

An Information Theoretic Framework for Exploratory Multivariate Market Segmentation Research*

Jamshid C. Hosseini

College of Business Administration, Marquette University, Milwaukee, WI 53233

Robert R. Harmon

School of Business Administration, Portland State University, Portland, OR 97207-0751

Martin Zwick

System Science Ph.D. Program, Portland State University, Portland, OR 97207-0751

ABSTRACT

State-of-the-art market segmentation often involves simultaneous consideration of multiple and overlapping variables. These variables are studied to assess their relationships, select a subset of variables which best represent the subgroups (segments) within a market, and determine the likelihood of membership of a given individual in a particular segment. Such information, obtained in the exploratory phase of a multivariate market segmentation study, leads to the construction of more parsimonious models. These models have less stringent data requirements while facilitating substantive evaluation to aid marketing managers in formulating more effective targeting and positioning strategies within different market segments. This paper utilizes the information-theoretic (IT) approach to address several issues in multivariate market segmentation studies. A marketing data set analyzed previously is employed to examine the suitability and usefulness of the proposed approach [12]. Some useful extensions of the IT framework and its applications are also discussed.

Subject Areas: Market Segmentation, Mathematical Programming, Minimum Discrimination Information (MDI), and Statistical Information Theory.

INTRODUCTION

Cross-fertilization of techniques and concepts has emerged as an important factor in developing an interdisciplinary posture in marketing research [24]. In contemporary market segmentation studies, it is often necessary to use multiple segmentation bases which include a variety of socioeconomic, demographic, psychographic, and behavioral variables [9] [12] [25]. Since many of these variables share commonalities, both conceptually and in practice, further analyses are needed in order to explicate the existence and the degree of overlap among the variables and identify those with the most (or least) contribution to this overlap. This analysis in the exploratory phase of a market segmentation study enables the marketing manager to construct more parsimonious descriptive and/or predictive models. The model involves less stringent data requirements, in terms of both the number of variables tracked and the number of observations required. Furthermore, calibration and validation of the model is easier, leading to increased usefulness of the model and, thereby, more effective marketing strategies.

*The authors wish to gratefully acknowledge the helpful comments of Professors W. R. Dillon, W. D. Perreault, Jr., and L. F. Feik on earlier work related to this study. The excellent reviews by Professor J. Craig Andrews, two anonymous reviewers, and an anonymous associate editor were invaluable in the presentation, flow, and focus of this paper.

Segmentation variables are usually categorical. When a number of them are considered simultaneously, multidimensional contingency tabulations are used to study their relationships. An approach to the analysis of multivariate segmentation problems known as segment congruence analysis [11] [12] [28] attempts to: (1) determine the extent of association among various segmentation variables; (2) select the variable (set) that exhibits the highest contribution to the (sub)set of variables which are mutually associated; and (3) assess the extent to which this subset is characterized by other variables. Segment congruence analysis usually progresses in the following manner:

1. Traditional dimension-reducing techniques such as factor and cluster analysis are used to identify a number of segmentation bases (batteries of variables).
2. These segmentation bases can then serve as categorical variables and a multidimensional contingency table is formed.
3. Various categorical data analysis tests are carried out on the multiway table to assess the nature and extent of associations among its dimensions.
4. A segmentation base is identified as the distinguished base and a model is developed for predicting this base from other (possibly external) variables.

Segment congruence analysis provides a relatively new approach to multivariate market segmentation studies where a segmentation base may be comprised of a composite cluster representing certain levels of several segmentation variables. These variables may include benefits, usage rates, psychographic, behavioral, socioeconomic, and/or demographic characteristics. Although marketing practitioners continue to choose a segmentation base by fiat [11], multivariate market segmentation analysis methodologies (such as segment congruence analysis) can provide new insights in terms of statistical support for the selected base as well as justification for the development of marketing strategy.

Log-linear modeling continues to be the most popular framework for the analysis of multivariate categorical data¹ [6] [7] [12] [22], and hence, multivariate market segmentation research. However, some researchers have expressed concern about the efficiency of this technique when (1) a large number of variables are involved and (2) the modeling process requires numerous analyses [6] [7]; furthermore, the log-linear modeling technique does not allow for (3) simultaneous hypothesis testing and estimation and (4) inclusion of external constraints [5] [21]. We believe that (1) and (2) are due to limitations of the currently available computational software and algorithms, while (3) and (4) may be considered potential limitations of the log-linear modeling technique. Nonetheless, in developing alternative decision-making and problem-solving methodologies, an existing technique does not have to be invalidated in order to justify the use of another method.

In this paper, we employ a framework based on information theory to analyze multiple segmentation variables with respect to their associations, strength of relations, and contributions to the prediction of membership in a target segment. This approach is closely related to the well-known entropy and minimum discrimination information (MDI) approaches that have been used to address a variety of marketing problems [2] [4] [5] [13] [18] [21] [29].

¹Recently, correspondence analysis [14] and its extensions (e.g., [8]) have received considerable attention in exploratory analysis of multivariate categorical data [24]. However, the main use of this technique has been in graphical representation and perceptual mapping of such data. Although the versatility of this technique in exploring patterns in the data has been demonstrated, it has not been extended to accommodate the issues addressed in the present study.

Many management and marketing scholars are already familiar with entropy models and their application to brand choice and switching behavior [2] [13] and market share analysis [29]. The Kullback-Liebler (KL) statistic used in the formulation of MDI models has also been used by some business and marketing researchers in brand choice and market segmentation [4] [5] [21]. Malhotra [23] described the use of the Akaike information criterion, a variation of the KL statistic, in evaluating logit models and expressed the need to further investigate the use of information theory in marketing research.

Charnes, Cooper, Learner, and Phillips [5] advocated the use of the IT approach in market structure analysis beyond the traditional maximum entropy models proposed by Herniter [13] and investigated by others [2] [18] [29]. Charnes et al. showed that the IT framework is more efficient than the classical approaches because of its capability to perform simultaneous hypothesis testing and estimation, and more effective because it allows the inclusion of external constraints based on management plans or policies. Recently, Kumar and Dillon [21] provided an excellent account of this claim by formulating several complex hypotheses with external constraints imposed by the manager's a priori knowledge about market conditions.

It is our belief that a general IT methodology is developing and gaining acceptance among marketing researchers and practitioners. We hope that this study contributes to the growing body of knowledge by demonstrating another area of usefulness for the IT framework; namely, multivariate market segmentation involving categorical variables. The following research questions provide the focus for the present study:

1. Given that a number of (a priori and/or post-hoc) segmentation variables are being considered as potential segmentation bases, is there a statistically significant overlap (or mutual association) among all or subsets of these variables?
2. Which segmentation variable makes the highest statistically significant contribution to the overall mutual association?
3. Of the segmentation variables under consideration, which one is best determined once we have the knowledge of the other (descriptor) variables?
4. What is the prioritized order of contributions of the set of descriptor variables to the determination of the criterion variable identified in 3?

The data analyzed in [12] are used to demonstrate the suitability of this approach and to compare the results obtained by this method with those obtained by the general log-linear and logit models. Additionally, some managerial implications of the results in the context of the example used are discussed.

THE INFORMATION THEORETIC FRAMEWORK

The most general IT formulation to date is the KL statistic (also known as cross-entropy) which has the following form [20]:

$$I(p : \pi) = \sum_{i=1}^n p_i \ln(p_i / \pi_i) \quad (1)$$

where \mathbf{p} and $\boldsymbol{\pi}$ are vectors with components $p_i \geq 0, \pi_i \geq 0$, and $\sum_i p_i = \sum_i \pi_i = 1, i=1, 2, \dots, n$. $I(\mathbf{p} : \boldsymbol{\pi})$ can be interpreted as the information distance between the actual cell probabilities (\mathbf{p}) and their hypothesized counterparts ($\boldsymbol{\pi}$). p_i and π_i can also represent prior and posterior probabilities, respectively, as in Bayesian decision theory. It can be shown that Shannon's entropy [27], which measures the distance between the observed (p_i) and the maximal expected probabilities ($\pi_i = 1/n, \forall_i$), can easily be derived from (1). For example, the entropy of a univariate distribution for a random variable X , $H(X)$, can be obtained by:

$$\begin{aligned}
 H(X) &= (\ln(1/n) - I(\mathbf{p} : \boldsymbol{\pi}_0)) / \ln(2) \\
 &= (\ln(1/n) - \sum_{i \in X} p_i \ln(p_i / \pi_0)) / \ln(2) \\
 &= (\ln(1/n) - \sum_{i \in X} p_i \ln p_i - \sum_{i \in X} p_i \ln(1/n)) / \ln(2) \\
 &= - \sum_{i \in X} p_i \log_2 p_i.
 \end{aligned} \tag{2}$$

Shannon's entropy is interpreted as a measure of variability (scatter, spread, uncertainty, behavioral freedom) of the variable [15]. This measurement is a unique, unambiguous criterion based on the intuitive notion that a broad distribution represents more *uncertainty* than does a narrow, sharply peaked distribution [17]. Reduction in the entropy of the variable X reflects the gain of information about the value of the phenomenon X . Entropy is a non-negative quantity applicable to metric and nonmetric variables. A zero value for $H(X)$ indicates the availability of only one state for X (i.e., $\log_2(1) = 0$); thus, X would have no choice or behavioral alternative. Maximum entropy is obtained when $H(X)$ is computed from a uniform distribution in which case $H(X) = \log_2 n$; n = the number of possible states for X .

It can also be shown that McGill's [26] transmission or transmitted information measure, which is the cross-entropy between the observed joint probabilities of two or more events ($P_{ijk\dots}$) and their expected probabilities at independence ($\pi_{ijk\dots} = P_i P_j P_k \dots$), can be obtained from the KL statistic. In a two-variable distribution, for example, the transmission measure, $T(X:Y)$, can be obtained by

$$\begin{aligned}
 T(X:Y) &= I(\mathbf{p} : \boldsymbol{\pi}) / \ln(2) \\
 &= \sum_{i \in X} \sum_{j \in Y} p_{ij} \log_2 (p_{ij} / p_i p_j) \\
 &= \sum_{i \in X} \sum_{j \in Y} p_{ij} \log_2 p_{ij} - \left(\sum_{i \in X} \sum_{j \in Y} p_{ij} \log_2 p_i + \sum_{i \in X} \sum_{j \in Y} p_{ij} \log_2 p_j \right) \\
 &= -H(X, Y) + H(X) + H(Y)
 \end{aligned} \tag{3}$$

where p and π are i by j matrices with $i \in X$ and $j \in Y$, and p_i and p_j are marginal probabilities. The above expressions measure the amount of statistical dependence (or constraint or relatedness) of a pair of variables X and Y . Transmission is symmetric, that is, $T(X:Y) = T(Y:X)$, and non-negative. Its range is given by $0 \leq T(X:Y) \leq \text{Min}\{H(X), H(Y)\}$, with a zero value denoting independence between X and Y , and $\text{Min}\{H(X), H(Y)\}$ occurring when the knowledge about the state of one variable completely determines the state of the other variable [26].

Multivariate measures of entropy and cross-entropy provide a convenient means of hypothesis testing and estimation regarding the relationships among the variables in a multidimensional contingency table [5], and thereby provide a new methodology for analyzing multivariate market segments in the context of segment congruence analysis. These measures are shown to have asymptotic χ^2 distributions with the likelihood-ratio χ^2 value (L^2) equal to $(2N)I(p:n) = (2N)\ln(2)(T(X:Y))$, where N is the number of observations (see, e.g., [1]). The degrees of freedom are computed based on the following formula:

$$df = \prod_i c(v_i) - \sum_i (c(v_i) - 1) - 1$$

where i = number of variables in the contingency table, and $c(v_i)$ = number of categories in variable i . Note that this formula reduces to $(c(v_1) - 1)(c(v_2) - 1)$ for a $V_1 \times V_2$ contingency table.

Information measures are completely decomposable and can be partitioned into variability between and variability within sets of variables, allowing for ANOVA-like analysis of variables [15] [26]. One decomposition algorithm developed by Krippendorff [19] can be used to prioritize the contributions of the elements in a set of descriptor variables to the reduction of uncertainty in a particular criterion variable. Let Y be a criterion variable with uncertainty $H(Y)$ and X_1 , X_2 , and X_3 be three descriptor variables which contribute to the reduction of uncertainty in Y . Based on Krippendorff's algorithm, bivariate transmissions are computed between the criterion variable and the remaining variables (e.g., $T(Y: j)$, $j = X_1, X_2, X_3$). These bivariate transmissions (i.e., $T(Y:X_1)$, $T(Y:X_2)$, and $T(Y:X_3)$), are then used to identify the variable that contributes the most to the reduction of uncertainty in the criterion variable. Next, the conditional bivariate transmissions between the criterion variable and the remaining descriptor variables, conditional upon the variable selected in the preceding step (say X_1), are computed (i.e., $T_{X_1}(Y:X_2)$ and $T_{X_1}(Y:X_3)$),² and the largest of these is selected (say X_3). Repeating this process for all of the independent variables results in some partitioning similar to the following:

$$H(Y) = T(Y:X_1) + T_{X_1}(Y:X_3) + T_{X_1X_3}(Y:X_2) + H_{X_1X_2X_3}(Y), \quad (4)$$

²By convention, IT conditional terms are represented by subscripting the variable which is given (or held constant). For example, $T_{i;j:k:l}$ refers to the transmission between the variables represented by j , k , and l , given the variable i . Computations for these terms are straightforward extensions of marginal IT measures, replacing the marginal probabilities with appropriate conditional probabilities.

where $H_{X_1 X_2 X_3}(Y)$ is the uncertainty in Y not accounted for by variables X_1 , X_2 , or X_3 . The sequence in which the descriptor variables are selected represents the prioritized contributions of these variables to reduction of uncertainty in the criterion variable. Thus, if a variable's contribution is negligible (as judged by its L^2 value and its corresponding degrees of freedom), it can be discarded from further analyses.

Table 1 presents the IT formulations of our research questions [16]. It is shown that (1) the transmission measure can be used to assess the existence and degree of mutual associations among the variables; (2) conditional transmissions can be employed to determine the contribution of each variable (or set of variables) to the total mutual association; (3) the normalized reduction in the uncertainty of each variable (or set of variables) can be used to determine the variable (set) best determined by the other variables; and (4) Krippendorff's decomposition method can be used to prioritize the contributions of each (descriptor) variable to the reduction of uncertainty in another (criterion) variable. A FORTRAN computer program that performs all of the above computations can be obtained from the authors.

AN ILLUSTRATIVE APPLICATION

In order to examine the suitability and usefulness of the proposed approach, a data set originally analyzed by Green and Carmone [12] is employed. This enables a comparison of the results obtained by the proposed methodology and those obtained by the log-linear/ logit modeling approach. Table 2 presents these data in a multidimensional contingency table format.

The source data were part of a larger data base on different attitudinal, psychographic, and behavioral traits of 534 respondents regarding automobile insurance along with some of their socioeconomic and demographic characteristics. The data were subsequently analyzed in [12] through factor analysis and hierarchical clustering, and four segmentation variables were identified and cross-classified. The procedures leading to the identification of the segmentation variables are standard procedures in post hoc segmentation studies and need not be repeated here. The four segmentation variables consisted of Insurance Supplier (S), Image (M), Psychographic (P), and Demographic (D), each composed of two categories.

Table 3 presents the amount of association among all possible combinations of variables as measured by their transmission (one with all four variables, four with three variables each, and six with two variables each), along with their associated L^2 , degrees of freedom, and p-values. As expected, the L^2 for each model (e.g., T(S:M:P:D), T(S:M:P)) would be identical to its equivalent log-linear model (i.e., {S} {M} {P} {D} and {S} {M} {P})³. Based on the p-values for different tests of independence, no statistical association exists between Insurance Supplier (S) and the Psychometric variable (P). Similarly, the association between Image (M) and Demographic (D) segments is not statistically significant. Although Green and Carmone [12] did not explore these refinements of associations among variables, these analyses could be performed by developing and testing various log-linear models.

³We use the general log-linear modeling nomenclature where (S) (M) (P) (D) represents the main-effects model, which excludes all two-way and higher order interactions; and (SMPD) represents the saturated model, which includes all possible interactions.

The additional findings indicate that the significance of associations in models with higher order relations are partially due to the lower order relations which have significant associations. Therefore, the significant two-way associations (i.e., S and M, S and *D*, *M* and P, and P and *D*) as well as the unique three-way interactions account for the significance of the four three-variable models. Similarly, the significance of the total mutual association is due to the significant two-way and three-way associations as well as the unique four-way association in the system.

Table 4 shows the segmentation variable (or set of variables) whose independence from the remaining variables is the subject of the second research question. First, Method 1 (described in Table 1) is used to determine the contribution of each variable set to the total mutual association (e.g., $T(i\ j:k:l)-T(j:k:l)$ or, equivalently, the log-linear model $\{i\} \{jkl\}$), its corresponding L^2 , and p-value for the four-variable system. Using the drop in L^2 (i.e., the smallest drop referring to the variable with the highest contribution to the mutual association), we can assess not only the contributions of each variable, but those of multiple variables to the total mutual association. The Image (*M*) variable with a drop of 14.81 (63.00-48.19) has the highest contribution among the four single variables, while the combination of Insurance Supplier and Psychographic segments (*SP*) has the highest contribution among the bivariate segmentation bases.

Also, in Table 4, Method 2 (described in Table 1) is used to compute associations due to each variable (e.g., $T_i(j:k:l)$), the L^2 values, and the p-levels. An examination of the two methods reveals that they address the same issue with respect to the contribution of a variable-set to the total mutual association in the system. The first method measures the drop in the distance between the observed probabilities and the expected probabilities at independence due to a particular variable-set. The smaller this drop is the more that variable-set contributes to the association in the system. On the other hand, the second method assesses whether, given a variable-set, the remaining variables become independent of each other. Here, the variable-set which causes the remaining variables to get closest to independence contributes the most to the total mutual association.

Using Method 2, the variable (set) with the smallest remaining L^2 contributes the most to the mutual association. Based on this interpretation, it can be verified that the results of the second approach are identical to the first approach.

Note that, although Image and Insurance Supplier are the variables with the highest individual contributions to the total mutual association, their combination (*SM*) ranks number four in the two-variable combinations of variables. This indicates that these two variables are highly associated with one another. This can be confirmed by referring to Table 3 where the association between S and *M* is the highest among the two-variable associations with $T(S:M)=.0450$ and $L^2=33.31$ on one degree of freedom.

Table 5 lists the marginal and conditional entropies (e.g., $H(i)$, $H(-1)$) for each possible combination of segmentation variable sets along with H_{det} , L^5 , degrees of freedom, and p-values. As shown in Table 1, H_{det} represents the entropy accounted for by other variables. Among the single variable segmentation bases, *M* and S appear to have the two highest reduced entropies, *SP* and *MD* are best accounted for among the two-variable bases, and *MPD* is best determined among the threevariable sets. The results are most meaningful with respect to a single variable segmentation base in this illustration since too few descriptor variables remain with

Table 1: IT characterization of the four research questions.

Research Question	IT Formulation	Remarks
Mutual association	$T(i;j:k;l)$ $T(i;j:k)$ $T(i;j:l)$ $T(i;k:l)$	The mutual association among variables cannot be fully addressed by simply developing and testing a log-linear independence (main-effects) model of all of the variables (e.g., $\{i\}\{j\}\{k\}\{l\}$). All main-effect models (i.e., $\{i\}\{j\}\{k\}$, $\{i\}\{j\}\{l\}$, etc.), must be constructed and tested. We use the transmission measure, not only for the four variable contingency table, but also for all possible compositions of three and two variable tables developed by summing over one or more variables.
Contributions of individual variables		
Method 1	$T(i;j:k;l)-T(j:k;l)$ $T(i;j:k;l)-T(i;k:l)$ $T(i;j:k;l)-T(i;j:l)$ $T(i;j:k;l)-T(k;l)$	Note that the first term concerns the distance between the four variable system and independence, while the second term measures the distance between the three variable system and independence. Here, we are testing the hypothesis that two-way and three-way associations are equal to zero (i.e., the main-effects models).
Method 2	$T_i(j;k;l)$ $T_j(i;k;l)$ $T_k(i;j;l)$ $T_l(i;j;k)$	If the knowledge of the state of one variable (e.g., that represented by i in the expression $T_i(j;k;l)$) renders the remaining variables statistically disjoint, then this variable accounts for a substantial amount of the mutual association in the system. We can, therefore, use the drop in the amount of total mutual association to determine the variable (or set of variables) with the highest contribution.
Determination of a best variable	$H_{det}^{(i)} = \frac{H(i) - H_{jk}(i)}{H_{max} - H(i,j,k,l)}$ $H_{det}^{(j)} = \frac{H(j) - H_{ik}(j)}{H_{max} - H(i,j,k,l)}$ $H_{det}^{(k,l)} = \frac{H(k,l) - H_j(k,l)}{H_{max} - H(i,j,k,l)}$	The numerator terms are the amount of uncertainty remaining in each variable (e.g., the variable represented by i in the first expression), given the knowledge the state of the remaining variables (i.e., those represented by j , k , and l). The denominator terms are the maximum amount of the uncertainty that can be accounted for. Therefore, the H_{det} is the normalized amount of uncertainty in the variable represented by i . The variable with the highest amount of reduced uncertainty, then, is designated as the variable best determined by the other variables.
Prioritization of the contributions of the descriptor variables	$H(i) = T(i;j) + T_j(i;k)$ $+ T_{jk}(i;l) + H_{jk}(i)$ $H(i,j) = T(i;j;k)$ $+ T_k(i;j;l) + H_k(i,j)$	Assuming that i represents the selected criterion variable; $T(i;j)$, $T_j(i;k)$, and $T_{jk}(i;l)$ represent the prioritized contributions by variables represented by j , k , and l to the determination of that variable, while $H_{jk}(i)$ measures the residual uncertainty.

Insurance Supplier Segments (S)	Image Segments (M)	Psychographic Segments (P)	Demographic Segments (D)	Frequency
1	1	1	1	5
1	1	1	2	18
1	1	2	1	11
1	1	2	2	34
1	2	1	1	12
1	2	1	2	29
1	2	2	1	9
1	2	2	2	37
2	1	1	1	52
2	1	1	2	59
2	1	2	1	46
2	1	2	2	111
2	2	1	1	22
2	2	1	2	42
2	2	2	1	16
2	2	2	2	31
				534

(S) Insurance Supplier Segments 1:

Sponsor's customers

2: Other firms' customers (M)

Image Segments

1: No specialization by age of insured

2: Caters especially to the elderly

(P) Psychographic Segments

1: Controlled, conservative, old fashion

2: Risk taker, fatalistic, willing to try new things (D)

Demographic Segments

1: Primarily married couples; young in age; larger number of household members

2: Greater incidence of widows and other single females; older in age; smaller number of household members

Note: Source is Green and Carmone [12].

the two- and three-variable segmentation bases. In general, this procedure enables the identification of the best set of variables as the composite segmentation base. Using the above results and assuming that a single-variable segmentation base is desired, we select Image (M) as the criterion variable. No equivalent analyses were done in either [12] or [28] with respect to predictability of potential segmentation bases. Instead, in both cases a criterion variable was selected on an a priori basis. With Image (M) as the criterion variable, we employed Krippendorff's [19] decomposition algorithm summarized in (4) to develop a prioritized list of the descriptor variables with respect to their contribution to the reduction of uncertainty in this variable. The following partitioning resulted:

$$H(M) = T(M:S) + T_S(M:P) + T_{SP}(M:D) + H_{SPD}(M),$$

$$.9513 = .0450 + .0150 \quad +.0051 \quad +.8862. \quad (5)$$

Table 3: Mutual associations among segmentation variables.

Segmentation Variables	Amount of Association	L^2	Degrees of Freedom	Risk
<i>S M P D</i>	.0851	63.00	11	.000
<i>S M P</i>	.0615	45.53	4	.000
<i>S M D</i>	.0554	41.01	4	.000
<i>S P D</i>	.0200	14.81	4	.005
<i>M P D</i>	.0239	17.69	4	.002
<i>S M</i>	.0450	33.31	1	.000
<i>S P</i>	.0014	1.04	1	.319
<i>S D</i>	.0101	7.48	1	.007
<i>M P</i>	.0118	8.74	1	.004
<i>M D</i>	.0013	.96	1	.339
<i>P D</i>	.0086	6.37	1	.012

$T(M:S)$, $T_S(M:P)$, and $T_{SP}(M:D)$ are the contributions of variables *S*, *P*, and *D*, respectively, to the determination of variable *M*, and $H_{SPD}(M)$ is the amount of variability not accounted for by the descriptor variables. Notice that the contribution of *S* to the reduction of uncertainty in *M* is statistically significant ($L^2=33.31$, $df=1$, $p<.001$). Likewise, the incremental contribution of *P* is significant ($L^2=11.10$, $df=2$, $p<.01$). However, the incremental contribution of *D* when the *S* and *P* variables are present in the model is not statistically significant ($L^2=3.78$, $df=4$, $p>.1$). This would indicate that this variable could be dropped from any subsequent descriptive/predictive models without a substantial decrease in the explanatory power of the model.

If the Insurance Supplier were selected a priori as the criterion variable, the prioritized contributions would have been Image ($T(S:M)=.0450$, $L^2=33.31$, $df=1$, $p<.001$), Demographic ($T_M(S:D)=.0091$, $L^2=6.73$, $df=2$, $p<.01$), and Psychographic ($T_{MD}(S:P)=.0071$, $L^2=5.26$, $df=4$, $p>.1$) variables, respectively, with the latter term being statistically insignificant. This sequence is identical to that derived in [12] using the coefficients of the logit model. However, this approach is practical only when a binomial main-effects logit model is used. Otherwise, the model would not have unique coefficients for each variable to be used in the prioritization.

EXTENSIONS AND CONCLUSIONS

Categorical data analysis involving multiple variables has received much attention in statistics, marketing, and psychology literature. The excellent work of Bishop, Feinberg, and Holland [3] focused on the use of log-linear modeling for this purpose. However, Gokhale and Kullback [10] provided the unifying rationale by using the K-L statistic and showed the log-linear model itself can be derived from this statistic. They demonstrated that the IT methodology can be used for hypothesis testing and estimation beyond that provided by log-linear modeling [5].

In the present paper, we demonstrated that the IT approach can be used in the exploratory analysis of multivariate market segments. We addressed four questions with respect to multivariate segmentation studies in the exploratory phase. The results obtained were consistent with those obtained through the log-linear modeling approach suggested by Green and Carmone [12]. Some critical extensions of their analyses not addressed previously were included.

Table 4: Contribution of each variable set to the total mutual association.

Segmentation Variables Set	Method 1				Method 2			
	Amount	L^2	df	Risk	Amount	L^2	df	Risk
S	.0612	45.31	7	.000	.0285	21.01	8	.008
M	.0651	48.19	7	.000	.0270	19.99	8	.011
P	.0297	21.99	7	.003	.0633	46.86	8	.000
D	.0236	17.47	7	.016	.0650	48.12	8	.000
S M	.0765	56.63	10	.000	.0132	9.77	4	.046
S P	.0838	62.04	10	.000	.0051	3.76	4	.444
S D	.0733	54.26	10	.000	.0198	14.66	4	.006
M P	.0750	55.52	10	.000	.0114	8.44	4	.071
M D	.0837	61.96	10	.000	.0071	5.26	4	.264
P D	.0401	29.69	10	.001	.0498	36.87	4	.000

results obtained were consistent with those obtained through the log-linear modeling approach suggested by Green and Carmone [12]. Some critical extensions of their analyses not addressed previously were included.

Although the analyses presented here primarily address the strength of associations, the variable contributions, and the predictability of each variable, the IT methodology is quite general. Not only can association and independence hypotheses be incorporated within this framework, but a priori knowledge and external constraints can be accommodated in the general MDI approach using mathematical programming concepts. For example, the question of mutual association among S, M, P, and D leads to the development of the following MDI model:

$$\text{Min } K(\mathbf{p};\boldsymbol{\pi}) = \sum_{i \in S} \sum_{j \in M} \sum_{k \in P} \sum_{l \in D} p_{ijkl} \ln(p_{ijkl} / \pi_{ijkl}) \quad (6)$$

subject to

$$\sum_{i \in S} \sum_{j \in M} \sum_{k \in P} \sum_{l \in D} p_{ijkl} = 1, \quad (7)$$

where, $\pi_{ijkl} = p_i p_j p_k p_l, i \in S, j \in M, k \in P, \text{ and } l \in D$. This model generates estimated cell probabilities for the four-way contingency table (P_{ijkl}) . A X^2 goodness-of-fit test between P_{ijkl} and p_{ijkl} would then lead to a decision regarding the rejection of the model and, hence, the hypothesis of independence among the four variables.

Other conditions such as independence of each variable from the remaining variables can be incorporated in the above model. For example, to test the independence of S from the remaining variables, the constraint set $p_{ijkl} = p_{2jkl}, j \in M, k \in P, l \in D$ must be included. Similarly, to include the hypothetical a priori knowledge that the Insurance Supplier (S) is not associated with the Psychographic variable (P), the constraint set $p_{1j1l} = p_{2j1l} = p_{1j2l} = p_{2j2l}, j \in M, l \in D$ must be added to (6) and (7). Kumar and Dillon [21] provided several examples of incorporating a priori or policy-based external constraints in the MDI models.

Table 5: Determination of each variable-set by the other variables.

Criterion Variable-Set	$H(i)$ Total	$H_{jkl}(i)$ Conditional	H_{det} Ratio*	L^2	df	Risk
<i>S</i>	.8691	.8079	.1681	45.31	7	.000
<i>M</i>	.9513	.8862	.1788	48.19	7	.000
<i>P</i>	.9921	.0297	.0816	21.99	7	.003
<i>D</i>	.9086	.0236	.0648	17.47	7	.016
<i>S M</i>	1.7753	1.7438	.0865	23.32	9	.006
<i>S P</i>	1.8597	1.7774	.2261	60.93	9	.000
<i>S M D</i>	1.7676	1.7044	.1736	46.79	9	.000
<i>M P</i>	1.9316	1.8684	.1736	46.79	9	.000
<i>M D</i>	1.8586	1.7763	.2261	60.93	9	.000
<i>P D</i>	1.8921	1.8606	.0865	23.32	9	.006
<i>S M P</i>	2.7509	2.7273	.0648	17.47	7	.016
<i>S M D</i>	2.6736	2.6439	.0816	21.99	7	.003
<i>S P D</i>	2.7498	2.6847	.1788	48.19	7	.000
<i>M P D</i>	2.8281	2.7669	.1681	45.31	7	.000

* $H_{\max} - \log_2(16) - 4$; and $H(i, j, k, l) - H(S, M, P, D) = 3.6360$.

Although we presented the case with all variables having two categories, our computer program allows for twelve variables, each having ten categories. These limits are set through parameters that can be changed easily. Furthermore, only one computer run is required to address all of the research questions as compared with numerous runs needed for equivalent log-linear analyses. Additionally, the IT measures presented in this paper are not limited to categorical data. All of the measures used here are based on the information distance between two distributions. These distributions are multinomial in the context of categorical data analysis. There is no conceptual difficulty with computing the information distance between two continuous distributions. In this case, the KL statistic (for continuous probabilities $p(x)$ and $n(x)$) is computed by

$$I(p : \pi) = \int p(x) \ln(p(x) / \pi(x)), \quad (8)$$

and the remaining equations are adjusted accordingly [10].

Furthermore, the proposed framework can easily be used for other business, economic, or public policy issues. These may include: how income is distributed among individuals in a particular population; how sales are distributed among brands, firms, or industries; how world trade is distributed among countries; how employment is distributed within or between geographic regions; how household purchases are distributed among product lines, product categories, or brands; how assets are distributed among the entries of a balance sheet; how much dispersion exists in rates of return for an individual's portfolio; how would the changes in the proportions of securities in the New York Stock Exchange affect the future proportions; and how decentralized is the decision-making process in an organization. All of these issues have already been addressed in the literature; our proposed framework allows for further investigation of these issues with respect to the identification and analysis of the most and least significant variables.

Further work in the IT approach to marketing problems should focus on the development of a unified body of techniques to analyze marketing data in exploratory, descriptive, or predictive modes. Although no conceptual barriers exist for such a development, many marketing practitioners are still relatively unfamiliar with information theory and its applications. Furthermore, no integrated and universally accepted computer package exists to perform the information-theoretic analyses. Mathematical programming based IT formulations have been solved using general nonlinear programming packages such as GRG2 and GINO [21]. In addition to our computer program, several other packages are available (e.g., SPECTRAL developed by Krippendorff [19]) which perform analyses similar to the SPSS-HILOGLINEAR routine.

Finally, more methodological and empirical applications using the IT framework would give this approach more credence. Thus far, information theory continues to remain an underutilized framework for business and marketing decision making. [Received: September 20, 1989. Accepted: June 12, 1990.]

REFERENCES

- [1] Attneave, F. *Applications of information theory to psychology: A summary of basic concepts, methods, and results*. New York: Holt, Rinehart and Winston, 1959.
- [2] Bass, F. M. The theory of stochastic preference and brand switching. *Journal of Marketing Research*, 1974, 11(1), 1-20.
- [3] Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. *Discrete multivariate analysis: Theory and practice*. Cambridge, MA: The MIT Press, 1975.
- [4] Charnes, A., Cooper, W. W., & Learner, D. B. Constrained information-theoretic characterizations in consumer purchase behavior. *Journal of the Operational Research Society*, 1978, 29(9), 833-42.
- [5] Charnes, A., Cooper, W. W., Learner, D. B., & Phillips, F. Y. An MDI model and an algorithm for composite hypothesis testing and estimation in marketing. *Marketing Science*, 1984, 3(1), 55-72.
- [6] Danaher, P. J. A log-linear model for predicting magazine audiences. *Journal of Marketing Research*, 1988, 25(4), 356-362.
- [7] Danaher, P. J. An approximate log-linear model for predicting magazine audiences. *Journal of Marketing Research*, 1989, 26(4), 473-479.
- [8] Fox, J. R. Perceptual mapping using the basic structure matrix decomposition. *Journal of the Academy of Marketing Science*, 1988, 16(1), 47-59.
- [9] Frank, R. E., Massy, W. F., & Wind, Y. *Market segmentation*. Englewood Cliffs, NJ: PrenticeHall, Inc., 1972.
- [10] Gokhale, D. V., & Kullback, S. *The information in contingency tables*. New York: Marcell Dekker, Inc., 1978.
- [11] Green, P. E. A new approach to market segmentation, *Business Horizons*. February 1977, 61-73. (12)
- [12] Green, P. E., & Carmone, F. J. Segment congruence analysis: A method for analyzing association among alternative bases for market segmentation. *Journal of Consumer Research*, 1977, 3(4), 217-222.
- [13] Herniter, J. D. An entropy model of brand purchase behavior. *Journal of Marketing Research*, 1973, 10(4), 361-75.
- [14] Hoffman, D. L., & Franke, G. R. Correspondence analysis: Graphical representation of categorical data in marketing research. *Journal of Marketing Research*, 1986, 23(3), 213-227.
- [15] Horowitz, A. R., & Horowitz, I. The real and illusory virtues of entropy-based measures for business and economic analysis. *Decision Sciences*, 1976, 7, 121-136.
- [16] Hosseini, J. C. *Segment congruence analysis: An information-theoretic approach*. Unpublished doctoral dissertation, Portland State University, 1987.
- [17] Jaynes, E. T. Information theory and statistical mechanics. *Physical Reviews*, 1957, 106(4), 620-630.

- [18] Kalwani, M. U., & Morrison, D. G. A parsimonious description of the Hendry system. *Management Science*, 1979, 23(5), 467-77.
- [19] Krippendorff, K. Information theory: Structural models for qualitative data. *Sage University Paper Series on Quantitative Applications in the Social Sciences*, 1986, 07-062, Beverly Hills, CA and London, England: Sage Publications.
- [20] Kullback, S., & Liebler, R. A. On information and sufficiency. *Annals of Mathematical Statistics*, 1951, 22, 79-86.
- [21] Kumar, A., & Dillon, W. R. Constrained discrimination via MDI estimation: The use of additional information in segmentation analysis. *Journal of Marketing Research*, 1987, 24(4), 396-403. [22] Kumar, A., & Sashi, C. M. Confirmatory analysis of aggregate hierarchical market structures: Inferences from brand switching behavior. *Journal of Marketing Research*, 1989, 26(4), 444-453. [23] Malhotra, N. K. The use of linear logit models in marketing research. *Journal of Marketing Research*, 1984, 21(1), 20-31.
- [24] Malhotra, N. K. Some observations on the state of the art in marketing research. *Journal of the Academy of Marketing Science*, 1988, 16(1), 4-24.
- [25] Malhotra, N. K. Segmenting hospitals for improved management strategy. *Journal of Health Care Marketing*, 1989, 9(3), 45-52.
- [26] McGill, W. J. Multivariate information transmission. *Psychometrika*, 1954, 19(2), 97-116.
- [27] Shannon, C. E., & Weaver, W. *The mathematical theory of communication*. Urbana, IL: University of Illinois Press, 1949.
- [28] Van Auken, S., & Lonial, S. C. Assessing mutual association between alternative market segmentation bases. *Journal of Advertising*, 1984, 13(1), 11-16.
- [29] Vanhonacker, W. R. Testing the exact order of an individual's choice process in an informationtheoretic framework. *Journal of Marketing Research*, 1985, 22(4), 377-87.

Jamshid Hosseini is Assistant Professor of Management Science at Marquette University. Dr. Hosseini received his Ph.D. in system science-management science from Portland State University. His research interests include the application of multivariate techniques, general systems methodology, and multi criteria decision-making tools to business problems. He has recent publications in several professional journals including *Decision Sciences*, *Journal of Business Research*, *Business Horizons*, and others. He is a member of TIMS, the Decision Sciences Institute, and the American Statistical Association.

Robert Harmon is a Professor and Chairman of the Marketing Department at Portland State University. He obtained his Ph.D. in marketing from Arizona State University. His research interests include marketing strategy, pricing strategy, marketing theory, and multivariate analysis. He has recent publications in the *Journal of Marketing Research*, *Journal of Marketing*, *Journal of Advertising*, *Journal of Advertising Research*, and others. Dr. Harmon is a member of AMA, ACR, Decision Sciences Institute, American Academy of Advertising, and Product Development and Management Association.

Martin Zwick is a Professor in the System Science Ph.D. Program at Portland State University. He obtained his Ph.D. in biophysics from the Massachusetts Institute of Technology (MIT). His research interests are in the application of general systems theory and methodology to the natural and social sciences, mathematical/computational methods in macromolecular structure analysis, and systems philosophy. He has recent publications in *General Systems*, *Annual Reviews of Biophysics and Biophysical Chemistry*, *Quarterly Review of Economics and Business*, and others. Dr. Zwick is a member of SGSR and Philosophy of Science Association.