obtaining a Composite Measure from a Number of Different Measures of the Same Attribute," *Psychometrika*, 1, 53-60.

Kaciak, Eugene and Jordan Louviere (1990), "Multiple Correspondence Analysis of Multiple Choice Data," *Journal of Marketing Research*, 27 (November), 455-65.

Lebart, Ludovic, Alain Morineau, and Kenneth M. Warwick (1984), *Multivariate Descriptive Statistical Analysis*. New York: John Wiley & Sons, Inc.

Nishisato, Shizuhiko (1980), *The Analysis of Categorical Data. Dual Scaling and its Application.* Toronto: University of Toronto Press.

Nishisato, Shizuhiko and Ira Nishisato (1986), "DUAL3 Users' Guide," MicroStats. Toronto, Canada.

Novak, Thomas P. and Donna L. Hoffman (1990), "Residual Scaling: An Alternative to Correspondence Analysis for the Graphical Representation of Residuals from Log-Linear Models," *Multivariate Behavioral Research*, 25(3), 351-370.

Pearson, Karl (1904), "On the Theory of Contingency and its Relation to Association and Normal Correlation," *Drapers Company Research Memoirs (Biometric Series)*, no. 1.

Rasch, G. (1966), "An Item Analysis Which Takes Individual Differences Into Account," *British Journal of Statistical Psychology*, 19, 49-57.

Richardson, Marion. and Kuder, G.F. (1933), "Making a Rating Scale that Measures," *Personnel Journal*, 12, 36-40.

SAS® Institute Inc. (1988), "SAS® Technical Report: P-179 Additional SAS/STAT™ procedures Release 6.03," Industryy, NC.

Smith, Scott M. (1988), PC-MDS: Multidimensional Statistics Package. Institute of Business Management, Brigham Young University, Provo, Utah.

SPSS Inc. (1989), "SPSS-X: CATEGORIES," Chicago, IL.

Ter Braak, Cajo J.F. (1986), "Canonical Correspondence Analysis: A New Eigenvector Technique for Multivariate Direct Gradient Analysis, *Ecology*, 67, 1167-79.

Tucker (1941)

Valette-Florence, Pierre and Bernard Rapacchi (1991), "Improvement in Means-End Chain Analysis Using Graph Theory and Correspondence Analysis," *Journal of Advertising Research*, February/March, 30-45.

van Rijckevorsel, Jan L.A. and Jan de Leeuw (1988), *Component and Correspondence Analysis: Dimension Reduction by Functional Approximation*. New York: John Wiley & Sons.

Wilks, Samuel S. (1938), "Weighting Systems for Linear Functions of Correlated Variables When There is No Independent Variable," *Psychometrika*, 3, 23-40.

Joint Map of Swedish Industries and Customer Satisfaction Variables

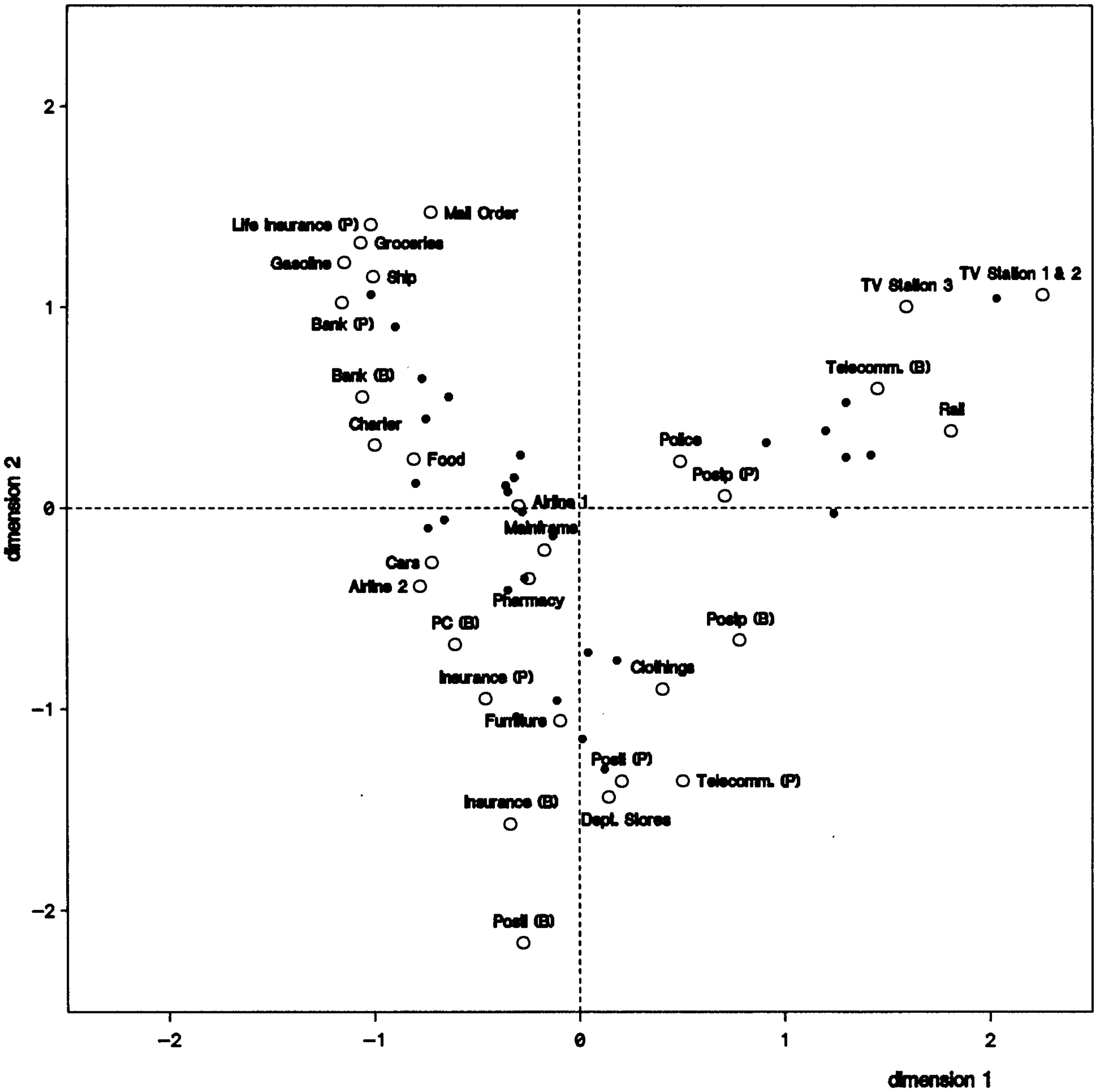

Figure 1

Map of Category Quantifications for Customer Satisfaction Variables

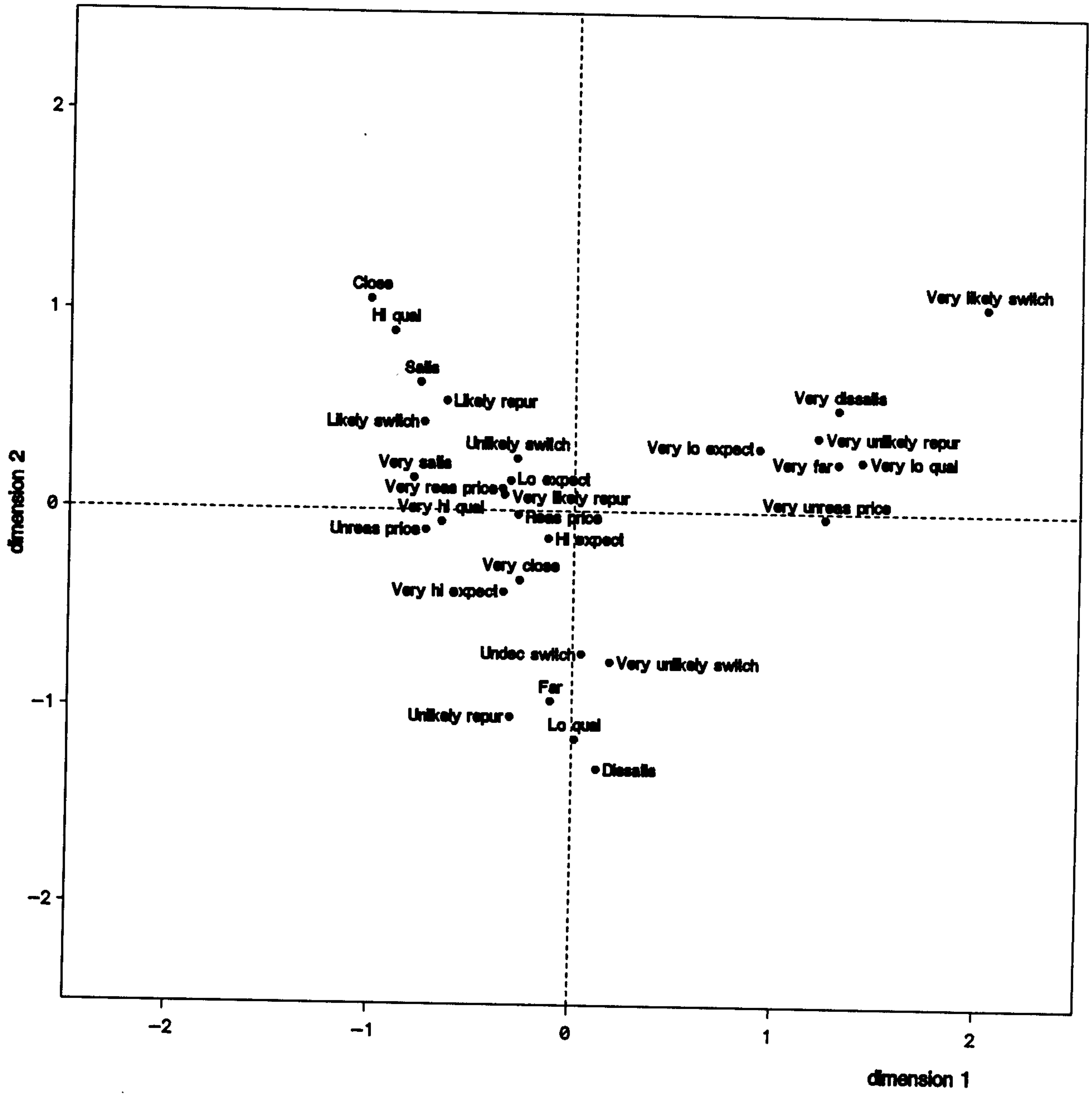

Figure 2

Map of 31 Swedish Industries

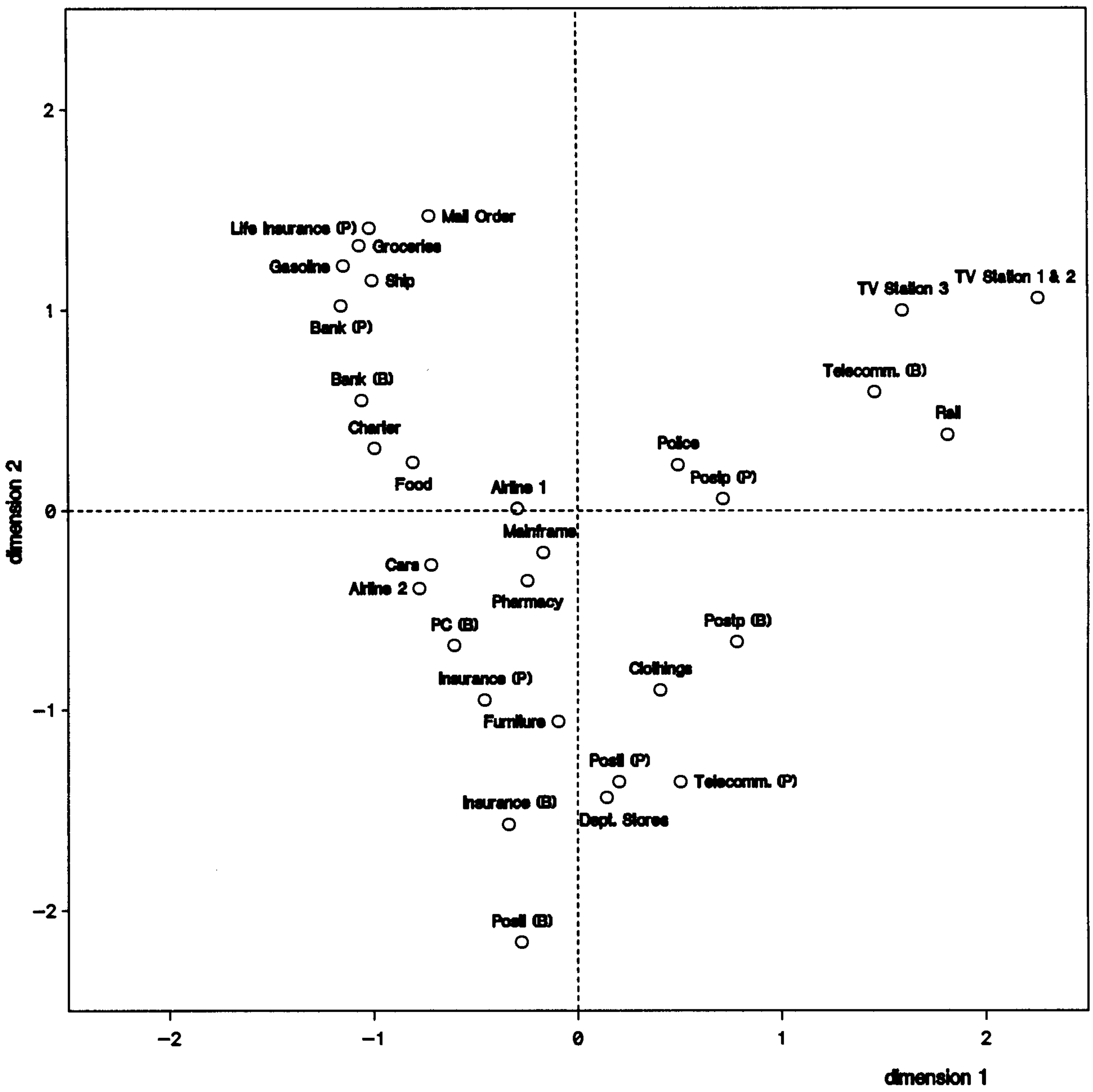

Figure 3

Star Plot for Quality

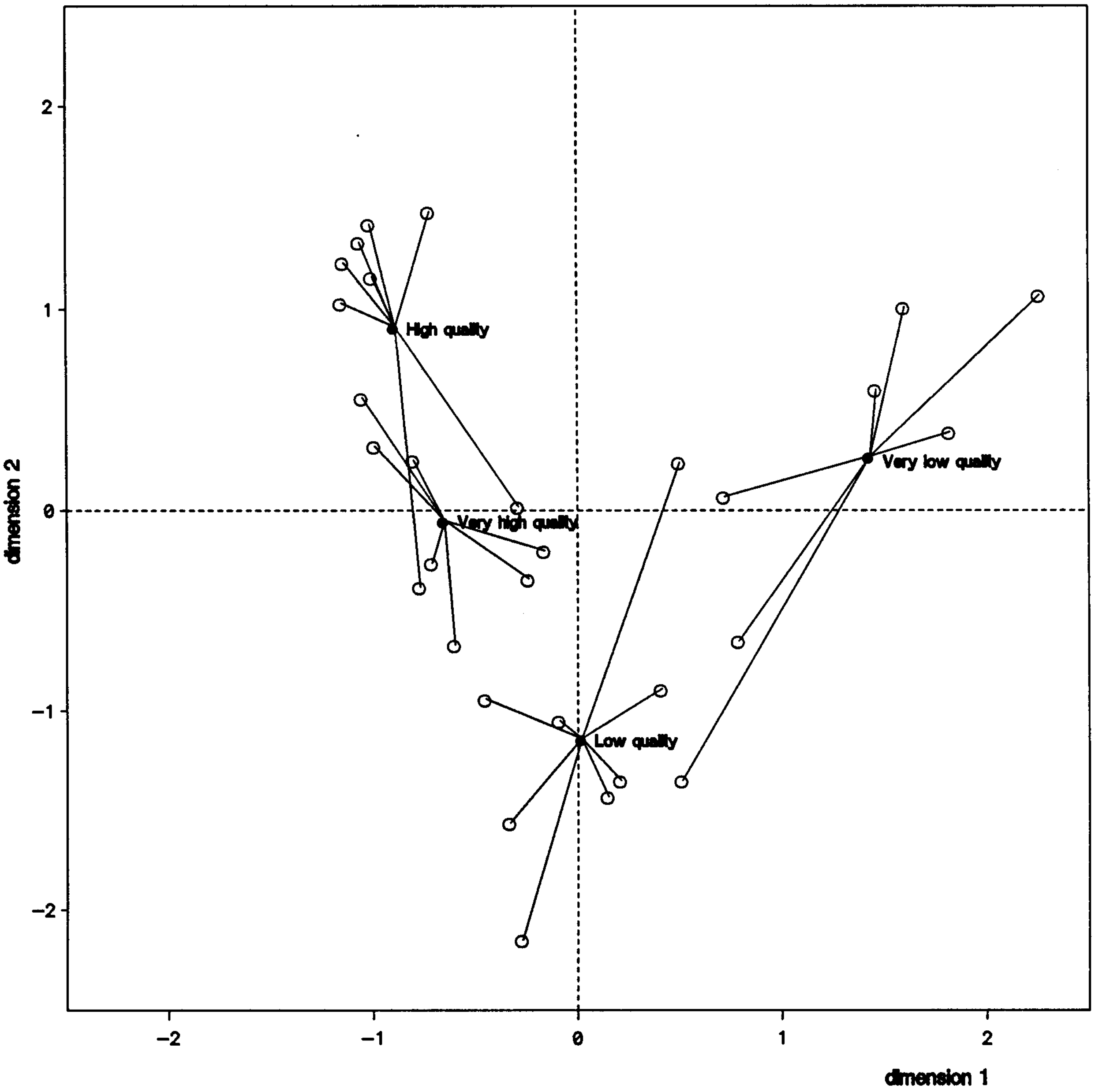

Figure 4

Star Plot for Customer Satisfaction

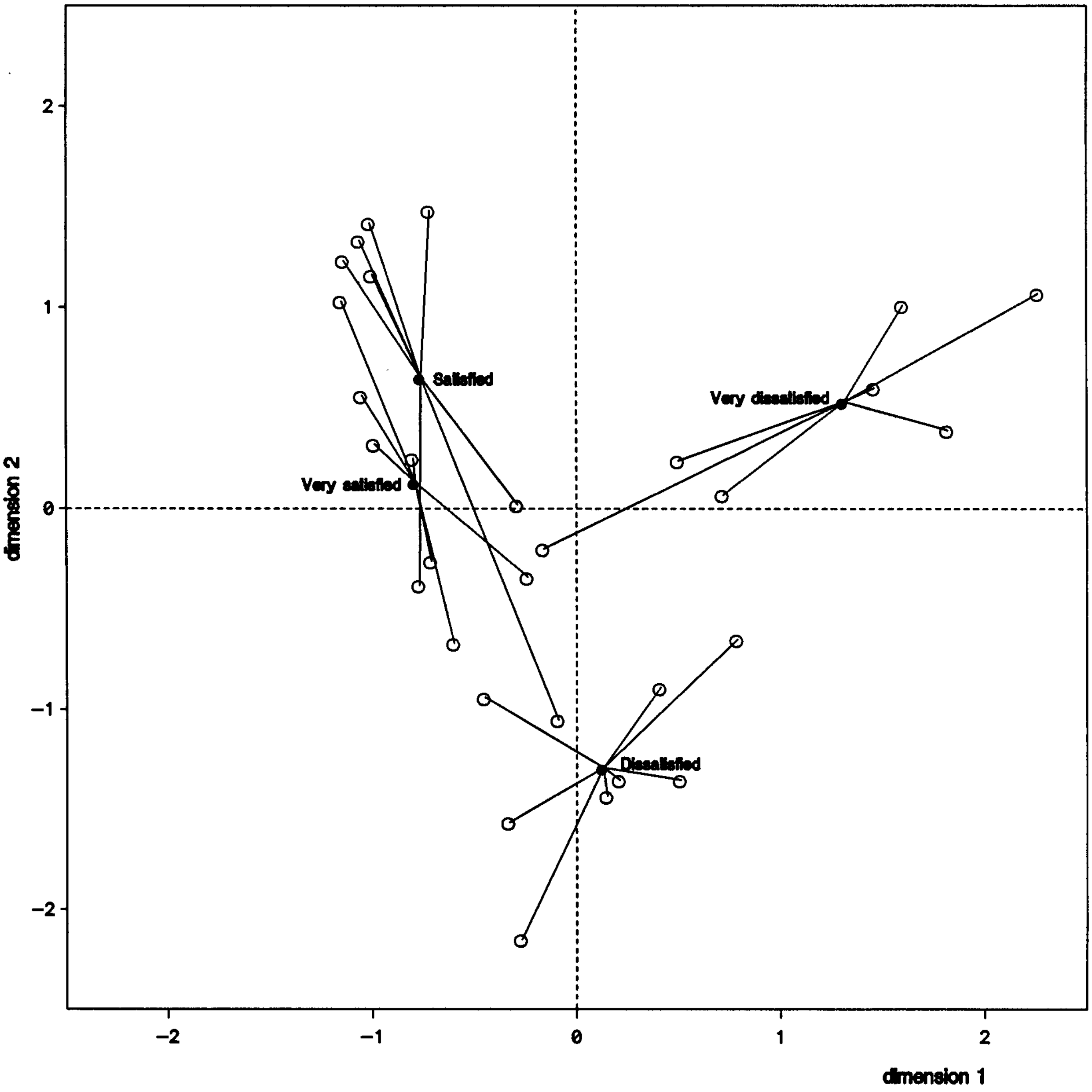

Figure 5

Star Plot for Distance to Ideal

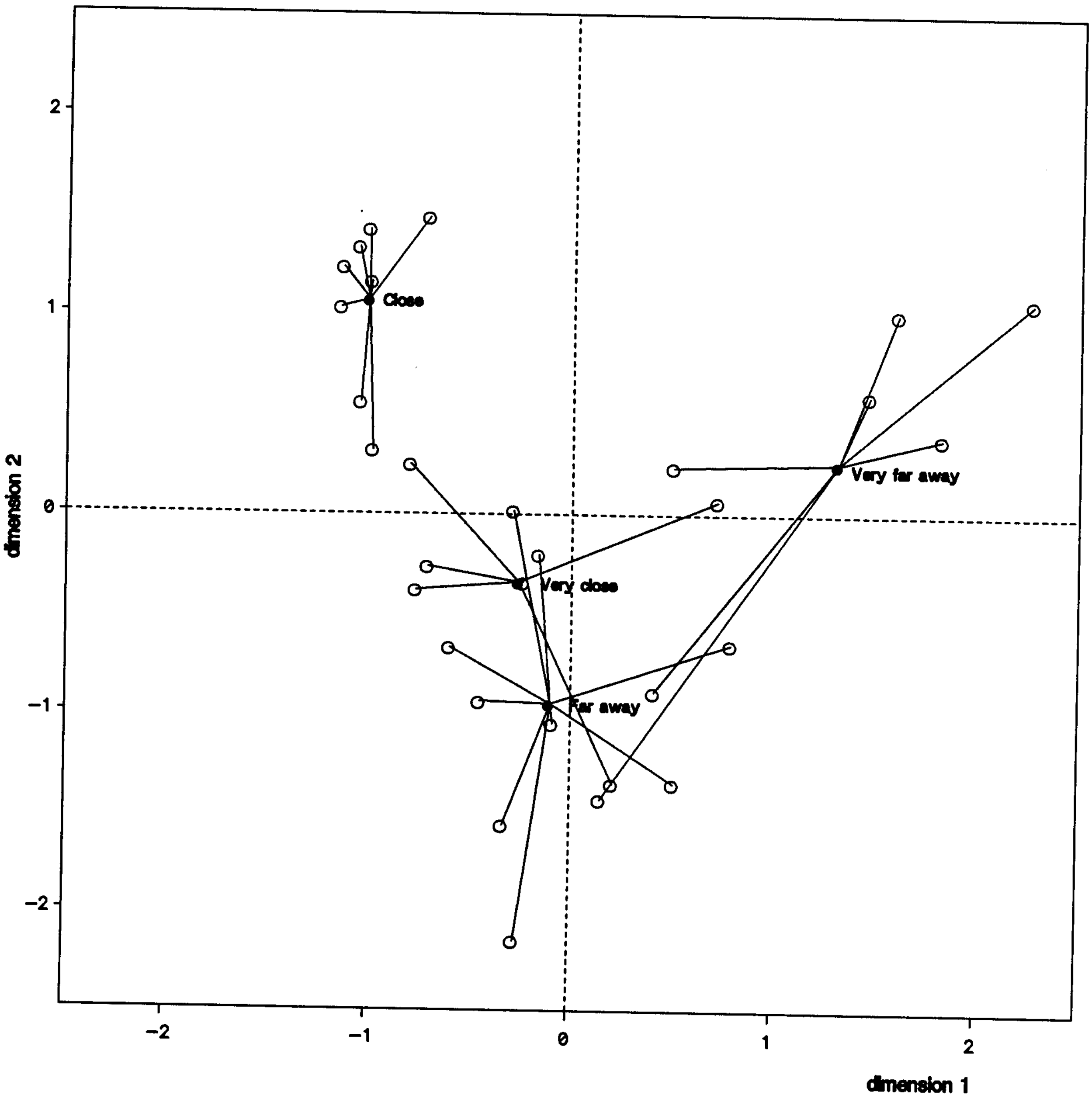

Figure 6

TABLE 1

CUSTOMER JUDGMENTS ON SEVEN SATISFACTION VARIABLES FOR 31 SWEDISH INDUSTRIES

Industry[1]	Price[2]	Quality[3]	Repurchase Intention[4]	Price Increase Tolerance[5]	Satisfaction[6]	Expectation of Quality[7]	Ideal[8]
Airline 1	unreasonable	high quality	very unlikely	very unlikely to switch	satisfied	high	far away
Airline 2	unreasonable	high quality	unlikely	neither likely nor unlikely to switch	satisfied	very high	very close
Bank (B)	unreasonable	very high quality	very likely	likely to switch	very satisfied	high	close
Insurance (B)	reasonable	low quality	unlikely	likely to switch	dissatisfied	high	far away
PC (B)	reasonable	very high quality	unlikely	unlikely to switch	very satisfied	high	far away
Postl (B)	unreasonable	low quality	unlikely	very unlikely to switch	dissatisfied	very high	far away
Postp (B)	very unreasonable	very low quality	very unlikely	unlikely to switch	dissatisfied	very high	far away
Telecommunication(B)	very unreasonable	very low quality	very unlikely	unlikely to switch e	very dissatisfied	high	very far away
Cars	very reasonable	very high quality	likely	neither likely nor unlikely to switch	very satisfied	very high	very close
Charter-travel	very reasonable	very high quality	unlikely	likely to switch	very satisfied	low	close

[1] Some industries are divided into two different customer types: (B)=business clients, (P)=public clients.

[2] Measured on the four-point scale: very unreasonable, unreasonable, reasonable, very reasonable.

[3] Measured on the four-point scale: very low quality, low quality, high quality, very high quality.

[4] Measured on the four-point scale: very unlikely, unlikely, likely, very likely.

[5] Measured on the five-point scale: very likely to switch, likely, neither likely or unlikely, unlikely, very unlikely to switch.

[6] Measured on the four-point scale: very dissatisfied, dissatisfied, satisfied, very satisfied.

[7] Measured on the four-point scale: very low, low, high, very high.

[8] Measured on the four-point scale: very far away, far away, close, very close.

Industry[1]	Price[2]	Quality[3]	Repurchase Intention[4]	Price Increase Tolerance[5]	Satisfaction[6]	Expectation of Quality[7]	Ideal[8]
Cloth	very reasonable	low quality	very likely	neither likely nor unlikely to switch	dissatisfied	very low	very far away
Department Stores	reasonable	low quality	unlikely	neither likely nor unlikely to switch	dissatisfied	low	very far away
Food	reasonable	very high quality	likely	unlikely to switch	very satisfied	----	very close
Food TR	reasonable	high quality	likely	likely to switch	satisfied	high	close
Furniture	very reasonable	low quality	unlikely	very unlikely to switch	satisfied	very low	far away
Mail	very reasonable	high quality	likely	unlikely to switch	satisfied	very low	close
Mainframe Computers	reasonable	very high quality	very likely	unlikely to switch	very dissatisfied	very high	far away
Bank (P)	unreasonable	high quality	very likely	likely to switch	very satisfied	low	close
Insurance (P)	unreasonable	low quality	likely	unlikely to switch	dissatisfied	low	far away
Gasoline	unreasonable	high quality	very likely	likely to switch	satisfied	low	close

[1] Some industries are divided into two different customer types: (B)=business clients, (P)=public clients.

[2] Measured on the four-point scale: very unreasonable, unreasonable, reasonable, very reasonable.

[3] Measured on the four-point scale: very low quality, low, high, very high quality.

[4] Measured on the four-point scale: very unlikely, unlikely, likely, very likely.

[5] Measured on the five-point scale: very likely to switch, likely, neither likely or unlikely, unlikely, very unlikely to switch.

[6] Measured on the four-point scale: very dissatisfied, dissatisfied, satisfied, satisfied.

[7] Measured on the four-point scale: very low, low, high, very high.

[8] Measured on the four-point scale: very far away, far away, close, very close.

Industry[1]	Price[2]	Quality[3]	Repurchase Intention[4]	Price Increase Tolerance[5]	Satisfaction[6]	Expectation of Quality[7]	Ideal[8]
Pharmacy	very reasonable	very high quality	very unlikely	very unlikely to switch	very satisfied	very high	very close
Life Insurance (P)	very reasonable	high quality	likely	unlikely to switch	satisfied	low	close
Police	very reasonable	low quality	very likely	likely to switch	very dissatisfied	very low	very far away
Postl (P)	very unreasonable	low quality	very likely	very unlikely to switch	dissatisfied	high	very close
Postp (P)	very unreasonable	very low quality	likely	neither likely nor unlikely to switch	very dissatisfied	high	very close
Telecommunication(P)	very unreasonable	very low quality	unlikely	neither likely nor unlikely to switch	dissatisfied	low	far away
Rail	very unreasonable	very low quality	very unlikely	very unlikely to switch	very dissatisfied	very low	very far away
Ship	reasonable	high quality	likely	unlikely to switch	satisfied	very high	close
TV Station 1	very unreasonable	very low quality	very unlikely	very likely to switch	very dissatisfied	very low	very far away
TV Station 2	very unreasonable	very low quality	very unlikely	very likely to switch	very dissatisfied	very low	very far away
TV Station 3	reasonable	very low quality	very unlikely	very likely to switch	very dissatisfied	low	very far away

[1] Some industries are divided into two different customer types: (B)=business clients, (P)=public clients.

[2] Measured on the four-point scale: very unreasonable, unreasonable, reasonable, very reasonable.

[3] Measured on the four-point scale: very low quality, low quality, high quality, very high quality.

[4] Measured on the four-point scale: very unlikely, unlikely, likely, very likely.

[5] Measured on the five-point scale: very likely to switch, likely, neither likely or unlikely, unlikely, very unlikely to switch.

[6] Measured on the four-point scale: very dissatisfied, dissatisfied, satisfied, very satisfied.

[7] Measured on the four-point scale: very low, low, high, very high.

[8] Measured on the four-point scale: very far away, far away, close, very close.

TABLE 2

THE SEVEN INDICATOR MATRICES CONSTRUCTED FROM THE DATA IN TABLE 1

	ATTRIBUTES						
Industry	Price	Quality	Repurch. Intention	Price Inc. Tolerance	Satis-faction	Expectation of Quality	Ideal
Airline 1	0 1 0 0	0 0 1 0	1 0 0 0	0 0 0 0 1	0 0 1 0	0 0 1 0	0 1 0 0
Airline 2	0 1 0 0	0 0 1 0	0 1 0 0	0 0 1 0 0	0 0 1 0	0 0 0 1	0 0 0 1
Bank (B)	0 1 0 0	0 0 0 1	0 0 0 1	0 1 0 0 0	0 0 0 1	0 0 1 0	0 0 1 0
Insurance (B)	0 0 1 0	0 1 0 0	0 1 0 0	0 1 0 0 0	0 1 0 0	0 0 1 0	0 1 0 0
PC (B)	0 0 1 0	0 0 0 1	0 1 0 0	0 0 0 1 0	0 0 0 1	0 0 1 0	0 1 0 0
Postl (B)	0 1 0 0	0 1 0 0	0 1 0 0	0 0 0 0 1	0 1 0 0	0 0 0 1	0 1 0 0
Postp (B)	1 0 0 0	1 0 0 0	1 0 0 0	0 0 0 1 0	0 1 0 0	0 0 0 1	0 1 0 0
Telecommunication(B)	1 0 0 0	1 0 0 0	1 0 0 0	0 0 0 1 0	1 0 0 0	0 0 1 0	1 0 0 0
Cars	0 0 0 1	0 0 0 1	0 0 1 0	0 0 1 0 0	0 0 0 1	0 0 0 1	0 0 0 1
Charter-travel	0 0 0 1	0 0 0 1	0 1 0 0	0 1 0 0 0	0 0 0 1	0 1 0 0	0 0 1 0
Cloth	0 0 0 1	0 1 0 0	0 0 0 1	0 0 1 0 0	0 1 0 0	1 0 0 0	1 0 0 0
Department Stores	0 0 1 0	0 1 0 0	0 1 0 0	0 0 1 0 0	0 1 0 0	0 1 0 0	1 0 0 0
Food	0 0 1 0	0 0 0 1	0 0 1 0	0 0 0 1 0	0 0 0 1	0 0 0 0	0 0 0 1
Food TR	0 0 1 0	0 0 1 0	0 0 1 0	0 1 0 0 0	0 0 1 0	0 0 1 0	0 0 1 0
Furniture	0 0 0 1	0 1 0 0	0 1 0 0	0 0 0 0 1	0 0 1 0	1 0 0 0	0 1 0 0
Mail	0 0 0 1	0 0 1 0	0 0 1 0	0 0 0 1 0	0 0 1 0	1 0 0 0	0 0 1 0
Mainframe Computers	0 0 1 0	0 0 0 1	0 0 0 1	0 0 0 1 0	1 0 0 0	0 0 0 1	0 1 0 0
Bank (P)	0 1 0 0	0 0 1 0	0 0 0 1	0 1 0 0 0	0 0 0 1	0 1 0 0	0 0 1 0
Insurance (P)	0 1 0 0	0 1 0 0	0 0 1 0	0 0 0 1 0	0 1 0 0	0 1 0 0	0 1 0 0
Gasoline	0 1 0 0	0 0 1 0	0 0 0 1	0 1 0 0 0	0 0 1 0	0 1 0 0	0 0 1 0
Pharmacy	0 0 0 1	0 0 0 1	1 0 0 0	0 0 0 0 1	0 0 0 1	0 0 0 1	0 0 0 1
Life Insurance (P)	0 0 0 1	0 0 1 0	0 0 1 0	0 0 0 1 0	0 0 1 0	0 1 0 0	0 0 1 0
Police	0 0 0 1	0 1 0 0	0 0 0 1	0 1 0 0 0	1 0 0 0	1 0 0 0	1 0 0 0
Postl (P)	1 0 0 0	0 1 0 0	0 0 0 1	0 0 0 0 1	0 1 0 0	0 0 1 0	0 0 0 1
Postp (P)	1 0 0 0	1 0 0 0	0 0 1 0	0 0 1 0 0	1 0 0 0	0 0 1 0	0 0 0 1
Telecommunication(P)	1 0 0 0	1 0 0 0	0 1 0 0	0 0 1 0 0	0 1 0 0	0 1 0 0	0 1 0 0
Rail	1 0 0 0	1 0 0 0	1 0 0 0	0 0 0 0 1	1 0 0 0	1 0 0 0	1 0 0 0
Ship	0 0 1 0	0 0 1 0	0 0 1 0	0 0 0 1 0	0 0 1 0	0 0 0 1	0 0 1 0
TV Station 1	1 0 0 0	1 0 0 0	1 0 0 0	1 0 0 0 0	1 0 0 0	1 0 0 0	1 0 0 0
TV Station 2	1 0 0 0	1 0 0 0	1 0 0 0	1 0 0 0 0	1 0 0 0	1 0 0 0	1 0 0 0
TV Station 3	0 0 1 0	1 0 0 0	1 0 0 0	1 0 0 0 0	1 0 0 0	0 1 0 0	1 0 0 0
	G_1	G_2	G_3	G_4	G_5	G_6	G_7

Table 3

THE DISCRIMINATION MEASURES PER VARIABLE PER DIMENSION

Variable	Dimension	
	One	Two
Price Given Quality	.58	.01
Quality Given Price	.83	.57
Repurchase Intention	.53	.40
Price Increase Tolerance	.56	.38
Satisfaction	.74	.61
Expectation of Quality	.25	.07
Ideal (Distance From)	.72	.60
λ	.60	.38