An MDS Approach to Multiple Correspondence Analysis

Donna L. Hoffman
School of Management
University of Texas at Dallas
JO 5.1
Richardson, TX 75083-0688
(214) 690-2545
E-mail: dlhoff@vm.utdallas.edu

Jan de Leeuw
Department of Mathematics
7228 MSB
UCLA
Los Angeles, CA 90024-1555
(213) 825-9550
E-mail: deleeuw@laplace.sscnet.ucla.edu

The authors thank J. Douglas Carroll, Don Lehmann, Donald Morrison, and two anonymous reviewers for their helpful comments on a previous version of this manuscript.

November 1991

Abstract

We formulate multiple correspondence analysis (MCA) as a nonlinear multivariate analysis method which integrates ideas from multidimensional scaling. MCA is introduced as a graphical technique that minimizes distances between connecting points in a "graph plot." We use this geometrical approach to show how questions posed of categorical marketing research data may be answered with MCA in terms of the notion of closeness. We introduce two new displays, the "star plot" and "line plot," which help illustrate the primary geometric features of MCA and enhance interpretation. Our approach, which extends Gifi (1981, 1990), emphasizes easy to interpret and managerially relevant MCA maps.

Key words: Graphical representation of Categorical Data, Homogeneity, Nonlinear Multivariate Analysis.

Multiple correspondence analysis (MCA) is well on its way to becoming a popular tool in marketing research (Hoffman & Franke, 1986). For MCA maps to be useful in marketing, rules for representation and interpretation must be explicit and unambiguous. Consider this classic, if somewhat hypothetical, example. Suppose we have a data matrix indicating which categories of various attributes a series of automobile makes falls into. For example, one attribute may be "body style" with categories two-door, four-door, and wagon. Thus, the data matrix has a 1 whenever a car falls into its category of the attribute and a 0 otherwise.

We wish to pose four questions of these data: 1) What are the similarities and differences among the automobiles with respect to the various attributes describing them?; 2) What are the similarities and differences among the attributes with respect to the automobiles?; 3) What is the interrelationship among the automobiles and the attributes?; 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 automobiles, two automobiles are close together if they share similar attributes, and between each attribute category, two categories are close if they occur in the same automobiles to the same degree; it also implies that an automobile is close to an attribute category if the automobile falls into that category.

One of the problems in graphically representing rectangular categorical data matrices is 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 has become increasingly more important, because the three major statistical packages now have MCA modules 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).

In this paper, we focus on a geometrical approach to MCA which provides for improved representation and enhanced interpretation of MCA maps. Our work extends the presentation in Gifi (1981, 1990) by placing increased emphasis on the geometry. Two new interpretive maps are introduced:

"star plots" and the variable "line plot." MCA is developed as an MDS method that minimizes the distances between connecting points a "graph plot." We feel that these additional geometric properties make MCA easier to understand.

1. Correspondence Analysis as a "Model"

What is multiple correspondence analysis? In this paper, we consider MCA as a mapping technique, although other approaches to optimal scaling (Hoffman and de Leeuw 1991) will also lead to the technique. The French literature (see, for example, Benzécri et. al. 1973) discusses it in the context of metric multidimensional scaling 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.

We formulate MCA in terms of connecting objects, automobiles say, with all the variable categories they are in and use a least squares loss function as the rule to do this. Then, interpretation stems not from terms of "chi-square distance" or "profiles" (cf. Hoffman and Franke 1986), but rather, follows from *le principe barycentrique*, the centroid principle, which says that automobiles close to each other are similar to each other. Our approach emphasizes the geometrical aspects of multiple correspondence analysis and, and we will demonstrate shortly, admits simple correspondence analysis as a special case of MCA.

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 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. It is precisely this lack of explicitness that leads to problems with interpretation.

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.

We view MCA as a nonlinear multivariate analysis method which integrates ideas from multidimensional scaling. MCA is the analysis of *interdependence* among a set of variables, as distinct from the analysis of dependence (with pre-defined sets of dependent and independent variables). Our geometric approach 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 that we are trying to produce not only an aesthetically pleasing map, but one that is also easy to interpret, and hence, managerially relevant. Using this framework, we shall have few interpretive difficulties with the joint maps produced from MCA.

2. MCA as an MDS Method

The concept of *homogeneity* serves as the basis for our development of multiple correspondence analysis (MCA). Homogeneity refers to the extent to which different variables measure the same characteristic or characteristics (Gifi 1981, 1990). Homogeniety 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.

Consider the arbitrary map displayed in Figure 1, which we constructed from Table 2 by connecting cars with categories of the variables they are in. Cars, represented by "X" points, are located randomly in the map, while variable categories, the "Y" points, are positioned at the centroid or average of all the cars in that category. Figure 1 contains the same information as the data matrix from Table 2, but is unappealing to the eye and difficult to interpret. There are $n*m=96$ lines and the figure is messy because many of the lines cross. In addition, the map gives the impression that cars are as far from the categories they occur in as they are from the categories they do not occur in. Thus, this arbitrary representation of the data is unsatisfactory.

-----Insert Figure 1 about here-----

Suppose we think of Figure 1 as a multivariable representation, i.e. as a joint map of the cars and the variable categories, in two-dimensional Euclidean space. The figure will be much less disorderly if the lines are as short as possible, that is, if cars 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 cars with similar profiles (i.e. cars that are often in the same categories) will be close, and categories containing roughly the same cars will be close, as well. The resultant map will capture the essence of Figure 1, but in a way that yields easier and better interpretation. We now formalize these ideas by defining a suitable loss function to be minimized.

2.1 Maximizing Variable Homogeneity

Let the data be m categorical variables on n objects, with the j^{th} variable taking on k_j different values, its categories. We code the variables using indicator matrices to allow for easy expression in matrix notation. Note that the indicator matrix for variable j, $\mathbf{G}_j$, is n × k_j and that each row of $\mathbf{G}_j$ sums to one. More specifically, consider the example in Table 1, with m=4, n=24, k_1=3, k_2=5, k_3=3, and k_4=3. Here, the objects are 24 small cars which Consumers Union judged with respect to degree of crash protection (Consumers Union, 1989). These judgments are based on Consumers Union's analysis of National Highway Traffic Safety Administration crash test data. The two "occupant protection" variables indicate how well the car protected a driver dummy and a passenger dummy during crash tests. "Structural integrity" indicates how well the passenger compartment held up to the forces of a crash; better performance is associated with a greater chance of avoiding injuries other than those caused by the immediate forces of a crash. The remaining categorical variable indicates car body style.

-----Insert Table 1 about here-----

The purpose of multiple correspondence analysis is to construct a joint map of the cars and variable categories in such a way that a car is 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 cars mostly occurring in the same categories tend to be close, while categories sharing mostly the same cars tend to be close, as well. The extent to which a particular representation **X** of the cars and particular representations $\mathbf{Y}_j$ of the categories, satisfy this is quantified by the *loss of homogeneity*, a least squares loss function:

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

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 cars and the categories they occur in, measures departure from perfect homogeneity or similarity. In words, Loss = Dist^2(Acura,2-door) + Dist^2(Daihatsu,2-door) + ... + Dist^2(Volkswagon,better) + Dist^2(Yugo,average).. A total of n*m = 24*4 = 96 squared distances are summed, and these squared distances correspond exactly to the 96 lines in Figure 1. Quite simply, *multiple correspondence analysis produces the map with the smallest possible loss*[1].

We link MCA to multidimensional scaling through the notion of distance[2]. 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 a car point and a variable category it occurs in is always smaller than the distance from that car 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 car points for those cars which "choose" that category, with the consequence that, overall, car points will be closer to the chosen variable categories than to the non-chosen variable categories. Interpretation of the cars is guided by the fact that we solve for **X** (with unit total sum of squared distances) such that the within-category squared Euclidean distances, are as small as possible (or, equivalently, that the between-category squared Euclidean distances, are as large as possible). Note that this amounts to the same thing as performing multidimensional scaling on a similarity matrix $S = \{s_{ij}\}$ with $s_{ij}=1$ if there is a "match" between the similarity of car i and variable category j, and 0 otherwise. The primary difference

[1]More precisely, MCA is the minimization of loss function (1) over all $\mathbf{Y}_j$ and over all normalized **X** (or over all normalized $\mathbf{Y}_j$ and all **X**). The reader is referred to Gifi (1981, 1990) for details.

[2]We can also reformulate the MCA problem in discriminant analysis or ANOVA terms. This development connects MCA to classical multivariate analysis. See Hoffman and de Leeuw (1990) for a fuller treatment.

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[3]. However, MDS methods for unfolding make weaker assumptions, but also tend to produce degenerate solutions[4].

The MCA algorithm, implemented in the SPSS-X program CATEGORIES (SPSS Inc. 1989), is exceedingly simple and relies on alternating least squares (or, equivalently, reciprocal averaging (Hirschfeld 1935; Horst 1935)). The *optimal* variable category coordinates are computed as the averages or *centroids* of the (optimal) coordinates of the cars in that category:

$$\mathbf{Y}_j = \mathbf{D}_j^{-1}\mathbf{G}_j'\mathbf{X} \tag{3}$$

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[5]. Similarly, the *optimal* car coordinates are the centroids of the (optimal) coordinates of the categories containing that car:

$$\mathbf{X} = m^{-1}\Sigma_j\mathbf{G}_j\mathbf{Y}_j \tag{4}$$

Equations (3) and (4) make clear the centroid principle.

2.2 Correspondence Analysis as a Special Case of MCA

Our approach allows us to *specialize* multiple correspondence analysis 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). First, consider that, in general, MCA can be formulated as a type of categorical PCA outlined by Guttman (1941) and Burt (1950):

---

[3]MCA approaches perfect fit, i.e. distance $d_{ij}(\mathbf{X},\mathbf{Y})$ between car 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 $d_{ij}(\mathbf{X},\mathbf{Y}) \geq d_{il}(\mathbf{X},\mathbf{Y})$.

[4]But see DeSarbo and Rao (1984, 1986) for a multidimensional unfolding solution, incorporating reparameterization, which avoids the degeneracy problem. DeSarbo and Hoffman (1986, 1987) extend the model to binary data and compare the solutions with correspondence analysis.

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

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

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[6]. $\mathbf{D}$=diag($\mathbf{C}$) is the diagonal super-matrix of univariate marginals. The variable category coordinates are given by the $\mathbf{Y}$ that satisfies (2), that is the p eigenvectors with corresponding eigenvalues of $\mathbf{C}$.

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.

3. Multiple Correspondence Analysis of the Car Data

3.1 Graph Plots

Now let us apply MCA to the car data of Table 2. We first present the set of "pre" and "post-treatment" *graph plots* from the analysis. Each "X" point in the plots represents a car and each "Y" point

---

[6] The Burt matrix has a block structure, with each off-diagonal submatrix $\mathbf{G}_{jl} = \mathbf{G}_j'\mathbf{G}_l$, $j \neq l$, the cross-table of variables j and l containing the bivariate marginals across the n objects. Each diagonal submatrix $\mathbf{D}_j = \mathbf{G}_j'\mathbf{G}_j$ is the kj × kj diagonal matrix with the univariate marginals of variable j.

a variable category. A graph plot connects all cars with the category points they belong to and has a line for every element in the super-indicator matrix **G** equal to one. Our graph plots illustrate the fundamental idea behind MCA.

Remember that the purpose of MCA is to produce a map with loss as small as possible and where the lines connecting cars to the variable categories they occur in are as short as possible. The graph plot in Figure 1, our original "arbitrary" solution, presents the initial (i.e. iteration 0) MCA solution, based on $\mathbf{Y}_j$ drawn in by the centroid principle and **X** arbitrarily normalized. Thus, Figure 1 is already a half-step in the right direction. The loss for this solution is 22.95. The initial solution is highly unsatisfactory, as the graph plot has many crossing lines and 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 line lengths are minimized.

After seventeen iterations, the *post-treatment graph plot* of Figure 2 is much more orderly as the lines connecting cars to their categories are as short as possible, and the fit is improved considerably (loss=4.25). Comparing these two plots shows clearly how much neater the map of the indicator super-matrix **G** is now presented in two-dimensional Euclidean space.

-----Insert Figure 2 about here-----

3.2 Geometrical Features of MCA Maps

In this section, we introduce star plots and line plots. These plots illustrate the primary geometric features of MCA and enhance interpretation. We use the car data 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.

The MCA of the car data produces dominant eigenvalues of .587, .389, and .347. Since the singular values from an MCA are canonical correlations, we interpret the eigenvalues (squared singular values) as squared canonical correlation coefficients. Cars and variable categories are represented as points in a joint low-dimensional map. The two-dimensional *joint map* is simply the optimal graph plot drawn without the lines and appears in Figure 3.

-----Insert Figure 3 about here-----

The horizontal direction in the map separates cars on the basis of occupant protection. Cars on the right are associated with injuries to the driver and passenger, while cars on the left are associated with no injuries to the passenger or driver and moderate injury to the driver. We might label dimension one "severity of injury" (but note that the relationship is distinctly nonlinear). This dimension also discriminates between cars on the basis of body style with two-door cars on the left and four-door and wagon styles on the right. The vertical dimension differentiates cars on the basis of how well the passenger compartment stood up to the forces of a crash. Structural integrity is best for cars at the top of the map and worsens (to "average") as we move to the lower left. The map clearly reveals nonlinearities among the variables (e.g. structural integrity, driver protection) that the MCA has "linearized" (see Hoffman and de Leeuw 1991).

The distance between two car points is related to the homogeneity (i.e. the similarity) of their profiles, or more generally, their response-patterns[7]. Cars with identical patterns are plotted as identical points. This is illustrated in Figure 3 for, among others, Plymouth Colt and Dodge Colt in the lower right, and Eagle Summit and Mitsubishi Mirage in the upper right, which have identical profiles in G. Two very similar cars are the two Hyundai Excels near the center right of Figure 3.

Another feature of the map is indicated by the positions, for example, of "average" structural integrity and "wagon" style. These point locations indicate that a category point with low marginal frequency will be plotted towards the edge of the map, while a category with high marginal frequency ("two door" style, "no injury" to passenger, and "better" structural integrity) will be plotted nearer to the origin of map. As a corollary, cars with response patterns similar to the "average" response pattern will be plotted more towards the origin (the two-door and four-door Hyundai Excels and Volkswagon Golf), while cars with "unique" patterns (for example, Mazda 323 and Yugo GV) appear near the edges[8].

A distinct view of the variables is afforded by the *variable line plot* in Figure 4, which depicts

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

[8]These statements, however, are only precisely true when considering all dimensions, and not necessarily the map for the first two dimensions only.

only the variables and their corresponding category coordinates. The line plot illustrates the spread of the category points for each variable. A variable *discriminates* better to the extent that its category points are further apart. The line plot thus show how well each variable disciminates, as visualized by the sum of the squared distances between the category points for a variable and the origin. Discrimination measures are quantified as the squared correlations between the car coordinates **X** and the optimally transformed variables, $\mathbf{G}_j\mathbf{Y}_j$, and are interpreted as squared factor loadings. We show them in Table 3. The larger the discrimination measure for a variable, the better the categories of that variable discriminate between the cars. Passenger protection (Figure 7) and Body Style (Figure 5) discriminate among cars on the first dimension, while driver protection (Figure 6) discriminates well on both dimensions. Structural Integrity (Figure 8) discriminates mostly along dimension two. The average over variables of the discrimination measures are the diagonals of Λ. For each dimension, MCA thus maximizes the sum total of the discrimination measures. This means that MCA assigns category coordinates for each variable that have the maximum spread.

-----Insert Figure 4 and Table 3 About Here-----

Intepretation of category points is guided by the centroid principle: category coordinates are the weighted average of car coordinates occurring in that category. A variable's *star plot* illustrates this principle. The *star plots* are displayed in Figures 5, 6, 7 and 8 for body style, driver protection, passenger protection, and structural integrity, respectively. Each star plot maps a particular variable's categories with all the car points and shows loss for each variable. Relative loss in the two-dimensional solution is the sum of the squared distances between car points in a cluster and their average, the category point[9]. We have drawn lines in the star plots to illustrate this. Since category points are the average of the car points that share the category, for each variable, categories of that variable divide the car points into clusters, and the category points are the means of the clusters. For example, Figure 5 depicts clearly the three different clusters of car body style, while Figure 8 displays the groups of cars classified according to their degree of structural integrity. Thus, the star plots visualize the homogeneity of the

[9]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.

cars, as the categories divide the cars into homogeneous subgroups.

-----Insert Figures 5, 6, 7 and 8 about here-----

With respect to between-set interpretation, cars are relatively close to categories they are in and relatively far apart from categories they are not in. For example, Subaru Justy, Honda Civic and Yugo GV are associated with average structural integrity, Mazda 323, Eagle Summit and Mitsubishi Mirage are associated with a high likelihood of severe injury to the driver, and the cars to the left of the map are associated with no or moderate injury to the driver. In terms of distance, cars in a particular category will (on average), have a smaller (squared) distance to that category than cars not in that category. Further, if a category applies uniquely to only one car, then the car point and this category point will coincide. The same is true when a category applies uniquely to a group of cars with identical response patterns. It follows that cars mostly occurring in the same category tend to be close to each other and categories sharing mostly the same cars tend to be close, as well[10].

4. Discussion

4.1 Graphical Representation

In our framework, distances corresponding to 1's in **G** must be small (compared to distances corresponding to 0's), 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. Unfortunately, this always leads to a one-dimensional solution. Thus, we require something stricter, so we impose dimension orthogonality and normalization constraints. Which way we choose to normalize (i.e. normalize the objects $\mathbf{X}$ and leave the variables $\mathbf{Y}_j$ free, or the reverse) is immaterial geometrically, since the problem is formulated in a joint space and interpretation is not affected. However, practical considerations require the researcher to make a choice, with substantive considerations the guide.

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

The centroid principle *defines* graphical representation and interpretation of the MCA map. The rationale for choosing it lies in the inherent asymmetry of multivariate data. Almost all the applications of MCA that we have seen, and consequently almost all interpretations, are inherently asymmetric as multivariate data are, by definition, row or column conditional. In other words, we usually treat rows and columns differently since each represents distinct entities we wish to characterize graphically. Put yet another way, in the context of a specific marketing analysis, 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 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 car data example, to a single map for the brands and a set of star plots for each separate variable.

Specifically, we can normalize the set of object coordinates $\mathbf{X}$ and leave the variable category coordinates $\mathbf{Y}_j$ free. We denote this *Case I*. This means that the variable categories $\mathbf{Y}_j$ are found by the centroid principle. In this case, the optimal scaling of a variable category (equation 3) satisfy $\mathbf{Y}_j'\mathbf{D}_j\mathbf{Y}_j = \mathbf{X}'\mathbf{G}_j\mathbf{D}_j^{-1}\mathbf{G}_j'\mathbf{X}$ and thus $\mathbf{Y'DY} = \mathbf{mX'P_*X} = \mathbf{m}\Lambda$, with $\mathbf{P}_*$ the average of the $\mathbf{G}_j\mathbf{D}_j^{-1}\mathbf{G}_j'$. 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.*

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 this normalization of MCA orders variables by these individuals. Then, we obtain a single map for all the variable categories and, if desired, a set of star plots for each individual with categories of all variables in the plots. As applications in this context often involve very large numbers of individuals, we may omit the star plots and focus research attention on the graph plot of variables only. Accordingly, we normalize the set of variables and leave the objects free. In this *Case II*, the **X** are found by the centroid principle and the optimal scaling of the objects (equation 4) satisfies $\mathbf{X'X} = m^{-1}\mathbf{Y'DY\Lambda} = \mathbf{\Lambda}$. In words, the optimal coordinate of an object is the centroid of the coordinates of the variable categories the object occurs in. Gifi (1981, 1990) calls this the *second centroid principle*.

It will almost always be the case that primary focus is on either the brands or the variables, but not both equally. However, should the latter situation arise, symmetric treatment of rows and columns, may be the most appropriate. Thus, we may normalize according to *Case III*, in which *both* the objects and the variable categories are normalized. This is sometimes 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[11], but no between-set interpretation is possible.

Greenacre (1989), for example, 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 (1986) recommend a variant of Case III, the so-called CGS scaling, which they argue provides for interpretation of all distances (but see Greenacre (1989) for a dissenting view). What should be clear from our discussion, however, is that *the aims of the investigation guides the researcher's choice of representation.*

4.2 Concluding Remark

Our geometric approach suggests that MCA is fruitfully thought of as a nonlinear multivariate analysis method which seeks to minimize the distance between lines connecting objects with all the categories they are in, rather than as a method to represent "chi-square" distances. Our bias suggests that simple CA may be more naturally thought of as a special case of MCA with the number of variables equal to two, than as a method to approximate within-set chi-square distance. Finally, we believe that

---

[11]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) $O - E^{l}/E^{l}$, observed minus expected counts under the log-linear model of independence as a proportion of expected counts under independence, and c) "profiles" (conditional probabilities).

if marketing researchers consider the relationship between correspondence analysis and multidimensional scaling through MCA (and the graph plots), rather than through CA (and the chi-square distance metric), they should have little difficulty applying this powerful multivariate methodology.

References

Benzécri, Jean-Paul, et. al. (1973), *L'Analyse des Données*. (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.

Consumers Union (1989), "Which Cars Do Better in a Crash?," *Consumer Reports*, April, 208-211.

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 (1986), "Simple and Weighted Unfolding Threshold Models for the Spatial Representation of Binary Choice Data," *Applied Psychological Measurement*, 10 (September), 247-264.

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.

DeSarbo, Wayne S. and Vithala Rao (1984), "GENFOLD2: A Set of Models and Algorithms for the GENeral unFOLDing Analysis of Preference/Dominance Data," *Journal of Classification*, 1 (Winter), 147-86.

DeSarbo, Wayne S. and Vithala Rao (1986), "A Constrained Unfolding Model for Product Positioning Analysis," *Marketing Science*, 5 (Winter), 1-19.

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

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.

Hayashi, C. (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 Jan de Leeuw (1990), Geometrical Aspects of Multiple Correspondence Analysis: Implications for the Coordinate Scaling Debate," UCLA Statistics Series, No. 49.

Hoffman, Donna L. and Jan de Leeuw (1991), "Linearizing Nonlinear Association with Optimal Scaling: Reducing Bias and Improving Stability in Multivariate Linear Models," paper presented at the ORSA/TIMS Joint National Meeting, Anaheim, CA, November 3-6.

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.

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

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.

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

Table 1

INDICATOR MATRICES CONSTRUCTED FROM CONSUMER UNION'S JUDGMENTS ON CRASH PROTECTION FOR 24 SMALL CARS

Car Make and Model	Body Style[a]	Occupant Protection[b] Driver	Occupant Protection[b] Passenger	Structural Integrity[c]
	1 2 3	1 2 3 4 5	1 2 3	1 2 3
Acura Integra	1 0 0	1 0 0 0 0	1 0 0	0 1 0
Daihatsu Charade*	1 0 0	0 1 0 0 0	1 0 0	0 1 0
Dodge Colt	0 0 1	0 0 0 0 1	0 1 0	0 1 0
Eagle Summit	0 1 0	0 0 0 1 0	0 1 0	1 0 0
Ford Escort*	1 0 0	1 0 0 0 0	1 0 0	1 0 0
Ford Festiva*	1 0 0	0 0 1 0 0	0 1 0	0 1 0
Honda Civic*	1 0 0	0 1 0 0 0	1 0 0	0 0 1
Hyundai Excel*	1 0 0	0 0 1 0 0	0 0 1	0 1 0
Hyundai Excel	0 1 0	0 0 1 0 0	1 0 0	0 1 0
Isuzu I-Mark	0 1 0	0 0 0 0 1	0 0 1	0 1 0
Mazda 323*	1 0 0	0 0 0 1 0	0 0 1	1 0 0
Mazda RX-7*	1 0 0	0 1 0 0 0	1 0 0	1 0 0
Mitsubishi Mirage	0 1 0	0 0 0 1 0	0 1 0	1 0 0
Mitsubishi Starion*	1 0 0	0 0 1 0 0	1 0 0	1 0 0
Nissan Pulsar NX*	1 0 0	0 0 0 1 0	1 0 0	0 1 0
Nissan Sentra	0 1 0	0 0 0 0 1	0 1 0	0 1 0
Nissan Sentra	0 0 1	0 0 0 1 0	1 0 0	1 0 0
Plymouth Colt	0 0 1	0 0 0 0 1	0 1 0	0 1 0
Pontiac LeMans	1 0 0	0 0 1 0 0	0 1 0	0 1 0
Subaru Justy*	1 0 0	1 0 0 0 0	1 0 0	0 0 1
Toyota Celica	1 0 0	1 0 0 0 0	1 0 0	1 0 0
Toyota Tercel*	1 0 0	0 0 1 0 0	1 0 0	1 0 0
Volkswagon Golf*	0 1 0	0 0 1 0 0	1 0 0	0 1 0
Yugo GV*	1 0 0	0 0 0 0 1	1 0 0	0 0 1
	G_1	G_2	G_3	G_4

[a]1=2-door, 2=4-door, 3=wagon; asterisk indicates a hatchback.
[b]1=no injury or minor injury, 2=possibly moderate injury, 3=certain injury-possibly severe, 4=high likelihood of severe or fatal injury, 5=severe or fatal injury virtually certain.
[c]1=much better than average, 2=better than average, 3=average, 4=worse than average, 5=much worse than average.

Table 2

THE DISCRIMINATION MEASURES PER VARIABLE PER DIMENSION

Variable	Dimension	
	One	Two
Body Style	.621	.082
Driver Protection	.714	.750
Passenger Protection	.642	.043
Structural Integrity	.373	.681
λ	.587	.389

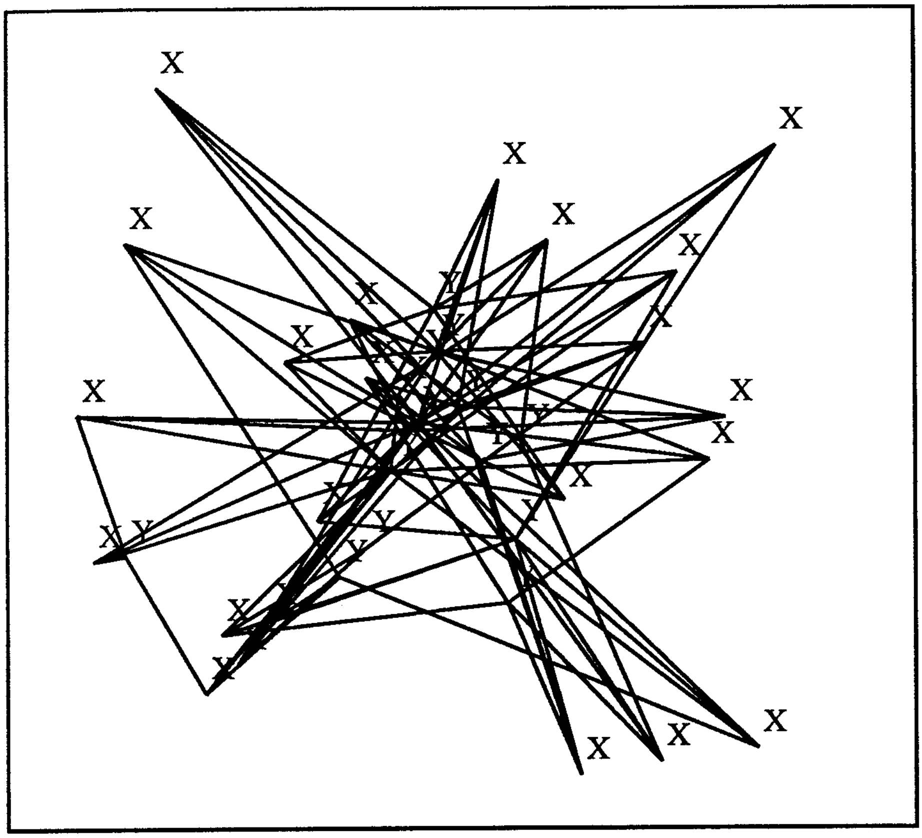

Figure 1 : Graph Plot Before Treatment

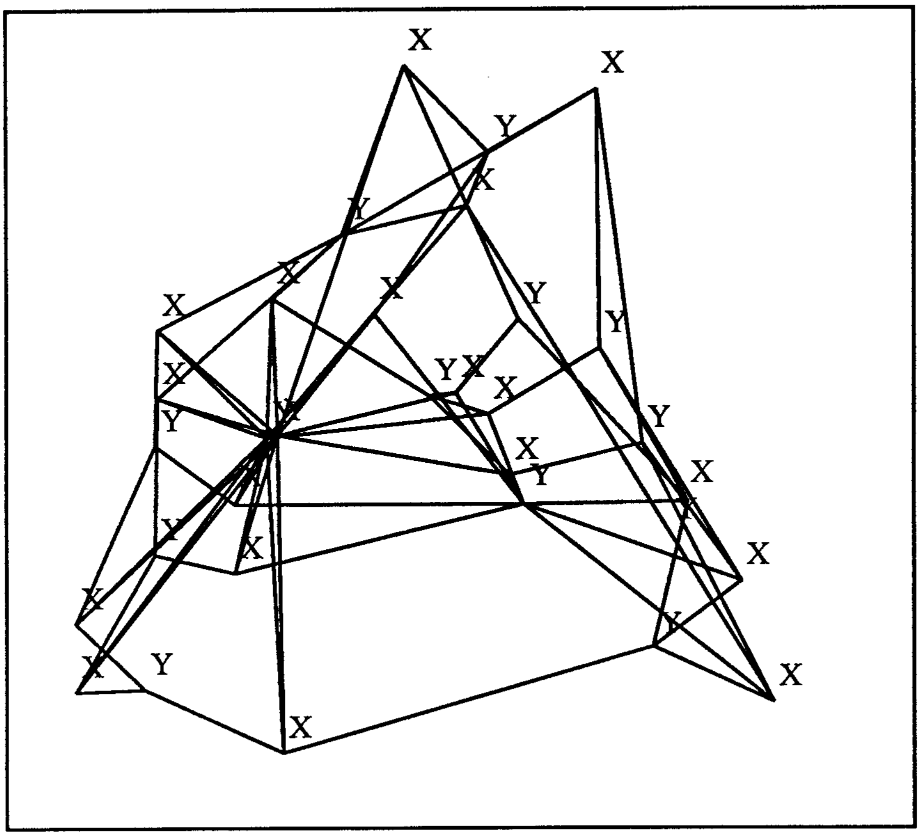

Figure 2 : Graph Plot After Treatment

FIGURE 3

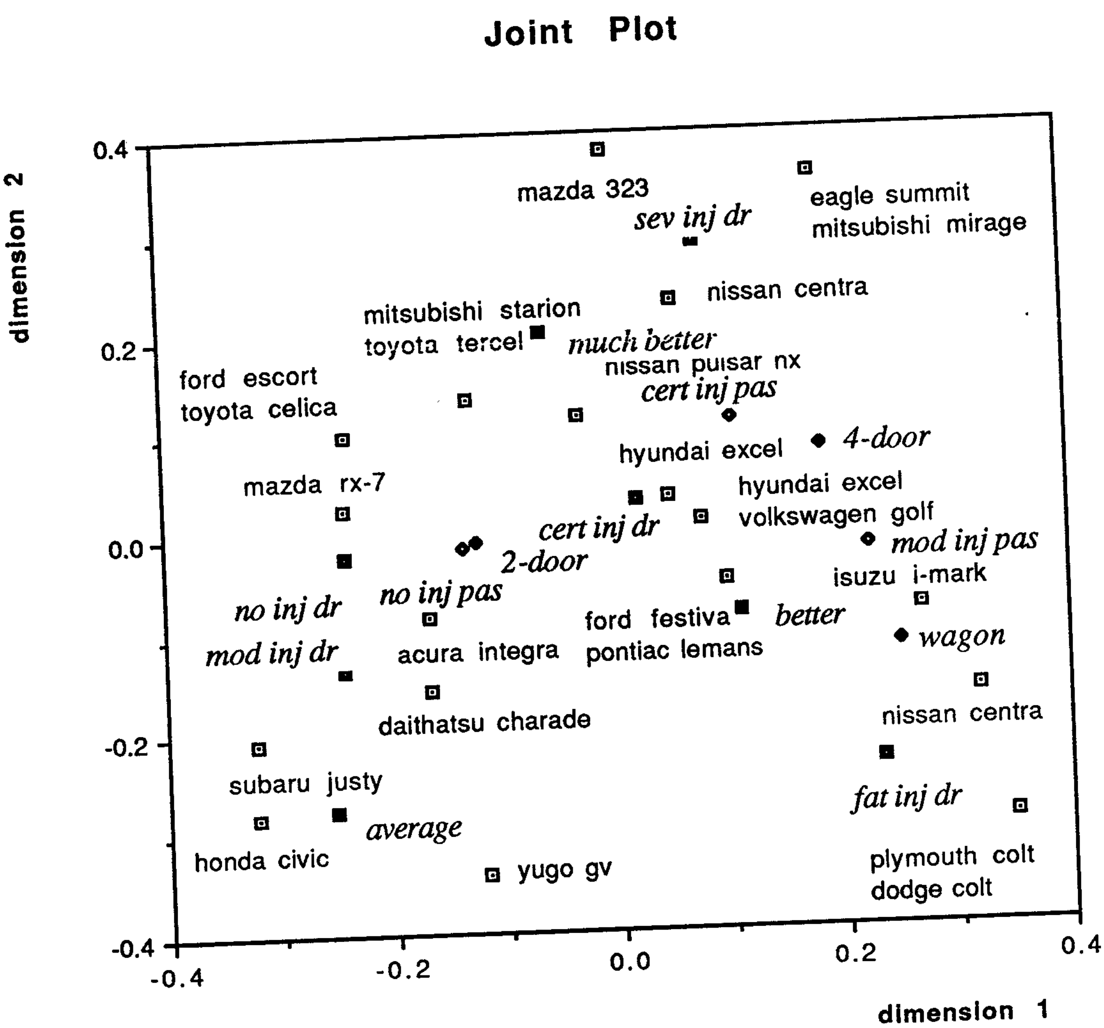
Joint Plot
dimension 2
dimension 1
0.4
0.2
0.0
-0.2
-0.4
-0.4
-0.2
0.0
0.2
0.4
mazda 323
sev inj dr
eagle summit
mitsubishi mirage
nissan centra
mitsubishi starion
toyota tercel
much better
nissan pulsar nx
cert inj pas
ford escort
toyota celica
hyundai excel
4-door
mazda rx-7
hyundai excel
volkswagen golf
cert inj dr
mod inj pas
2-door
isuzu i-mark
no inj dr
no inj pas
ford festiva
better
wagon
mod inj dr
acura integra
pontiac lemans
nissan centra
daithatsu charade
subaru justy
fat inj dr
average
honda civic
yugo gv
plymouth colt
dodge colt

Figure 4

Line plot Category Quantifications

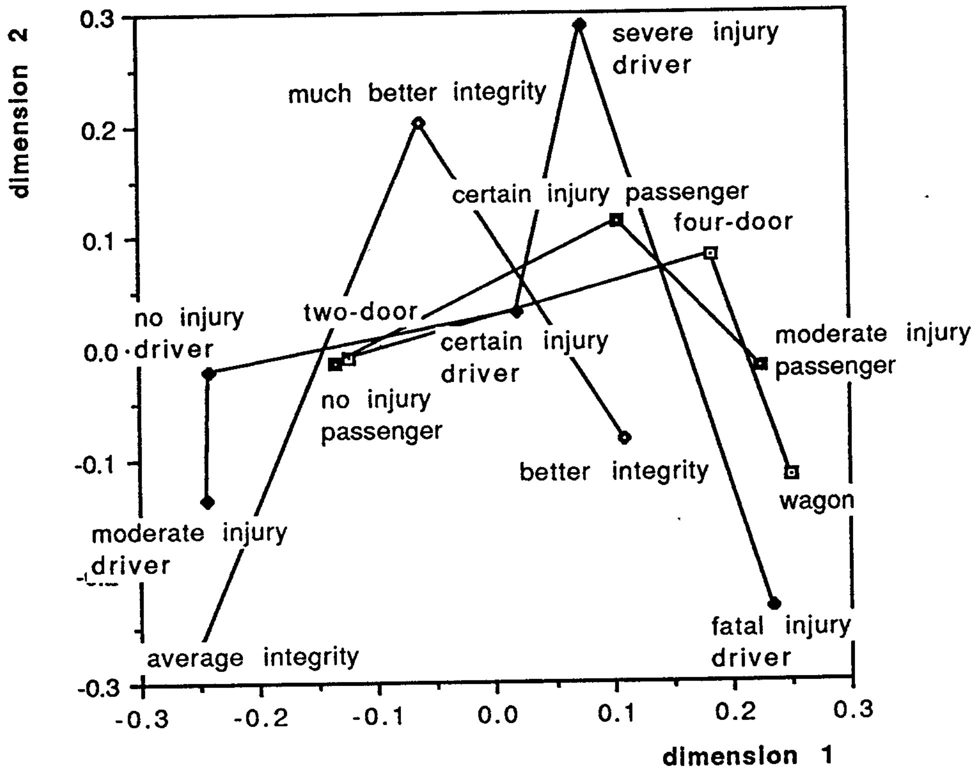

Figure 5

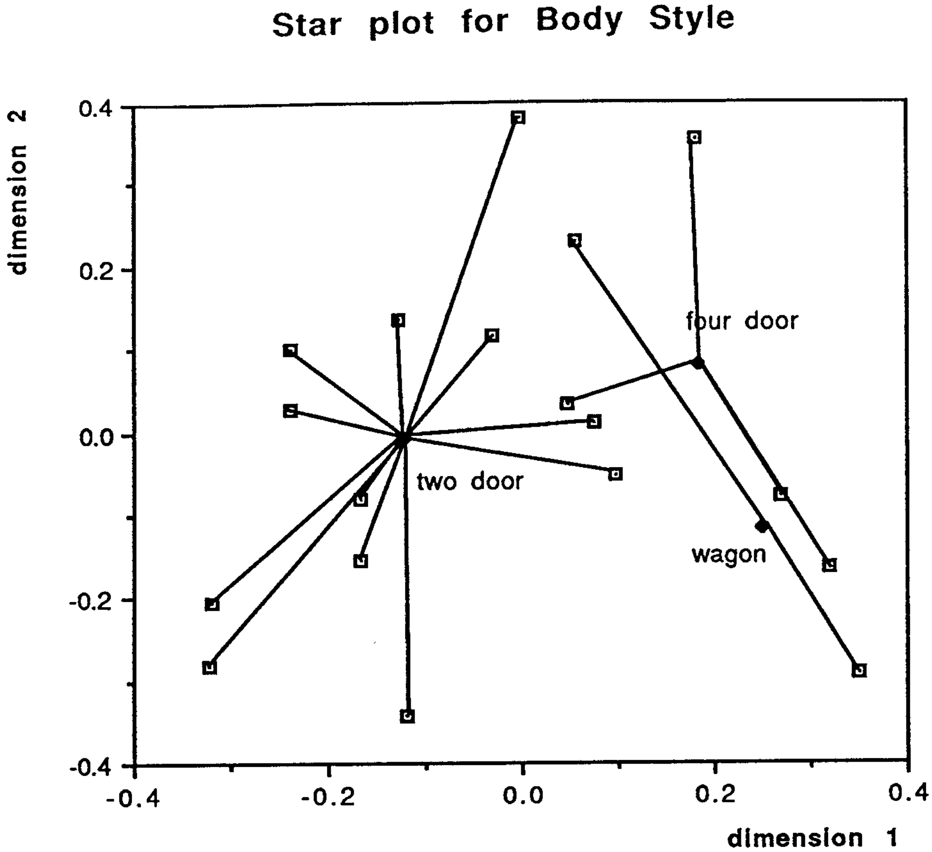

Figure 6

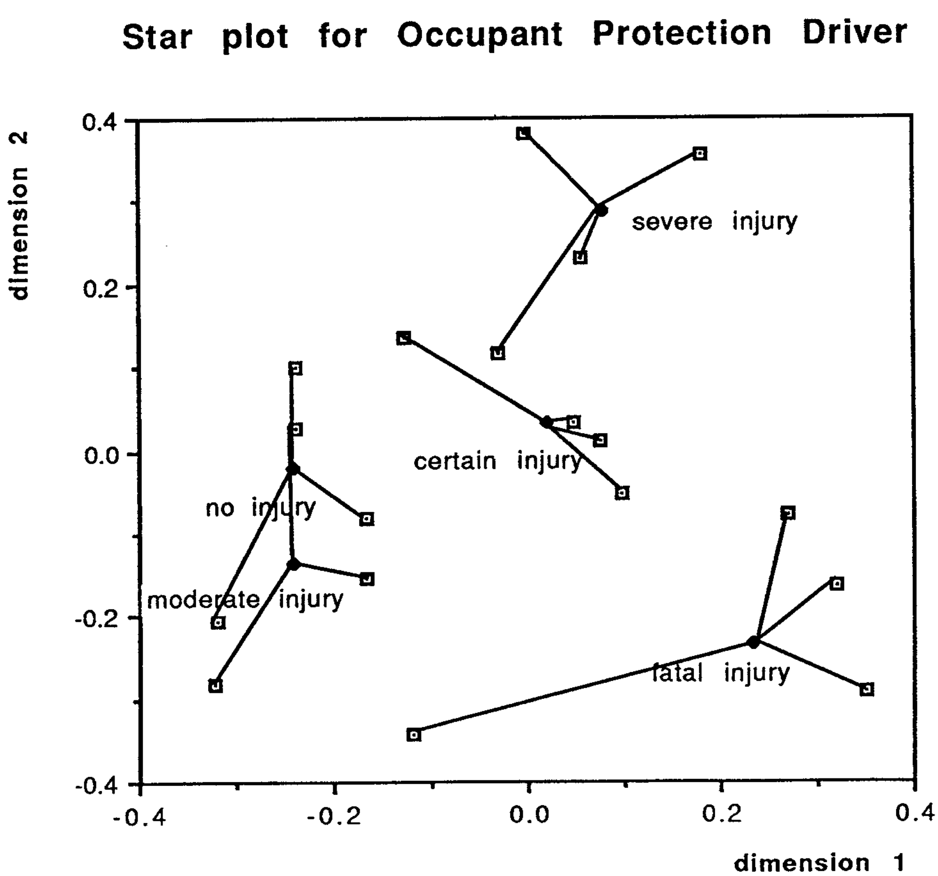

Figure 7

Star plot for Occupant Protection Passenger

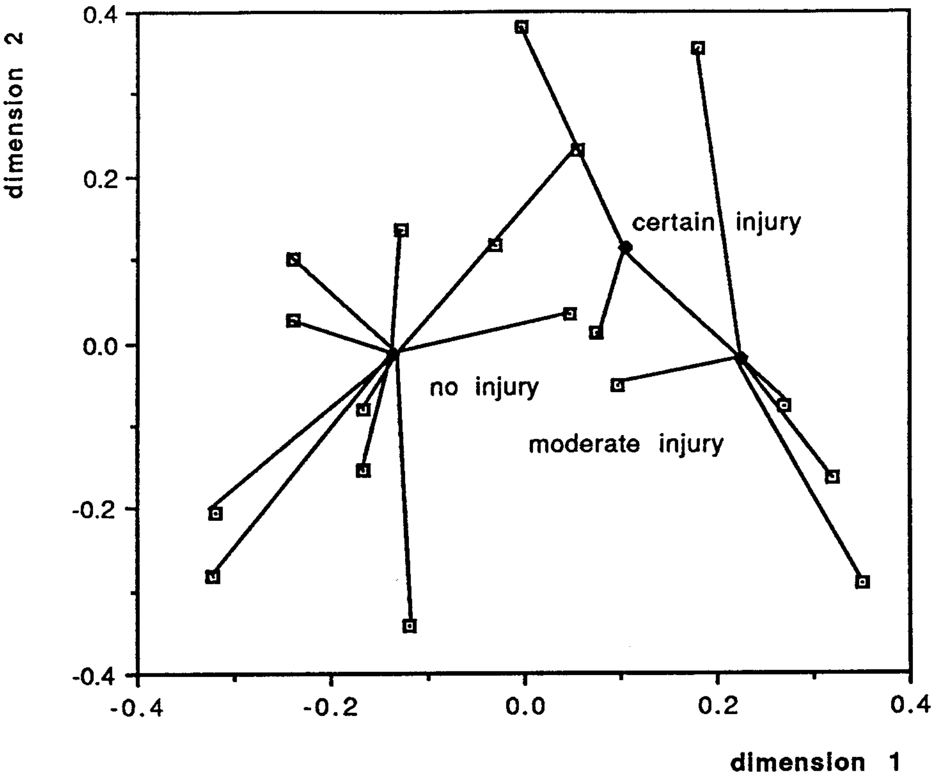

Figure 8

Star plot for Structural Stability