1-1-1986

Confirmatory analysis of market segments : an information theoretic approach.

Ajith Kumar
University of Massachusetts Amherst

CONFIRMATORY ANALYSIS OF MARKET SEGMENTS:

AN INFORMATION THEORETIC APPROACH

A Dissertation Presented

By

AJITH KUMAR

Submitted to the Graduate School of the
University of Massachusetts in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

February 1986

School of Management

CONFIRMATORY ANALYSIS OF MARKET SEGMENTS:

AN INFORMATION-THEORETIC APPROACH

A Dissertation Presented

by

Ajith Kumar

Approved as to style and content by:

William R. Dillon, Chairperson

Anna Nagurney, Member

Thomas Madden, Member

Neil Firtle, Outside Member

Thomas Schneeweis
Acting Doctoral Program Director
School of Management

ACKNOWLEDGEMENTS

This dissertation is the natural consequence of an abiding interest in the application of quantitative methods to marketing which was fostered in me by William R. Dillon. My debt to Bill is both intellectual and personal.

Anna Nagurney's timely and expert supervision of the mathematical programming aspects of the dissertation is gratefully acknowledged. Her knowledge of mathematical programming made it possible for me to feel less uneasy about my relative ignorance of the same.

Although Tom Madden's constant words of encouragement and prodding to "get on with the program" were not always received with the proper enthusiasm at the times they were proffered, they are, in retrospect, gratefully acknowledged as being a contributory factor in the timely completion of this dissertation.

A words of thanks to an unofficial committee member--Narendra Mulani, whose programming expertise and knowledge of the CDC CYBER 175 system were always generously made available to the author. Against Naren, it could truly be said, even the CYBER didn't stand a chance.

And finally, an acknowledgement of gratitude to an anonymous benefactor whose generosity reached halfway across the globe and made possible the realization of a long-cherished dream--the ordinary taxpayer of Massachusetts who made it all possible.

ABSTRACT

Confirmatory Analysis of Market Segments:
An Information Theoretic Approach
(February 1986)

Ajith Kumar, B.Sc. (Hons.), Indian Institute of Technology, Kharagpur
P.G.D.M., Indian Institute of Management, Calcutta
Ph.D., University of Massachusetts
Directed by: Dr. William R. Dillon

This dissertation develops a model-based framework for the analysis/identification of market segments. Following earlier work, the segmentation problem is conceptualized as the specification of two sets of variables--a basis set used to form segments and a descriptor set used to discriminate among segments. The problem of using descriptor sets consisting of categorical variables is the focus of research. The evaluation of descriptor sets, in terms of their performance in discriminating among segments, is conceptualized as a sequence of tests of nested models.

The information theoretic approach is shown to be a suitable one for estimating model parameters and simultaneously assessing the goodness-of-fit of the model to the data. The methodology is implemented using the variable metric method of minimization.

A simulation study establishes the satisfactory performance of both the methodology and the algorithm used in implementation. An example illustrating the use of the methodology is also presented.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

This dissertation focuses on the development of a methodology to evaluate sets of variables (descriptors) in terms of their efficacy in discriminating among previously defined market segments. The choice of variables is restricted to those which are categorical in nature.

In Chapter II, a review of the literature on segmentation reveals important gaps in approaches to segmenting markets especially with regard to the specification of descriptor sets. An alternate approach to segmentation is then presented which treats the segmentation problem as a dual problem of clustering and discrimination.

Chapter III introduces and describes the problem of evaluating descriptor sets consisting of categorical variables. The evaluation problem is reformulated as a set of models/hypotheses posited to hold in the population under study, and the estimation of certain unknown population parameters. Under certain specified conditions, the estimation problem is shown to reduce to a nonlinear programming problem. The use of the Minimum Discrimination Information statistic provides for the simultaneous estimation of parameters and the assessment of the goodness-of-fit of the model under which the estimation is carried out. Illustrative examples of possible models/hypotheses which could be tested in the context of the segmentation problem are also provided.

The implementation of the nonlinear programming problem is described in Chapter IV. The method of sequential unconstrained minimization is briefly presented, with details of the exterior penalty function method. The large number of variables involved and the needed computational requirements dictate the use of the Davidon-Fletcher-Powell variable metric method for minimization.

Chapter V describes the results of a simulation study carried out to assess the performance of the methodology and the algorithm used to implement it. In addition an example illustrating the use of the methodology in a segmentation context is presented.

The final chapter contains a summary of the conclusions which emerged from the empirical investigations and provides recommendations for extensions of the methodology, both within and outside the marketing discipline.

CHAPTER II

SEGMENTATION: THEORY AND PRACTICE

Segmentation has long been recognized by academics and practitioners alike as being a dominant concept of marketing. In one of the earliest articles on segmentation, Wendell Smith (1956) sought to draw a distinction between product differentiation and market segmentation as alternative strategies available to the firm. Both strategies implicitly assume the existence of several demand curves for a single product, where each curve graphs the response of a subset of consumers. The strategy of product differentiation describes attempts by the firm to bring about convergence on the demand side to a single product while the strategy of market segmentation requires several product offerings, with each product meeting the requirements of a specific sub-group of consumers rather than the total market.

Changes in the marketplace, both on the supply side and the demand side, have made market segmentation the dominant strategy and in many instances, the strategy of product differentiation is no longer a viable alternative. While Smith (1956) provides an elegant conceptualization of segmentation, no framework or model is offered which could be used to develop a theory or methodology of segmentation.

In examining the literature, one is struck by the gap between academically oriented research and managerial applications. As noted by Wind (1978), theories of segmentation that have been proposed are normative (e.g., Claycamp and Massy, 1968; Mahajan and Jain, 1978; Tollefson and Lessig, 1978) and virtually ignore the practical difficulties involved in implementation.

Before attempting to analyze the theoretical literature on segmentation in greater detail, it is necessary to establish a general framework within which the analysis can be carried out. This framework is provided by the segmentation model, which, following Wind (1978), requires the specification of two sets of variables--one set forming a basis for segmentation and the other set consisting of variables which serve to describe the segments. It should be noted that there exists no consensus in the segmentation literature on the use of terminology. Sometimes the terms dependent variables and independent variables are used to denote the basis set and the descriptor set respectively (e.g., Frank, Massy and Wind, 1972). The terms basis and descriptor variables will be used herein since the labeling of variables as dependent and independent typically tends to imply the existence of structural or causal relationships between the two sets.

A review of the literature shows that most segmentation studies do not maintain the distinction between basis variables and descriptors. Indeed, as noted by Wind, "the variables used as basis for and descriptors of segments have included all variables suggested in the consumer behavior literature" (1978, p. 319). This suggests one of two possibilities--that the distinction is vacuous or that it needs to be explicated in greater detail if it is to prove useful. The position taken herein is that the distinction is useful, and as will be shown later, conceptually and methodologically important.

Studies attempting to provide a perspective on segmentation have, for the most part, adopted what might be termed a taxonomic approach. The objective of such approaches was to classify segmentation studies into two or more groups on the basis of some criterion variable. Thus Wind's (1978) review of segmentation research classifies segmentation studies as a priori or clustering based segmentation designs, which appears to differ very little from Green's (1977) dichotomy of a priori and post hoc segmentation designs.

An alternative classification of approaches to segmentation was provided by Assael and Roscoe (1976), where two dichotomous variables were used simultaneously to cross-classify segmentation studies. One dichotomy was the definition of response behavior as univariate or multivariate. The other dichotomy was the specification of the behavioral criterion as response level at a given point in time versus response elasticity over time.

Another distinction which has been made is between behaviorist and decision oriented schools of market segmentation research (Frank, Massy and Wind, 1972). Behavioral research seeks to identify and document group differences, searches for predictors of such differences and attempts to provide a theoretical explanation for the existence of such group differences. Decision oriented research, on the other hand, presupposes the existence of group differences and focuses on forming meaningful segments. As with behavioral research, predictors of group differences are specified, and in addition, procedures are sought to be developed for the allocation of marketing resources to various segments.

The essential difference between the two schools is the presence or absence of a theoretical framework which postulates the existence of structural/causal relationships (in contradistinction to the theory characterizing normative approaches). However, a review of the segmentation and consumer behavior literature shows the absence of any acceptable, unifying theoretical framework and it may be safely assumed that any theoretical developments in segmentation would probably occur on the normative side, making the above classification unnecessary.

Critique

As pointed out earlier, most of the previous analyses of segmentation studies took the form of classification of the studies on the basis of some criterion variable. Wind's (1978) classification of studies as a priori or clustering based segmentation designs focuses on one half of the segmentation model--the delineation of segments using some set of variables as a basis. The same comment holds for Green's (1977) dichotomy of a priori and post hoc segmentation designs.

While the need to take cognizance of customer characteristics is pointed out, descriptor variables are not explicitly incorporated into the two way classification of segmentation studies by Assael and Roscoe (1976).

A fundamental gap in the above approaches to segmentation research is that they tend to highlight the basis part of the segmentation model and virtually ignore the problems of descriptor set specification. A notable departure from this trend is found in the more recent work of

Green and his colleagues who have introduced flexible and componential segmentation designs which incorporate both product and consumer characteristics (Green and Wind, 1973; Green, Carroll and Carmone, 1977).

From a managerial standpoint effective segmentation requires the specification of both the basis variables and the descriptor set. Situations may arise where the choice of a suitable basis set yields well-defined and meaningful segments. At the same time, the lack of a suitable descriptor set permitting the decision maker to discriminate among those segments may result in the basis set being rejected and alternative bases being examined. In addition to segmenting the market, it is important to evaluate each segment to ascertain the feasibility of marketing to a particular segment. It is in this context that descriptors play an important role in segmentation. These managerial considerations dictate an alternative approach to the segmentation problem.

An Alternative Approach

In simple terms the segmentation problem can be described as follows. It is assumed that the total market for a product is composed of sub-groups where each sub-group is characterized by similar needs and wants. This is a precondition for segmentability of the market. Thus one can specify variables such that for each variable every consumer has a preference for some level of that variable. If the preferred level for every variable is known, then each consumer can be represented as a point in the joint space with the variables as coordinates. Then the problem of specifying a basis set reduces to a problem of selecting a set of variables which yield clusters of consumers in the joint space. Once a suitable basis yielding meaningful clusters has been found, the task becomes one of selecting a set of descriptors which enables the decision maker to discriminate among the clusters, that is, to identify the segments.

From a managerial standpoint, the basis set should be chosen from the set of controllable variables in the marketing mix, that is, those variables whose levels can be varied freely by the decision maker, or variables which are surrogates for the controllable variables.

Similarly, the descriptor set should be chosen from variables which help to identify consumers or from surrogates of these variables. Thus market segmentation is a dual problem of clustering and discrimination. In this framework, the previously proposed classifications of segmentation studies are seen to be classifications of clustering procedures depending on the choice of variables and/or methods. In addition, this framework preserves the conceptual distinction between basis variables and descriptors, and a particular variable can belong to only one of the two sets.

CHAPTER III

THE PROBLEM OF SEGMENT EVALUATION USING CATEGORICAL DATA

Any product offering can be viewed as embodying a bundle of attributes. From a practical standpoint, it makes sense to consider only those attributes which are elements of the marketing mix. Therefore any product can be represented as a vector whose elements are the levels of various attributes.

The set of products in a market can be represented in terms of attribute vectors. However, each attribute vector may not represent a distinct market segment. It is possible that two or more attribute vectors may be similar enough to represent the same market segment.

Here it is assumed that the set of products has been partitioned such that each member of the partition represents a distinct market segment. It is necessary to characterize the consumers in the different segments in much the same way as the segments themselves can be characterized by attribute vectors.

Just as products are characterized as attribute vectors, individual consumers can be represented by measurements on a predetermined set of descriptor variables. Ideally, the set of descriptor variables should be chosen such that each vector of these variables can be uniquely assigned to one (and only one) market segment; that is, all consumers with identical measurements on the descriptor variables set should belong to the same market segment. The set of all descriptor variable vectors assigned to a particular market segment serve to define the consumers belonging to that segment.

In practice, however, it is found that such unique assignments are not possible, that is, consumers with the same measurements on the set of descriptors may not belong to the same segment. In such instances, the assignments of descriptor variable vectors have to be made probabilistically. In other words, given a particular vector of measurements on the descriptor set, there is a probability that a consumer with those measurements on the descriptor variables will belong to any particular market segment. The probability may be zero for some market segments.

The probabilistic assignment reflects the fact that the set of descriptor variables are not perfect indicators of the market segments. In the absence of a theory linking consumer preferences with a set of descriptors, the choice of a descriptor set tends to be somewhat ad hoc and therefore necessarily imperfect.

In the case where the variables constituting the descriptor set are continuous, and the market segments are specified a priori, the technique of discriminant analysis can be used to classify consumers into distinct market segments. Typically a linear discriminant function is employed to effect the classification. When the variables in the descriptor set are categorical, a variety of methods have been adopted to develop classification schemes. These include: (1) treating the categorical variables as if they were continuous and using Fisher's linear discriminant function or some variant thereof; (2) reparameterization of the full multinomial model to achieve a more parsimonious representation of the data (e.g., loglinear and logit models); and (3) the use of procedures based on distributional distances (Matusita, 1954).

The purpose of this study is to develop a method by which sets of descriptor variables can be evaluated in terms of their effectiveness in discriminating among consumers belong to different segments. The study is restricted to variables which are categorical in nature. Prior to the elucidation of the method proposed, it is necessary to redefine the problem for the special case of categorical variables.

Notation

Let X_g denote the categorical variable that indicates an individual's membership of a market segment

X_i, X_j, X_k and X_ℓ denote an individual's "measurements" on the variables in the descriptor set

$g=1,2,\ldots,G$
$i=1,2,\ldots,I$
$j=1,2,\ldots,J$
$k=1,2,\ldots,K$
$\ell=1,2,\ldots,L$

For expository purposes, only four variables are included in the descriptor set. The extension of the method to descriptor sets of larger or smaller sizes is straightforward.

With reference to all the variables mentioned above, it is assumed that an individual is assigned to only one category of each variable, that is, the classification is mutually exclusive and collectively exhaustive. The total number of profiles generated by the descriptor variable set is IxJxKxL.

Let $X_{ijk\ell}$ = (i,j,k,ℓ) denote the typical profile with respect to the descriptor set alone

$(i,j,k,\ell) = (1,1,1,1), (1,1,1,2), \ldots, (I,J,K,L)$

$X_{gijk\ell}$ = (g,i,j,k,ℓ) denote the individual's profile with respect to the descriptor set and the market segment indicator

$(g,i,j,k,\ell) = (1,1,1,1,1), (1,1,1,1,2), \ldots, (G,I,J,K,L)$

Given the a priori specification of market segments, the segmentation problem becomes one of choosing a descriptor set such that each profile in the descriptor set, that is, each (i,j,k,ℓ) can be uniquely assigned to one (and only one) market segment. Mathematically the problem becomes one of choosing variables X_i, X_j, X_k, and X_ℓ such that

$$P(X_g|X_{ijk\ell}) = \begin{cases} 1 & \text{for only one value of } X_g \\ 0 & \text{for other values of } X_g \end{cases}$$

for all IxJxKxL profiles. In the above, $P(X_g|X_{ijk\ell})$ represents the conditional probability of being in the gth category of X_g given that the individual has profile (i,j,k,ℓ).

However, in practice such unique assignments are not possible, and one finds that the conditional probability tends to be non-zero for more than one value of g. The question then arises as to what classification schemes might be optimal in such situations. One approach would be to assign each profile in the descriptor set to that category of X_g for which the conditional probability is the highest among all categories of X_g. Should there be more than one such category, the assignment is to be done randomly to one category from among those for which the tie occurs.

If the descriptor set profile and the market segment membership of every individual in the relevant population is known, then the approach outlined above can be implemented in a fairly straightforward manner and some estimate can be obtained of the errors in classification. In terms of discriminating among segments, alternative descriptor sets can be compared on the basis of their relative error rates.

In most practical situations, however, the population probabilities have to be estimated from observations made on a random sample. In the case where no assumptions are made about the descriptor set profiles and their relationships to the market segments, the sample-based probabilities are taken to be the estimates of the corresponding population probabilities, and the assignments of descriptor set profiles to market segments are made accordingly, and estimates of classification errors are obtained.

However, situations could arise where additional information is available to the decision maker which could be utilized in conjunction with the sample observations for estimating the population parameters. The additional information takes the form of relationships hypothesized to hold among the descriptor set profiles and the market segments.

The method to be proposed can utilize any information which can be expressed as a linear combination of the population probabilities. This approach differs from the approach of estimating the population probabilities from the corresponding sample probabilities in that the population values are estimated subject to one or more linear constraints.

In a particular segmentation problem, the set of linear constraints can be viewed as a model or underlying mechanism generating the data.

The constraints used in the estimation process fall into two categories--those which are known to hold in the population and those representing hypotheses postulated by the decision maker. An example of the former is a situation where the market shares of all the brands concerned are known. This information can usefully be incorporated into the estimation process in the form of certain equality constraints on some marginal probabilities, as will be shown later.

As an example of the second category, the decision maker might hypothesize that a particular variable in the descriptor set does not discriminate among the market segments given the other variables in the descriptor set. As with the previous example, this hypothesis can be translated into a set of constraints on certain population probabilities.

Although a model in a typical problem would consist of constraints belonging to both categories, the distinction is important when evaluating the adequacy of the models in terms of how well they fit the data. The estimation of population probabilities involves the minimization of a certain function subject to the constraints implied by the model. The estimation process, in addition to providing estimates of population probabilities, also provides a test statistic which can be used to assess the goodness-of-fit between the model and the data. While acceptance of the null hypothesis (the model) would imply empirical support for the entire model, rejection of the null hypothesis would only imply rejection of the constraints representing the untested hypotheses.

Estimation

The problem as described previously requires the selection of a set of probabilities, satisfying the constraints imposed by the model, as estimates of the population values. In general, many different sets of probabilities are feasible solutions, that is, more than one set of probabilities will satisfy the set of constraints implied by the model. Therefore, a criterion is required by which a solution can be chosen from the feasible set.

The criterion proposed to be used is the discrimination information function (Gokhale and Kullback, 1978) defined by

$$I(\Pi:p) = \sum_{\Omega} \Pi(\omega)\ln(\Pi(\omega)/p(\omega))$$

where $p(\omega)$ are the observed sample cell probabilities, $\Pi(\omega)$ is any set of probabilities satisfying the model constraints, and the summation is carried out over all cells in the multiway contingency table. The set of probabilities chosen as the estimate of the population values is that which minimizes the function described above. In other words, the solution chosen from the feasible set is that which is "nearest" to the observed sample probabilities.

In the special case where no constraints are placed on the population probabilities, the estimates minimizing the function will be equal to the corresponding sample probabilities, and the value of the function will be zero since $\ln(\Pi(\omega)/p(\omega)) = \ln\ 1 = 0$, for all terms in the summation.

Another special case is the situation where all the constraints in the model equate some of the population probabilities to the corresponding sample marginals. The estimates obtained in such cases would be identical to those obtained using the equivalent loglinear models.

Hypothesis Testing

If in the expression

$$I(\Pi:p) = \sum_{\Omega} \Pi(\omega) \ln(\Pi(\omega)/p(\omega))$$

cell frequencies/counts are substituted for the corresponding probabilities, the function can be alternatively expressed as

$$I(X^*:X) = \sum_{\Omega} X^*(\omega) \ln(X^*(\omega)/X(\omega))$$

where $X^*(\omega) = N\Pi(\omega)$

$X(\omega) = Np(\omega)$

N = sample size

The function $2I(X^*:X)$ is distributed asymptotically as a central chi-square random variable with degrees of freedom equal to the number of linearly independent equality constraints in the model. This does not include the equality constraint which specifies the probabilities to sum to one. Large values of the test statistic would lead to the rejection of the model. In this approach, parameter estimation and hypothesis testing are carried out simultaneously.

Some Illustrative Examples

Notation

Let $\Pi(gijk\ell)$ be the probability (in the population) that an individual belongs to the gth category of X_g, the ith category of X_i, the jth category of X_j, the kth category of X_k, and the ℓth category of X_ℓ.

$\Pi(gi/jk\ell)$ be the probability that the individual belongs to the gth category of X_g and the ith category of X_i given that he/she belongs to the jth, kth, and ℓth categories of X_j, X_k, and X_ℓ respectively.

$P(gijk\ell)$ and $P(gi/jk\ell)$ be the sample based probabilities corresponding to the population probabilities described above.

Example 1: The Test that a Specified Descriptor Set Does Not Discriminate Among Market Segments

Consider a typical descriptor set profile $(ijk\ell)$. The profile cannot be assigned, except randomly, to any group g if

$$\Pi(g/ijk\ell) = 1/G \text{ for all categories of } X_g$$

This implies that

$$\Pi(1/ijk\ell) = \Pi(2/ijk\ell) = \ldots = \Pi(G/ijk\ell)$$

However

$$\Pi(g/ijk\ell) = \Pi(gijk\ell)/\Pi(ijk\ell)$$

Therefore

$$\Pi(1ijk\ell) = \Pi(2ijk\ell) = \ldots = \Pi(Gijk\ell)$$

is an equivalent hypothesis.

The above hypothesis is reformulated in terms of linear constraints as follows.

$$\Pi(1ijk\ell) - \Pi(2ijk\ell) = 0$$

$$\Pi(2ijk\ell) - \Pi(3ijk\ell) = 0$$

$$\cdots$$

$$\Pi[(G-1)ijk\ell] - \Pi(Gijk\ell) = 0$$

For each $(ijk\ell)$ we have G-1 linearly independent restrictions, yielding a total of IxJxKxLx(G-1) restrictions in all.

Example 2: Improving Estimates of Population Probabilities when Additional Information is Available

Let the market shares of various brands be known and for illustrative purposes let each brand represent a distinct segment, that is, a distinct category of X_g. Letting c_g denote the market share of the gth brand, the following restrictions can be imposed.

$$\Pi_g = \sum_i\sum_j\sum_k\sum_\ell \Pi_{gijk\ell} = c_g \quad g=1,2,\ldots,G$$

In the above case, if the numerical value of the test statistic turns out to be significantly large, then the sample has to be rejected as being unrepresentative of the population since the model only contains constraints known to hold in the population.

Example 3: Detection of "Significant" Profiles

In certain situations, when a model is rejected, the decision maker might wish to ascertain the extent to which a subset of the profiles contributes to the rejection of the model. Here a model is fitted (M1) where the restrictions are applied to all profiles. Then another model (M2) is fitted where the restrictions are applied to all profiles other than those belonging to the specified subset. Each model yields a chi-square test statistic. Since the models are nested, the difference in chi-square values is itself distributed as a chi-square random variate with degrees of freedom equal to the difference in the degrees of freedom for the two models.

CHAPTER IV

ESTIMATION BY SEQUENTIAL UNCONSTRAINED MINIMIZATION

For the sake of clarity in exposition, a different notation will be used in this chapter.

Let s be the total number of cells in the complete multiway table

p_i be the observed probability (in the sample) for the ith cell

Π_i be the population probability for the ith cell

$i = 1,2,\ldots,s$

First the estimation process is described for the single sample case. The extension to the multi-sample case, which is fairly straightforward, is then briefly presented.

The estimation of parameters under the hypotheses given in Chapter III can be subsumed under the mathematical programming problem given below.

Minimize

$$I(\Pi:p) = \sum_{i=1}^{s} \Pi_i \ln(\Pi_i/p_i)$$

subject to

$$\sum_{i=1}^{s} \Pi_i = 1$$

$$\sum_{i=1}^{s} c_{ji}\Pi_i = \theta_j \quad j=1,2,\ldots,m \quad m<s$$

$$\sum_{i=1}^{s} d_{\ell i}\Pi_i \geq \eta_\ell \qquad \ell=1,2,\ldots,q$$

$$\Pi_i \geq 0 \qquad i=1,2,\ldots,s$$

where c_{ji}, θ_j, $d_{\ell i}$, and η_ℓ are known constants.

The objective function is strictly convex for $\Pi_i>0$, while the constraints are linear. Hence the estimation process simplifies to a convex programming problem, guaranteeing the existence of a unique minimum.

Rewriting the constraints as follows

$$\sum_{i=1}^{s} \Pi_i - 1 = 0$$

$$\sum_{i=1}^{s} c_{ji}\Pi_i - \theta_j = 0 \qquad j=1,2,\ldots,m \qquad m<s$$

$$\eta_\ell - \sum_{i=1}^{s} d_{\ell i}\Pi_i \leq 0 \qquad \ell=1,2,\ldots,q$$

$$\Pi_i \leq 0 \qquad i=1,2,\ldots,S$$

and let

$$<(-\Pi_i)> = \max\{(-\Pi_i),0\} \qquad i=1,2,\ldots,s$$

$$<(\eta_\ell-\sum_{i=1}^{s} d_{\ell i}\Pi_i)> = \max\{(\eta_\ell-\sum_{i=1}^{s} d_{\ell i}\Pi_i),0\} \qquad \ell=1,2,\ldots,q$$

The objective function and the constraints are utilized to form the following auxiliary function

$$Q(\underset{\sim}{\Pi}_i,\gamma_k) = \sum_{i=1}^{s} \Pi_i \ell n(\Pi_i/p_i) + \gamma_k \sum_{\ell=1}^{q} <(n_\ell - \sum_{i=1}^{s} d_{\ell i}\Pi_i)>^2$$

$$+ \gamma_k \sum_{i=1}^{s} <(-\Pi_i)>^2 + \gamma_k (\sum_{i=1}^{s} \Pi_i - 1)^2$$

$$+ \gamma_k \sum_{j=1}^{m} (\sum_{i=1}^{s} c_{ji}\Pi_i - \theta_j)^2$$

where γ_k is an increasing sequence of positive real numbers (k=1,2,3,...) and is termed the penalty parameter.

Following Rao (1984) the estimation proceeds as follows.

(i) Set k=1. Start with a set of values for the Π_i's and a suitable value for γ_i.

(ii) Find the vector $\underset{\sim}{\Pi}^*$ that minimizes $Q(\underset{\sim}{\Pi},\gamma_k)$.

(iii) Test whether the point $\underset{\sim}{\Pi}^*$ satisfies all the constraints. If $\underset{\sim}{\Pi}^*$ is indeed feasible, then it is the desired minimum. Otherwise, set k=2 and choose the next value of the penalty parameter which satisfies the relation

$$\gamma_{k+1} > \gamma_k$$

(iv) Go to step (ii).

The choice of the exterior penalty function method (over the interior penalty function method) is made on grounds of expediency. The use of the interior penalty function method requires the specification of a vector from the feasible set as start values. In problems with a large number of constraints, finding an appropriate vector of start values itself becomes a mathematical programming problem.

The constraints specifying nonnegativity of parameter estimates, that is, $\Pi_i \geq 0$, may be redundant since the objective function is defined only for positive values of the Π_i's. Another point to be noted is that the function $\Pi_i \ell n(\Pi_i/p_i)$ is not defined when Π_i is exactly equal to zero. In implementing the optimization problem, insufficiency of arithmetic precision may cause some parameter to be estimated at zero (say Π_k). In such cases, it is proposed to set the corresponding summand in the objective function ($\Pi_k \ell n(\Pi_k/p_k)$) to zero, consistent with the limiting behavior of the function as $\Pi_k \to 0^+$.

For each value of the penalty parameter γ_k, the unconstrained minimization is to be carried out using the Davidon-Fletcher-Powell variable metric method (Davidon, 1959; Fletcher and Powell, 1963). This method is preferred in cases where the number of variables in the objective function is large. In the present problem, the number of variables is equal to the number of cells in the multiway contingency table. The Davidon-Fletcher-Powell method does not require the evaluation of the matrix of second order partial derivatives of the auxiliary function. Further, being a conjugate gradient method, it is quadratically convergent.

Following Rao (1984) the iterative procedure of the method is as follows.

(i) Start with an initial vector $\underset{\sim}{\Pi}_i$ and an sxs positive definite symmetric matrix $\underset{\sim}{H}_1$, where s is the number of parameters to be estimated, that is, the number of cells in the contingency table. Usually $\underset{\sim}{H}_1$ is taken as the identity matrix I. Set iteration number n=1.

(ii) Compute the gradient of the function (∇Q_n) and at the point $\underset{\sim}{\Pi}_n$ and set

$$\underset{\sim}{S}_n = -\underset{\sim}{H}_n \nabla Q_n$$

(iii) Find the optimal step length λ_n^* in the direction $\underset{\sim}{S}_n$ and set

$$\underset{\sim}{\Pi}_{n+1} = \underset{\sim}{\Pi}_n + \lambda_n^* \underset{\sim}{S}_n$$

(iv) Test the new point $\underset{\sim}{\Pi}_{n+1}$ for optimality. If $\underset{\sim}{\Pi}_{n+1}$ is optimal, terminate the iterative process. Otherwise go to step (v).

(v) Update the H matrix as

$$\underset{\sim}{H}_{n+1} = \underset{\sim}{H}_n + \underset{\sim}{M}_n + \underset{\sim}{N}_n$$

where

$$\underset{\sim}{M}_n = \lambda_n^* \frac{\underset{\sim}{S}_n \underset{\sim}{S}_n^T}{\underset{\sim}{S}^T \underset{\sim}{Q}}$$

$$\underset{\sim}{N}_n = \frac{-(\underset{\sim}{H}_n \underset{\sim}{Q}_n)(\underset{\sim}{H}_n \underset{\sim}{Q}_n)^T}{\underset{\sim}{Q}_n^T \underset{\sim}{H}_n \underset{\sim}{Q}_n}$$

and

$$\underset{\sim}{Q}_n = \nabla Q_{n+1} - \nabla Q_n$$

(vi) Increase the iteration number by one unit and go to step (ii).

The computation of the gradient vector, the search direction, and the matrices $\underset{\sim}{H}$, $\underset{\sim}{M}$, and $\underset{\sim}{N}$ is straightforward. However, the efficient use of the method requires the accurate determination of the optimal step length λ_n^* at each iteration. The optimal step length is to be determined as follows.

Let $\underset{\sim}{S}_n^T = (S_{1n}, S_{2n}, \ldots, S_{sn})$ where s is the number of cells in the table

Then $\quad \underset{\sim}{\Pi}_{n+1} = \underset{\sim}{\Pi}_n + \lambda_n^* \underset{\sim}{S}_n$

and $\quad \underset{\sim}{\Pi}_{n+1}^T = (\Pi_{1n} + \lambda_n^* S_{1n}, \Pi_{2n} + \lambda_n^* S_{2n}, \ldots, \Pi_{sn} + \lambda_n^* S_{sn})$

Then the problem of finding the optimal step length reduces to finding the value of λ_n which minimizes $Q(\underset{\sim}{\Pi}_n + \lambda_n \underset{\sim}{S}_n)$, for fixed $\underset{\sim}{\Pi}_n$ and $\underset{\sim}{S}_n$.

Since λ_n is the only variable, and since the function has continuous first and second order partial derivatives with respect to λ_n, a Newton-Raphson procedure can be employed to determine the optimal step length.

Extension to n Samples

It is assumed that the samples are drawn independently.

Let s_k denote the total number of cells in the kth sample (multi-way table)

$k = 1, 2, \ldots, n$

p_{ik} denote the observed probability (in the kth sample) for the ith cell)

$i = 1, 2, .., s$

Π_{ik} denote the population probability (in the kth population) for the ith cell

w_k denote a set of known weights, that is,

$$\sum_{k=1}^{n} w_k = 1 \text{ and } 0 \leq w_k \leq 1$$

The estimation problem then becomes

Minimize

$$I(\Pi:p) = \sum_{k=1}^{n} w_k \sum_{i=1}^{s_k} \Pi_{ik} \ln(\Pi_{ik}/p_{ik})$$

subject to

$$\sum_{i=1}^{s_k} \Pi_{ik} = 1 \qquad k=1,2,\ldots,n$$

$$\sum_{i=1}^{s_k} c_{jik}\Pi_{ik} = \theta_{jk} \qquad j=1,2,\ldots,m_k \qquad m_k<s_k$$

$$\sum_{i=1}^{s_k} d_{\ell ik}\Pi_{ik} \geq \eta_{\ell k} \qquad \ell=1,2,\ldots,q_k$$

$$\Pi_{ik} \geq 0 \qquad i=1,2,\ldots,s_k;\ k=1,2,\ldots,n$$

where c_{jik}, θ_{jk}, $d_{\ell ik}$, and $\eta_{\ell k}$ are known constants.

The auxiliary function is

$$Q(\underset{\sim}{\Pi}_{ik},\gamma_h) = \sum_{k=1}^{n} w_k \sum_{i=1}^{s_k} \Pi_{ik} \ln(\Pi_{ik}/p_{ik})$$

$$+ \gamma_h \sum_{k=1}^{n} \sum_{\ell=1}^{q_k} <(\eta_{\ell k} - \sum_{i=1}^{s_k} d_{\ell ik}\Pi_{ik})>^2 + \gamma_h \sum_{k=1}^{n} \sum_{i=1}^{s_k} <(-\Pi_{ik})>^2$$

$$+ \gamma_h \sum_{k=1}^{n} (\sum_{i=1}^{s_k} \Pi_{ik} - 1)^2 + \gamma_h \sum_{k=1}^{n} \sum_{j=1}^{m} (\sum_{i=1}^{s_k} c_{jik}\Pi_{ik} - \theta_{jk})^2$$

where γ_h is the penalty parameter at the hth interation.

CHAPTER V

SIMULATION AND DATA ANALYSIS

This chapter describes a simulation which was conducted to assess the practical utility of the methodology, and presents an illustrative example of how the methodology can be applied in a segmentation context.

Simulation

Although the nature of the non-linear programming problem and the use of the variable metric minimization method assures theoretical convergence to the global minimum, it is still necessary to assess the methodology from a practical standpoint. Numerical errors and the arbitrary specification of the parameters of the algorithm (e.g., specifications of convergence and termination criteria) may lead to lack of convergence or convergence to a sub-optimal feasible solution. For example, efficient use of the variable metric minimization method requires accurate determination of the step length. However, too much accuracy may result in convergence to sub-optimal solutions (Box, 1966). An important practical consideration is the rate of convergence. If the rate of convergence is inadequate even for problems of reasonable size, then alternative algorithms (minimization methods) should be studied. For any particular problem, the parameters of the algorithm can be specified by trial and error to provide reasonably accurate solutions. However, the assessment of a methodology (in its implementation) requires that its performance be monitored over a wide variety of problems which form a representative sample drawn from the domain of application of the methodology. This requirement provided the rationale for the simulation design and the analysis of simulation results.

Simulation Design

Although the methodology can be applied to solve a variety of problems in different disciplines, the focus of this dissertation is on the analysis of market segments. Therefore, the domain of application was restricted to segmentation issues in designing the simulation study.

In the segmentation area alone several models can be hypothesized and tested. The appropriateness of a particular model or subset of models is a function of the specific problem situation and the managerial requirements, if any. However, one model which is of interest in almost all situations is an assessment of the extent to which a specified set of descriptor variables serve to discriminate among segments. This assessment is carried out by estimating parameters (cell probabilities) under a model which hypothesizes that the descriptors jointly provide no discrimination among market segments. Tests of and estimation under other models are meaningful only if the above hypothesis is rejected. Should the hypothesis not be rejected, the decision maker has to specify an alternative descriptor set for segment identification/evaluation. Therefore, the model which tests the hypothesis of no discrimination among market segments was chosen for the simulation study.

Given the above model, there exists infinitely many population structures (discrete probability distributions) from which random samples can be drawn for use in the simulation. However, from the segmentation standpoint, these distributions can be placed on a bipolar continuum with one endpoint consisting of distributions which provide absolutely no discrimination and the other consisting of probability distributions which discriminate perfectly among segments. Mathematically, the endpoints can be described as follows.

Let $\Pi(g/ijk\ell)$ be the conditional probability of being in the gth group (i.e., gth category of X_g) given the ith, jth, kth and ℓth levels of descriptor variables X_i, X_j, X_k and X_ℓ respectively.

$$g = 1,2,\ldots,G$$

$$i = 1,2,\ldots,I$$

$$\ell = 1,2,\ldots,L$$

Then the end-point consisting of distributions which provide absolutely no discrimination is the set of all distributions which satisfy the condition

$$\Pi(1/ijk\ell) = \Pi(2/ijk\ell) = \ldots = \Pi(g/ijk\ell) = \ldots = \Pi(G/ijk\ell) \quad \forall \quad (ijk\ell)$$

Similarly, the end-point consisting of distributions which discriminate perfectly among segments is the set of all distributions which satisfy the condition

$$\Pi(g/ijk\ell) = \begin{cases} 1 \text{ for one category of } X_g \\ 0 \text{ for other categories of } X_g \end{cases} \quad \forall \ (ijk\ell)$$

In order to generate random samples for the simulation it was decided to specify two population structures/distributions. The first distribution provided absolutely no discrimination among groups. The number of groups was fixed at two. The second distribution was specified by the condition

$$\Pi(1/ijk\ell) = .8$$

$$\Pi(2/ijk\ell) = .2$$

$$\forall\ (ijk\ell)$$

which provides reasonably good discrimination. It was not considered necessary to vary the number of groups for reasons given later.

Given the above models, the size of the problems can be varied by changing the number of descriptor variables or the number of categories associated with each descriptor variable. Insofar as the models are concerned, the method by which the problem size is altered is immaterial since the only effect is that the cell probabilities decrease as the number of cells in the complete multiway table increase. Therefore, rather than consider the number of descriptor variables or the number of categories for each variable, the problem was reformulated in terms of varying the total number of cells. It follows from the above that varying the number of groups would only serve to change the total number of cells; hence the decision not to incorporate the number of groups as a factor in the study. Thus the effect of varying the problem size was incorporated by specifying three levels of problem size-64 cells, 32 cells, and 16 cells. Assuming that the number of groups is fixed at two and that all descriptor variables are dichotomous, these cell sizes correspond to problems having 5, 4, and 3 descriptor variables respectively. While an upper limit of 64 cells might appear small, it appears unlikely that a confirmatory approach such as the one adopted in this dissertation can be implemented with large problems since the decision-maker has to specify a model by positing constraints on cell probabilities. Except possibly for the initial hypothesis of no discrimination among groups most models incorporate constraints which are essentially derived from the intuitions and knowledge of the decision-maker. Consequently the cognitive strain of model specification increases rapidly with increasing problem size. On the other hand, if the problem is large and the set of constraints relatively few in number, the methodology is unlikely to yield estimates which differ meaningfully from those observed in the sample.

In addition to the above, the effect of varying sample size was explicitly incorporated by specifying three levels of sample size--100, 500, and 1,000 respectively. The upper limit reflects what is usually observed in practice. This factor was incorporated to examine the effect of variability due to sampling. As the sample size increases, the distribution observed in the sample can be expected to conform more closely to the underlying population structure. Thus the simulation design used was a three-way layout with eighteen cells--two population structures (uniform and discriminant), three levels of problem size (16, 32, and 64 cells), and three levels of sample size (100, 500, and 1,000), completely crossed with one another. The number of replications in each of the eighteen cells was set at 100. For all cells, the estimation of cell probabilities was carried out under the model that the descriptor variable set did not provide for any discrimination between the two groups.

Specification of Dependent Measures

As mentioned earlier, an important practical consideration is the rate of convergence. There exist several criteria which could be used to assess this. The implementation of the methodology required three different types of iterations (each type being nested in the type immediately following)--a one-dimensional minimization using the cubic interpolation method to determine the optimal step length, the unconstrained minimization of the penalty function itself for a predetermined value of the penalty parameter, and iteration of the above for a sequence of penalty parameter values. The third type of iteration, that is, the number of times the penalty function was minimized (KITER) was chosen as the measure of the rate of convergence. The choice was a logical consequence of the program implementation. For the sake of efficiency the maximum number of iterations allowed for the unconstrained minimization of the penalty function was set equal to the number of variables (i.e., the number of cells in the multiway table). Therefore this number varied with the problem size and could not be used as a measure of the rate of convergence. Since the iterations of the optimal step-length determination were nested within the above, that measure could not be used either.

A second measure, designed to assess the methodology (i.e., the information-theoretic approach) rather than the algorithm used, was the average absolute deviation of the cell probability estimates (ABSDEV) from a predesignated probability distribution. The averaging was done over the number of cell probabilities estimated. The baseline distribution (from which absolute deviations were computed) was taken to be the population model of no discrimination between groups, for all cells in the simulation design. The rationale for choosing this measure is described below. The constraints in the model do not require the algorithm to recover the population structure. However, the model constraints would require that probability estimates be close to the above baseline distribution at least to the extent the sample mimics the population structures used for sample generation.

In summary, two dependent measures were specified--the number of times the penalty function was minimized (hereinafter referred to as "KITER") and the average absolute deviation of the cell probability estimates from the baseline distribution (hereinafter referred to as "ABSDEV").

Simulation Results

The analysis of results obtained from simulation are reported for the two dependent measures separately. In both cases, the data were analyzed in an ANOVA framework. Some additional analysis was carried out using the second measure (ABSDEV) to investigate its distribution about the baseline distribution. The alpha level was set at .10 for all statistical tests of significance.

Results for KITER (the Measure of the Rate of Convergence

Table 1 provides the ANOVA table for the dependent measure KITER. Although the table provides results of all main effects and interactions, any explanation of the findings should begin with the highest order interaction which is found to be significant and focus on the simple main effects associated with that particular interaction term. Interpretation of lower order interactions and main effects is not meaningful and provides no additional information. In the present case, the three-way interaction is not statistically significant. However all two-way interactions are significant. Hence the analysis is based on the simple main effects associated with these interaction terms. Table 2 provides the cell means necessary to analyze these simple main effects.

The general framework used for the analysis of simple main effects is as follows. If a kth order interaction is the highest significant effect in an n-way ($n \geq k$) layout (which implies that the nth order interaction is the highest possible) then variations in cell means across levels of one factor are analyzed for fixed levels of the remaining k-1 factors appearing in that interaction term. This is repeated for each of the k factors. Thus the results provided in Table 2 can be interpreted as follows. For every level of sample size, the discriminant model required more iterations than the uniform model. This is as expected since the estimation in all cases was carried out under the model hypothesizing no discrimination between groups. The samples

TABLE 1

ANOVA RESULTS FOR KITER

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Main effects				
(i) Sample size	321.01	2	160.51	1085.07
(ii) Model	246.42	1	246.42	1665.86
(iii) Problem size	14.06	2	7.03	47.52
(iv) Total [(i)+(ii)+(iii)]	581.49	5	116.30	786.21
Two-way interactions				
(i) Sample size x Model	5.52	2	2.76	18.67
(ii) Sample size x Problem size	142.78	4	35.70	241.31
(iii) Model x Problem size	100.05	2	50.03	338.19
(iv) Total [(i)+(ii)+(iii)]	248.35	8	31.04	209.87
Three-way interactions	0.273	4	.07	0.46*
Residual	263.60	1782	.15	

*Not significant at the pre-specified alpha level of .10

TABLE 2

CELL MEANS FOR THE ANALYSIS OF SIMPLE MAIN EFFECTS (KITER)

(a)

Sample size	Model: Uniform	Model: Discriminant
100	8.11	9.01
500	7.34	8.00
1000	7.33	7.99

(b)

	Problem Size		
Sample size	16	32	64
100	8.11	8.52	9.05
500	8.00	7.50	7.50
1000	8.00	7.50	7.49

(c)

	Problem Size		
Model	16	32	64
Uniform	8.00	7.31	7.47
Discriminant	8.07	8.37	8.56

generated using the uniform model would, on average, be more similar to the hypothesized model, and therefore be expected to converge more readily. For both levels of the model KITER decreases with increasing sample size. While the result is not unexpected for the uniform model, the plausible explanation for the occurrence of the same in the case of the discriminant model is that with increasing sample size, the probability of observing cells with zero counts (given the population structure used to generate the samples) decreases. It is likely that the logarithmic component of the objective function affects the rate of convergence for samples with cell probabilities in the neighborhood of zero.

In examining the simple main effects associated with the other two interactions certain anomalies manifest themselves. In the case of the interaction between sample size and the problem size KITER decreases with increasing problem size for sample sizes of 500 and 1,000 and increases with increasing problem size when the sample size is 100. This suggests that a sample size of 500 is adequate at least for problems of sizes incorporated into the simulation design. This is further substantiated by the observation that for every level of problem size KITER decreases as the sample size is increased from 100 to 500 and remains fairly stable thereafter. An inexplicable anomaly occurs in the interaction of problem size and model variations. While KITER increases with increasing problem size for the discriminant model, in the case of the uniform model it decreases as problem size increases from 16 to 32 cells and then increases. In the absence of any other probable cause, this can only be interpreted as a sampling artifact. However, for all levels of problem size KITER is higher for the discriminant model.

Results for ABSDEV (the Measure of Deviation from the Baseline Distribution)

Table 3 presents the ANOVA table for the dependent measure ABSDEV. Since the three-way interaction is significant, the analysis is done for the associated simple main effects. The appropriate cell means are provided in Table 4.

The analysis of simple main effects yields the following general conclusions. ABSDEV is less, on average, for the uniform model compared to the discriminant model although the differences tend to diminish with increasing sample size. With smaller sample sizes one can expect to find more zero cells (i.e., cells with zero counts) in samples generated from the discriminant population structure. In samples generated from either population structure ABSDEV tends to decrease with increasing sample size. Contrary to expectations, for fixed levels of sample size and model, ABSDEV did not vary as problem size was varied in three of the six comparisons. This may well be a sampling artifact. Another counterintuitive observation is that for sample size of 500 ABSDEV increases with decreasing problem size for both uniform and discriminant models. The same phenomenon occurs to a lesser extent with sample size of 1000.

Distribution of ABSDEV. In another attempt to assess the performance of the methodology, the distribution of ABSDEV in each cell of the simulation design was studied. Within each cell the 100 replications

TABLE 3

ANOVA RESULTS FOR ABSDEV

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Main effects				
(i) Sample size	0.057	2	0.029	3436.14
(ii) Model	0.004	1	0.004	534.11
(iii) Problem size	0.002	2	0.001	127.24
(iv) Total [(i)+(ii)+(iii)]	0.064	5	0.013	1532.17
Two-way interactions				
(i) Sample size x Model	0.004	2	0.002	223.39
(ii) Sample size x Problem size	0.000	4	0.000	1.65*
(iii) Model x Problem size	0.000	2	0.000	4.33
(iv) Total [(i)+(ii)+(iii)]	0.004	8	0.000	57.76
Three-way interactions	0.000	4	0.000	3.49
Residual	0.015	1782	0.000	

*Not significant at the pre-specified alpha level of .10

TABLE 4

CELL MEANS FOR THE ANALYSIS OF SIMPLE MAIN EFFECTS (ABSDEV)

(a) Problem size = 16

	Model	
Sample size	Uniform	Discriminant
100	0.01	0.02
500	0.01	0.01
1000	0.00	0.01

(b) Problem size = 32

	Model	
Sample size	Uniform	Discriminant
100	0.01	0.02
500	0.00	0.01
1000	0.00	0.00

(c) Problem size = 64

	Model	
Sample size	Uniform	Discriminant
100	0.01	0.02
500	0.00	0.00
1000	0.00	0.00

available were used to compute the standard deviation. Each value of ABSDEV was then compared to the standard deviation corresponding to that cell to assess the nature of the distribution. Using the standard deviation as the unit of measurement, the frequency of occurrence of ABSDEV in different intervals (ranges) was computed. The results are presented in Table 5 for each of the eighteen cells.

From the correspondence that exists between each sample (replication) and the set of cell probability estimates associated with that sample it is clear that the distribution of the solutions would bear a direct relation to the distribution of the samples. However, the measure ABSDEV is an average deviation where the averaging is done over the cell probabilities. Therefore, it is reasonable to expect the Central Limit Theorem to hold and consequently about 68% of the values of ABSDEV can be expected to lie within the range of one standard deviation. On average this expectation is largely fulfilled. The two most serious aberrations in this regard are the two cells characterized by the treatment combinations of (1) Uniform model, problem size = 64, sample size = 100 and (2) Discriminant model, problem size = 16, sample size = 1000. Overall, the results suggest that the variations in the solutions obtained are a direct consequence of sampling variations. It should be noted that the standard deviation used above was a sample-based estimate rather than the true population value since the latter was unknown.

TABLE 5

DISTRIBUTION OF ABSDEV

Model	Problem Size	Sample Size	Range*				
			a	b	c	d	e
Uniform	16	100	35	11	15	16	23
Uniform	16	500	31	18	20	13	18
Uniform	16	1000	30	16	17	17	20
Uniform	32	100	34	6	25	18	17
Uniform	32	100	34	6	25	18	17
Uniform	32	1000	29	18	11	22	20
Uniform	64	100	42	10	15	14	19
Uniform	64	500	30	15	13	21	21
Uniform	64	1000	32	14	16	30	18
Discriminant	16	100	28	13	16	16	27
Discriminant	16	500	33	15	16	17	19
Discriminant	16	1000	41	9	13	15	22
Discriminant	32	100	37	7	19	13	24
Discriminant	32	500	32	11	19	15	23
Discriminant	32	1000	37	9	13	16	25
Discriminant	64	100	27	22	29	0	22
Discriminant	64	500	35	12	20	20	13
Discriminant	64	1000	34	17	12	21	16

*a - greater than one standard deviation
b - between 75% and 100% of standard deviation
c - between 50% and 75% of standard deviation
d - between 25% and 50% of standard deviation
e - within 25% of standard deviation

Summary of Simulation Results

Two measures of performance were used--one to assess the performance of the algorithm (KITER) and another to assess the performance of the methodology (ABSDEV). The analysis of simulation results shows that on both counts the performance was satisfactory, and with few exceptions, in accordance with expectations. While the ANOVAs showed significant effects due to varying levels of different factors, the tests should be interpreted with some caution. An important assumption in the analysis of variance is that of variance homogeneity across treatments. This assumption is clearly violated at least in treatments with differing sample sizes and has implications for how the F-tests for the statistical significance of various main and interaction effects should be interpreted. On the other hand, the relatively large number of observations (1800 in all) may have a countervailing effect since the F-test is relatively robust to variance heterogeneity when sample sizes are large. However, large sample sizes make the F-test relatively powerful with the result that differences in means which are found to be statistically significant may have no practical significance whatsoever. At least with one measure (KITER) this appears to be the case especially considering that only integer valued differences are meaningful from a practical standpoint. Also no direct correspondence should be made between variations (or lack thereof) in KITER and computer time required since the number of iterations in the middle loop (i.e., the iterations to minimize the penalty function for a fixed value of the penalty parameter) was allowed to vary in accordance with the requirements of the algorithm.

An Illustrative Example

Description of Data

The data for the illustrative example are taken from Goldstein and Dillon (1978) and previously reported in an abridged form by Dash, Schiffman and Berenson (1977). Data on information-seeking activities were used to discriminate between two groups--shoppers who patronized a full-line department store and shoppers who patronized an audio equipment specialty store. The descriptor set consisted of four dichotomous variables related to information-seeking activities and is described below (Goldstein and Dillon, 1978).

Variable 1: (Information Seeking)

x_1 = 1 if the individual sought information from friends and/or neighbors before purchase; 0 otherwise

Variable 2: (Information Transmitting)

x_2 = 1 if the individual has recently been asked for an opinion about buying any audio product; 0 otherwise

Variable 3: (Prior Shopping Experience)

x_3 = 1 if the individual has shopped in any stores for audio equipment before making a decision; 0 otherwise

Variable 4: (Catalog Experience)

x_4 = 1 if the individual had sought information from manufacturers' catalogs before purchase; 0 otherwise

The complete cross-classification of all respondents in the sample as reported by Goldstein and Dillon (1978, p. 16) is given in Table 6. Goldstein and Dillon (1978) used the data to illustrate the similarities/differences in the classification of states to one of the two groups using different methods such as the full-multinomial, nearest neighbor and first-order independence rules.

<u>Reanalysis of the Data</u>

The first step in the reanalysis was to test the hypothesis (H_1) that the four descriptor variables did not provide any discrimination between the two groups. Tests of other hypotheses are meaningful only if the above hypothesis is rejected. To test the hypothesis of no discrimination between groups, the estimation of the population cell probabilities was done subject to the following constraints. Letting $\Pi(gijk\ell)$ denote the joint probability of being in the gth group and the ith, jth, kth, and ℓth categories of the descriptor set X_i, X_j, X_k, X_ℓ respectively, the constraints are given by

$$\Pi(11111) - \Pi(21111) = 0$$
$$\Pi(11110) - \Pi(21110) = 0$$
$$\vdots \qquad \vdots \qquad \vdots$$
$$\Pi(10000) - \Pi(20000) = 0$$

In general

$$\Pi(1ijk\ell) - \Pi(2ijk\ell) = 0 \quad \forall (ijk\ell)$$

TABLE 6

DATA PRESENTED BY GOLDSTEIN AND DILLON (1978)

State (X_1	X_2	X_3	X_4)	Full-line Department Store	Audio Equipment Specialty Store
1	1	1	1	5	86
1	1	1	0	2	22
1	1	0	1	15	23
1	1	0	0	4	11
1	0	1	1	3	3
1	0	1	0	3	4
1	0	0	1	3	4
1	0	0	0	5	3
0	1	1	1	14	33
0	1	1	0	8	6
0	1	0	1	26	30
0	1	0	0	12	5
0	0	1	1	2	8
0	0	1	0	3	6
0	0	0	1	32	8
0	0	0	0	17	6

The value of the objective function at the minimum was .2230, which, when multiplied by twice the sample size, is distributed as a central chi-square with degrees of freedom equal to the number of linearly independent equality constraints in the model. This excludes the equality constraint which requires the cell probabilities to sum to unity. The number of relevant equality constraints in the model is 16. The value of the chi-square random variable is approximately 183.75 (2x412x.2230) and therefore the hypothesis of no discrimination between groups is unambiguously rejected. In the present case this finding is hardly surprising since visual examination of the sample data would serve to indicate such an outcome. Given the rejection of the model, the estimates of cell probabilities are not meaningful and therefore are not reported. Having established that the descriptor set provides discrimination between the two groups, the logical step is to determine whether there exists some managerially meaningful structure/model underlying the data. In the present case, closer examination of the sample data reveals that not all profiles (that is, states described by the descriptor set alone) contribute to discrimination between the two groups. Thus from a managerial standpoint it would be useful to separate the profiles which discriminate well between the two groups from those which do not. A framework which provides for such a partition of the profiles is described below.

It is clear even from a cursory examination of the audio equipment market that consumers exhibit varying degrees of involvement. Given the nature of the variables in the descriptor set, the degree of involvement can be characterized by the number of variables to which the individual responds positively. For example, the profiles denoted by (1111) and (0000) describe individuals with the highest and lowest degrees of involvement respectively. If it is assumed that all four variables describe the degree of involvement equally well (i.e., they are equally weighted), then any two profiles can be ordered (by the degree of involvement) by comparing the number of ones appearing in each profile. Thus profiles (1100) and (0101) would imply the same degree of involvement whereas (1100) implies a lesser degree of involvement compared to (0111). In this framework the midpoint is characterized by profiles with two zeros and two ones.

The above framework is used to develop a sequence of nested hypotheses as follows. The first hypothesis (H2) in the sequence is that only states (1111) and (0000) serve to discriminate between segments. This implies the set of constraints

$$\Pi(1ijk\ell) - \Pi(2ijk\ell) = 0$$

for all $(ijk\ell)$ except (1111) and (0000). There are fourteen constraints in all and the model comprising these is nested in the model described earlier, that is, the hypothesis that no profile in the descriptor set discriminates between segments. The minimum value of the objective function was .757 yielding a chi-square value of approximately 62.38 with 14 degrees of freedom which leads to rejection of the hypothesis.

The next hypothesis, (H3), nested in the previous two, is that only states with at least three zeros (or ones) provide discrimination between the two segments. The constraints are

$$\pi(1ijk\ell) - \pi(2ijk\ell) = 0$$

for all $(ijk\ell)$ with exactly two ones (or equivalently, exactly two zeros). This gives a model with six constraints. The value of the objective function at the minimum was .0102 and the chi-square value was 8.4 with 6 degrees of freedom. The critical value at the .10 level is 10.645. Therefore the hypothesis is supported. The estimates of cell probabilities corresponding to the hypothesis are given in Table 7.

The table shows certain interesting features. All unconstrained probabilities are close to but higher than the corresponding observed values. The same phenomenon occurred with the cell probability estimates corresponding to the previous hypothesis (H2). The effect of sample size on the rejection or acceptance of hypotheses is highlighted by the fact that even though most probability estimates are close to the observed values, the chi-square value is fairly high. However, the utility of the estimates can best be illustrated in using them to allocate the states to one of the two groups using the following rule. Assign state $(ijk\ell)$ to Π_1 if $\pi(1ijk\ell)>\pi(2ijk\ell)$ and to Π_2 if $\pi(1ijk\ell)<\pi(2ijk\ell)$. The assignment is to be made randomly if equality holds. The assignments are given in Table 8. For comparison purposes the allocation according to the full multinomial model (see Goldstein and Dillon, 1978) is also presented.

Except for those states for which assignments are to be made at random and the state (1101) the two rules are in agreement. The allocation using the estimated cell probabilities appears to be more conservative. With both rules an anomaly occurs with state (0010). Contrary

TABLE 7

ESTIMATES OF CELL PROBABILITIES UNDER HYPOTHESIS H3

State	Full-line Department Observed	Store (Π_1) Estimated	Specialty Observed	Store (Π_2) Estimated
1111	.0121	.0123	.2087	.2108
1110	.0048	.0049	.0534	.0539
1101	.0364	.0368	.0558	.0564
1100[a]	.0097	.0163	.0267	.0163
1011	.0073	.0074	.0073	.0074
1010[a]	.0073	.0085	.0097	.0085
1001[a]	.0073	.0085	.0097	.0085
1000	.0121	.0123	.0073	.0074
0111	.0340	.0343	.0801	.0809
0110[a]	.0194	.0170	.0146	.0170
0101[a]	.0631	.0685	.0728	.0685
0100	.0291	.0294	.0121	.0123
0011[a]	.0048	.0098	.0194	.0098
0010	.0073	.0074	.0146	.0147
0001	.0777	.0784	.0194	.0196
0000	.0413	.0417	.0146	.0147

[a]indicates states with equality constraints

The probabilities may not sum to one due to rounding error.

TABLE 8

ALLOCATION OF STATES TO POPULATIONS

State	Allocation	Allocation by full multinomial model
1111	Π_2	Π_2
1110	Π_2	Π_2
1101	Π_2	Π_1
1100	Random	Π_2
1011	Random	Π_1
1010	Random	Π_1
1001	Random	Π_1
1000	Π_1	Π_1
0111	Π_2	Π_2
0110	Random	Π_1
0101	Random	Π_1
0100	Π_1	Π_1
0011	Random	Π_2
0010	Π_2	Π_2
0001	Π_1	Π_1
0000	Π_1	Π_1

to expectation, the state is assigned to population Π_2, and may represent a sampling artifact.

Summary of Reanalysis

The reanalysis of the data presented by Goldstein and Dillon (1978) illustrates the potential of the methodology for model building and hypothesis testing. The methodology is sufficiently flexible for a wide variety of models to be hypothesized and tested. In addition the methodology permits the development of models which provide for conservative allocation rules in discrimination problems.

CHAPTER VI

CONCLUSIONS

This dissertation develops a model-based approach to the analysis and evaluation of market segments. The segmentation problem was formulated as one of specifying two sets of variables--a basis set and a descriptor set. The basis set consists of variables which enable the decision maker to form managerially meaningful segments. The descriptor set serves to discriminate among segments.

From a methodological standpoint, the segmentation task was conceptualized as a dual problem of clustering and discrimination. A normative framework for discriminating among segments using descriptor sets consisting of categorical variables was developed. The use of the information theoretic approach made it possible to perform statistical estimation and hypothesis testing simultaneously.

A simulation study was designed to assess the performance of the methodology and the efficiency of the algorithm used to implement the methodology. The results showed that the methodology performed satisfactorily in uncovering any underlying structure. For problems of reasonable size, the algorithm was found to be reasonably efficient. The methodology was applied to a particular data set to show how the problem of discriminating among segments could be specified as a sequence of tests of nested models/hypotheses.

While the methodology was developed in the context of discriminating among segments, it is applicable in a wide variety of problem settings. There are two general conditions under which the methodology can be used. First, the variables used are categorical in nature. Second, a model can be specified, that is, a set of relationships posited to hold among population parameters/probabilities. The estimation of population parameters is then carried out under the null hypothesis that the model holds in the population concerned. A significant value of the chi square test statistic would indicate lack of support for the null hypothesis, that is, the model.

There are many other areas in marketing itself where this methodology could be fruitfully applied. One important application is to the analysis of brand switching data and the inference of market structure. Another possible extension is to latent class analysis which is presently modeled as a special case of the general framework of loglinear models.

SELECTED BIBLIOGRAPHY

Assael, Henry and A. Marvin Roscoe, Jr. "Approaches to Market Segmentation Analysis." Journal of Marketing, October 1976, 67-76.

Box, M.J. "A Comparison of Several Current Optimization Methods and the Use of Transformations in Constrained Problems." The Computer Journal, 9, 1966, 67-77.

Claycamp, H.J. and W.F. Massy. "A Theory of Market Segmentation." Journal of Marketing Research, Vol. V, November 1968, 388-394.

Dash, J.F., L. Schiffman, and C. Berenson. "Information Search and Store Choice." Journal of Advertising Research, 16, 1976, 35-40.

Davidon, W.C. "Variable Metric Method of Minimization." Argonne National Laboratory Report No. ANL-5990, 1959.

Fletcher, R. and M.J.D. Powell. "A Rapidly Convergent Descent Method for Minimization." Computer Journal, Vol. 6, No. 2, 1963, 163-168.

Frank, R., William Massy, and Yoram Wind. Market Segmentation. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1972.

Goldstein, Matthew and William R. Dillon. Discrete Discriminant Analysis. New York: John Wiley & Sons, Inc., 1978.

Green, P.E. "A New Approach to Market Segmentation." Business Horizons, 20, February 1977, 61-73.

Green, P.E., J.D. Carroll, and F.J. Carmone. "Some New Types of Fractional Factorial Designs for Marketing Experiments." In J.N. Sheth, ed., Research in Marketing, Vol. I. Greenwich, Connecticut: JAI Press, 1977.

Green, P.E. and Yoram Wind. Multiattribute Decisions in Marketing: A Measurement Approach. Hinsdale, Illinois: Dryden Press, 1973.

Mahajan, V. and A.K. Jain. "An Approach to Normative Segmentation." Journal of Marketing Research, 15, August 1978.

Matusita, K. "On Estimation by the Minimum Distance Method." Ann. Inst. Stat. Math., 7, 67-77.

Rao, S.S. Optimization Theory and Applications, 2nd ed. New York: John Wiley and Sons, Inc., 1984.

Smith, W. "Product Differentiation and Market Segmentation as Alternative Marketing Strategies." Journal of Marketing, 21, July 1956, 3-8.

Tollefson, J.O. and P. Lessig. "Aggregation Criteria in Normative Market Segmentation Theory." Journal of Marketing Research, 15, August 1978.

Wind, Y. "Issues and Advances in Segmentation Research." Journal of Marketing Research, 15, August 1978, 317-337.