Marketing Bulletin, 1993, 4, 30-42, Article 4

A Multi-dimensional Exploration of the Decision Process Using Correspondence Analysis

Michael F. Fox

A method of analysing rank-order contributions to a decision process using Multiple Correspondence Analysis (MCA) is presented. This procedure is compared to the traditional methods of analysing this type of data, the main results of which are replicated in the MCA study. The proposed MCA procedure however provides a much richer understanding of the relationship between the decision factors and the decision outcome through the higher dimensionality of the solution. Also, an indication of the completeness of the solution is obtained. Some limitations of the data remain, notably the treatment of unranked items.

Keywords: correspondence analysis, dual scaling, optimal

Introduction

Marketing practitioners often express a desire to understand the driving force behind consumers' decisions. There is a belief that an understanding of what motivates consumers in their decisions to buy, will lead to better marketing decisions by marketing managers. This contrasts with the view proposed by many academic researchers, that improved decision-making requires decision oriented research. Nevertheless, the fact remains that a great deal of commercial research is conducted that aims to "understand the consumer".

A common vehicle used for research of this nature is a ranking of "factors which contribute to the decision". Respondents are asked about a particular purchase decision, e.g. did they use a certain type of product, and then are asked to rank a number of factors which contributed to their decision process. There are many variations in the exact data collection procedure. In a telephone or personal interview, the response can be unprompted or prompted. In mail surveys, respondents are shown a list of factors deemed (by the researcher) to be relevant. Respondents may be asked to rank the top 3, the top 5 (out of the list), or as many as were influential in their own individual decisions.

The analysis of this type of data presents a number of difficulties, quite apart from the question whether or not the results could provide managerially relevant information at all. Three methods of analysis are commonly applied, namely a frequency tabulation of the first ranked items, a frequency tabulation of all ranks, and a univariate scaling of some sort, usually by determining a "mean rank position" for each of the factors.

Each of these methods of analysis has several drawbacks. First of all, there are several different treatments possible for all unranked items, none of which is theoretically superior. This is particularly problematic when respondents have not selected a constant number of factors. Secondly, all three methods are essentially one-dimensional, and therefore restricted in the information they can detect. For example, a factor which is infrequently mentioned, but which is the most important factor for a small subsample, will score poorly with all three methods. Finally, there is no indication of how well the list of factors matches with respondents actual reasons for making their decision. A category such as "Other (please specify)" is often used to elicit from respondents important factors that have not been included in the questionnaire design, however there is no guarantee that this category adequately compensates for any missing factors.

The root of the difficulties in the analysis lies in the nature of the data. Because ranks are self-referential data, that is the value of the rank depends of the value of other items in the list, they can not be analyzed with most of the common research tools. For example, it is well known that a principal components analysis of such data will merely produce artefacts of this ipsativity (Nishisato & Gaul 1988). Similar problems are encountered with other techniques.

This paper proposes a new method of analysing this type of data using Multiple Correspondence Analysis (MCA), which extracts a greater amount of information from the data. Because MCA summarises both the rows and columns of a data matrix simultaneously, it avoids the problems associated with self-referential data. The result is a multi-dimensional map, showing the relationships between the various factors and the decision outcome in more detail than is possible with the univariate analysis procedures. In addition, there is an indication of the strength of the relationship between the listed factors and the decision.

Multiple Correspondence Analysis

Multiple Correspondence Analysis (also known as dual scaling, optimal scaling, and method of reciprocal averages) is an exploratory technique for examining patterns in a data matrix. It is a multivariate extension of Correspondence Analysis to the joint analysis of multi-way tables, developed primarily by the French researcher Benzecri (1969). Good theoretical treatments of the procedure can be found in Lebart, Morineau and Warwick (1984), Greenacre (1984) and Nishisato (1980).

Essentially, MCA summarises the response patterns in both the rows and the columns of a data matrix simultaneously. It defines a (typically low dimensional) space in which a graphical representation is possible of both rows and columns together. The space is defined in such a way that the axes represent the greatest amount of variance possible of both columns and rows. Because several criteria are maximised simultaneously, there is no question of a rotation of axes, such as in principal component analysis.

While MCA was originally used to analyze categorical data, it is actually extremely flexible in handling many types of data. It has been applied to contingency tables, sorting and multiple choice data, paired comparisons, ranking tasks, and associative data (see Lebart, Morineau & Warwick 1984, for a listing of applications). Depending on the exact formulation of the problem, it can behave like a principal components analysis of categorical data, while on other occasions it is more like an extended form of Thurstonian scaling.

In recent years there has been an increasing awareness of this technique in the anglophone marketing community, and a corresponding increase in the number of applications published in the literature. For example, Vasserot (1976) used MCA to choose a brand name for a new product. Franke (1983) explored the causes of, and solutions to, the energy crisis in the United States. The measurement of brand image and positioning require frequent monitoring, and can be accomplished quickly and cheaply using MCA (Hoffman & Franke 1986). An overview of possible applications to marketing problems is provided by Nishisato and Gaul (1988).

Some of the more interesting work relates to the combination of MCA with other techniques, as a means of enhancing the interpretation of both analyses. Van der Heijden and de Leeuw (1985) show that MCA is essentially equivalent to a decomposition of the difference between two matrices, each of which follows a log-linear model. As such, MCA allows a geometrical representation of the results of a log-linear analysis. Kaciak and Louviere (1990) use MCA with discrete choice conjoint experiment data to assist in the determination of market segments. They also suggest an exploratory analysis with MCA may detect significant interactions between attribute variables and co-variates, thus potentially enhancing the specification of the choice model.

The purpose of this paper is to demonstrate the application of MCA to the analysis of rank-order data. An example is given where respondents rank order a set of influences on their decision to use, or not to use a certain product. The results from this analysis are compared to the traditional method of analysis, which consists essentially of tabulating the frequency with which each item is ranked in first position and various univariate scaling procedures.

Method

The application of MCA to rank data requires a transformation of the matrix of ranks into a matrix of scores. Following Nishisato and Gaul (1988), the scores can be calculated by equation (1), where N is equal to the number of items in the list.

$$\text{Score} = N - \text{Rank} \quad (1)$$

The treatment of unranked items presents a problem when a varying number of items are ranked. To improve the stability of the solution, we have found that unranked items are best ranked tied at last place. While there is no theoretical justification that this is the best strategy to adopt, it is consistent with the suggestion in Nishisato (1978) for the treatment of missing data. Furthermore, it has the beneficial effect of maintaining constant row totals, and therefore a constant weight of each of the individuals in the analysis. In several data sets where this situation arises, we have found that it tends, on the whole, to produce more interpretable maps.

When the number of ranked items is equal to a constant, R, across respondents, unranked items can arbitrarily be given a score of 1. The value of N can be varied, from R+2 to the number of items on the list. Varying N has the effect of changing the relative weight of the unranked items, relative to the ranked items. The exploratory nature of MCA suggests that an appropriate strategy in this case is to run several analyses, with several different values for N. The most interpretable of the resulting maps can then be chosen.

The scoring procedure produces a matrix of positive values. Each respondent has "scored" each of the factors on the list. The column marginal totals should be checked at this point to eliminate any factors that are not ranked by a sufficiently large number of respondents. Factors which are included as influential in the decision only by 10% of the sample or less have a destabilizing effect on the solution, and are best eliminated from the analysis. This may require a minor readjustment of some of the scores, to maintain constant row totals. The reduced matrix is then submitted to MCA.

Any eliminated variables, and a categorical decision outcome variable (normally coded 0/1) can be included as supplementary variables. Individual difference variables can be categorized, and also included. Supplementary variables are projected into the solution space, but do not contribute to the determination of the space. The projection process is effectively a regression on the axes of the space, and can assist with the interpretation of the space. It also allows the researcher to observe the association between various decision outcomes and the active variables. Since supplementary variables do not affect the space that is obtained from the active variables, there is no limit on the number of such variables that can be included.

The number of axes to retain for the interpretation is determined primarily by the eigenvalues of the square matrix obtained by multiplying the input data matrix with its transpose, similar to principal components analysis. Depending on the number of factors and respondents, this will usually be just two or three. The positions of the decision factors can be plotted in this reduced space, and the resulting map interpreted. The locations of the decision outcomes in this space show which factors are primarily associated with each outcome. In addition, individual difference variables, for example, demographics, can also be plotted in the same space, to assist in the interpretation and to search for differences in the decision process for different segments of the population.

Procedure

The data was collected in a survey of 242 cattle farmers in New Zealand, conducted during the first six months of 1992 for an animal health company. Respondents were asked whether or not they used a certain type of animal health product in the last 12 months, and then to rank in order of importance, from a list of 15, as many factors as they felt contributed to their decision. The 15 decision factors are shown in Table 1, together with the abbreviations used in subsequent maps and tables (some of the factors are disguised to protect the confidentiality of the manufacturer).

Table 1. List of decision factors

NEED	Have always had a need for this product
NEVER	Have never had a need for this product
LOOK	I can tell by the 'look' of my cattle
VET	Veterinarian advice
A_SCI	Scientific test A
B_SCI	Scientific test B
C_SCI	Scientific test C
D_SCI	Scientific test D
TIMES	Cattle need this product at certain times of the year
COSTS	Costs outweigh the benefits
BENEF	Benefits justify the costs
DKNOW	I don't know enough about the product
NEIGH	Neighbourly advice
FARMAD	Advice of a farm advisor
OTHER	Other

Results

MCA Analysis

Since all 15 decision factors were considered influential by a sufficient number of respondents, all 15 were included in the analysis. Ties were assigned the mean rank, and all unranked items were ranked tied at last place. The conversion to scores used equation (2):

$$\text{Score} = 15 - \text{rank} \quad (2)$$

The decision to use the product (coded 1/0) or not to use the product (coded 0/1) were added to the matrix as a 2-category supplementary decision outcome variable. The resulting matrix was then analysed using an MCA program written by the author, which has been tested against the routines in the SPAD analysis package (Lebart & Morineau 1982). On the first three axes, significant deviations from the origin were observed. All three axes were interpretable, and were retained in the discussion. The eigenvalues of the axes were 0.0232, 0.0193 and 0.0121, corresponding respectively to 19.2%, 16.0% and 10.0% of the variation in the data. A plot of the first two axes is shown in Figure 1, including the decision outcome variable.

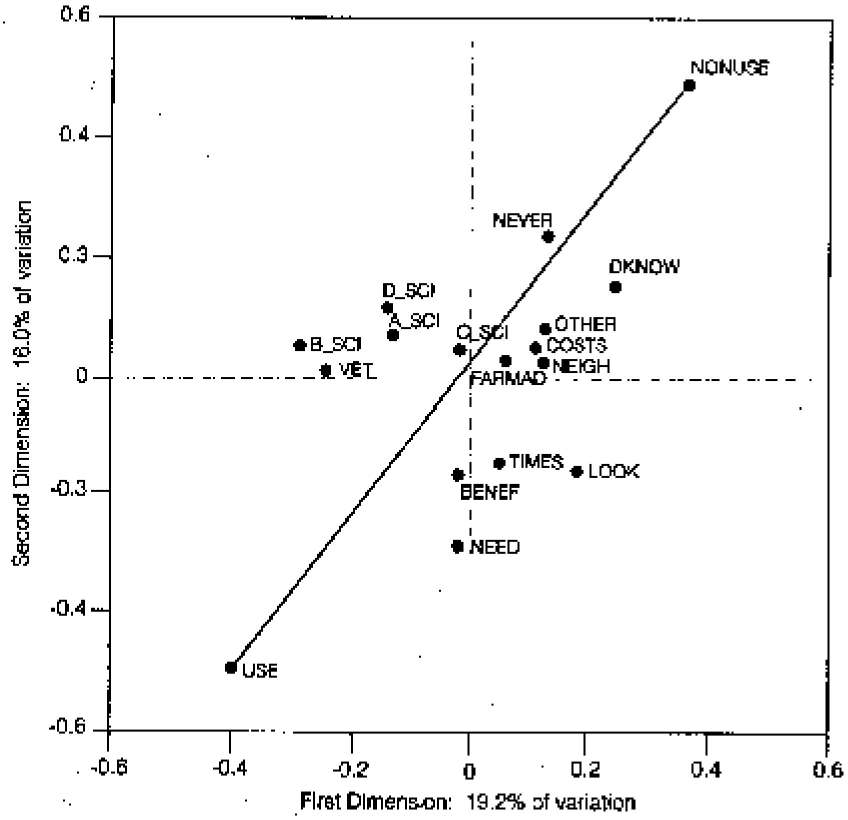

Figure 1. MCA map for first two dimensions

(see Table 1 for an explanation of the symbols)

The points in an MCA map are centred on the origin, so that the origin corresponds to an "*average*" response, or an "*average*" column profile. Elements of the data matrix that differ from this average response in a similar manner lie in the same direction from the origin. The further from the origin, the greater the extent of the deviation from the average response. Thus points that lie close to the origin deviate only slightly from the average response. Two points that lie at some distance from the origin, but on two orthogonal axes, deviate quite strongly from the average, in two very different ways.

Focusing primarily on <u>direction</u> from the origin then, the map in Figure 1 shows that the decision factors split into 3 groups, indicating that there are only 3 distinct influences on the decision. The three influences can be characterized by the terms Scientific (VET, A_SCI, B_SCI, D_SCI and to a lesser extent C_SCI), Habit (NEED, BENEF, TIMES) and Never Use (NEVER, DKNOW, OTHER, COSTS, NEIGH and to a lesser extent FARMAD). The decision factor LOOK is an outlier (see below). The factors that lie furthest from the origin (VET, B_SCI, NEED, NEVER, KDNOW) are those that have the greatest influence, while those at the centre (C_SCI, FARMAD, COSTS, NEIGH) have little weight.

The projection of USE and NONUSE into the space shows that the decision influences represented by the categories Habit and Scientific tend to support the use of this product, whereas the category Never Use is associated with nonuse of the product. Note that this does NOT imply that the decision influences Habit and Scientific are related in a *causal* way to the use of the product. It simply indicates a greater association between the occurrence of these influences, and product usage.

There is one factor in the map, LOOK, which is located at some distance from the other factors (this is especially apparent when higher dimensions are included). This indicates that the factor in question is an outlier. Two possible explanations are that LOOK was not deemed influential by a sufficient number of respondents or that it is simply not related to any of the other factors. An examination of the frequency with which LOOK is cited as an influential factor indicates the latter explanation is likely to be correct in this case. The fact that LOOK is located at quite some distance from the origin, but in a direction perpendicular to the USE/NONUSE axis suggests that this decision factor is associated equally often with USE as with NONUSE.

The Scientific influence lies along the first dimension, indicating that this influence has a strong effect on the decision. That is, that farmers generally abide by the results of scientific testing. USE and NONUSE however, do not lie along the first axis, indicating that at times, the scientific tests support NONUSE. Since not all animals require the product, this is not an unexpected result. Because of the qualitative, exploratory nature of MCA, it is not possible to determine quantitatively the probability that a farmer using a Scientific testing approach, would use the product.

The second axis can be described by the Habit influence. The fact that Habit lies along the second dimension indicates that there is less variation in this influence than in the Scientific influence, but the closer alignment with the USE/NONUSE axis shows that Habit has a stronger effect in explaining this particular behaviour pattern. The opposite of Habit is Never Use, which does not lie exactly opposite in the MCA map. In fact, Never Use is also (partially) associated with the opposite of the Scientific influence, and lies exactly along the NONUSE vector. The location of the three influences in the first two dimension suggests that NONUSE of the product is primarily associated with the belief that the product is not required, whereas USE is associated with a belief in the need for the product OR with the results of scientific tests.

The 3rd dimension, although not shown, is also interpretable, and shows a split in the Scientific influence. Certain tests, usually performed by the veterinarian, are more associated with use, while other tests, usually performed by the farm advisor or by the farmer himself, are more closely associated with non-use. Apart from the outlier LOOK, these are the only significant deviations in the third dimension. The small size of the deviations suggests however, that the relationship is quite weak.

The degree to which these decision factors "explain" the usage decision can be estimated by the correlation between the USE/NONUSE variables, and the axes determined by MCA. This information is routinely printed out by most MCA software. In the example, the correlations with the first and second axes are 0.10 and 0.175 respectively. Correlations with higher axes are 0.01 or less. Clearly, the evidence suggests that these decision factors alone cannot completely explain the behaviour of respondents.

Traditional Analysis

The traditional approach, as mentioned above, is (a) to tabulate the frequency each factor is ranked as the most important factor, (b) to tabulate the frequency each factor is ranked at position M or above (especially when only a partial ranking is available), and (c) to perform a univariate scaling on the ranks. Each analysis is normally broken down by the decision outcome variable, in this case USE and NONUSE. Since in this case we asked each individual to provide a full ranking, we shall only consider options (a) and (c).

Table 2. Mean score of each decision factor, by USE/NONUSE

USERS		NONUSERS	
VET	11.1	A_SCI	9.6
NEED	10.6	VET	9.4
B_SCI	9.9	LOOK	9.3
LOOK	9.1	NEVER	9.1
A_SCI	9.0	DKNOW	8.7
TIMES	8.9	B_SCI	8.5
BENEF	8.8	NEED	8.1
D_SCI	7.9	D_SCI	7.4
C_SCI	6.8	TIMES	7.4
FARMAD	6.8	FARMAD	7.3
NEVER	6.3	BENEF	7.2
DKNOW	6.3	OTHER	7.2
OTHER	6.3	NEIGH	7.1
COSTS	6.1	COSTS	7.0
NEIGH	6.1	C_SCI	6.8

Table 2 shows the result of a univariate scaling procedure. The procedure used here was to define a score according to equation (2), and then to find the mean score over all individuals for each decision factor. Although there are more complicated procedures available, these usually require additional assumptions, such as a normal distribution of the ranks along an interval scale at the individual level, which may not be valid given the nature of the data. The table shows that VET, NEED and B_SCI are the most important items for the users, while A_SCI, VET, LOOK and NEVER score highly for the non-users. In both cases, COSTS, NEIGH, OTHER, and C_SCI seem to be unimportant.

Table 3. Frequency distribution of first-ranked factors, by USE/NONUSE

USERS		NONUSERS	
NEED	33	NEVER	23
B_SCI	30	A_SCI	22
VET	26	LOOK	21
LOOK	17	NEED	13
D_SCI	13	VET	11
A_SCI	12	B_SCI	10
TIMES	12	OTHER	8
BENEF	9	D_SCI	6
C_SCI	7	FARMAD	5
FARMAD	5	DKNOW	4
NEVER	2	TIMES	3
DKNOW	2	BENEF	2
NEIGH	1	COSTS	1
OTHER	1	NEIGH	1
COSTS	0	C_SCI	0

Table 3 shows the frequency with which each factor was ranked in the top position, again broken down by decision outcome. The order of the various decision factors is slightly different from that in Table 2. However, for the users there is generally good agreement with Table 1, with NEED, B_SCI and VET at the top of the list again. The agreement is not so good for the non-users, where NEVER, A_SCI, LOOK and NEED are the top four items. In particular, the appearance of NEED in this list is puzzling, since this item emphatically supports use, rather than non-use. The agreement in both categories is better at the other end of the scale: COSTS and NEIGH are not important at all.

Discussion

The MCA map contains all of the information contained in the traditional analyses. To demonstrate this, Table 4 shows the projection of all the factors in the MCA space onto the axis formed by the USE/NONUSE poles. The projection is performed in the first three dimensions, which were interpreted earlier. The result of the projection is to put the factors onto a single axis, with USE/NONUSE at the coordinates +1 and -1 respectively. The order of the factors in this table replicates well the order observed in Tables 2 and 3 for the users. However, the reverse order does not agree well with the previous lists for non-users. Since the order of factors for non-users also does not agree well between the two univariate (traditional) techniques tested above, it is perhaps significant that the MCA replicates the more robust user results.

Table 4. Projection of decision factors along USE/NONUSE axis in the first 3 dimensions of the MCA space

USE	**1.0000**
NEED	0.2316
BENEF	0.1599
VET	0.1493
B_SCI	0.1445
TIMES	0.0776
LOOK	0.0158
D_SCI	0.0029
C_SCI	-0.0213
A_SCI	-0.0436
NEIGH	-0.0760
FARMAD	-0.0836
COSTS	-0.0977
OTHER	-0.1358
DKNOW	-0.2433
NEVER	-0.2611
NONUSE	**-1.0000**

The question of whether the MCA map is more accurate than the traditional method of analysis remains to be answered, and cannot be determined with this data. The point that can be made, however, is that the MCA map contains significantly more information than the univariate scales can provide. All four scientific tests appear in the MCA map in the same direction, thus indicating a similar influence in the decision. The variability in their position in Tables 2 and 3 is simply a reflection of the frequency with which they occur. In the MCA map, this is reflected in the distance from the origin. For instance, C_SCI is the most extreme case, which appears to be a completely unimportant decision factor in Tables 2 and 3. However, the MCA map shows that, while C_SCI has only a small effect (it is close to the origin), it is in the same direction as the other scientific tests. The reason for this is that C_SCI is an extremely accurate, but expensive test. As a result, it is only performed rarely.

The strong showing of certain decision factors, specifically LOOK, VET, A_SCI and B_SCI, in both the USE and NONUSE groups simultaneously is also consistent with the MCA results. The map in Figure 1 shows that these items are strong on the first axis, i.e. there is a considerable amount of variation in these factors. The USE/NONUSE axis, however, is inclined to the first axis, indicating that the factors on this axis are associated to a degree with both ends of the USE/NONUSE scale. What is not apparent in the traditional analysis, is that the effect of LOOK is almost opposite to VET, A_SCI and B_SCI. Thus, the MCA map shows that while all four factors are associated to a degree with both USE and NONUSE, they do not do so in the same manner.

A third difficulty with the traditional analysis techniques, is the complete absence of an indication of how well the list of factors captures respondents' motivations. As reported earlier, the MCA procedure provides a measurement of how much of the variation in the USE/NONUSE variable is explained by the active variables in the MCA map. The correlations between these outcomes and the axes are 0.10 and 0.175 for the first two axes, and less than 0.01 for higher axes. In addition to the non-causal nature of the analysis in general, the MCA correlation results further warn against too great a reliance on this information to predict behaviour.

Finally, the visual presentation of a map allows a much easier interpretation than the lists of numbers shown in Tables 2 and 3. When several difficulties of the traditional analyses are also avoided, it is clear that MCA provides a much richer interpretation of the data than the univariate analysis techniques currently used.

Conclusions

Subject to the caveat that MCA is a qualitative and exploratory technique, the following four conclusions are possible in this study. These results are consistent with the results of the traditional analysis techniques also. (1) The main factor associated with use of the product is a general belief that cattle need the product. (2) Scientific testing can provide evidence to the farmer to confirm/change his decision, and generally is associated with higher usage. (3) The cost of the product is not seen as an obstacle to use. (4) All factors listed and studied here only account for a small percentage of farmers' motivation. These conclusions suggest a number of hypotheses which may lead to strategies to increase sales. However, these hypotheses should be tested, and alternative strategies evaluated in terms of their quantitative effect on sales before implementation.

The example presented in this paper demonstrates that MCA produces similar results to traditional methods of analysis, but allows the researcher to gain a better understanding of the respondents' decision process than traditional analyses do. Also, some of the problems inherent in the traditional methods are avoided. Therefore, the MCA method, where possible, should be used in preference to the traditional univariate scaling techniques. However, it should only form the basis of further quantitative research, because as an exploratory tool it is not capable of providing definitive answers to questions about decision influences, nor can it predict the (aggregate) decision in any given circumstance.

A further advantage of the MCA method is that it can be used for small groups of people as well as on large samples, and this is where it is probably most useful. Small groups are normally used to form the hypotheses for more detailed quantitative work to be carried out later. The opportunity also exists for the researcher to improve the relevance of the method to the managerial decision making process by including such decision factors as "I liked the advertising", "My local store doesn't carry this brand" etc. which can be directly related to the decision at hand. Again, any implications drawn from this should be tested with appropriate quantitative research. However, the preliminary MCA results could be very useful in defining alternative courses of action.

Some limitations of the data remain, and further research should be carried out to determine the sensitivity of the method to changes in the conversion formula and the effect of different treatments of any unranked items. Also, the omission of important decision factors from the list will undoubtedly affect the pattern of responses to the remaining items. The method has been applied to 5 different data sets, and the highest correlation observed between the decision outcome and an axis was 0.35. It therefore seems likely that missing factors will be a common occurrence, and the effect this may have on the interpretation of the map should therefore be investigated.

In conclusion, this method of analysis is a major improvement on existing methods for this type of data, because (a) it is multidimensional, providing a richer understanding of the relationship between decision factor and decision outcome, and (b) there is some indication of how severe the missing information problem is. It can not overcome the difficulty of how to treat unranked items. On a more philosophical note, it remains an exploratory technique, and can not be used to evaluate alternative options. Therefore it is not a decision making tool. However, an appropriate use for the technique and this type of data in general may be in the strategy formulation phase.

References

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

Franke GR (1983). Dual scaling: A model for interpreting and quantifying categorical data. In W.R. Darden, K.B. Monroe & W.R. Dillon (Eds.). *Research methods and causal modelling in marketing* (pp.111-114). Chicago: American Marketing Association.

Greenacre MJ (1984). *Theory and application of Correspondence Analysis*. London: Academic Press, Inc.

Hoffman DL & Franke GR (1986). Correspondence Analysis: Graphical representation of categorical data in marketing research. *Journal of Marketing Research*, *23*, 213-227.

Kaciak E & Louviere J (1990). Multiple Correspondence Analysis of multiple choice experiment data. *Journal of Marketing Research*, *27*, 455-465.

Lebart L & Morineau A (1982). *SPAD: Systeme Portable pour L'Analyse desDonnees*, CESIA, Paris.

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

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

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

Nishisato S & Gaul W (1988). Marketing data analysis by dual scaling. *International Journal of Research in Marketing*, *5*, 151-170.

van der Heijden PGM & de Leeuw J (1985). Correspondence analysis used complementary to log-linear analysis. *Psychometrika*, *50*, 429-447.

Vasserot G (1976). L'analyse des correspondances appliquee au marketing: choix d'un nom d'engrais. *Les Cahiers des Donnees*, *1*, 319-333.

Michael F Fox is a Lecturer in the Department of Marketing, Massey University.