Division of Research
Graduate School of Business Administration
The University of Michigan

April 1973

MULTIVARIATE METHODS IN MARKETING RESEARCH

Working Paper No. 72

by

Thomas C. Kinnear
The University of Western Ontario
and
James R. Taylor
Associate Professor of Marketing

BACKGROUND

This paper is an update of a paper by the authors titled "Multivariate Methods in Marketing Research: A Further Attempt at Classification," Journal of Marketing Vol. 35, No. 4 (October, 1971). Three new multivariate methods are discussed and classified using a scheme previously developed by the authors.

Because this paper has been submitted for publication in the Journal of Marketing, it follows the style prescribed by that journal rather than that prescribed by the Division of Research.

CONTENTS

TABLES

FIGURES

MULTIVARIATE METHODS IN MARKETING RESEARCH: AN UPDATE

In 1971, Sheth proposed a system for the classification of multivariate methods.[1/] In the same year, Kinnear and Taylor expanded on this classification scheme.[2/] Since that time, several new multivariate methods have been developed which appear to have exciting marketing applications. The purpose of this article is to describe three new multivariate methods and classify them using the Kinnear - Taylor scheme.

The methods to be described are Conjoint Measurement (CM), Multivariate Nominal Scale Analysis (MNA), and Theta Automatic Interaction Detector (THAID).

Conjoint Measurement

Conjoint measurement (CM) can be viewed as an interdependent method of analysis. That is, it does not require the designation of dependent and independent variables. It was first described by Luce and Tukey for use in mathematical psychology.[3/] The basic idea of

---

[1/] Jagdish N. Sheth, "The Multivariate Revolution in Marketing Research," Journal of Marketing, Vol. 34 (January, 1971), pp. 13-19.

[2/] Thomas C. Kinnear and James R. Taylor, "Multivariate Methods in Marketing Research: A Further Attempt at Classification," Journal of Marketing, Vol. 34 (October, 1971), pp. 56-59.

[3/] R. Duncan Luce and John W. Tukey, "Simultaneous Conjoint Measurement," Journal of Mathematical Psychology, 1 (February, 1964), pp. 1-27.

conjoint measurement is that interval-scale utility values can be determined for two or more variables from the ordered joint effects of the variables. The technique is of practical value in situations where the direct assignment of numerical (interval-scaled) values is procedurally difficult and/or where the validity of direct assignment is questioned. An example will illustrate the purpose and procedure more clearly.

An aircraft executive is interested in determining the utility of various price levels and cruising speed options available to the buyers of business aircraft. For the purpose of illustration, assume that in evaluating a business aircraft, only the variables of price and cruising speed are salient to the purchase decision. The executive is interested in studying cruising speeds of 300, 400, and 500 mph and corresponding price levels of $400,000, $600,000, and $800,000. Table 1 presents this situation as a two-variable matrix with each variable having three levels.

The input data are collected by having potential buyers rank order the nine combinations of price and cruising speed in terms of preference. A hypothetical rank order is presented in the nine cells of Table 1. The combination of 500 mph and a $400,000 price level is preferred first, and the combination of 300 mph and an $800,000 price level is preferred last. The conjoint-measurement algorithm searches for utility-scale values for each variable level so that when combined these values will maintain as nearly as possible the original preference rank order.

Table 2 presents utility-scale values for each level of the price and cruising-speed variables. When these utility values are added for each of the nine combinations, the resulting joint scale maintains the ranked (monotonic) relationship observed in the original preference judgments. Although the usual procedure is to add or multiply the utilities, more complex combining methods are possible.[4/] Given the combining method, the CM algorithm searches for variable-level utility values that when combined, maintain as closely as possible the respondents' original rank order judgments.

Figure 1 presents the nature of the response function for the price-level variable. It appears that the disutility of increasing price is most dramatic between the $600,000 and $800,000 levels. Figure 2 presents corresponding data for the cruising-speed variable. The utility increase between 300 to 400 mph is much more pronounced than at the next higher level.

The aircraft executive might conclude from the data that the disutility of higher price levels is less dramatic than the corresponding utility increase associated with higher cruising speeds. Although other marketing considerations would enter into the decision of optimal marketing mix of price and cruising speed, it does appear that substantial utility is added to the aircraft by moving from the 300 to 400 mph cruising speed and that minimal disutility is associated with the increased price level.

4/ Douglas Davidson, "Forecasting Demand for a New Mode of Transportation," in Proceedings 3rd Annual Conference of the Association for Consumer Research, ed. by M. Venkatesan, pp. 294-303, 1972.

The previous example assumed that only two variables were salient to the purchase decision. Of course, more than two variables are likely to be involved in real purchase decisions. If seating capacity and operating cost per mile are also relevant to the decision, the researcher would have to collect preference judgments for a four-variable matrix. The respondents would have to form preference judgments concerning the joint effects of the four variables. In addition, the number of variable combinations to be ranked by the respondent would increase multiplicatively as the number of variables increased. Consequently, the nature of the respondents' judgment task would be very difficult in multivariable problem situations.[5/]

MONANOVA, the most commonly used CM algorithm, requires input data that involve a complete ordering of the joint effects.[6/] Consequently, its application has been restricted to problem situations involving two or three variables because of the data-collection problem just discussed. A new procedure developed by Johnson somewhat overcomes this problem by allowing the respondent to make judgments on two attributes at a time.[7/]

5/ Paul E. Green and Vithala R. Rao, "Conjoint Measurement for Quantifying Judgment Data," Journal of Marketing Research, 8 (August, 1971), pp. 355-63.

6/ Joseph B. Kruskal, "Analysis of Factoral Experiments by Estimating Monotone Transformations of the Data," Journal of the Royal Statistical Society, Series B, 27 (March, 1965), pp. 251-63.

7/ John A. Fiedler, "Condominium Design and Pricing: A Case Study in Consumer Trade-off Analysis," in Proceedings 3rd Annual Conference of the Association for Consumer Research, 1972, pp. 279-303.

This simplification of input format allows the researcher to study multi-variable problem situations without overburdening the respondent's judgement task.

There are many examples of the successful application of conjoint measurement to marketing problems. It has been used to study problems in condominium design and pricing,[8/] air travel,[9/] menu selection,[10/] financial services, and government regulation.[11/] The future of conjoint measurement in marketing appears very encouraging.

MNA and THAID

Both Multivariate Nominal Scale Analysis (MNA)[12/] and Theta Automatic Interaction Detector (THAID)[13/] are dependence methods of analysis. That is, they require the researcher to distinguish between dependent and independent variables prior to analysis. They allow prediction to one

---

8/ Same reference as footnote 7.

9/ Same reference as footnote 4.

10/ Paul E. Green, Yoram Wind, and Arun K. Jain, "Preference Measurement of Item Collections," Journal of Marketing Research, 9 (November, 1972), pp. 371-7.

11/ Same reference as footnote 7.

12/ Frank M. Andrews and Robert C. Messenger, Multivariate Nominal Scale Analysis (unpublished monograph, Institute for Social Research, University of Michigan, Ann Arbor, Michigan, August, 1972).

13/ James Morgan and Robert Messenger, THAID (unpublished monograph, Institute for Social Research, University of Michigan, Ann Arbor, Michigan, November, 1972).

nominally scaled dependent variable from a set of nominally scaled independent variables. The ability to perform this type of analysis represents a major methodological advancement. Prior to the advent of these methods, a nominally defined dependent variable had to be dichotomized (i.e., coded 0 or 1) if the researcher was using nominally defined independent variables because discriminant analysis, which can handle a multicategory dependent variable, requires intervally scaled independent variables. Also, Dummy Variable Regression (DVR), Automatic Interaction Detector (AID), and Multiple Classification Analysis (MCA), which all accept nominal independent variables, require an interval dependent variable. By dichotomizing the dependent variable the researcher was able to create a single interval on the dependent variable and thus to create an interval scale. Consequently, the scale-level assumptions could be met for AID, MCA, and DVR. However, the problem of taking full advantage of a multicategory nominal dependent variable with nominal independent variables still remained. MNA and THAID were developed to meet this need. Although it is possible to use Dummy Variable Discriminant Analysis in this situation, MNA and THAID have significant input and output advantages over this method. They also provide better conceptual information on the effect of predictor categories.

MNA

MNA is a regression-type routine. It involves a series of dummy-variable multiple regressions. MNA proceeds by dummyizing (coding 0 or 1) every category of the dependent variable and every category of each independent variable. A dummy-variable regression run is made on the first category of the dependent variable. A similar run is made on the second category, the third category, and so on, until all categories have been analyzed. Since each dependent-variable category is coded 0 or 1, each of the runs yields a probability value associated with each of the dependent-variable categories for each respondent. For example, if there were five dependent variable codes, each subject would have five probability values. These values would represent the predicted probability that the respondent would fall into each category. MNA then predicts that an individual would appear in the dependent-variable category for which he has the highest predicted probability. It assigns people to these categories and then determines the percentage of subjects that it is able to correctly classify with the independent variables.

MNA yields a coefficient for each category of each predictor. The researcher can examine the pattern of relationship between the dependent variable and categories of the independent variables. This examination allows the researcher to gain conceptual data interpretation beyond that possible with other nominal-level procedures.

MNA appears to have many marketing applications for problems dealing with classification data (brand choice, consumer typologies, and so on). These types of problems can now be examined in a regression type of analysis. To date, only two marketing studies have used MNA.[14 and 15] As the MNA computer program becomes more widely available, its use in marketing studies should substantially increase.

THAID

THAID performs functions similar to those performed by the popular AID routine.[16] That is, it can be used to examine the characteristics of sample subgroups or to search for data interactions. THAID proceeds in the same manner as AID. It first divides the sample into two groups, then divides each of these groups into two new groups, and so on. The result is a "tree diagram."

The difference between AID and THAID relates to the criterion on which the groups are split. AID splits on the interval-level measure of maximizing between groups sums of squares on the dependent variable. Since THAID has a nominal-dependent variable, this splitting criterion is inappropriate. As an alternative, THAID uses the theta statistic as the splitting criterion. Theta is defined as the percentage of respondents who are correctly classified. THAID splits on those categories of an

---

14/ Kenneth L. Bernhardt and Thomas C. Kinnear, "Using Multivariate Nominal Scale Analysis to Identify Demand Segments for Interracial Housing," Working Paper, University of Western Ontario, December, 1972.

15/ Kenneth L. Bernhardt and Thomas C. Kinnear, "Who Wants Vacation Housing," Working Paper, University of Western Ontario, January, 1973.

16/ John A. Sonquist and James N. Morgan, The Detection of Interaction Effects (Ann Arbor, Michigan: Institute for Social Research, The University of Michigan, 1967).

independent variable that give the maximum increase in the percentage of respondents who are correctly classified in their proper dependent variable category.

The researcher may also specify another criterion for THAID to split on. This is the delta statistic. Delta is like the Chi-square statistic in that it measures differences in distributions. The program would split on the categories of the independent variable that maximize the difference in the distribution of subjects across the dependent variable, as measured by delta.

THAID offers the researcher the same potential as MNA. THAID may be used to search for interaction prior to the use of MNA or used by itself to examine the characteristics of nominally defined consumer subgroups. Although THAID has not been used in any published marketing study, its potential is very promising.

Classification Scheme

Figure 3 is an update of the multivariate-classification scheme presented in this Journal by Kinnear and Taylor.[17/] Added to the original scheme are the methods of Conjoint Measurement (CM), Multivariate Nominal Scale Analysis (MNA), and Theta Automatic Interaction Detector (THAID).

The method of CM is an interdependence method for use with nonmetric data. The input data and the numeric procedures they use do not distinguish between independent and dependent variables. However, it should be noted

---

17/ Same reference as footnote 2.

that the results of a CM analysis are often used to predict to a dependent variable, such as respondent purchasing behavior.

The methods of MNA and THAID are dependence methods for use with one nonmetric dependent variable and a set of nonmetric independent variables.

The proper selection and use of multivariate analysis techniques are important for the marketing manager and researcher. The purpose of this article is to acquaint the reader with three new and very promising techniques and to explain the data circumstances appropriate for their use.

TABLE 1

RANK ORDERED JOINT EFFECT INPUT DATA

Cruising Speed	Price Levels $400,000	$600,000	$800,000
300 mph	7	8	9
400 mph	3	4	6
500 mph	1	2	5

TABLE 2

INTERVAL SCALE UTILITY VALUES FOR VARIABLES

Cruising Speed and Utility Values	Price Levels and Utility Values $400,000	$600,000	$800,000
	(.52)	(.45)	(.30)
300 mph (.20)	.72	.65	.50
400 mph (.61)	1.13	1.06	.91
500 mph (.75)	1.27	1.20	1.05

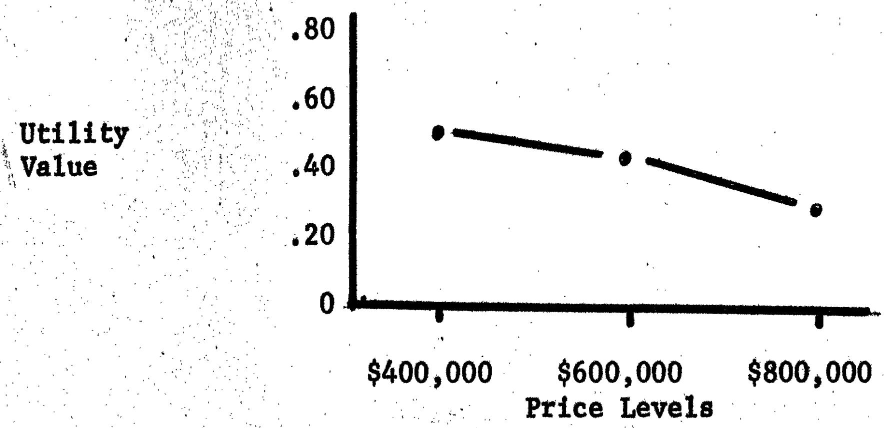

FIGURE 1. Utility Value of Price Levels

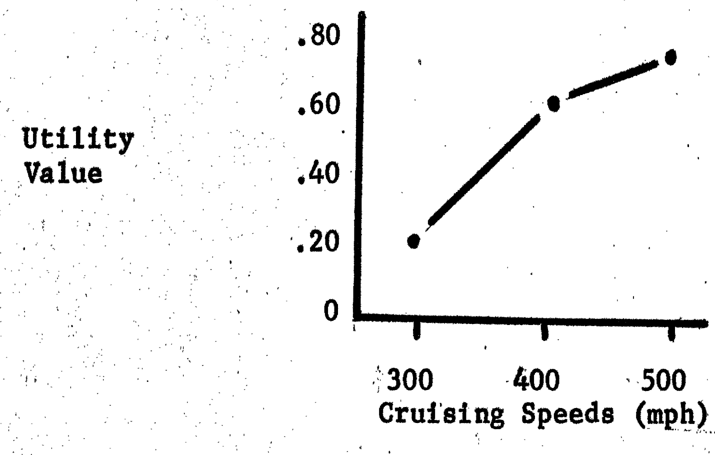

FIGURE 2. Utility Value of Cruising Speed Levels

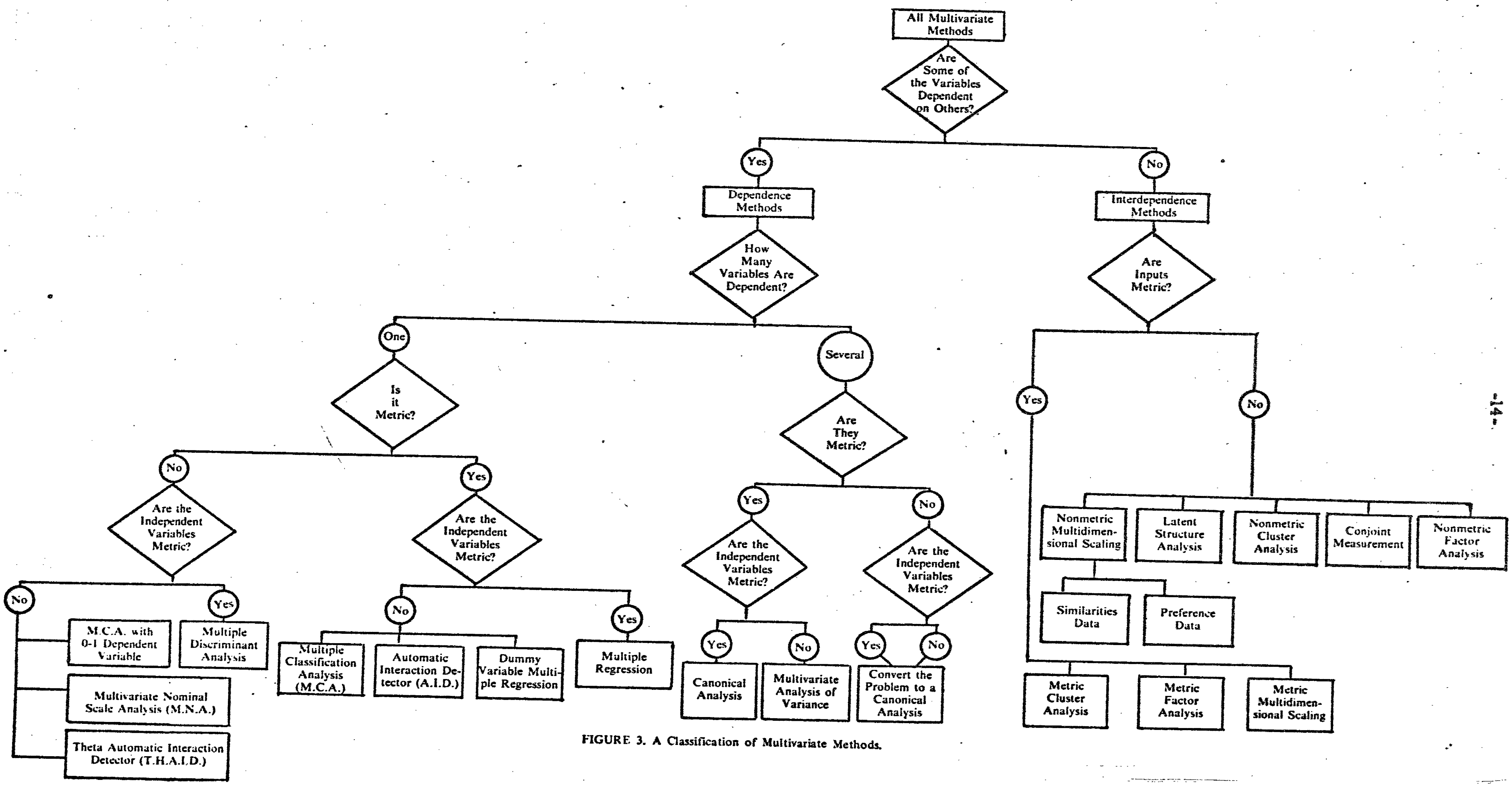

FIGURE 3. A Classification of Multivariate Methods.